MATH 5310: Probability
Fall 2024
Homework 2
Due before 11:59pm on Wednesday 9/25

Part1:

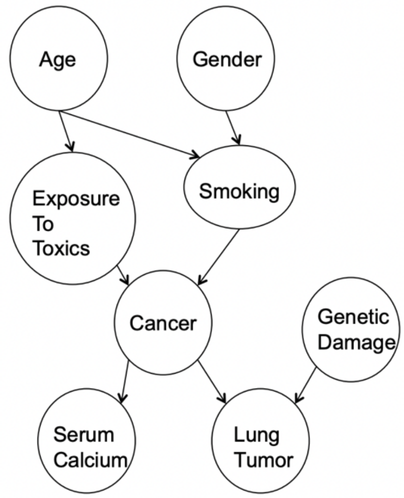
1. Give only a short answer: Which of the CDF, PDF, and/or the PMF is used to show random variables have the same probability distribution? Cite a theorem from the text to support your answer.
2. An urn contains 5 red balls and 2 green balls. A single ball is randomly drawn from the urn. If the ball is green, then a red ball is added to the urn. If the ball is red, then a green ball is added to the urn. The original ball is not returned to the urn no matter if it is green or red. Now, a second ball is randomly drawn from the urn.
   (i) In probability notation, write the P(second ball is red) as a function of the possible outcomes of the first draw (that is, the two events "first ball is red" and "first ball is green"). Hint: think of the first draw as a partition and apply the law of total probability.
   (ii) Calculate the probability that the second ball is red.

3. Independence and dependence (conditional probabilities) for two events are critical concepts. However, real data is usually much more complicated and involves dependencies and independencies between a network of more than two events. A key concept to describe the complex relationships between multiple events is <u>conditional independence.</u>

   Conditional independence occurs when two (or more) events can themselves be dependent *but* they become independent when each is conditioned on a third (or more) event. In more practical language: knowing the value of a third event breaks the dependency between the other two events. Here are two mathematical definitions:

   Two event A and B are conditionally independent given an event C with P(C)>0 if $P(A \cap B|C) = P(A|C) * P(B|C)$ *or equivalently* $P(A|B,C) = P(A|C)$

   The figure below illustrates the concept in a real data example. The circles represent variables (their measured value being the "event") and the arrows represent one variable having an effect on another. For example, the arrow starting at Gender and pointing to Smoking indicates that an individual's gender has an effect on whether they are a smoker or not. In the figure, Cancer is *conditionally independent* of both Age and Gender given (conditioned on) the Exposure To Toxics and Smoking variables. This is because Age and Gender affect Cancer only through the Exposure To Toxics and Smoking variables. If Exposure To Toxics and Smoking are known, there is no more information in Age and Gender that impacts Cancer. If there had been another arrow from Age or Gender directly to Cancer, then I couldn't make this claim.



Source: https://cedar.buffalo.edu/~srihari/CSE674/Chap3/3.6-ConditionalIndependence.pdf

Here is a question for you to practice this concept:

A box contains two coins: (1) a *regular fair coin* that when flipped has P(Head)=P(Tail)=0.5 and (2) a f*ake two-headed coin* where both sides are a head (there is no tail), meaning P(Head)=1

Now, a single coin is chosen at random and then flipped twice. Define the following events:

A = First coin toss results in a Head
B = Second coin toss results in a Head
C = The regular (not fake) coin was the one that was randomly chosen coin

(a) Explain in practical language, why A and B are not independent, but are conditionally independent given C.

(b) Calculate $P(A|C)$, $P(B|C)$, $P(A \cap B|C)$, $P(A)$, $P(B)$, and $P(A \cap B)$. Hint: calculate these probabilities in the order they are listed here.

Part 2: Casella and Berger Problems:
1.39
1.41 – use a Bayes approach for your answer to (a)
1.47 (a) and (d) only
1.52
1.53
1.54 (b) only

You do not have to do 1.49, but I have seen this concept becoming more prevalent in methods that integrate statistics and computation mathematics (e.g. uncertainty quantification). It is a reasonably short problem that would be a quick introduction to the area if research using a stats and comp math hybrid is of interest to you.