4.2: Analyzing Discrete Random Variables

Learning Objectives

- Draw connections between a relative frequency distribution of a data set and a probability distribution of a random variable
- Define, compute, and interpret the expected value of a discrete random variable
- Define and compute the variance and standard deviation of a discrete random variable
- Interpret the standard deviation of a discrete random variable

 Section 4.2 Excel File (contains all of the data sets for this section)

Random Variables and Data

As we seek to understand the world around us, there are times when we are unsure what could happen in a particular situation; we do not necessarily know all the possible outcomes or the likelihood of every outcome happening. That is okay. We learn by observation, data collection, and experimentation. When studying a quantitative variable, we collect data by measuring a value for each observation. Once our data is collected, we can analyze the data using relative frequency distributions to show the proportion of the observations in a particular class and provide an empirical estimate of the probability of that class occurring. We realize that our relative frequency distributions are connected with the probability distributions of our random variables (classes corresponding to values the variable takes on and relative frequencies corresponding to probabilities). There is, however, a significant difference to keep in mind. Relative frequency distributions describe data that has been collected. Random variables describe the possibilities and probabilities of what can happen. Relative frequency distributions are descriptive, while random variables are predictive.

To analyze and understand data, we developed methods of visualization, measures of centrality, and measures of dispersion. For these methods with random variables, we use the connection to relative frequency distributions. Recall how we treated these three concepts when given a relative frequency distribution for a discrete quantitative variable.

Visualization: We visualized the relative frequency distribution using bar graphs.

Measures of Centrality: We discovered that the mean of a data set could be calculated by "weighing" our class values (x_j) by their relative frequencies $(P(x_j))$

$$\mu=\frac{\sum\left[x_j\cdot P(x_j)\right]}{\sum P(x_j)}=\sum\left[x_j\cdot P(x_j)\right]$$

Measures of Dispersion: We discovered that the variance of a population data set could be calculated by "weighing" the squared deviations from the mean $\left((x_j-\mu)^2\right)$ by their relative frequencies $(P(x_j))$

$$\sigma^2=\frac{\sum\left[(x_j-\mu)^2\cdot P(x_j)\right]}{\sum P(x_j)}=\sum\left[(x_j-\mu)^2\cdot P(x_j)\right]$$

From the variance, we can also compute the standard deviation of the data set: $\sigma=\sqrt{\sigma^2}$.

We will visualize our discrete random variables with bar graphs and develop measures of centrality and dispersion similar to the mean and variance of a population data set.

Analyzing Discrete Random Variables

The primary challenge in connecting our analyses of relative frequency distributions and random variables is interpretation. As mentioned above, relative frequency distributions are constructed from collected data and describe what happened; random variables are constructed from all the possible values resulting from a random experiment and the associated probabilities of these values occurring. Even in this distinction, the best predictor of future behavior is past behavior; we often use relative frequencies as estimates of probabilities (the empirical method of probability). So, how can we understand the measures of centrality and dispersion for a random variable? They would be estimates for the actual measures of centrality and dispersion if we were to repeatedly run the random experiment, collect data, and analyze it. Because of the Law of Large Numbers, we would expect the estimates and computed measures to converge if we conducted the random experiment repeatedly. With this in mind, we will study discrete random variables.

Graphical Visualization: Bar Graphs

The method of constructing bar graphs remains the same for discrete random variables. The values of the random variable are listed on the horizontal axis, and bars are formed with the height indicating the probability of that particular value occurring. The gaps between the bars indicate the discrete nature of the random variable. Here, we produce bar graphs for the random variables X, Y, and D from our last section (the sum of two fair dice rolled, the max of two fair dice rolled, and the number of days adults exercise in a week, respectively).

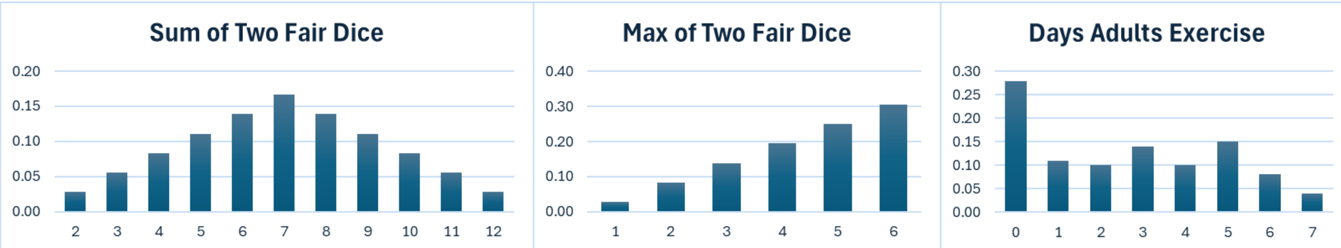

Figure $4.2.1$: Bar graph representations of the probability distributions of the random variables X, Y, and D (left to right)

