EE5907/EE5027 Week 1: Probability Review Problems

The following questions are from Kevin Murphy's (KM) book "Machine Learning: A Probabilistic Perspective".

Exercise 2.6: Conditional independence

(a) Let $H \in \{1, \cdots, K\}$ be a discrete random variable, amd let e_1 and e_2 be the observed values of two other random variables E_1 and E_2. Suppose we wish to calculate the vector

$$\vec{P}(H|e_1, e_2) = (P(H = 1|e_1, e_2), \cdots, P(H = K|e_1, e_2)) \tag{1}$$

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

i. $P(e_1, e_2), P(H), P(e_1|H), P(e_2|H)$
ii. $P(e_1, e_2), P(H), P(e_1, e_2|H)$
iii. $P(e_1|H), P(e_2|H), P(H)$

(b) Now suppose we now assume $E_1 \perp E_2|H$ (i.e., E_1 and E_2 are conditionally independent given H). Which of the above 3 sets are sufficent now?

Show your claculations as well as giving the final result. Hint: use Bayes rule.

Exercise 2.7: Pairwise independence does not imply mutual independence

We say that two random variables are pairwise independent if

$$p(X_2|X_1) = p(X_2) \tag{2}$$

and hence

$$p(X_2, X_1) = p(X_1)p(X_2|X_1) = p(X_1)p(X_2) \tag{3}$$

We say that n random variables are mutually independent if

$$p(X_i|X_S) = p(X_i) \; \forall S \subseteq \{1, \cdots, n\} \setminus \{i\} \tag{4}$$

and hence

$$p(X_{1:n}) = \prod_{i=1}^{n} p(X_i) \tag{5}$$

Show that pairwise independence between all pairs of variables does not necessarily imply mutual independence. It suffices to give a counter example.

Exercise 2.8: Conditional indepence iff joint factorizes

In the text we said $X \perp Y|Z$ iff

$$p(x, y|z) = p(x|z)p(y|z) \tag{6}$$

for all x, y, z such that $p(z) > 0$. Now prove the following alternative definition: $X \perp Y|Z$ iff there exist function g and h such that

$$p(x, y|z) = g(x, z)h(y, z) \tag{7}$$

for all x, y, z such that $p(z) > 0$

Exercise 2.10: Deriving the inverse gamma density

Suppose $Y = g(X)$, where g is monotonic. If the pdf of X is given by $f(x)$, then the pdf of Y is given by $\left|\frac{d}{dy}(g^{-1}(y))\right| f(g^{-1}(y))$. This is known as the change of variables formula (wikipedia link)

Now, let $X \sim Ga(a, b)$, i.e.

$$Ga(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-xb} \tag{8}$$

Let $Y = 1/X$. Show that $Y \sim IG(a, b)$, i.e.,

$$IG(x|shape = a, scale = b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} e^{-b/x} \tag{9}$$

Hint: use the change of variables formula

Exercise 2.12: Expressing mutual information in terms of entropies

Show that

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X), \tag{10}$$

where $I(X,Y)$ = mutual information, $H(X)$ = entropy, $H(Y|X)$ = conditional entropy:

$$I(X,Y) = \mathrm{KL}(p(X,Y)||p(X)p(Y)) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

$$H(X) = -\sum_x p(x) \log_2 p(x)$$

$$H(Y|X) = \sum_x p(x) H(Y|X=x) = \sum_x p(x) \left(-\sum_y p(y|X=x) \log_2 p(y|X=x) \right)$$

Exercise 2.16: Mean, mode, variance for the beta distribution

Suppose $\theta \sim Beta(a,b) = \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1}$, where $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ and $\Gamma(n) = (n-1)!$ if $n \in \mathbb{Z}^+$. Derive the mean, mode and variance.