Manuel Schwarz
2022-10-20
Due 10/20/22 3pm

1.3 Conditioning on Background Evidence

(a) Product Rule

Can be derived of conditional probability

$P(X,Y|E) = \frac{P(X,Y,E)}{P(E)}$

$P(X,Y,E) = P(X,E)P(Y|X,E)$ product rule

$P(X,E) = P(E)P(X|E)$ product rule

$P(X,Y|E) = \frac{P(E)P(X|E)P(Y|X,E)}{P(E)}$ product rule / simplify

$P(X,Y|E) = P(X|Y,E)P(Y,E)$

(b) Bayes Rule

Starting with the LHS

$P(X|Y,E) = \frac{P(X,Y,E)}{P(Y,E)}$ conditional def

$= \frac{P(Y|X,E)P(X,E)}{P(Y,E)}$ product rule

$= \frac{P(Y|X,E)P(X|E)P(E)}{P(Y,E)}$ product rule

$= \frac{P(Y|X,E)P(X|E)P(E)}{P(Y|E)P(E)}$ conditional prob def

$= \frac{P(Y|X,E)P(X|E)}{P(Y|E)}$ product rule / simplify

$P(X|Y,E) = \frac{P(Y|X,E)P(X|E)}{P(Y|E)}$

(c) Marginalization

Start $P(Y|E) = \sum_X P(X,Y|E)$ product rule, marginalization

$P(X,Y) = P(X|Y)P(Y)$ conditional prob

$= P(X)\sum_Y P(X,Y,Y=y)$ relevant, $P(X)$ leave $\sum_Y$

$P(Y|E) = P(Y)\sum_X P(X|Y=y,E)$ product rule, simplify

$P(Y|E) = \sum_X P(X|Y=y,E)P(Y|E)$

1.6 Conditional Independence

Looking at the LHS of (1) we can express it using the conditional form of Bayes rule denoted in (b)(b)

$P(X|Y,E) = \frac{P(Y|X,E)P(X|E)}{P(Y|E)}$

If (1) is true, we can consider $P(Y|X,E) = P(Y|E)$:

$P(X|Y,E) = \frac{P(Y|E)P(X|E)}{P(Y|E)}$

$P(X|Y,E) = P(X|E)$

Thus (1) implies (2)

From the conditional form, version of the product rule proved in 1.5a

$P(X,Y|E) = P(Y|X,E)P(X|E)$

If (2) is true, we can rewrite this as

$P(X,Y|E) = P(Y|E)P(X|E)$

Thus (2) implies (1)

From the product rule proved in 1.5a

$P(X,Y|E) = P(Y|X,E)P(X|E)$

If (3) is true, this can be transformed as done above

(3) $P(X,Y|E) = P(Y|X,E)P(X|E)$

$P(Y|X,E)P(X|E) = P(X|Y,E)P(Y|E)$

$P(X|Y,E) = P(X|E)$

Therefore, (3) implies (2)

We have now proven all the assumptions are equivalent as (1)$\Rightarrow$(2)$\Rightarrow$(3)$\Rightarrow$(1)

1.5 Counting Coins

(a) Frequency evidence

$P(E=1,\sum_i X_i = 5) = P(E=1|\sum_i X_i = 5)P(\sum_i X_i = 5)$

E: get an "A" on ESSA (0 not at $\pm$ 1 A)

$\sum_i X_i$: sum obtaining coins (5 about, 5 not about)

Y: not doing FBs (0: do this + 1: don't)

The probability of getting an A is equal than the probability of getting 5 heads given 5 heads. Given 5 heads, we can with 100% chance get an A. Thus the probability of getting an A is 100% given 5 heads.

(b) Frequency evidence

$P(Y=1|E=1) = P(Y=1)$

$P(Y=1|E=1) = P(Y=1)$ ← $P(Y=1|E=1)$ = Y: Variable in question, 1: Value in question, E: Evidence, 1: Value in question

Y and E independent given X (common dependent)

Y: get attempts

E: Standard Deviation

X: Cost

The get attempts and standard deviation are directly not related but rather has a direct effect on the cost.