Measure of Centrality: Expected Value

Recall that the mean of a data set is the balancing point of that data set, which we could visualize from our graphs. Given this same manner of representation, the mean of a random variable X, also referred to as the **expected value of X, $\mathrm{E}(X)$,** is also the balance point of the bar graph. From the symmetry of X in the figure above, we conclude that $\mathrm{E}(X)=7$. Determining the center of mass for Y and D is a little more complicated, but by looking at the distributions, we can see that $\mathrm{E}(Y)>4$ and $\mathrm{E}(D)<3$. A more precise calculation is in order. By understanding the relative frequency as an estimate of that particular event's probability, we define the **expected value of a discrete random variable X.**

$$\mu=\mathrm{E}(X)=\frac{\sum\left[x_j \cdot P\left(X=x_j\right)\right]}{\sum P\left(X=x_j\right)}=\sum\left[x_j \cdot P\left(X=x_j\right)\right]$$

In computing the expected value of a discrete random variable, we consider each possible value of the random variable and weigh it according to the associated probability; the higher the probability, the heavier the weight. Since the sum of probabilities of all possible values adds to 1, the denominator consisting of the "total" weight simplifies. Let us confirm that the expected value for the sum of two fair dice is indeed 7.

Table $4.2.1$: Computation of the expected value of the sum of two fair dice

$X=x_j$	$P(X=x_j)$	$x_j \cdot P(X=x_j)$
2	$\frac{1}{36}$	$2 \cdot \frac{1}{36}=\frac{2}{36}$
3	$\frac{2}{36}$	$3 \cdot \frac{2}{36}=\frac{6}{36}$
4	$\frac{3}{36}$	$4 \cdot \frac{3}{36}=\frac{12}{36}$
5	$\frac{4}{36}$	$5 \cdot \frac{4}{36}=\frac{20}{36}$
6	$\frac{5}{36}$	$6 \cdot \frac{5}{36}=\frac{30}{36}$
7	$\frac{6}{36}$	$7 \cdot \frac{6}{36}=\frac{42}{36}$
8	$\frac{5}{36}$	$8 \cdot \frac{5}{36}=\frac{40}{36}$
9	$\frac{4}{36}$	$9 \cdot \frac{4}{36}=\frac{36}{36}$
10	$\frac{3}{36}$	$10 \cdot \frac{3}{36}=\frac{30}{36}$

$X = x_j$	$P(X = x_j)$	$x_j \cdot P(X = x_j)$
11	$\frac{2}{36}$	$11 \cdot \frac{2}{36} = \frac{22}{36}$
12	$\frac{1}{36}$	$12 \cdot \frac{1}{36} = \frac{12}{36}$
$\mu = \mathrm{E}(X) = \frac{2}{36} + \frac{6}{36} + \ldots + \frac{12}{36} = \frac{252}{36} = 7$		

After adding all of the entries $(x_j \cdot P(X = x_j))$ in the third column, we arrived at a total of 7, confirming our visual estimation. $\mathrm{E}(X) = 7$ means that as we roll a pair of fair dice repeatedly, we expect the mean of all of the sums to be close to 7. Consider rolling two fair dice using a simulator and calculating the mean of the dice sums. We (the authors) ran the simulation with 20 trials, producing a data set of 20 dice sums, and found the mean to be $\frac{142}{20} = 7.1$; this is fairly close to our expected value of 7. Run the simulation yourself and compute the mean. How close to 7 is your computed mean?

? Text Exercise 4.2.1

1. Recall the discrete random variable Y which describes the maximum value of two fair dice when rolled. Compute and interpret the expected value $\mathrm{E}(Y)$ using its probability distribution reproduced below.

Table $4.2.2$: Probability distribution for the random variable Y

$Y = y_j$	$P(Y = y_j)$	$y_j \cdot (P(Y = y_j)$
1	$\frac{1}{36}$	
2	$\frac{3}{36}$	
3	$\frac{5}{36}$	
4	$\frac{7}{36}$	
5	$\frac{9}{36}$	
6	$\frac{11}{36}$	

Answer

Table $4.2.3$ Table of computation

$Y = y_j$	$P(Y = y_j)$	$y_j \cdot (P(Y = y_j)$
1	$\frac{1}{36}$	$1 \cdot \frac{1}{36} = \frac{1}{36}$
2	$\frac{3}{36}$	$2 \cdot \frac{3}{36} = \frac{6}{36}$
3	$\frac{5}{36}$	$3 \cdot \frac{5}{36} = \frac{15}{36}$
4	$\frac{7}{36}$	$4 \cdot \frac{7}{36} = \frac{28}{36}$
5	$\frac{9}{36}$	$5 \cdot \frac{9}{36} = \frac{45}{36}$
6	$\frac{11}{36}$	$6 \cdot \frac{11}{36} = \frac{66}{36}$
$\mu = \mathrm{E}(Y) = \frac{1}{36} + \frac{6}{36} + \ldots + \frac{66}{36} = \frac{161}{36} = 4 + \frac{17}{36} \approx 4.4722$		

The expected value of the random variable Y is approximately 4.4722, meaning after repeatedly rolling two fair dice and examining the maximum values from each pair, the mean maximum value would be about 4.4722. We can verify this experimentally; see for yourself. Roll two dice, record the highest number, repeat 20 times, and then take the mean of those 20 numbers. We (the authors) obtained a mean of 4.7 when doing this. Notice that this seems less accurate than our empirical estimate obtained for X. Some random variables require more trials than others to accurately estimate the expected value. Note that our expected value cannot occur for any single trial of our random experiment.

