Sampling with replacement

When sampling *with replacement*, you put back what you just drew.

- Imagine you have a bag with 5 red, 3 blue and 2 orange chips in it. What is the probability that the first chip you draw is blue?

  5 ● , 3 ● , 2 ●

$$Prob(1^{st} \text{ chip } B) = \frac{3}{5+3+2} = \frac{3}{10} = 0.3$$

- Suppose you did indeed pull a blue chip in the first draw. If drawing with replacement, what is the probability of drawing a blue chip in the second draw?

  1^{st} draw: 5 ● , 3 ● , 2 ●
  2^{nd} draw: 5 ● , 3 ● , 2 ●

$$Prob(2^{nd} \text{ chip } B | 1^{st} \text{ chip } B) = \frac{3}{10} = 0.3$$

Sampling with replacement (cont.)

- Suppose you actually pulled an orange chip in the first draw. If drawing with replacement, what is the probability of drawing a blue chip in the second draw?

$$1^{st}\text{ draw: } 5 \bullet, 3 \bullet, 2 \bullet$$
$$2^{nd}\text{ draw: } 5 \bullet, 3 \bullet, 2 \bullet$$

$$Prob(2^{nd}\text{ chip } B|1^{st}\text{ chip } O) = \frac{3}{10} = 0.3$$

- If drawing with replacement, what is the probability of drawing two blue chips in a row?

$$1^{st}\text{ draw: } 5 \bullet, 3 \bullet, 2 \bullet$$
$$2^{nd}\text{ draw: } 5 \bullet, 3 \bullet, 2 \bullet$$

$$Prob(1^{st}\text{ chip } B) \cdot Prob(2^{nd}\text{ chip } B|1^{st}\text{ chip } B) = 0.3 \times 0.3$$
$$= 0.3^2 = 0.09$$

Sampling with replacement (cont.)

- When drawing with replacement, probability of the second chip being blue does not depend on the color of the first chip since whatever we draw in the first draw gets put back in the bag.

$$Prob(B|B) = Prob(B|O)$$

- In addition, this probability is equal to the probability of drawing a blue chip in the first draw, since the composition of the bag never changes when sampling with replacement.

$$Prob(B|B) = Prob(B)$$

- *When drawing with replacement, draws are independent.*

Sampling without replacement

When drawing *without replacement* you do not put back what you just drew.

- Suppose you pulled a blue chip in the first draw. If drawing without replacement, what is the probability of drawing a blue chip in the second draw?

$$1^{st} \text{ draw: } 5\ {\color{red}\bullet}\ ,\ 3\ {\color{blue}\bullet}\ ,\ 2\ {\color{orange}\bullet}$$

$$2^{nd} \text{ draw: } 5\ {\color{red}\bullet}\ ,\ 2\ {\color{blue}\bullet}\ ,\ 2\ {\color{orange}\bullet}$$

$$Prob(2^{nd} \text{ chip } B | 1^{st} \text{ chip } B) = \frac{2}{9} = 0.22$$

- If drawing without replacement, what is the probability of drawing two blue chips in a row?

$$1^{st} \text{ draw: } 5\ {\color{red}\bullet}\ ,\ 3\ {\color{blue}\bullet}\ ,\ 2\ {\color{orange}\bullet}$$

$$2^{nd} \text{ draw: } 5\ {\color{red}\bullet}\ ,\ 2\ {\color{blue}\bullet}\ ,\ 2\ {\color{orange}\bullet}$$

$$Prob(1^{st} \text{ chip } B) \cdot Prob(2^{nd} \text{ chip } B | 1^{st} \text{ chip } B) = 0.3 \times 0.22$$

$$= 0.066$$

Sampling without replacement (cont.)

- When drawing without replacement, the probability of the second chip being blue given the first was blue is not equal to the probability of drawing a blue chip in the first draw since the composition of the bag changes with the outcome of the first draw.
$$Prob(B|B) \neq Prob(B)$$
- *When drawing without replacement, draws are not independent.*
- This is especially important to take note of when the sample sizes are small. If we were dealing with, say, 10,000 chips in a (giant) bag, taking out one chip of any color would not have as big an impact on the probabilities in the second draw.

Practice

In most card games cards are dealt without replacement. What is the probability of being dealt an ace and then a 3? Choose the closest answer.

(a) 0.0045
(b) 0.0059
(c) 0.0060
(d) 0.1553

Random variables

- A *random variable* is a numeric quantity whose value depends on the outcome of a random event
  - We use a capital letter, like X, to denote a random variable
  - The values of a random variable are denoted with a lowercase letter, in this case x
  - For example, $P(X = x)$
- There are two types of random variables:
  - *Discrete random variables* often take only integer values
    - Example: Number of credit hours, Difference in number of credit hours this term vs last
  - *Continuous random variables* take real (decimal) values
    - Example: Cost of books this term, Difference in cost of books this term vs last

Expectation

- We are often interested in the average outcome of a random variable.
- We call this the *expected value* (mean), and it is a weighted average of the possible outcomes

$$\mu = E(X) = \sum_{i=1}^{k} x_i \, P(X = x_i)$$

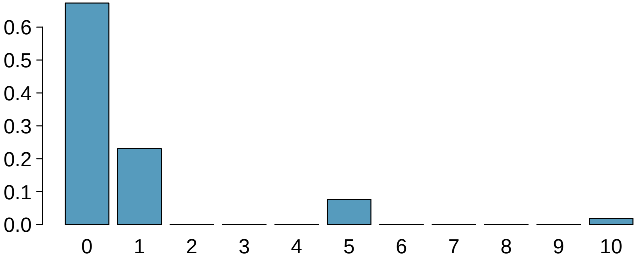
Expected value of a discrete random variable

In a game of cards you win \$1 if you draw a heart, \$5 if you draw an ace (including the ace of hearts), \$10 if you draw the king of spades and nothing for any other card you draw. Write the probability model for your winnings, and calculate your expected winning.