1.9 Bayes Rule

C: has COVID $\in \{T,S\}$

T: test positive $\in \{T,S\}$

(a)

$P(C=T)=0.01$	C	$P(T=1\|C)$	$P(T=0\|C)$
$P(C=S)=0.99$	T	0.98	0.02
	S	0.05	0.95

(b) $P(C=0|T=1)$

$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$

$P(C|T) = \frac{P(T=1|C)P(C)}{P(T)}$

$P(C=0|T=1) = \frac{P(T=1|C=0)P(C=0)}{P(T=1)}$

We know

$P(T=1|C=S) = 0.05$

$P(C=S) = 0.99$

We need to know

$P(T=1)$

$P(T=1) = \sum_{C=\{T,S\}} P(T=1,C=i)$ marginalization

$= \sum_{i \in \{T,S\}} P(T=1|C=i)P(C=i)$ product rule

$= P(C=T)P(T=1|C=T) + P(C=S)P(T=1|C=S)$

$= 0.99(0.05) + (0.01)(0.98)$

$= 0.0595$

$P(C=0|T=1) = \frac{0.05 \cdot 0.99}{0.0595} = 0.8319$

(c) $P(C=1|T=0)$

$P(C=1|T=0) = \frac{P(T=0|C=1)P(C=1)}{P(T=0)}$

$P(T=0) = \sum_{i \in \{T,S\}} P(C=i,T=0)$

$= \sum_{i \in \{T,S\}} P(C=i)P(T=0|C=i)$

$= P(C=T)P(T=0|C=T) + P(C=S)P(T=0|C=S)$

$= 0.99(0.95) + 0.01(0.02)$

$= 0.9304$

$P(C=1|T=0) = \frac{0.02 \cdot 0.01}{0.9304}$

$= 2.15 \times 10^{-4}$

1.5 Entropy

(a) Binary entropy

$H(X) = -\sum_i f(x_i)\log_2 \quad$ where $\sum f_i = 1$

$\frac{d}{dp}\left[-\sum f(x_i)\log_2 - \lambda\left(\sum f_i - 1\right)\right]$

$= \frac{d}{dp}\left(-\log(p) - 1\right) - \lambda \frac{d}{dp}(1)$

$= \frac{d}{dp}\left(-\log_2 - 1\right) - \lambda p_i$

$0 = \frac{d}{dp}\log_2 - \lambda n$

$0 = \frac{d}{dp}\ln(p) - \lambda n$

$\frac{d}{dp}\ln(p) = \frac{d}{dp}(x)\lambda n$

$\frac{d}{dp}\ln(p) = \lambda$

$\frac{d}{dp}\ln(p) = 1$

$e^{\ln(p)} = 1$

$e^{\ln(p)} = 1/n$

ONCE $e^{\ln(p)} = p$.

$p = 1/n$

(b) Joint entropy

Show that $H(X_1, X_2 \ldots X_n) = \sum_i H(X_i)$

$P(x_1, x_2 \ldots x_n) = \prod_{i=1}^{n} P(x_i)$

$H(X_1, X_2 \ldots X_n) = -\sum_{i=1}^{n} \sum_{x_i} P(x_1, x_2 \ldots x_n)\log P(x_1, x_2 \ldots x_n)$

$= -\sum_{i=1}^{n} \sum_{x_i} P(x_1 \ldots x_n)\log\left(\prod_{i=1}^{n} P(x_i)\right)$ (indep)

$= -\sum_{i=1}^{n} \sum_{x_i} P(x_1 \ldots x_n)\log P(x_1 \ldots x_n) + \log(P(x_i))$ (log)

$= -\sum_{i=1}^{n} \underbrace{\sum_{x_i} P(x_1 \ldots x_n)\log P(x_1 \ldots x_n)}_{H(x_1 \ldots x_n)} + \underbrace{\sum_{x_i} P(x_i)}_{1}$