2. Recall the discrete random variable D which describes the number of days adults exercise per week. Compute and interpret the expected value $\mathrm{E}(D)$ using its probability distribution reproduced below.

Table $4.2.4$: Probability distribution of the random variable D

$\boldsymbol{D = d_j}$	$\boldsymbol{P(D = d_j)}$	$\boldsymbol{d_j \cdot (P(D = d_j)}$
0	0.28	
1	0.11	
2	0.10	
3	0.14	
4	0.10	
5	0.15	
6	0.08	
7	0.04	

Answer

Table $4.2.5$ Table of computation

$D = d_j$	$P(D = d_j)$	$d_j \cdot (P(D = d_j)$
0	0.28	$0 \cdot 0.28 = 0$
1	0.11	$1 \cdot 0.11 = 0.11$
2	0.10	$2 \cdot 0.10 = 0.20$
3	0.14	$3 \cdot 0.14 = 0.42$
4	0.10	$4 \cdot 0.10 = 0.4$
5	0.15	$5 \cdot 0.15 = 0.75$
6	0.08	$6 \cdot 0.08 = 0.48$
7	0.04	$7 \cdot 0.04 = 0.28$
$\mu = \mathrm{E}(D) = 0 + 0.11 + \ldots + 0.28 = 2.64$		

The expected value of the random variable D is 2.64 days, meaning that after repeatedly asking random adults how many days they work out per week, we would predict the mean number of days to be about 2.64.

Measures of Dispersion: Variance and Standard Deviation

A question arises from our discussion of expected value: how much variation should there be as the random experiment is repeated? Will most values be close to the expected value, or will a wide variety of values occur? We can answer these questions intuitively using bar graphs, but the intuition is rarely straightforward. Use Figure $4.2.1$ to determine which random variable has the most spread before continuing with your reading.

Using the notion of relative frequency as an estimate for a value's probability and utilizing our expected value as our measure of center, we define the **variance of a discrete random variable X.**

$$\begin{aligned}\sigma^2=\operatorname{Var}(X) &= \frac{\sum\left[\left(x_j-\mathrm{E}(X)\right)^2\cdot P(X=x_j)\right]}{\sum P(X=x_j)}=\sum\left[\left(x_j-\mathrm{E}(X)\right)^2\cdot P(X=x_j)\right] \\ &= \frac{\sum\left[\left(x_j-\mu\right)^2\cdot P(X=x_j)\right]}{\sum P(X=x_j)}=\sum\left[\left(x_j-\mu\right)^2\cdot P(X=x_j)\right]\end{aligned}$$

In computing the variance of a discrete random variable, we "weigh" the values that we are averaging, namely the squared deviations from the mean, by their probabilities. Since the sum of the probabilities of all the possible values is 1, the denominator reduces. As before, we define the **standard deviation of a discrete random variable X** as the square root of the random variable's variance. Let us determine the dispersion of the discrete random variables, X, Y, and D by computing their variances and standard deviations.

Table $4.2.6$: Table of computation

$X=x_j$	$P(X=x_j)$	$(x_j-\mu)^2$	$(x_j-\mu)^2\cdot P(X=x_j)$
2	$\frac{1}{36}$	$(2-7)^2=25$	$25\cdot\frac{1}{36}=\frac{25}{36}$
3	$\frac{2}{36}$	$(3-7)^2=16$	$16\cdot\frac{2}{36}=\frac{32}{36}$
4	$\frac{3}{36}$	$(4-7)^2=9$	$9\cdot\frac{3}{36}=\frac{27}{36}$
5	$\frac{4}{36}$	$(5-7)^2=4$	$4\cdot\frac{4}{36}=\frac{16}{36}$
6	$\frac{5}{36}$	$(6-7)^2=1$	$1\cdot\frac{5}{36}=\frac{5}{36}$
7	$\frac{6}{36}$	$(7-7)^2=0$	$0\cdot\frac{6}{36}=\frac{0}{36}$
8	$\frac{5}{36}$	$(8-7)^2=1$	$1\cdot\frac{5}{36}=\frac{5}{36}$
9	$\frac{4}{36}$	$(9-7)^2=4$	$4\cdot\frac{4}{36}=\frac{16}{36}$
10	$\frac{3}{36}$	$(10-7)^2=9$	$9\cdot\frac{3}{36}=\frac{27}{36}$
11	$\frac{2}{36}$	$(11-7)^2=16$	$16\cdot\frac{2}{36}=\frac{32}{36}$
12	$\frac{1}{36}$	$(12-7)^2=25$	$25\cdot\frac{1}{36}=\frac{25}{36}$
$\sigma^2=\operatorname{Var}(X)=\frac{25}{36}+\frac{32}{36}+\ldots+\frac{25}{36}=\frac{210}{36}=5\frac{5}{6}\approx 5.8333$			
$\sigma=\sqrt{\operatorname{Var}(X)}=\sqrt{\frac{35}{6}}\approx 2.4152$			

