XX50215 Statistics for Data Science

Problems 2 - Solutions

1. Give an example of a discrete random variable and a contiguous random variable.

   Discrete have a fixed set of outcomes. E.g. roll of a die.

   Contiguous have an infinitely dividable range of outcome. E.g. height of a tree.

2. If B \ A is the set of elements in B but not in A, known as the difference, verify the following identity:

   $$A \setminus B = A \setminus (A \cap B) = A \cap B^c$$

   We need to show containment in both directions.

   $x \in A\setminus B \Leftrightarrow x \in A$ and $x \notin B \Leftrightarrow x \in A$ and $x \notin A \cap B \Leftrightarrow x \in A \setminus (A \cap B)$

   Also, $x \in A$ and $x \notin B \Leftrightarrow a \in A$ and $x \in B^c \Leftrightarrow x \in A \cap B^c$

   Note: A ⇔ *B* means *A* is true if *B* is true and *A* is false if *B* is false.

3. The Smiths have two children. At least one of them is a boy. What is the probability that both children are boys?

   Enumerating the sample space gives S = {(B,B), (B,G),(G,B),(G,G)} with each outcome equally likely.

   P(at least one boy) = 3/4 and P(both are boys) = 1/4. Then using conditional probabilities

   $P(A|B) = P(A\cap B)/P(B)$ Equivalent to restricting the sample space to B.

   P(both are boys | at least one boy) = 1 /3

   An ambiguity may arise if we don't acknowledge order.

   The space then becomes S = {(B,B, (B,G), (G,G)}

4. In the game of dominoes, each piece is marked with two numbers. The pieces are symmetrical so that the number pair is not ordered (so, for example, (2, 6) = (6, 2)). How many different pieces can be formed using the numbers 1, 2, …, n?

   There are $\binom{n}{2}$ = n(n – 1)/2 pieces on which the two numbers do not match. (Choose 2 out of n numbers without replacement.) There are n pieces on which the two numbers match. So the total number of different pieces is n + n(n − 1)/2 = n(n + 1)/2.

5. Suppose that 5% of men and 25% of women are colour-blind. A person is chosen at random and that person is colour-blind. What is the probability that the person is male? You can assume there are equal number of males and females.

Using Bayes rule

M Male
F Female
CB Colour blind

Equal numbers of male + female means P(M) and P(F) are 1/2.

P(A|B) = P(B|A)P(A) / P(B)

P(M|CB) = $\frac{P(CB|M)P(M)}{P(CB|M)P(M)+P(CB|F)P(F)} = \frac{0.05 *1/2}{0.05 *1/2+ 0.25 * 1/2}$ = 0.166667

I had really intended P(CB|F) to be 0.25%, but it's the process not the numbers that matter.

6. Prove the following functions are cdfs.

   a. $\frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x)$, $x \in (-\infty,\infty)$

   b. $e^{-e^{-x}}$, $x \in (-\infty,\infty)$

Both functions are continuous, hence right-continuous. So we only need to check the limit, and that they are non-decreasing.

$\lim_{x\to-\infty} \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x) = \frac{1}{2} + \frac{1}{\pi}\left(\frac{-\pi}{2}\right) = 0$, $\lim_{x\to\infty} \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x) = \frac{1}{2} + \frac{1}{\pi}\left(\frac{\pi}{2}\right) = 1$, and $\frac{d}{dx}\left(\frac{1}{2}+\frac{1}{\pi}\tan^{-1}(x)\right) = \frac{1}{1+x^2} > 0$, so $F(x)$ is increasing.

$\lim_{x\to-\infty} e^{-e^{-x}} = 0$, $\lim_{x\to\infty} e^{-e^{-x}} = 1$, $\frac{d}{dx}e^{-e^{-x}} = e^{-x}e^{-e^{-x}} > 0$.

7. A particular powerstations generating load peaks each day. Suppose that the low load is set at 1 and the peak load Y has distribution function

$$F_Y(y) = P(Y <= y) = 1 - \frac{1}{y^2}, \ 1 <= y <= \infty$$

   a. Verify that $F_Y(y)$ is a cdf.
   b. Find $f_Y(y)$, the pdf of Y.
   c. If the low load is reset to 0 and we use a unit of measurement that is 1/10th of that given previously, the peak load becomes Z = 10(Y-1). Find $F_Z(z)$.

a. $\lim_{y\to-\infty} F_Y(y) = \lim_{y\to-\infty} 0 = 0$ and $\lim_{y\to\infty} F_Y(y) = \lim_{y\to\infty} 1 - \frac{1}{y^2} = 1$. For $y \leq 1$, $F_Y(y) = 0$ is constant. For $y > 1$, $\frac{d}{dy}F_Y(y) = 2/y^3 > 0$, so F_Y is increasing. Thus for all y, F_Y is nondecreasing. Therefore F_Y is a cdf.

b. The pdf is $f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} 2/y^3 & \text{if } y > 1 \\ 0 & \text{if } y \leq 1. \end{cases}$

c. $F_Z(z) = P(Z \leq z) = P(10(Y-1) \leq z) = P(Y \leq (z/10)+1) = F_Y((z/10)+1)$. Thus,

$$F_Z(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ 1 - \left(\frac{1}{[(z/10)+1]^2}\right) & \text{if } z > 0. \end{cases}$$

8. The Monty Hall Problem

“Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?”

This is a rather famous problem that has generated much debate. If you’ve not see it before think about your own answer before you Google the solution.

It may seem counter intuitive, but the answer is to switch. Your odds will go from 1/3 to 2/3. Think of it this way: Treat doors no2and no3 as a single option. Initially you choose door no1. It’s got a 1/3 chance. The combined no2 and no3 has a 2/3 chance, but if you select one or the other the chance drops to 1/3 for each. Once the host rules out no2 or no3, there’s still a 2/3 chance of it being one of those and you can now select the one the host did n’t open and maintain the 2/3 chance. If you need further convincing, imagine there was a million doors and the host will open all but one of the doors you did n’t check…