STA 571 HW1

Ryan Tang

September 9th 2020

1 Excerise 2.6

(a) Let $H \in \{1, ..., K\}$ be a discrete random variable, and let e1 and e2 be the observed values of two other random variables E1 and E2. Suppose we wish to calculate the vector

$$\vec{P}(H|e_1, e_2) = [P(H = 1|e_1, e_2), ..., P(H = K|e_1, e_2)]^T$$

Which of the following sets of numbers are sufficient for the calculation?

1. $P(e_1, e_2), P(H), P(e_1|H), P(e_2|H)$
2. $P(e_1, e_2), P(H), P(e_1, e_2|H)$
3. $P(e_1|H), P(e_2|H), P(H)$

Answer Without any more underlying assumptions, only point (2) is sufficient.
The conditional probability of H can be expanded based off the Bayes rule.

$$P(H|e_1, e_2) = \frac{P(e_1, e_2|H)P(H)}{P(e_1, e_2)}$$

(b) Now suppose we now assume $E_1 \perp E_2|H$

Answer Now all 3 are sufficient.
For (1), given the conditional independence, $P(e_1, e_2|H) = P(e_1|H)P(e_2|H)$.
For (2),

$$\begin{aligned}
P(e_1, e_2) &= \int P(e_1|H)P(e_2|H)P(H)\,dH \\
&= \int P(e_1, e_2|H)P(H)\,dH \\
&= \int P(e_1, e_2, H)\,dH \\
&= P(e_1, e_2)
\end{aligned}$$

2 Excerise 2.7

Pairwise independence does not imply mutual independence

Answer Suppose a simple example $X_1, X_2, X_3 \in 0, 1$ three random binary variables.

Mutual independence $\rightarrow P(X_1, X_2, X_3) = P(X_1)P(X_2)P(X_3)$.

Pairwise independence $\rightarrow P(X_1, X_2, X_3) = P(X_1|X_2, X_3)P(X_2)P(X_3)$ instead.

If we suppose *mutual independence* $\rightarrow$ *pairwise independence*, then we have

$$P(X_1) = P(X_1|X_2, X_3)$$

However, the statement is violated given the following example. Say X_2 and X_3 are two fair binary random variables $\in \{0, 1\}$ and X_1 is, instead, generated in the following way where $P(X_1 = 0|X_2 = X_3) = 1$ and vice versa.

X2	X3	X1
0	0	0
1	1	0
1	0	1
0	1	1

3 Excerise 2.8

Conditional independence iff joint factorizes $X \perp Y|Z$ iff $p(x, y|z) = p(x|z)p(y|z)$

Proofs $X \perp Y|Z$ iff $p(x, y|z) = g(x, z)h(y, z)$

Answer Let's assume the above statement is true, then we can write

Proof.

$$\begin{aligned}
\iint p(x, y|z)\, dxdy &= \iint g(x, z)h(y, z)\, dxdy \\
p(x|z) &= \int_Y g(x, z)h(y, z)\, dy \propto g(x, z) \\
p(y|z) &\propto h(y, z) \\
\therefore p(x|z)p(y|z) &= c(z)g(x, z)h(y, z)
\end{aligned}$$

Because both side are proper probability and the intergral of the entire space should sum up to 1. Hence, $c(z)$, the constant has to be 1 which implies

$$g(x, z)h(y, z) \Rightarrow p(x|z)p(y|z)$$

$\square$

4 Excerise 2.12

Proof.

$$
\begin{aligned}
\mathbb{MI}(X,Y) &= KL[p(x,y)||p(x)p(y)] \\
&= \sum_X \sum_Y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= \sum_X \sum_Y p(x,y) \log \frac{p(x,y)}{p(x)} - \sum_X \sum_Y p(x,y) \log p(y) \\
&= \sum_X p(x) \sum_Y p(y|x) \log p(y|x) - \sum_Y p(y) \log p(y) \sum_X p(x|y) \\
&= -\sum_X p(x) H(Y|X=x) - \sum_Y p(y) \log p(y) \\
&= H(Y) - H(Y|X) \\
&= H(X) - H(X|Y)
\end{aligned}
$$

□

5 Excerise 2.15

Proof.

$$
\begin{aligned}
\hat{\theta} &= \arg\max_\theta q(X|\theta) && X \in R^{N\times D} \\
&= \arg\max_\theta \prod_{i\in X} q(x_i|\theta) \\
&= \arg\max_\theta \frac{1}{N} \sum_{i\in X} \log q(x_i|\theta) \\
&= \arg\min_\theta \mathop{\mathbb{E}}_X[-\log q(x|\theta)] && \text{MLE result}
\end{aligned}
$$

$$
\begin{aligned}
KL(p||q(X|\theta)) &= \sum_X p(x) \log \frac{p(x)}{q(x|\theta)} \\
&= \mathop{\mathbb{E}}_X[\log p(x) - \log q(x|\theta)] \\
&= \arg\min_\theta \mathop{\mathbb{E}}_X[-\log q(x|\theta)] && \text{remove constant} \\
&= \text{MLE } \hat{\theta}
\end{aligned}
$$

□

6 Excerise 3.20