DATA 1010
In-class exercises
Samuel S. Watson
26 September 2018

Problem 1

Devise a principle of inclusion-exclusion for *three* sets. In other words, write $\mathbb{P}(A \cup B \cup C)$ in terms of probabilities of A, B, C, and their intersections.

Solution

In computing the sum $\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$, the probability masses in exactly one of the three sets contribute once to the total, while the probability masses in exactly two of the sets contribute twice, and any probability mass in the intersection contributes three times. Meanwhile, each probability mass in $A \cup B \cup C$ contributes only once to $\mathbb{P}(A \cup B \cup C)$.

To correct for the double- and triple-counting, we subtract off the probabilities of the pairwise intersections:

$$\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C).$$

In this sum, the probability masses in either one or two of the sets contribute once to the total. However, the probability masses in all three sets contribute 3 times and -3 times for a net contribution of 0. Therefore, we should add those back in:

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Problem 2

Find the probability of getting two pairs (like 3, 4, 5, 4, 3) with a roll of five dice. Express your answer as an unreduced fraction.

Solution

There are 6^5 rolls of five dice, and they are all equally likely. Therefore, our goal is to find the number of rolls which satisfy the two-pair criterion.

There are many ways of thinking about how to build a two-pair tuple. Here's one: (1) we choose the three numbers which will appear in the tuple, (2) we choose which of these numbers appears only once, (3) we choose the slot for the number which appears only once, and (4) we choose a way of filling in the two pairs in the remaining four slots. So the number of ways of making all of these decisions is

$$\binom{6}{3}\binom{3}{1}\binom{5}{1}\binom{4}{2} = 1800.$$

So the probability is $1800/6^5$. We can check our answer as follows.

```julia
"""
Returns an array which contains all k-tuples of
elements of 1:n.
"""
function tuples(n,k)
    if k == 1
        [[i] for i in 1:n]
    else
        [[i;t] for i in 1:n for t in tuples(n,k-1)]
    end
end

using StatsBase
"""
Determines whether the tuple t is a two-pair tuple
"""
function twopair(t)
    sort(collect(values(countmap(A)))) == [1,2,2]
end
twopair.(tuples(6,5)) # returns 1800
```

Problem 3

A problem on a test requires students to match molecule diagrams to their appropriate labels. Suppose there are three labels and three diagrams and that student guesses a matching uniformly at random. Let X denote the number of diagrams the student correctly labels. What is the probability mass function of the distribution of X?

Solution

The number of correctly labeled diagrams is an integer between 0 and 3 inclusive. Suppose the labels are A, B, C, and suppose the correct labeling sequence is ABC (the final result would be the same regardless of the correct labeling sequence). The sample space consists of all six possible labeling sequences, and each of them is equally likely since the student applies the labels uniformly at random. So we have

$$\begin{aligned}\Omega &= \{\mathrm{ABC}, \mathrm{ACB}, \mathrm{BAC}, \mathrm{BCA}, \mathrm{CAB}, \mathrm{CBA}\}, \\ \{X = 0\} &= \{\mathrm{BCA}, \mathrm{CAB}\}, \\ \{X = 1\} &= \{\mathrm{ACB}, \mathrm{CBA}, \mathrm{BAC}\}, \\ \{X = 2\} &= \{\}, \text{ and} \\ \{X = 3\} &= \{\mathrm{ABC}\}.\end{aligned}$$

The probability mass function of the distribution of X is therefore

$$m_X(0) = \frac{1}{3}$$

$$m_X(1) = \frac{1}{2}$$

$$m_X(2) = 0$$

$$m_X(3) = \frac{1}{6}$$

All together, we have

$$m_X(x) = \begin{cases} \frac{1}{3}, & \text{if } x = 0 \\ \frac{1}{2}, & \text{if } x = 1 \\ \frac{1}{6}, & \text{if } x = 3 \\ 0, & \text{otherwise.} \end{cases}$$

Problem 4

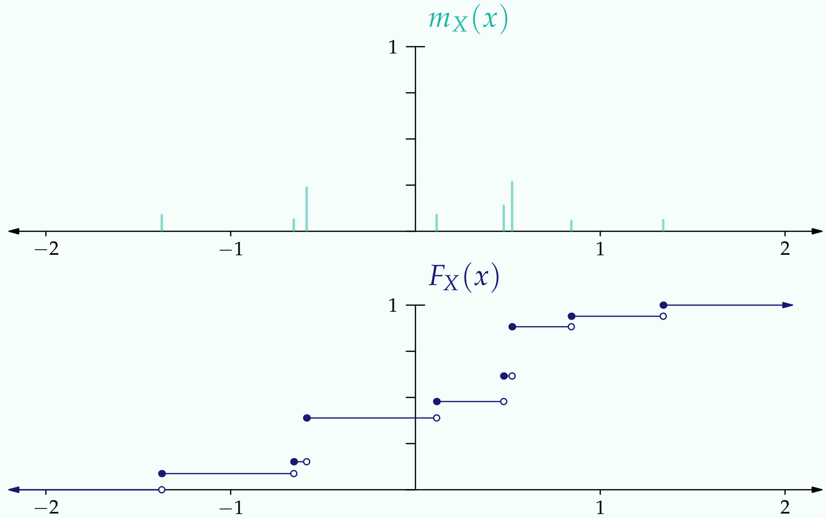
Consider a random variable X whose distribution is the one shown in the figure below. Identify each of the following statements as true or false.

(a) $\mathbb{P}(-1 < X < 1)$ is greater than $\frac{3}{5}$

(b) $\mathbb{P}(X \geq 2) = 0$

(c) $\mathbb{P}\left(-\frac{1}{2} < X < 0\right)$ is greater than $\frac{1}{100}$

(d) $\mathbb{P}(100X < 1)$ is greater than $\frac{1}{2}$



Solution

(a) is true, since the CDF goes from about 0.1 at -1 to about 0.9 at $+1$. The difference, about 0.8 is larger than 0.6.

(b) is also true, since there is no probability mass past 2.

(c) is false: there is no probability mass in the interval from $-\frac{1}{2}$ to 0.

(d) $\mathbb{P}(100X < 1)$ is equivalent to the probability that X is less than $\frac{1}{100}$, which (reading the graph of the CDF) we see is between 0.25 and 0.5. Therefore, (d) is false.