As with the mean, we understand the standard deviation as an estimate for the computed standard deviation as we repeatedly conduct the random experiment. As repetitions increase, the computed standard deviation approaches the expected standard deviation value. We understand the mean of a data set as its center and can loosely understand the standard deviation as the typical distance our observations are from the mean. If we translated this in terms of random variables, we could say that a typical value is within a standard deviation from the expected value. Within the context of summing the rolls of two fair dice, we would expect a typical value to be within 2.4152units from 7, those sums being 5, 6, 7, 8, or 9. If we were playing a board game based on the sum of the rolls of two fair dice, we could expect these 5 numbers to occur somewhat "regularly" and plan our strategy accordingly.

Text Exercise 4.2.2

1. Determine the probability that the sum of the rolling of two fair dice would be within one standard deviation of the expected value: $P(|X-\mu|<\sigma)$.

Answer

Since the mean is 7 and the standard deviation is about 2.4152, we are looking for the probability that our random variable is between the values between $7-2.4152=4.5848$ and $7+2.4152=9.4152$. Our question reduces to finding $P(X=5,6,7,8,\text{ or }9)=\frac{4}{36}+\frac{5}{36}+\frac{6}{36}+\frac{5}{36}+\frac{4}{36}=\frac{24}{36}=\frac{2}{3}\approx 66.6667\%$.

2. Determine the probability that the sum of the rolling of two fair dice would be within two standard deviations of the expected value: $P(|X-\mu|<2\sigma)$.

Answer

We are now looking for the probability that our random variable takes on the values between $7-2\cdot 2.4152=2.1696$ and $7+\cdot 2.4152=11.8304$. So our question reduces to finding $P(3\leq X\leq 11)=1-P(X=2\text{ or }12)=1-(\frac{1}{36}+\frac{1}{36})=1-\frac{2}{36}=\frac{34}{36}=\frac{17}{18}\approx 94.4444\%$. Notice that all possible values of our random variable lie within 3 standard deviations of the expected value.

Text Exercise 4.2.3

1. Recall the discrete random variable Y that describes the maximum value of two fair dice when rolled. In text exercise $4.2.1.1$, we computed that $\mathrm{E}(Y)=\frac{161}{36}$.
   a. Compute the variance and standard deviation of Y using its probability distribution reproduced below.
   b. Compute $P(|Y-\mu|<\sigma)$.
   c. Compute $P(|Y-\mu|<2\sigma)$.

Table $4.2.7$: Probability distribution of the random variable Y

$Y=y_j$	$P(Y=y_j)$	$(y_j-\mu)^2\cdot P(Y=y_j)$
1	$\frac{1}{36}$	
2	$\frac{3}{36}$	
3	$\frac{5}{36}$	
4	$\frac{7}{36}$	
5	$\frac{9}{36}$	
6	$\frac{11}{36}$	

Answer

Table $4.2.8$ Table of computation

$Y=y_j$	$P(Y=y_j)$	$(y_j-\mu)^2\cdot P(Y=y_j)$
1	$\frac{1}{36}$	$\left(1-\frac{161}{36}\right)^2\cdot\frac{1}{36}\approx 0.3349$
2	$\frac{3}{36}$	$\left(2-\frac{161}{36}\right)^2\cdot\frac{3}{36}\approx 0.5093$

$Y=y_j$	$P(Y=y_j)$	$(y_j-\mu)^2\cdot P(Y=y_j)$
3	$\frac{5}{36}$	$\left(3-\frac{161}{36}\right)^2\cdot\frac{5}{36}\approx 0.3010$
4	$\frac{7}{36}$	$\left(4-\frac{161}{36}\right)^2\cdot\frac{7}{36}\approx 0.0434$
5	$\frac{9}{36}$	$\left(5-\frac{161}{36}\right)^2\cdot\frac{9}{36}\approx 0.0696$
6	$\frac{11}{36}$	$\left(6-\frac{161}{36}\right)^2\cdot\frac{11}{36}\approx 0.7132$
$\sigma^2=\mathrm{Var}(Y)\approx 0.3349+0.5093+0.3010+0.0434+0.0696+0.7132\approx 1.9715$		
$\sigma=\sqrt{\mathrm{Var}(Y)}\approx\sqrt{1.9715}\approx 1.4041$		

a. Note: when conducting such involved and messy calculations, be sure to round only at the very last stage. We provide intermediate approximations along the way to help facilitate readability, but we do not use the rounded values in the actual computations for final answers. The variance of Y is approximately $1.9715$and the standard deviation is about 1.4041
b. We are tasked with determining the probability that the value produced in running the random experiment will fall within one standard deviation of the expected value.