Event	X	$P(X)$	$X\ P(X)$
Heart (not ace)	1	$\frac{12}{52}$	$\frac{12}{52}$
Ace	5	$\frac{4}{52}$	$\frac{20}{52}$
King of spades	10	$\frac{1}{52}$	$\frac{10}{52}$
All else	0	$\frac{35}{52}$	0
Total			$E(X) = \frac{42}{52} \approx 0.81$

Expected value of a discrete random variable (cont.)

Below is a visual representation of the probability distribution of winnings from this game:



Variability

We are also often interested in the variability in the values of a random variable.

$$\sigma^2 = Var(X) = \sum_{i=1}^{k} (x_i - E(X))^2 P(X = x_i)$$

$$\sigma = SD(X) = \sqrt{Var(X)}$$

Variability of a discrete random variable

For the previous card game example, how much would you expect the winnings to vary from game to game?

X	$P(X)$	$X\ P(X)$	$(X - E(X))^2$	$P(X)\ (X - E(X))^2$
1	$\frac{12}{52}$	$1 \times \frac{12}{52} = \frac{12}{52}$	$(1 - 0.81)^2 = 0.0361$	$\frac{12}{52} \times 0.0361 = 0.0083$
5	$\frac{4}{52}$	$5 \times \frac{4}{52} = \frac{20}{52}$	$(5 - 0.81)^2 = 17.5561$	$\frac{4}{52} \times 17.5561 = 1.3505$
10	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	$\frac{1}{52} \times 84.0889 = 1.6242$
0	$\frac{35}{52}$	$0 \times \frac{35}{52} = 0$	$(0 - 0.81)^2 = 0.6561$	$\frac{35}{52} \times 0.6561 = 0.4416$
		$E(X) = 0.81$		

Linear combinations

- A *linear combination* of random variables X and Y is given by

$$aX + bY$$

  where a and b are some fixed numbers.
- The average value of a linear combination of random variables is given by

$$E(aX + bY) = a \times E(X) + b \times E(Y)$$

Calculating the expectation of a linear combination

On average you take 10 minutes for each statistics homework problem and 15 minutes for each chemistry homework problem. This week you have 5 statistics and 4 chemistry homework problems assigned. What is the total time you expect to spend on statistics and physics homework for the week?

Linear combinations

- The variability of a linear
  combination of two independent random variables is calculated as

$$V(aX + bY) = a^2 \times V(X) + b^2 \times V(Y)$$

- The standard deviation of the linear combination is the square root of the variance.

Note: If the random variables are not independent, the variance calculation gets a little more complicated and is beyond the scope of this course.

Calculating the variance of a linear combination

The standard deviation of the time you take for each statistics homework problem is 1.5 minutes, and it is 2 minutes for each chemistry problem. What is the standard deviation of the time you expect to spend on statistics and physics homework for the week if you have 5 statistics and 4 chemistry homework problems assigned? Suppose that the time it takes to complete each problem is independent of another.

Practice

A casino game costs \$5 to play. If the first card you draw is red, then you get to draw a second card (without replacement). If the second card is the ace of clubs, you win \$500. If not, you don't win anything, i.e. lose your \$5. What is your expected profits/losses from playing this game? Remember: profit/loss = winnings - cost.

(a) A profit of 5¢

(b) A loss of 10¢

(c) A loss of 25¢

(d) A loss of 30¢

Fair game

A *fair* game is defined as a game that costs as much as its expected payout, i.e. expected profit is 0.

Do you think casino games in Vegas cost more or less than their expected payouts?

Simplifying random variables

Random variables do not work like normal algebraic variables:

$$X + X \neq 2X$$

$$\begin{aligned} E(X+X) &= E(X) + E(X) \\ &= 2E(X) \end{aligned} \qquad \begin{aligned} Var(X+X) &= Var(X) + Var(X) \text{ (assuming independence} \\ &= 2\ Var(X) \end{aligned}$$

$$E(2X) = 2E(X) \qquad \begin{aligned} Var(2X) &= 2^2\ Var(X) \\ &= 4\ Var(X) \end{aligned}$$

$E(X+X) = E(2X)$, but $Var(X+X) \neq Var(2X)$.

Adding or multiplying?

A company has 5 Lincoln Town Cars in its fleet. Historical data show that annual maintenance cost for each car is on average \$2,154 with a standard deviation of \$132. What is the mean and the standard deviation of the total annual maintenance cost for this fleet?

Note that we have 5 cars each with the given annual maintenance cost ($X_1 + X_2 + X_3 + X_4 + X_5$), not one car that had 5 times the given annual maintenance cost ($5X$).

$$
\begin{aligned}
E(X_1 + X_2 + X_3 + X_4 + X_5) &= E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5) \\
&= 5 \times E(X) = 5 \times 2,154 = \$10,770 \\
Var(X_1 + X_2 + X_3 + X_4 + X_5) &= Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4) + Va \\
&= 5 \times V(X) = 5 \times 132^2 = \$87,120 \\
SD(X_1 + X_2 + X_3 + X_4 + X_5) &= \sqrt{87,120} = 295.16
\end{aligned}
$$