**Review Questions: SOLUTIONS**

1. Lesson 7 Slides 65 and 66
2. Lesson 7 Slide 71
3. Lesson 6 Slide 56
4. Lesson 6 Slide 40
5. Lesson 7 Slide 61
6. a). What theoretical probability is this code estimating?

   **SOLUTION:**

   $\boxed{P(\text{pick purple ball})}$

   b). Calculate the exact probability for the quantity you listed in part (a). i.e. what number should this code output approach as you increase NumSamples?

   **SOLUTION**:

   b). Calculate the exact probability for the quantity you listed in part (a). i.e. what number should this code output approach as you increase NumSamples? $\boxed{\frac{5}{21} + \frac{5}{36}}$

   let A: Choose a purple ball

   $$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$$
   $$= (5/14)(2/3) + (5/12)(1/3) = \frac{5}{21} + \frac{5}{36}$$

---

7. .

   7ai). Histogram, since this is asking for a distribution of a quantitative variable.

   7aii). Scatter plot, since we want to see how 2 quantitative variables are associated. We wouldn't choose a line plot here because the data points won't necessarily pass the vertical line test ( it's possible for either column to contain repeat values, so wecan't think of one as a function of the other).

   7aiii). This is asking how a qualitative variable (outcome) varies with the average of a quantitative variable. We could show this with either a Bar Chart or a Line Plot.

8. (a) You choose a coin at random from the box. What is the probability that it is not a fair coin? (Give your answer as a fraction). **SOLUTION**: $\frac{5}{6}$

   (b) You choose a coin at random and flip it. What is the probability that the coin comes up heads? Give your answer as a single fraction, fully simplified.

   $$P(H) = P(H|F)P(F) + P(H|BH)P(BH) + P(H|BT)P(BT)$$
   $$= \left(\tfrac{1}{2}\right)\cdot\left(\tfrac{1}{6}\right) + \left(\tfrac{3}{4}\right)\cdot\left(\tfrac{1}{3}\right) + \left(\tfrac{1}{4}\right)\cdot\left(\tfrac{1}{2}\right)$$
   $$= \frac{11}{24}$$

   (c) Suppose you pick a coin at random and flip it. Are the events "flip comes up heads" and "you picked a fair coin" independent? Justify your answer using the **mathematical definition** of independence.

H and F are dependent.

Justification: option 1

$\left(P(H|F) = \frac{1}{2}\right) \neq \left(P(H) = \frac{11}{24}\right)$

OR: option 2

$P(F|H) = \frac{P(F \cap H)}{P(H)} = \frac{P(H|F) \cdot P(F)}{11/24} = \frac{(\frac{1}{2})(\frac{1}{6})}{\frac{11}{24}} = \frac{2}{11}$

$P(F) = \frac{1}{6}$ not equal

OR option 3

$P(H \cap F) = P(H|F)P(F) = (1/2)(1/6)$

not equal

and

$P(H) \cdot P(F) = \left(\frac{11}{24}\right)\left(\frac{1}{6}\right)$

(d) You choose a coin at random and flip it. It comes up heads. Given this information what is the probability that the coin you chose was one of the BH coins?

$P(BH|H) = \frac{P(BH,H)}{P(H)}$

$= \frac{P(H|BH)P(BH)}{P(H)}$ (using Bayes' Theorem)

We already calculated $P(H) = \frac{11}{24}$ using the Law of Total Probability in part (b).

Thus

$P(BH|H) = \frac{(\frac{3}{4})(\frac{2}{6})}{\frac{11}{24}} = \boxed{\frac{6}{11}}$

(e) You choose a coin at random and flip it **three times in a row**. It comes up tails all 3 times. Given this information, what is the probability that the coin that you chose was a fair coin? (You may leave your answer unsimplified).

**SOLUTION:**

Notice: $P(3T|F)$ is the probability that given the coin is Fair, all 3 flips are tails. Given the coin is fair, the probability of tails is $(1/2)$.

