10-606: Mathematical Foundations for Machine Learning **Fall 2021**

Recitation 4: Probability Exercises

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

4.1 The Chain Rule

One of the most useful tools in probability is the so-called *Chain Rule*. This rule allows you to express a joint probability in terms of only conditional probabilities. Suppose that A and B are random variables. Then the chain rule states that

$$P(A, B) = P(A|B)P(B)$$

This can also be applied to more than one variable. For example, with three random variables A_1, A_2, A_3:

$$P(A_1, A_2, A_3) = P(A_1|A_2, A_3)P(A_2, A_3) = P(A_1|A_2, A_3)P(A_2|A_3)P(A_3)$$

The chain rule can be applied in the same way for some arbitrary N random variables.

4.2 Exercises

Question 1: Intuition

Let T be a binary random variable indicating that any given student is tall (i.e. $T = 0$ means the student is not tall and $T = 1$ means the student is tall), and let V be another binary random variable indicating whether the given student likes volleyball. Suppose that if we randomly choose a student, the probability that they like to play volleyball is $P(V = 1) = 0.01$. Also suppose that if we randomly choose a student that likes to play volleyball, the probability that they are tall is 0.3, i.e. $P(T = 1|V = 1) = 0.3$.

Question 1.1 Intuitively, would you expect the probability that a student likes to play volleyball *and* is tall to be lower or higher than 0.01?

Question 1.2 Justify your above answer by applying the chain rule.

Question 2

Consider discrete random variables A, B, C that can take values $a_1, \ldots, a_N$; $b_1, \ldots, b_N$; and $c_1, \ldots, c_N$, respectively. Which of the following sums to 1?

Question 2.1 $\sum_{i=1}^{N} P(A = a_i | B = b_2)$

Question 2.2 $\sum_{i=1}^{N} P(A = a_2 | B = b_i)$

Question 2.3 $\sum_{i=1}^{N} P(A = a_i | B = b_2, C = c_2)$

Question 2.4 $\sum_{k=1}^{N} \sum_{j=1}^{N} \sum_{i=1}^{N} P(A = a_i | B = b_j, C = c_k)$

Question 2.5 $\sum_{j=1}^{N} \sum_{i=1}^{N} P(A = a_i, B = b_j | C = c_2)$

Question 2.4 $\sum_{k=1}^{N} \sum_{j=1}^{N} \sum_{i=1}^{N} P(A = a_i | B = b_j, C = c_k) P(B = b_j, C = c_k)$

Question 3 Bayes' Theorem

One of the most important theorems in probability is *Bayes Theorem.* This theorem states that for two random variables, A and B, that

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Show that this theorem is true. *Hint: use the definition of conditional probability and then the chain rule.*

Question 4: Weather Forecasting

You have decided to become CMU's new weather forecaster. The last person in this position kept probabilities regarding the following random variables:

- C: Binary variable for if is cloudy outside, i.e. $C = 1$ means it is cloudy and $C = 0$ means it is not.
- R: Binary variable for if it is raining.
- W: Binary variable for if it is windy.

Unfortunately, your predecessor was not very organized and is missing some probabilities in their notes. Calculate the below probabilities for your next weather report!

$P(W = 1 \mid R = 1, C = 0)$	0.6
$P(W = 0 \mid R = 1, C = 0)$	0.4

$P(C = 1 \mid R = 1)$	0.9
$P(C = 0 \mid R = 1)$	0.1
$P(C = 1 \mid W = 1)$	0.733
$P(C = 0 \mid W = 1)$	0.267

$P(R = 1)$	0.8
$P(R = 0)$	0.2
$P(W = 1)$	0.292
$P(W = 0)$	0.708

Question 4.1 $P(W = 1, C = 0 \mid R = 1) =$

Question 4.2 $P(W = 1, C = 0, R = 1) =$

Question 4.3 $P(R = 1 \mid C = 0, W = 1) =$