$$\begin{aligned}P\left(|Y-\mu|<\sigma\right) &= P\left(4.4722-1.4041<Y<4.4722+1.4041\right)=P\left(3.0681<Y<5.8763\right)\\ &= P\left(Y=4\text{ or }5\right)\\ &= \frac{7}{36}+\frac{9}{36}=\frac{16}{36}\\ &= \frac{4}{9}\approx 44.4444\%\end{aligned}$$

c. We now determine the probability that the value produced in running the random experiment will fall within two standard deviations of the expected value.

$$\begin{aligned}P\left(|Y-\mu|<2\sigma\right) &= P\left(4.4722-2\cdot 1.4041<Y<4.4722+2\cdot 1.4041\right)=P\left(1.664<Y<7.2804\right)\\ &= P\left(Y=2,3,4,5,\text{ or }6\right)\\ &= 1-P\left(Y=1\right)=1-\frac{1}{36}\\ &= \frac{35}{36}\approx 97.2222\%\end{aligned}$$

2. Recall the discrete random variable D that describes the number of days adults exercise per week. In text exercise $4.2.1.2$, we computed that $\mathrm{E}(D)=2.64$ days.
   a. Compute the variance and standard deviation of D using its probability distribution reproduced below.
   b. Compute $P\left(|D-\mu|<\sigma\right)$.
   c. Compute $P\left(|D-\mu|<2\sigma\right)$.

Table $4.2.9$: Probability distribution of the random variable D

$D=d_j$	$P(D=d_j)$	$(d_j-\mu)^2\cdot P(D=d_j)$
0	0.28	
1	0.11	
2	0.10	
3	0.14	
4	0.10	
5	0.15	

$D=d_j$	$P(D=d_j)$	$(d_j-\mu)^2 \cdot P(D=d_j)$
6	0.08	
7	0.04	

Answer

Table 4.2.10 Table of computation

$D=d_j$	$P(D=d_j)$	$(d_j-\mu)^2 \cdot P(D=d_j)$
0	0.28	$(0-2.64)^2 \cdot 0.28 \approx 1.9515$
1	0.11	$(1-2.64)^2 \cdot 0.11 \approx 0.2959$
2	0.10	$(2-2.64)^2 \cdot 0.10 \approx 0.0410$
3	0.14	$(3-2.64)^2 \cdot 0.14 \approx 0.0181$
4	0.10	$(4-2.64)^2 \cdot 0.10 \approx$ \0.1850
5	0.15	$(5-2.64)^2 \cdot 0.15 \approx 0.8354$
6	0.08	$(6-2.64)^2 \cdot 0.08 \approx 0.9032$
7	0.04	$(7-2.64)^2 \cdot 0.04 \approx 0.7604$
$\sigma^2 = \text{Var}(D) \approx 1.9515+0.2959+\cdots+0.7604 \approx 4.9904$		
$\sigma = \sqrt{\text{Var}(D)} \approx \sqrt{4.9904} \approx 2.2339$		

a. The variance of D is approximately 4.9904days^2 and the standard deviation is about 2.2339days.

b. We determine the probability that the value produced in running the random experiment will fall within one standard deviation of the expected value.

$$\begin{aligned} P(|D-\mu|<\sigma) &= P(2.64-2.2339<D<2.64-2.2339) = P(0.4061<D<4.8739) \\ &= P(D=1,2,3,\text{ or }4) \\ &= 0.11+0.10+0.14+0.10 = 45\% \end{aligned}$$

c. We determine the probability that the value produced in running the random experiment will fall within two standard deviations of the expected value.

$$\begin{aligned} P(|D-\mu|<2\sigma) &= P(2.64-2\cdot 2.2339<D<2.64-2\cdot 2.2339) = P(-1.8278<D<7.1078) \\ &= P(D=0,1,2,3,4,5,6\text{ or }7) \\ &= 100\% \end{aligned}$$

As we have seen with the last several text exercises, the probability that a discrete random variable is within a standard deviation or two of the expected value varies depending on the distribution. This, hopefully, does not come as a surprise (if it did, recall our discussion in section 2.7 about Chebyshev's Inequality and the Empirical Rule). For this reason, we must take our understanding of a typical value rather loosely at this stage. It will get better. We end this section with text exercises exploring games of chance with cash prizes.

Note: Equivalent Formula for Variance

We may further reduce the formula for the variance of a random variable and produce an equivalent formulation with some computational benefits. The two formulas produce the same values; they are equivalent. The first formula emphasizes the underlying logic of what the measure means; we, therefore, recommend its use primarily, but we provide this second formula regardless.

$$\sigma^2 = \text{Var}(X) = \left(\sum\left[x_j^2 \cdot P(X=x_j)\right]\right) - \mu^2$$