Thus the probability of getting 3 Tails from a fair coin is $P(3T|F) = (1/2)(1/2)(1/2)$ since the results of repeatedly flipping the same coin are independent.

$$P(F|3T) = \frac{P(3T|F)P(F)}{P(3T)} = \frac{(\frac{1}{2})^3(\frac{1}{6})}{P(3T|F)P(F) + P(3T|BH)P(BH) + P(3T|BT)P(BT)}$$

$$= \frac{(\frac{1}{2})^3(\frac{1}{6})}{(\frac{1}{2})^3(\frac{1}{6}) + (\frac{1}{4})^3(\frac{1}{3}) + (\frac{3}{4})^3(\frac{1}{2})}$$

$$= \frac{\frac{1}{48}}{\frac{1}{48} + \frac{1}{192} + \frac{27}{128}}$$

**=8/91**

9. Lesson 11 Slide 41

10. Lesson 7 Slide 53

11. Lesson 8 Slide 46

12. Suppose the scores on a college entrance exam have a mean 75 and and a variance of 81.

(a) At least 75% of the scores lie between what two values?

**SOLUTION**:

By Chebyshev's Bounds: (Lesson 12 Slide 31) at least 75% of the scores lie within 2 standard deviations of the mean. That is $75 \pm 2(9)$.

Thus $57 \leq s \leq 93$

(b) Suppose you are told the scores have a normal distribution. 68% of the scores lie between what two values?
**SOLUTION**
For normal distributions, 68% of the data lie within one standard deviation of the mean: Lesson 15 Slide 49).
That is $75 \pm 9$.
Thus $66 \leq s \leq 84$

13. You roll two fair six-sided dice, one red and one blue. Define two random variables:

$X$: the outcome of the red die.

$Y$ : 1 if the sum of the two dice is greater than 7, and 0 otherwise.

Are the random variables

X and Y independent or dependent? Justify your answer using the definition of independent random variables.

**SOLUTION**

To prove two discrete random variables are independent you must show $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all $x, y$.

We will show they are dependent by finding one set of events $X = x$ and $Y = y$ that are dependent.

Consider the events $X = 1$ and $Y = 1$.

$X = 1 : \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\}$

Thus $P(X = 1) = \frac{6}{36} = \frac{1}{6}$

$Y = 1 : \{(2, 6), (3, 5), (3, 6), (4, 4), (4, 5), (4, 6), (5, 3), (5, 4), (5, 5), (5, 6), (6, 2)(6, 3), (6, 4), (6, 5), (6, 6)\}$

Thus $P(Y = 1) = \frac{15}{36} = \frac{5}{12}$

Notice the set $X = 1, Y = 1 := \{\}$ It is empty!

Thus $P(X = 1, Y = 1) = \frac{0}{36} = 0$

Thus $P(X = 1, Y = 1) \neq P(X = 1)P(Y = 1)$.

Since we have found one set of events that are dependent, we can conclude that the random variables are not independent (they are dependent).

14. Lesson 9 Slide 26

15. Lesson 7 Slide 69

16. Lesson 14 Slide 13

17. (a) A basketball player makes 70% of her free throws. The results of each free throw are independent. You want to model the number of successful free throws she makes out of 10 attempts.

    **SOLUTION**

    i). X = number of successful free throws she makes out of 10 attempts.

    ii). $X \sim Bin(10, 0.70)$

    iii). $P(X = k) = \binom{10}{k}(0.70)^k(.30)^{10-k}$ for $k$ in $\{0, 1, 2, 3, ...10\}$

    (b) A factory machine produces bolts, and each bolt can either pass or fail a quality inspection. You are interested in modeling whether the next bolt produced will be defective. From past data, 5% of the bolts produced by the machine are defective. Whether or not bolts pass the inspection are independent of the results of previously produced bolts.

    **SOLUTION**

    i). X = whether or not the next bolt is defective (1 for yes, 0 for no)

    ii). $X \sim Ber(0.05)$

    iii). $P(X = k) = (0.05)^k(.95)^{1-k}$ for $k$ in $\{0, 1\}$

    (c) Phone calls at a call center are received independently and at a constant average rate of 10 calls per hour. You want to model the time (in minutes) until the next phone call.

    **SOLUTION**

    i). X = time (in minutes) until the next phone call

    ii). $X \sim Exp(1/6)$

    iii). $f(x) = \frac{1}{6}e^{-\frac{x}{6}}$ for $x \geq 0$

    (d) A person is equally likely to arrive at a bus stop at any time between 2:00 PM and 3:00 PM. You want to model the time of arrival within this interval.

    **SOLUTION**

    i). X = time (in minutes starting at 2pm) until the bus arrives

    ii). $X \sim Uni(0, 60)$

    iii). $f(x) = \frac{1}{60}$ for $0 \leq x \leq 60$

    (e) The salaries of people at a large company have a mean of \$70000 and standard deviation of \$15000. You want to model the salary of a randomly chosen employee from this company.

