In-Term #1: Solutions

Milica Cudina

2022-09-28

Problem 1. (5 points)

Write down the definition of the *cumulative distribution function* of a random variable.

Solution: See your class notes.

Problem 2. (5 points)

Write down the definition of the *independence* of two events.

Solution: See your class notes.

Problem 3. (15 points)

Consider the following two-phase experiment defining the outcome of a random variable X:

First a fair coin is tossed. If the outcome of the coin toss is heads, then a fair six-sided die is rolled. The sides of the die have numbers 1, 2, ..., 6 on them. Whichever number comes up is the value of the random variable X. If the outcome of the coin toss is tails, then a fair tetrahedron die is rolled. The sides of the tetrahedron have numbers 1, 2, 3, 4 on them. Whichever number comes up is the value of the random variable X.

Write down the *probability mass function* of the random variable X.

Solution: The **support** of the random variable X is $\{1, 2, 3, 4, 5, 6\}$. It *probability mass function* is

$$\begin{aligned}
p_X(1) &= \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{4}\right) = \frac{5}{24}, \\
p_X(2) &= \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{4}\right) = \frac{5}{24}, \\
p_X(3) &= \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{4}\right) = \frac{5}{24}, \\
p_X(4) &= \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{4}\right) = \frac{5}{24}, \\
p_X(5) &= \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) = \frac{1}{12}, \\
p_X(6) &= \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) = \frac{1}{12}.
\end{aligned}$$

Problem #4. (5 points)

In a class there are four first-year Thunderbirds, six first-year Wampuses, and six second-year Thunderbirds. How many second-year Wampuses must be present if house (Thunderbirds and Wampuses) and year (first and second) are to be independent when a student is selected at random? There are no other students in this class!

Solution: Let the total number of students be n and let the number of second-year Wampuses be k. Then, $n = 16 + k$. One condition for independence reads as (in obvious notation)

$$\mathbb{P}[T \cap I] = \mathbb{P}[T] \times \mathbb{P}[I] \quad \Leftrightarrow \quad \frac{4}{n} = \frac{10}{n} \times \frac{10}{n} \quad \Leftrightarrow \quad n = 25.$$

We conclude that $k = 25 - 16 = 9$. One can easily check that all the other independence conditions hold as well:

$$\begin{aligned}
\mathbb{P}[T \cap II] = \mathbb{P}[T] \times \mathbb{P}[II] \quad &\Leftrightarrow \quad \frac{6}{25} = \frac{10}{25} \times \frac{15}{25} \quad \Leftrightarrow \quad TRUE, \\
\mathbb{P}[W \cap I] = \mathbb{P}[W] \times \mathbb{P}[I] \quad &\Leftrightarrow \quad \frac{6}{25} = \frac{15}{25} \times \frac{10}{25} \quad \Leftrightarrow \quad TRUE, \\
\mathbb{P}[W \cap II] = \mathbb{P}[W] \times \mathbb{P}[II] \quad &\Leftrightarrow \quad \frac{9}{25} = \frac{15}{25} \times \frac{15}{25} \quad \Leftrightarrow \quad TRUE.
\end{aligned}$$

Problem #5. (5 points)

Let $\Omega = \{a_1, a_2, a_3, a_4\}$ be a sample space, and let $\mathbb{P}$ be a probability on Ω. Assume that $\mathbb{P}[\{a_2, a_3\}] = 2/3, \mathbb{P}[\{a_2, a_4\}] = 1/2$ and $\mathbb{P}[\{a_2\}] = 1/3$. Then we have that $\mathbb{P}[\{a_1\}]$ equals the following value:

a. 1/12
b. 1/6
c. 1/3
d. 1/2
e. None of the above

Solution: From the given values of $\mathbb{P}$ on certain events, we conclude that

$$\mathbb{P}[\{a_3\}] = \mathbb{P}[\{a_2, a_3\}] - \mathbb{P}[\{a_2\}] = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}\,,$$
$$\mathbb{P}[\{a_4\}] = \mathbb{P}[\{a_2, a_4\}] - \mathbb{P}[\{a_2\}] = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}\,.$$

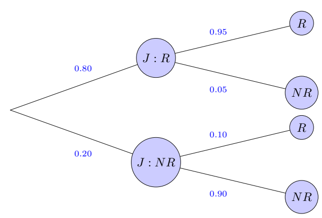
So,

$$\mathbb{P}[\{a_1\}] = 1 - (\mathbb{P}[\{a_2\}] + \mathbb{P}[\{a_3\}] + \mathbb{P}[\{a_4\}]) = \frac{1}{6}\,.$$

Problem #6. (10 points)

Most mornings, Bertie Wooster asks Jeeves whether it is going to rain that day. It being England, Jeeves forecasts rain 80% of the time and dry weather the remaining 20% of the time. If Jeeves forecasts rain, the chance if it actually raining is 95%. If Jeeves forecasts no rain, the chance of it not raining is 90%. Suppose that one day Bertie forgot to ask Jeeves if it would rain. It did not rain. What is the probability that Jeeves would have predicted no rain?

