Computer Science 384 | Monday, March 21, 2022
St. George Campus | University of Toronto

Take-Home Quiz #2: Uncertainty
Due: March 28, 2022 by 11:59 PM

Late Policy: You may use grace days to hand in this Take-Home Quiz after the due date, without penalty. (Recall you were given 5 grace days at the beginning of term.) Quizzes may not be handed in late otherwise. You do NOT need to notify us about using your grace days.

What to hand in on paper: Nothing.

What to submit electronically: **Submit written answers in a file called** `answers.pdf` **as well as** `acknowledgment_form.pdf` using MarkUs. Assignments must be completed ***individually***, and you must submit a single document. Handwritten submissions are acceptable as long as they are written ***neatly*** and ***legibly*** (typed submissions are preferred but **not** required).

How to submit: You will submit your assignment using MarkUs. Your login to MarkUs is your teach.cs username and password. It is your responsibility to include all necessary files in your submission. You can submit a new version of any file at any time, following the above noted restrictions on late submissions. For the purposes of determining the number of grace days required for a late submission, the submission time is considered to be the time of your latest submission. More detailed instructions for using MarkUs are available at: `https://markus.teach.cs.toronto.edu/2022-01`.

Clarifications: Important corrections (hopefully few or none) and clarifications to the assignment will be posted on Quercus.

Questions: Questions about the assignment should be posted to Piazza.

1. **(15 marks)** There is a large box containing various bags of jelly beans. Each bag contains 10 jelly beans, where the distribution over flavours within a bag varies according to the "type" of bag. 75% of the jelly bean bags are Type A bags, which contain 5 strawberry-flavoured (your favourite!) and 5 grass-flavoured (the absolute worst!) jelly beans. 25% of the bags are Type B bags, which contain 6 strawberry-flavoured and 4 chocolate-flavoured jelly beans. You grab a bag at random and eat a jelly bean.

   (For each sub-question below, please show your work, then clearly indicate your final answer, which you may round to three decimal places.)

   (a) **[4]** What is the probability that you will eat a strawberry-flavoured jelly bean?

   (b) **[4]** What is the probability that you will eat a chocolate-flavoured jelly bean?

   (c) **[7]** After eating the jelly bean, you conclude that it was strawberry-flavoured (yum!). Given this observation...

      i. ...what is the probability that you grabbed a Type A bag?

      ii. ...what is the probability that the next jelly bean you eat (chosen at random from the same bag, which now contains 9 beans) will be grass flavoured?

Preliminary info for Question 2

Terminology: "posterior" Bayesian statistics is an approach to using probabilities that relies heavily on Bayes' rule. Recall the expression for Bayes' rule:

$$P(Y|X) = \frac{P(X,Y)}{P(X)} \tag{1}$$

$$= \frac{P(Y)P(X|Y)}{\sum_{Y'} P(X,Y')} \tag{2}$$

where (1) is the basic expression, and (2) makes some simple substitutions on the numerator and denominator using the product and sum rules of probability.

In many situations, we would like to simultaneously model observed and unobserved random variables. As an example, X could represent the number of Red-winged Blackbirds that Elliot saw at Queen's Park today (which he can measure), while Y could represent the total number of Red-winged Blackbirds in Toronto (which is too big to measure, but will be statistically related to the Queen's Park measurement). When we use Bayes rule to make an inference about the probability of an unobserved event Y given an observed event X, we often refer to the resulting conditional distribution $P(Y|X)$ as a posterior distribution.

2. **[15 marks]** A box contains 5 coins and each has a different probability of showing heads. Let $p_1, \ldots, p_5$ denote the probability of heads on each coin. Suppose that

$$p_1 = 0,\ p_2 = \frac{1}{4},\ p_3 = \frac{1}{2},\ p_4 = \frac{3}{4} \text{ and } p_5 = 1.$$

Let H denote "heads is obtained" and let C_i denote the event that coin i is selected.

(Please show your work for all of the questions asked below.)

(a) **[4]** If a coin is selected at random and then flipped, write down a joint distribution $P(C_i, H)$ over the resulting outcome. Here, $C_i \in \{0,1\}$ represents which coin was chosen (with $i \in \{1,2,3,4,5\}$ indicating the coin index), and $H \in \{0,1\}$ represents that a head was obtained. You may answer this by either writing down a mathematical expression for this joint distribution (some function of C_i and H that computes the joint probability), or by writing down a probability table, which can be formatted as a two-column table where the first column enumerates different outcomes (for example, $(C_i = 1, H = 0)$ and the second column enumerates their probabilities $(P(C_i = 1, H = 0))$.