    **SOLUTION**

    Need more information. We can't assume anything about the distribution of a random variable given only the mean and standard deviation.

    (f) Phone calls at a call center are received independently and at a constant average rate of 10 calls per hour. You want to model the number of calls in the next 30 minutes.

    **SOLUTION**

    i). X = number of calls in the next 30 minutes

    ii). $X \sim Poi(5)$

    iii). $P(X = k) = \frac{e^{-5}5^k}{k!}$, for $k$ in $\{0, 1, 2, 3, ....\}$

---

18. Lesson 7 Slide 61

---

19. Lesson 8 Slides 53-54

---

20. An emergency room at a particular hospital gets an average of five patients per hour. A doctor wants to know the probability that the ER gets more than five patients in the next hour.

    **SOLN**:

    Let $X$ be the number of patients per hour. If we assume the arrival of patients is independent of previous patients and that the average rate is constant over time we can model this with a Poisson distribution.

    $\lambda = \frac{\textbf{number of patients}}{perhour} = \frac{5}{1}$

    $P(X > 5) = 1 - P(X \leq 5) = 1 - \sum_{k=0}^{5} \frac{e^{-5}5^k}{k!} \approx 0.384039$

---

21. Lesson 8 Slide 65

    And/or see video explanation here: Link to Video Explanation of Solution

---

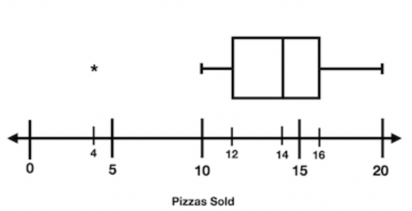
22. **SOLUTION**

$sum(x > np.average(x))/len(x) < 0.5$

Explanation:

The question is asking for the proportion of the data that is greater than the mean in this distribution (i.e. the number of data points greater than the mean divided by the total number of data points) , and whether that proportion is more than half, less than half or exactly equal to half.

Exactly half of the data is always less than the median, and in this case, the median¡mean (due to the **right skew of the distribution**). Since the mean is greater than the median, we know that the proportion of data points that are greater than the mean will have to be less than 1/2.




23.

Determine which of the following questions you have enough information to answer. If you don't have enough information, state "need more info."

(a) What is the mean?
**SOLUTION**: Need more information (can't determine a mean from a boxplot)

(b) What is the median?
**SOLUTION**: 14

(c) What is the mode?
**SOLUTION**: Need more information (can't determine a mean from a boxplot)

(d) What is the IQR?
**SOLUTION**: IQR = 3rd quartile - 1st quartile $= 16 - 12 = 4$

(e) Is this distribution right skewed?
**SOLUTION**: No, it is left-skewed, because the median is closer to the 3rd quartile.

(f) What is the smallest data point within 1.5*IQR of the 25th percentile?
**SOLUTION**: 10. This occurs at the end of the left whisker.

24. According to Baydin, an email management company, an email user gets, on average, 147 emails per day.

a). What is the probability that an email user receives more than 160 emails per day?

**SOLUTION:**. Let $X$ be the number of emails in a day. If we assume emails arrive independently and at a constant average rate, $X \sim Poi(147)$.

Thus $P(X > 160) = 1 - P(X <= 160) = 1 - \sum_{k=0}^{160} \frac{e^{-147}147^k}{k!}$

In Python this is:

$1 - \mathit{stats.poisson.cdf}(160, 147) \approx 0.13337$

b). What is the standard deviation of the number of emails a user gets per day?

If $X \sim Poi(\lambda)$ we showed in the lectures slides that $Var(X) = \lambda$. Thus $SD(X) = \sqrt{\lambda}$

Thus in our example above, $SD(X) = \sqrt{147} \approx 12.1244$ emails.

---

25. Lesson 12 Slide 65

---

26. Suppose you have a group of 100 people and 5 people in the group are ambidextrous (can write with both hands).

(a) P(Not ambidextrous) = 0.95

(b) Suppose you randomly sample 20 individuals from this group **with replacement**. What is the probability that 2 out of the 20 people in your sample are ambidextrous?
**SOLUTION:**
OPTION 1:
Notice, when sampling with replacement we can model this with a Binomial distribution.