Optional derivation for the mathematically inclined

$$\begin{aligned}\sigma^2=\mathrm{Var}(X) &=\sum\left[(x_j-\mu)^2\cdot P(X=x_j)\right]\\ &=\sum\left[\left(x_j^2-2x_j\mu+\mu^2\right)\cdot P(X=x_j)\right]\\ &=\sum\left[x_j^2\cdot P(X=x_j)-2x_j\mu\cdot P(X=x_j)+\mu^2\cdot P(X=x_j)\right]\\ &=\sum\left[x_j^2\cdot P(X=x_j)\right]-\sum\left[2x_j\mu\cdot P(X=x_j)\right]+\sum\left[\mu^2\cdot P(X=x_j)\right]\\ &=\sum\left[x_j^2\cdot P(X=x_j)\right]-2\mu\sum\left[x_j\cdot P(X=x_j)\right]+\mu^2\sum P(X=x_j)\\ &=\sum\left[x_j^2\cdot P(X=x_j)\right]-2\mu^2+\mu^2\\ &=\left(\sum\left[x_j^2\cdot P(X=x_j)\right]\right)-\mu^2\end{aligned}$$

Games of Chance

Text Exercise 4.2.4

1. Consider entering a raffle with a single cash prize of $\$1,000$. Each player may only purchase one ticket for $\$25$, and only 100 raffle tickets will be sold.
   a. Represent our financial net gains or losses with the discrete random variable R and construct its probability distribution.
   b. Compute and interpret $\mathrm{E}(R)$.

Answer

a. The random experiment is entering the raffle; either we win or we lose. If we lose, we are out the $\$25$. If we win, our net gain is $\$975$ because we spent $\$25$ on the raffle ticket. Our random variable R takes on two values: -25 and 975. Each ticket is typically equally likely to be drawn. Therefore, $P\left(X=-25\right)$ (the probability that we lose), is $\frac{99}{100}$. Its complement, the probability that we win, is $\frac{1}{100}$.

Table $4.2.11$: Probability distribution for the random variable R

$R=r_j$	$P(R=r_j)$
-25	$\frac{99}{100}$
975	$\frac{1}{100}$

b. Table $4.2.12$ Table of computation

$R=r_j$	$P(R=r_j)$	$r_j\cdot P(R=r_j)$
-25	$\frac{99}{100}$	$-25\cdot\frac{99}{100}=-\frac{99}{4}$
975	$\frac{1}{100}$	$975\cdot\frac{1}{100}=\frac{39}{4}$
$\mu=\mathrm{E}(R)=-\frac{99}{4}+\frac{39}{4}=-\frac{60}{4}=-15$		

We have found that the expected value for this raffle is $-\$15$. When we play, we either lose $\$25$ or gain $\$975$, but on average we should expect to lose $\$15$. It is important to note that a negative expected value does not necessarily mean we will lose money. Someone will win the raffle. What it means is that, if we played this raffle many times, we would expect to lose money in the long run. Imagine if we did the raffle 100 times; we might expect to win 1 time and lose 99 times. In such a situation, we won $\$1000$, but we also lost $\$25\cdot100=\$2500$ in purchasing tickets. Hence, we see a net loss of $\$1500$ for an average loss of $\$15$ each time.

2. Consider entering a raffle with one $\$1,000$ cash prize and one $\$500$ cash prize. Each player may only purchase one ticket for $\$50$, and 200 raffle tickets will be sold.
   a. Represent our financial net gains or losses with the discrete random variable R and construct its probability distribution.

b. Compute and interpret $\mathrm{E}(R)$.

Answer

a. We will set up our random variable similarly. However, this time, there are three possible values since there are three different outcomes: winning the big prize, winning the small prize, and losing altogether.

Table $4.2.13$ Probability distribution for the random variable R

$R = r_j$	$P(R = r_j)$
-50	$\dfrac{198}{200}$
450	$\dfrac{1}{200}$
950	$\dfrac{1}{200}$

b. Table $4.2.14$ Table of computation

$R = r_j$	$P(R = r_j)$	$r_j \cdot P(R = r_j)$
-50	$\dfrac{198}{200}$	$-50 \cdot \dfrac{198}{200} = -\dfrac{198}{4} =$
450	$\dfrac{1}{200}$	$450 \cdot \dfrac{1}{200} = \dfrac{9}{4}$
950	$\dfrac{1}{200}$	$950 \cdot \dfrac{1}{200} = \dfrac{19}{4}$
$\mu = \mathrm{E}(R) = -\dfrac{198}{4} + \dfrac{9}{4} + \dfrac{19}{4} = -\dfrac{170}{4} = -42.50$		

We have found that the expected value for this raffle is \-$\$42.50$.Despite more prize money being available, the higher price of entering the raffle and the larger number of tickets sold made the expected value decrease starkly.

3. A family is hosting an extended family reunion and thought having a raffle with a single cash prize would be fun. The dad does not want to make or lose any money in running the raffle, but his wife would like the raffle to help cover the cost of hosting the reunion and wants $\$5$ per ticket to help cover the costs. If 80 family members signed up for the raffle at a ticket price of $\$15$, determine the cash prize and compute the expected value when

   a. the dad gets his desire.
   b. the mom gets her desire.