(b) **[5]** Now select a coin at random and toss it (in other words, draw a sample from $P(C_i, H)$, where C_i is not directly observed). Suppose a head is obtained. What is the "posterior" probability that coin i was selected for each value of i in $1, \ldots, 5$? In other words, find $P(C_i|H)$ for $i \in \{1, \ldots, 5\}$.

(c) **[3]** Toss the coin again. What is the probability of another head? in other words find $P(H_2|H_1)$ where H_j is "heads on toss j".

(d) **[3]** Now suppose that the experiment is carried out as follows: We select a coin at random and toss it until a head is obtained. Find $P(C_i|B_4)$ where B_4 is "first head is obtained on toss 4".

Preliminary info for Question 3

Factorizations Recall that we can always write a joint distribution as a product of factors. For example the factorization $P(X,Y,Z) = P(X)P(Y|X)P(Z|X,Y)$ will always hold, as it represents an application of the product rule of probability. However, more "compact" factorizations can be found whenever independence (or conditional independence) relationships can be established for a particular distribution. For example if the variables X, Y, and Z are independent, we can write $P(X,Y,Z) = P(X)P(Y)P(Z)$.

This question will ask you to explore how different independence relationships imply different factorizations, and to think about the computational benefits of finding a "compact" representation.

3. **[10 marks]** Imagine you run a book store called Bayes' books. Books in your store can be sorted using the following three properties:

   - Genre $G \in \{History, Romance, Philosophy, Music\}$;
   - Publication date $D \in \{1800s, 1900s, 2000s\}$;
   - Cover type $C \in \{Hardback, Paperback\}$.

   Note that each book has exactly one of each property. This means, for example, that there is no book that belongs to two or more genres, and that every book has one publication date.

   Choosing a book at random from the store, then determining its properties (G,D,C) can be thought of as taking a sample from the joint distribution $P_{Bookstore}(G,D,C)$. Note that while in this case the distribution over books would be uniform, the distribution over (G,D,C) may be non-uniform, since some genres may have more books than others (likewise for the other book properties: publication date and cover type).

   In most instances, however, shoppers do not choose a book uniformly at random, but instead "sample" a book according to their own preferences. For the following sub-questions, we hope to formulate personalized distributions over (G,D,C) for our shoppers. Specifically, we want to store these in the shop log by writing down probability tables over the factors of each shopper's distribution.

   (a) **[2]** Imagine that you want to write a probability table for the distribution $P_{bookstore}(G,D,C)$ described above. This is a table with two columns, where the left column enumerates the possible (G,D,C) (for example, $(History, 1800s, Hardback)$) while the right column stores the associated probabilities ($P(G = History, D = 1800s, C = Hardback)$). How many rows will this table contain?
   Note: we are not actually asking you to write this probability table here (or any any following sub-questions); you only need to report how many rows it would have.

   (b) **[4]** Shopper 1 is a regular at your store who shops in the following peculiar way: First, she always chooses a publication date first (this value might be different during every visit, but she always starts by choosing D). Second, *based on the chosen publication date*, she chooses a genre G independently from the cover type C. With the values of (G,D,C) fully specified, she selects a book that has the chosen properties.

      i. Let $P_{S1}(G,D,C)$ denote the joint distribution for this shopper. Write down a factorization of $P_{S1}(G,D,C)$ that captures the behaviours of Shopper 1.

ii. If we wanted to write down a probability table for *each* of the factors you identified, what would be the total number of rows in all of these tables?

(c) **[4]** Shopper 2 comes in every weekend and shops according to the following pattern: He begins by choosing the genre G and publication date D together according to his specific tastes: whenever choosing a book from the 1800s, he chooses *Music*; whenever choosing a book from the 2000s, he chooses Romance; and for his 1900s books he can choose either $History$ or $Philosophy$. Since he doesn't care about cover type, he chooses this separately from the other factors.

i. Let $P_{S2}(G,D,C)$ denote the joint distribution for this shopper. Write down a factorization of $P_{S2}(G,D,C)$ that captures the behaviours of Shopper 2.

ii. If we wanted to write down a probability table for *each* of the factors you identified, what would be the total number of rows in all of these tables?

HAVE FUN and GOOD LUCK!