Let X = number of people in your sample who are ambidextrous. Thus $X \sim Bin(20, .05)$.

Thus $P(X = 2) = \binom{20}{2}(0.05)^2(.95)^{18} = 190(0.05)^2(0.95)^{18}$

OPTION 2:

Another way of seeing this. Let $A_i$ be the event that the $i$th person sampled was ambidextrous and $B_j$ be the event that the $j$th person was not.

Consider the specific scenario where the first 2 people you selected were ambidextrous and the rest weren't.

The joint probability of this is given by:

$P(A_1, A_2, B_3, B_4, B_5, ....B_{20}) = P(A_1)P(A_2|A_1)P(B_3|A_1, A_2)P(B_4|A_1, A_2, B_3).....P(B_{20}|A_1, A_2, B_1, B_2, ...B_{19})$

(Using the multiplication rule)

But $A_i = 0.05$ for all $i$ and $B_j = 0.95$ for all $j$ and all events $A_i$ and $B_j$ are independent (since we're replacing people after we sample).

Thus this becomes $= (P(A_i))^2(P(B_j))^{18} = (0.05)^2(0.95)^{18}$

But this was just the ordering that assumed the first 2 people we selected were ambidextrous. Since there are $\binom{20}{2}$ different ways we can select 2 ambidextrous people in a group of 20, and each of them have the same probability $(0.05)^2(0.95)^{18}$ the overall probability is

$\binom{20}{2}(0.05)^2(0.95)^{18} = 190(0.05)^2(0.95)^{18} \approx 0.18867680$

(c) **SOLUTION**:

Since we aren't replacing people after sample, the probabilities of choosing one ambidextrous person is not independent from the probability of choosing a 2nd ambidextrous person.

Option 1:

Consider the specific scenario where the first 2 people you selected were ambidextrous and the rest weren't.

The probability of this is given by the following (using the multiplication rule):

$P(A_1, A_2, B_3, B_4, B_5, ....B_{20}) = P(A_1)P(A_2|A_1)P(B_3|A_1, A_2)P(B_4|A_1, A_2, B_3).....P(B_{20}|A_1, A_2, B_1, B_2, ...B_{19})$

$\left(\frac{5}{100}\right)\left(\frac{4}{99}\right)\left(\frac{95}{98}\right)\left(\frac{94}{97}\right)\left(\frac{93}{96}\right)\cdots\left(\frac{78}{81}\right)$

Writing this in terms of factorials we get:

$\frac{\frac{5!}{3!}\frac{95!}{77!}}{\frac{100!}{80!}}$

But this was just the ordering that assumed the first 2 people we selected were ambidextrous. Since there are $\binom{20}{2}$ different ways we can select 2 ambidextrous people in a group of 20, and each of them have the same probability as above, the overall probability is given by:

$\binom{20}{2}\frac{\frac{5!}{3!}\frac{95!}{76!}}{\frac{100!}{80!}} = 190\frac{\frac{5!}{3!}\frac{95!}{77!}}{\frac{100!}{80!}} \approx 0.2073437$

Option 2:

There are $\binom{100}{20}$ ways to choose 20 people out of a group of 100.

There are $\binom{5}{2}\binom{95}{18}$ ways to choose 2 people who are ambidextrous and 18 who aren't.

Thus P(two ambidextrous) $= \frac{\binom{5}{2}\binom{95}{18}}{\binom{100}{20}} \approx 0.2073437$

---

27. **SOLUTION**:

X counts the number of 1's out of 6 Bernoulli trials where the probability of a choosing a 1 is $p = 0.30$.

Thus $X \sim Bin(6, 0.30)$

---

28. .

(a) Box Plots and Histogram (other options are violin plots, but that wasn't given as a choice).

(b) Both variables are quantitative, but number of homes is discrete. So we could use a scatter plot (with number of homes on the x-axis), as long as there aren't too many datapoints that we have overplotting. Otherwise we could use side-by-side box plots.