$= H(x_1 \ldots x_n)$

We can see the same logic the prove completes independence $\sum H(x_i)$

1.6 Kullback-Leibler distance

(a) Sketch graphs

$\frac{d}{dx}\left(\log(x) - (x-1)\right) = 0$

$\frac{1}{x} - 1 = 0$

$\frac{1}{x} = 1$

$x = 1$ ✓

(b) Show $KL(p,q) \geq 0$

Know $\ln(x) \leq x-1$

$-\ln(x) \geq 1-x$ mult by (-1)

let $x = \frac{q_i}{p_i}$ $\rightarrow -\ln\left(\frac{q_i}{p_i}\right) \geq 1 - \frac{q_i}{p_i}$

$KL(p,q) = \sum p_i \log\left(\frac{p_i}{q_i}\right)$ mult by (-1)

$= -\sum p_i \log\left(\frac{q_i}{p_i}\right)$ bring down sum

$KL(p,q) \geq -\sum p_i\left(1 - \frac{q_i}{p_i}\right)$

$\geq \sum p_i - \sum p_i$

$\geq \sum p_i - \sum q_i$

(c) Show tighter lower bound.

$\ln x \leq x-1$ (a)

$\ln(2x^2+x) \leq 2x^2+x-1$ (a) (b)

$\ln 2 \leq 2(1-\frac{1}{2})$ (a) (b)

$\sum 2x^2 + \sum x - \sum 1 \geq \sum \ln 2$ (a) + (b) + (c)

$\sum p_i \ln\left(\frac{p_i}{q_i}\right) \geq \sum 2p_i(1-\frac{q_i}{p_i}) + \sum \frac{p_i}{q_i} - \sum p_i$ (a) + (b) + (c)

$\geq \sum 2p_i - \sum 2\frac{q_i}{p_i} - \sum p_i$ (a) + (b) + (c)

$\geq \sum p_i - \sum q_i$

$\geq \sum p_i \ln\left(\frac{p_i}{q_i}\right) \geq \sum (\sqrt{p_i} - \sqrt{q_i})^2$

(d) Symmetry Cauchy-Schwarz inequality

For a distribution on where

	x	
p	0.1	0.9
q	0.9	0.3

$KL(p,q) = \sum p_i \log(p_i/q_i)$

$= 0.1\log\left(\frac{0.1}{0.9}\right) + 0.9\log\left(\frac{0.9}{0.3}\right)$

$= 0.549$

$KL(q,p) = \sum q_i \log(q_i/p_i)$

$= 0.9\log\left(\frac{0.9}{0.1}\right) + 0.3\log\left(\frac{0.3}{0.9}\right)$

$= 0.549$

$\Rightarrow KL \neq KL$

1.8 Mutual Information

(a) $I(X,Y) = \sum_x \sum_y P(x,y)\log\left[\frac{P(x,y)}{P(x)P(y)}\right]$

(b) Homogeneous

$\frac{\ln(x)}{\ln(x)} \leq \frac{x-1}{x-1}$ l'hospital

$\sum_x \sum_y P(x,y)\left[\frac{P(x,y)}{P(x)P(y)}\right] \leq \sum_x \sum_y P(x,y)\left[1 - \frac{P(x)P(y)}{P(x,y)}\right]$

$= \sum_x \sum_y P(x,y)\left[\frac{P(x,y)}{P(x)P(y)}\right] - 0$

(c) ME variables iff X, Y independent

IF X and Y are independent then

$\frac{P(x,y)}{P(x)P(y)} = 1$ by definition

Thus

$I(X,Y) \sum_x \sum_y P(x,y)\log\left[\frac{P(x,y)}{P(x)P(y)}\right]$

$= \sum_x \sum_y P(x,y)\ln(1) = 0$

Thus in this only circumstances where this would occur

1.8 Conjunct & Sentence

(a) Yes, Sunny implies I has a conditional dependence $P(I|S)$ that is not product of S P and I P and S are correlated by I thus they become conditionally independent - which outlook is but not S