Answer

a. Since the dad does not want the raffle to make or lose any money on the raffle, the cash prize will need to be all the ticket revenue. Since there are 80 tickets being sold at $\$15$ a ticket. The total ticket revenue is $\$1200$.We can construct a random variable similar to the variables above representing the game from the player's perspective.

Table $4.2.15$ Probability distribution for the random variable R along with computation

$R = r_j$	$P(R = r_j)$	$r_j \cdot P(R = r_j)$
-15	$\dfrac{79}{80}$	$-15 \cdot \dfrac{79}{80}$
1185	\(\dfrac{1}{80}\\)	$1185 \cdot \dfrac{1}{80}$
$\mu = \mathrm{E}(R) = -\dfrac{15 \cdot 79}{80} + \dfrac{1185}{80} = \dfrac{-1185 + 1185}{80} = 0$		

If the dad gets his desire and all the ticket revenue is given as the cash prize, the expected value would $\$0$.

b. If the mom gets her desire to help cover the cost of hosting the reunion, $\$10$ of every ticket will go to the cash prize, and $\$5$ will go towards the reunion. With 80 tickets sold, the cash prize would be $80 \cdot 10 = 800$ dollars.

Table $4.2.16$: Probability distribution for the random variable R along with computation

$R = r_j$	$P(R = r_j)$	$r_j \cdot P(R = r_j)$
-15	$\dfrac{79}{80}$	$-15 \cdot \dfrac{79}{80}$
785	\(\dfrac{1}{80}\\)	$785 \cdot \dfrac{1}{80}$
$\mu = \mathrm{E}(R) = -\dfrac{15 \cdot 79}{80} + \dfrac{785}{80} = \dfrac{-1185 + 785}{80} = -\dfrac{400}{80} = -5$		

If the mom gets her desire, the expected value would $-\$5$. We notice that the dad did not want to make any money off the raffle and the expected value was $\$0$, while the mom wanted to make $\$5$ per ticket and the expected value was $-\$5$. In general, if the host wanted to make $\$x$ off of each ticket, then the expected value from the participant's perspective would be $-\$x$.

Having worked through these text exercises, we understand the expected value as a measure of who the game of chance favors: the players or the house/host. Under the mom's desires, the host family making $\$5$ a ticket is represented in the expected value being $-\$5$. If we then apply this idea to the earlier raffles, where the expected values were $-\$15$ and $-\$42.50$, we understand that the hosts were raising money at $\$15$ and $\$42.50$ a ticket. This is reasonable because raffles are generally run as fundraisers for some organizations.

? Text Exercise $4.2.5$

Over the years, there have been big lottery prizes, as in millions and billions of dollars. Consider a simplified version of the Powerball game. While in reality, there are 9 different ways to win a cash prize from purchasing a $\$2$ ticket, we will only consider winning the grand prize without any additional options and assume there is only one winner. Recall that the grand prize is won when a player matches all 5 white balls (order does not matter) and the red Powerball. There are 69 white balls labeled 1 to 69 and 26 red balls labeled 1 to 26. The grand prize increases until a winner is found. Determine how large the grand prize must be for the expected value (from the player's perspective) to be nonnegative.

Answer

We are trying to find the smallest grand prize so that the expected value is nonnegative. Let us refer to the grand prize amount as p and set up a discrete random variable R modeling this simplified version of the Powerball. Refer to this previous text exercise for additional aid in determining the probabilities.

Table $4.2.17$: Probability distribution for the random variable R along with computation

$R = r_j$	$P(R = r_j)$	$r_j \cdot P(R = r_j)$
-2	$1 - \dfrac{1}{292,201,338}$	$-2 + \dfrac{2}{292,201,338}$
$p - 2$	$\dfrac{1}{292,201,338}$	(\dfrac{p-2}{292,201,338}\)
$\mu = \mathrm{E}(R) = -2 + \dfrac{2}{292,201,338} + \dfrac{p}{292,201,338} - \dfrac{2}{292,201,338} = -2 + \dfrac{p}{292,201,338}$		

We want the expected value to be nonnegative, so we are looking to solve $\mathrm{E}(R) \geq 0$.

$$\begin{aligned} \mathrm{E}(R) &\geq 0 \\ -2 + \frac{p}{292,201,338} &\geq 0 \\ \frac{p}{292,201,338} &\geq 2 \\ p &\geq 2 \cdot 292,201,338 = 584,402,676 \end{aligned}$$

Only after the grand prize grows beyond $\$584,402,676$ will the expected value of the Powerball be positive. Note that this analysis does not consider taxation on the winnings, which is quite hefty.

Given the fact that expected value is negative (unless the pot is enormous), we know that playing the lottery is not a good way to make money.

? Text Exercise 4.2.6

1. Consider a biased die, weighted so that rolling a one is 5 times as likely as the other sides, but each of the other sides are equally probable. Determine the probability distribution for the discrete random variable Z defined to be the value landing up after rolling this biased die.