Solution: This probability tree describes the situation in the problem:



We use the Bayes theorem here.

$$\mathbb{P}[\text{Jeeves would have said no rain} \,|\, NR] = \frac{\mathbb{P}[NR \,|\, \text{Jeeves would have said no rain}]\mathbb{P}[\text{Jeeves would have said no rain}]}{\mathbb{P}[NR]}.$$

Using our tree, we get

$$\mathbb{P}[NR] = 0.8(0.05) + 0.2(0.90) = 0.22.$$

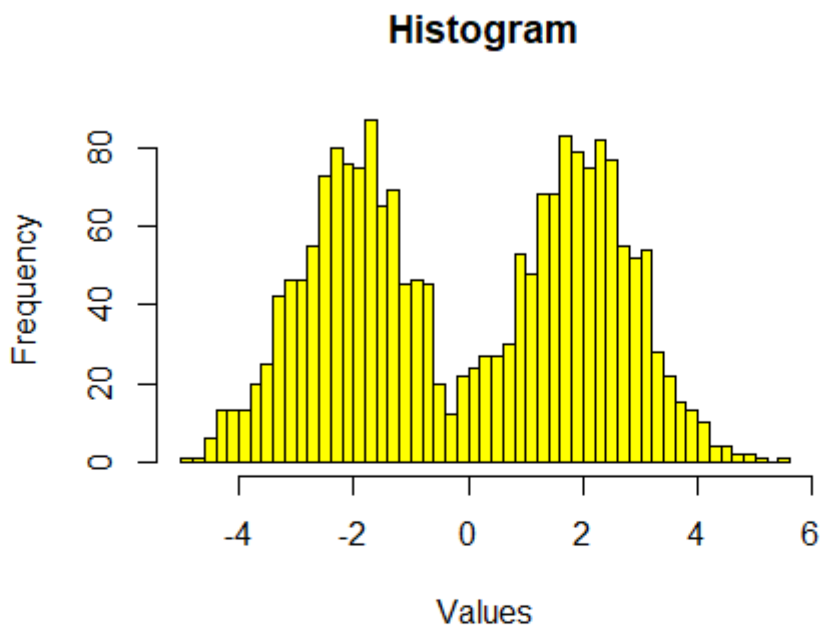
So,

$$\mathbb{P}[\text{Jeeves would have said no rain} \,|\, NR] = \frac{0.2(0.9)}{0.22} = \frac{9}{11} = 0.82..$$

Problem #7. (5 points)

Consider the following histogram:

```
knitr::include_graphics("hist-yellow.png")
```



The histogram is ...

a. unimodal.
b. bimodal, skewed.
c. bimodal, symmetric.
d. trimodal.
e. uniform.

Solution: The correct solution is **c.**

Problem 8. (5 points)

Source: AMC8, 2013.

Hammie is in the 6^{th} grade and weighs 106 pounds. His quadruplet sisters are tiny babies and weigh $5, 5, 6$, and 8 pounds. Which is greater, the average (mean) weight of these five children or the median weight, and by how many pounds?

a. The median, by 60 pounds.
b. The median, by 20 pounds.
c. The mean, by 5 pounds.
d. The mean, by 15 pounds.
e. The mean, by 20 pounds.

Solution: The correct solution is **e.**

Lining up the weights in ascending order $(5, 5, 6, 8, 106)$, we see that the median weight is 6 pounds. The mean weight is

$$\frac{5+5+6+8+106}{5} = 26.$$

Problem 9. (5 points)

Below are some summary statistics from the `score` variable indicating employee satisfaction.

```
min Q1 median   Q3 max   mean       sd   n missing
 30 57   69.5   77  99 65.075 16.09361 200       0
```

Which of the following is **true**?

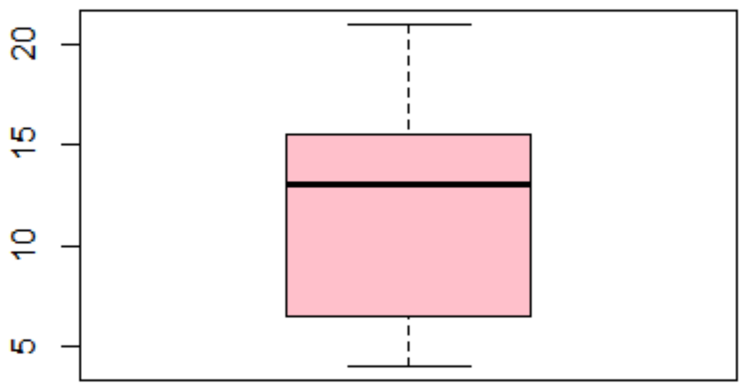
a. The standard deviation estimate is not possible because `score` is a whole number.
b. There is evidence that the distribution of `score` is right-skewed.
c. The minimum value of 30 would be identified as an outlier in a box plot.
d. There were more survey respondents who reported job satisfaction scores less than 57 than survey respondents who reported job satisfaction scores greater than 77.
e. None of the above are true.

Solution: The correct solution is **e.**

Problem 10. (5 points)

Consider the following box plot:

```r
knitr::include_graphics("box-pink.png")
```



What do you suspect to be true about the data set (circle **all** that apply)?

a. The distribution is symmetric.
b. The maximum observation is 22.
c. The minimum observation is 3
d. The distribution is skewed.
e. None of the above.

Solution: The correct solution is **b., c., and d.**

Look at p. 49 from the book to see how the box plot is constructed and how to interpret it.