(b) No additional independence

(c) Yes, given in X_2 Y and Z are conditionally independent when conditioned on X The independence is not product = 0.3

1.9 Phenomena

(a) Chance

correctly guessed	incorrectly guessed	best next guess f	$P(f_i = 1$ for some $i \in \{1,2,3,4\})$
[]	[]	$\frac{1}{2}$	0.5625
[1]	[1]	$\frac{1}{2}$	0.6328
[1,2]	[]	$\frac{1}{2}$	0.8125
[1,2,3]	[]	$\frac{1}{2}$	0.9375
[1,2,3,4]	[]	$\frac{1}{2}$	0.9375
[1,2,3,4]	[1]	$\frac{1}{2}$	0.9375
[1,2,3,4]	[1,2]	$\frac{1}{2}$	0.9375
[1,2,3,4]	[1,2,3]	$\frac{1}{2}$	0.9375
[1,2,3,4]	[1,2,3,4]	$\frac{1}{2}$	0.9375
[1]	[]	1	0.7031
[1,2]	[1]	1	0.8125
[1,2,3]	[1,2]	1	0.9375
[1,2,3,4]	[1,2,3]	1	0.9375
[1,2,3,4]	[1,2,3,4]	1	0.9375
[1,2]	[]	2	0.7031
[1,2,3]	[1]	2	0.8125
[1,2,3,4]	[1,2]	2	0.9375
[1,2,3,4]	[1,2,3]	2	0.9375
[1,2,3,4]	[1,2,3,4]	2	0.9375
[1,2,3]	[]	3	0.7031
[1,2,3,4]	[1]	3	0.8125
[1,2,3,4]	[1,2]	3	0.9375
[1,2,3,4]	[1,2,3]	3	0.9375
[1,2,3,4]	[1,2,3,4]	3	0.9375
[1,2,3,4]	[]	4	0.6328
[1,2,3,4]	[1]	4	0.7031
[1,2,3,4]	[1,2]	4	0.8125
[1,2,3,4]	[1,2,3]	4	0.9375
[1,2,3,4]	[1,2,3,4]	4	0.9375

(b) Optimal

$P(C=T)=0.01$	C	$P(T=1\|C)$	$P(T=0\|C)$
$P(C=S)=0.99$	T	0.98	0.02
	S	0.05	0.95

(c) $P(C=0|T=1)$

$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$

$P(C|T) = \frac{P(T=1|C)P(C)}{P(T)}$

$P(C=0|T=1) = \frac{P(T=1|C=0)P(C=0)}{P(T=1)}$

We know

$P(T=1|C=S) = 0.05$

$P(C=S) = 0.99$

We need to know

$P(T=1)$

$P(T=1) = \sum_{C=\{T,S\}} P(T=1,C=i)$ marginalization

$= \sum_{i \in \{T,S\}} P(T=1|C=i)P(C=i)$ product rule

$= P(C=T)P(T=1|C=T) + P(C=S)P(T=1|C=S)$

$= 0.99(0.05) + (0.01)(0.98)$

$= 0.0595$

$P(C=0|T=1) = \frac{0.05 \cdot 0.99}{0.0595} = 0.8319$

(d) $P(C=1|T=0)$

$P(C=1|T=0) = \frac{P(T=0|C=1)P(C=1)}{P(T=0)}$

$P(T=0) = \sum_{i \in \{T,S\}} P(C=i,T=0)$

$= \sum_{i \in \{T,S\}} P(C=i)P(T=0|C=i)$

$= P(C=T)P(T=0|C=T) + P(C=S)P(T=0|C=S)$

$= 0.99(0.95) + 0.01(0.02)$

$= 0.9304$

$P(C=1|T=0) = \frac{0.02 \cdot 0.01}{0.9304}$

$= 2.15 \times 10^{-4}$