Answer

We know that $P(Z=2)=P(Z=3)=P(Z=4)=P(Z=5)=P(Z=6)$, $P(Z=1)=5P(Z=2)$, and that the sum of all the probabilities needs to be 1. We have $1=P(Z=1)+P(Z=2)+P(Z=3)+P(Z=4)+P(Z=5)+P(Z=6)=5P(Z=2)+P(Z=2)+P(Z=2)+P(Z=2)+P(Z=2)+P(Z=2)=10P(Z=2)$. Thus $P(Z=2)=\frac{1}{10}$ and $P(Z=1)=\frac{5}{10}=\frac{1}{2}$.

Table 4.2.18 Probability distribution for the random variable Z

$Z=z_j$	$P(Z=z_j)$
1	$\frac{1}{2}$
2	$\frac{1}{10}$
3	$\frac{1}{10}$
4	$\frac{1}{10}$
5	$\frac{1}{10}$
6	$\frac{1}{10}$

2. Compare the expected value and standard deviation of Z with those of rolling a fair die.

Answer

Table 4.2.19 Table of computation for the two random variables

$Z=z_j$	$P(Z=z_j)$	$z_j\cdot P(Z=z_j)$	$(z_j-\mu)^2\cdot P(Z=z_j)$	$F=f_j$	$P(F=f_j)$	$f_j\cdot P(F=f_j)$	$(f_j-\mu)^2\cdot P(F=$ [illegible]
1	$\frac{1}{2}$	$\frac{1}{2}$	$\left(1-\frac{5}{2}\right)^2\cdot\frac{1}{2}=\frac{9}{8}$	1	$\frac{1}{6}$	$\frac{1}{6}$	$\left(1-\frac{7}{2}\right)^2\cdot\frac{1}{6}=$ [illegible]
2	$\frac{1}{10}$	$\frac{1}{5}$	$\left(2-\frac{5}{2}\right)^2\cdot\frac{1}{10}=\frac{1}{40}$	2	$\frac{1}{6}$	$\frac{1}{3}$	$\left(2-\frac{7}{2}\right)^2\cdot\frac{1}{6}=$ [illegible]
3	$\frac{1}{10}$	$\frac{3}{10}$	$\left(3-\frac{5}{2}\right)^2\cdot\frac{1}{10}=\frac{1}{40}$	3	$\frac{1}{6}$	$\frac{1}{2}$	$\left(3-\frac{7}{2}\right)^2\cdot\frac{1}{6}=$ [illegible]
4	$\frac{1}{10}$	$\frac{2}{5}$	$\left(4-\frac{5}{2}\right)^2\cdot\frac{1}{10}=\frac{9}{40}$	4	$\frac{1}{6}$	$\frac{2}{3}$	$\left(4-\frac{7}{2}\right)^2\cdot\frac{1}{6}=$ [illegible]
5	$\frac{1}{10}$	$\frac{1}{2}$	$\left(5-\frac{5}{2}\right)^2\cdot\frac{1}{10}=\frac{5}{8}$	5	$\frac{1}{6}$	$\frac{5}{6}$	$\left(5-\frac{7}{2}\right)^2\cdot\frac{1}{6}=$ [illegible]
6	$\frac{1}{10}$	$\frac{3}{5}$	$\left(6-\frac{5}{2}\right)^2\cdot\frac{1}{10}=\frac{49}{40}$	6	$\frac{1}{6}$	1	$\left(6-\frac{7}{2}\right)^2\cdot\frac{1}{6}=$ [illegible]

$Z = z_j$	$P(Z = z_j)$	$z_j \cdot P(Z = z_j)$	$(z_j - \mu)^2 \cdot P(Z = z_j)$	$F = f_j$	$P(F = f_j)$	$f_j \cdot P(F = f_j)$	$(f_j - \mu)^2 \cdot P(F =$
$\mu = \mathrm{E}(Z) = \frac{1}{2} + \frac{1}{5} + \ldots + \frac{3}{5} = \frac{25}{10} = \frac{5}{2} = 2.5$				$\mu = \mathrm{E}(F) = \frac{1}{6} + \frac{1}{3} + \ldots + 1 = \frac{21}{6} = \frac{7}{2} = 3.5$			
$\sigma^2 = \mathrm{Var}(Z) = \frac{9}{8} + \frac{1}{40} + \cdots + \frac{49}{40} = \frac{130}{40} = \frac{13}{4} = 3.25$				$\sigma^2 = \mathrm{Var}(F) = \frac{25}{24} + \frac{3}{8} + \cdots + \frac{25}{24} = \frac{70}{24} = \frac{35}{12} \approx 2.9167$			
$\sigma = \sqrt{3.25} \approx 1.8028$				$\sigma = \sqrt{\frac{35}{12}} \approx 1.7078$			

In comparing the two random variables, we notice that the bias brought the expected value closer to the biased value. This is relatively intuitive because one would expect the biased value to occur more often. The bias also increased the standard deviation a little. This may be less intuitive because it seems natural that the spread would decrease. After all, we expect ones to appear quite frequently. With a heavy enough bias, this can happen. The degree to which the expected value is pulled towards the biased value and the bias's impact on the standard deviation depends on the amount of bias and the probabilities of the other values.