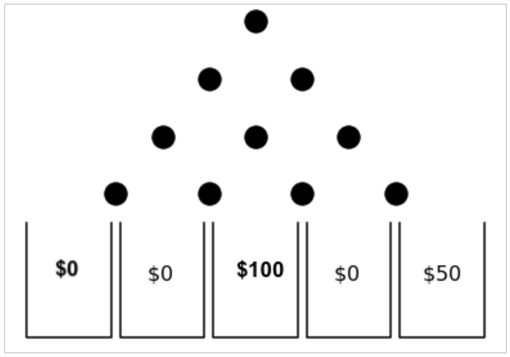
---

29. (a) Correct answer is $P(X > 3|X \text{ is even})$

(b)

$$P(X > 3 \mid X\text{is even }) = \frac{X > 3 \cap X\text{is even}}{X\text{is even}} = \frac{3/8}{1/2} = \boxed{\frac{3}{4}}$$

30. Lesson 12 Slides 60 and 61




31. A game of **Plinko** is to be played on the board shown below. The pegs **are biased such that a disc is twice as likely to move to the right than the left** at each peg. Furthermore, the disc can only be dropped from directly above the top-most peg. Answer the following questions about this Plinko game. Be sure to show your work.

(a)

| $k$ | $P(Y=k)$ |
|---|---|
| 0 | 41/81 |
| 50 | 16/81 |
| 100 | 8/27 |

(b) .

b). You can get \$100 in two ways:

You either get $\underbrace{\{100, 0\}}_{\text{Event Q}}$ or $\underbrace{\{50, 50\}}_{\text{Event R}}$

Each disc result is independent of previous result

Q: There are $\binom{2}{1}$ ways of arranging the outcomes (ie (100, 0) or (0, 100))

Thus $P(Q) = \binom{2}{1} g(100) g(0) = \binom{2}{1}\left(\frac{8}{27}\right)\left(\frac{41}{81}\right)$

R: $P(R) = g(50) \cdot g(50) = \left(\frac{16}{81}\right)^2$

Thus $P(Q \cup R) = 2\left(\frac{8}{27}\right)\left(\frac{41}{81}\right) + \left(\frac{16}{81}\right)^2 = \frac{2224}{6561}$

(c) .

c). You can get \$200 in three ways:

You either get $\underbrace{\{100,100,0,0,0\}}_{\text{Event Q}}$ or $\underbrace{\{100,50,50,0,0\}}_{\text{Event R}}$ or $\underbrace{\{50,50,50,50,0\}}_{\text{Event S}}$

Each disc result is indeoendent of previous result

Q: There are $\binom{5}{2}$ ways of arranging the outcomes

$P(Q) = \binom{5}{2}[g(100)]^2[g(0)]^3 = \binom{5}{2}\left(\frac{8}{27}\right)^2\left(\frac{41}{81}\right)^3 \approx .1139$

R: There are $\binom{5}{2}\binom{3}{2}$ ways of arranging the outcomes

choose 50's choose 0's

$P(R) = \binom{5}{2}\binom{3}{2} g(100)[g(50)]^2[g(0)]^2 = \binom{5}{2}\binom{3}{2}\left(\frac{8}{27}\right)\left(\frac{16}{81}\right)^2\left(\frac{41}{81}\right)^2 \approx .0889$

S: There are $\binom{5}{1}$ ways of arranging the outcomes

$P(S) = \binom{5}{1} g(0)[g(50)]^4 = \binom{5}{1}\left(\frac{41}{81}\right)\left(\frac{16}{81}\right)^4 \approx .0039$

Thus $P(Q \cup R \cup S) \approx .207$

(d) .

The pmf for Y is:

$$\boxed{g(y) = \begin{cases} 41/81 & y = 0 \\ 16/81 & y = 50 \\ 8/27 & y = 100 \end{cases}}$$

Thus $E[Y] = \left(\frac{41}{81}\right)(0) + \left(\frac{16}{81}\right)(50) + \left(\frac{8}{27}\right)(100)$

$= \frac{3200}{81} \approx 39.5$

So with 3 discs, the expected winnings are:

$E[Y_1 + Y_2 + Y_3] = E[Y_1] + E[Y_2] + E[Y_3]$

$= 3E[Y] = 3(39.5)$

$\approx \$118.5$

32. Lesson 14 Slide 78

**SOLUTION:**

$= 1 - 0.92 \approx 0.08$

33. Lesson 14 Slide 75

34. Lesson 14 Slide 65

35. **SOLUTION:**

For a binomial we know the Expected Value (i.e. mean) is $np$. We can use that to help us match the graphs. In addition, we can see that only histogram $B$ has a support of 0 through 5, so it's the only one with $n = 5$ (the others all have a support of 0 to 10).

1. $A$
2. $C$
3. $D$
4. $B$