What is the probability you get a sum of 5 on two dice? Now assume one die is a 4. Does that affect the probability the sum is 5?

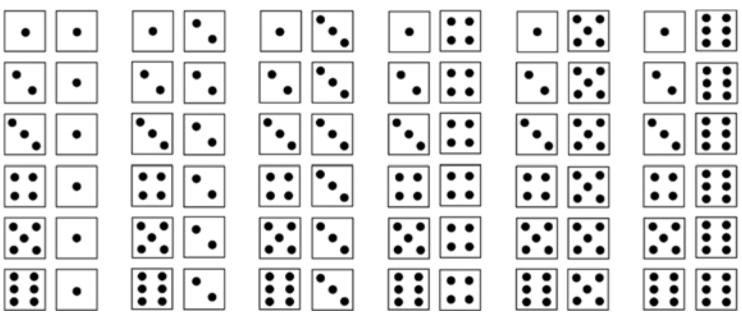
General Education Statistics
Class Notes
Conditional Probability (Section 5.4)

Events often do *not* influence another's probability, such as with tossing a coin and a die. This independence allows us to find probabilities using the formula $P(E \text{ and } F) = P(E) \cdot P(F)$.

But what if the events are *not* independent? Are the events "dies from heart disease" and "is diabetic" independent? In fact, the probability of dying from heart disease is 2 to 4 times higher in a person with diabetes. (source: www.webmd.com) Would you call that independent? We study a way to deal with these probabilities in this section.

Definition: Conditional Probability: The notation $P(F \mid E)$ is read "the probability of event F given event E". It is the probability that the event F occurs *given* that event E has occurred.

expl 1: Consider rolling two distinguishable, six-sided dice. Here is the sample space. Answer the questions that follow.



How many successes? How many possibilities?

a.) What is the probability that the sum of the two dice is 5? Circle those successes.

b.) What is the probability that the first die is a 4? Circle those successes.

c.) Given that the first die is a 4, what is the probability that the sum is 5? What are the possibilities now? How many successes are there?

How did the sample space change?

Conditional Probability Rule:

The fact that the first die is known to be 4 changes the probability that the sum is 5 because it reduces the number of possibilities we'll consider. Only those outcomes where the first die is 4 are possibilities now. This takes our familiar formula which was always

$$\text{Probability of an event} = \frac{\text{number of successes}}{\text{number of total possibilities}}$$

and makes it more like $P(F \mid E) = \dfrac{N(E \text{ and } F)}{N(E)}$.

Now the possibilities are just those that are in event E.

Another version of this formula, using the probabilities, tells us $P(F \mid E) = \dfrac{P(E \text{ and } F)}{P(E)}$.

expl 2: Let's verify this formula with our dice example. We know already that $P(\text{sum is } 5) = \frac{4}{36}$, $P(\text{first die is } 4) = \frac{6}{36}$, and $P(\text{sum is } 5 \mid \text{first die is } 4) = \frac{1}{6}$. Find $P(\text{sum is 5 AND first die is } 4)$ by looking at the sample space given earlier.

Now, notice that $P(\textit{sum is } 5 \mid \textit{first is } 4) = \dfrac{P(\textit{sum is 5 AND first is } 4)}{P(\textit{first is } 4)}$. Fill this equation out with the appropriate probabilities to verify it.

You can do the fraction math by hand or on a calculator.

Another way to state this rule is $P(E \text{ and } F) = P(E) \cdot P(F \mid E)$. This is reminiscent of the formula we had in the last section (for independent events), $P(E \text{ and } F) = P(E) \cdot P(F)$.

So if the events are *not* independent, we use this formula with conditional probabilities.

There are some great real-world examples of conditional probabilities where we will use contingency (or two-way) tables.

expl 3: The data below show the marital status of males and females 15 years old or older in the US in 2013. Answer the following questions.

Marital Status	**Males (in millions)**	**Females (in millions)**	**Total (in millions)**
Never married	41.6	36.9	**78.5**
Married	64.4	63.1	**127.5**
Widowed	3.1	11.2	**14.3**
Divorced	11.0	14.4	**25.4**
Separated	2.4	3.2	**5.6**
Total	**122.5**	**128.8**	**251.3**

(Source: U.S. Census Bureau, Current Population Reports)

a.) What is the total number of males?

How many successes? How many possibilities?

b.) Find the probability that a person selected at random would be widowed *given* that the person is a male.

c.) Find the probability that a person selected at random would be a male *given* that the person is widowed.

How many successes? How many possibilities?

d.) Return to questions *b* and *c* to write a sentence for each that helps show the difference between what you found.

expl 4: Again, the data below show the marital status of males and females 15 years old or older in the U.S. in 2013. Answer the following questions.

Marital Status	**Males (in millions)**	**Females (in millions)**	**Total (in millions)**
Never married	41.6	36.9	**78.5**
Married	64.4	63.1	**127.5**
Widowed	3.1	11.2	**14.3**
Divorced	11.0	14.4	**25.4**
Separated	2.4	3.2	**5.6**
Total	**122.5**	**128.8**	**251.3**

(Source: U.S. Census Bureau, Current Population Reports)

a.) Find the probability that a randomly selected person is separated. Round to the nearest whole percent.

How many successes? How many possibilities?

b.) Find the probability that a randomly selected person is separated *given* the person is a male. Round to the nearest whole percent.

c.) Notice these two (rounded) probabilities are equal. What do you think that means about the events "separated" and "male"?

Being male does not affect the probability that the person is separated.

We have a rule, in fact. You can think of it as an extension to the definition of independent events.

Definition: Two events are **independent** if $P(E \mid F) = P(E)$ [or equivalently, $P(F \mid E) = P(F)$].

expl 5: According to the U.S National Center for Health Statistics, 0.15% of deaths in the U.S. are 25-to-34-year-olds whose cause of death is cancer. In addition, 1.71% of all those who die are 25-34 years old. What is the probability that a randomly selected death is the result of cancer if the individual is known to have been 25-34 years old? Would this be unusual?

> Which formula do we need?
>
> $$P(F \mid E) = \frac{P(E \text{ and } F)}{P(E)} \quad \text{or}$$
>
> $$P(E \text{ and } F) = P(E) \cdot P(F \mid E)$$

expl 6: Suppose that a computer chip company has just shipped 5,000 computer chips to a computer company. Unfortunately, 30 of the chips are defective.

a.) Compute the probability that two randomly selected chips are defective using conditional probability.

> One chip is chosen out of 5,000 chips. Assume it is defective and then select another. What is the probability that it too is defective?

b.) Compute the probability that two randomly selected chips are defective under the assumption of independent events.

> Here, use the same probability of "defective" for each chip.

c.) What can you say about these two probabilities?

> If a small random sample is taken from a large population *without* replacement, it is reasonable to assume independence of the events. If n is less than 5% of N, we can assume independence.

expl 7: We return to the computer chips. Suppose that a computer chip company has just shipped 5,000 computer chips to a computer company. Unfortunately, 30 of the chips are defective.

The computer company randomly tests two chips out of every shipment. If both chips work, they will accept the shipment. If one or both tested chips are defective, they will reject the shipment. Use a tree diagram (with conditional probability) to organize the possibilities (and probabilities) for the two chips. What is the probability that the shipment will be rejected? Would that be unusual?

To get us started, for the two chips, what is the sample space? Use D for defective and G for good (not defective) to write out the four possibilities for these two chips.

Alternatively, we could have found

P(at least 1 defective)

= 1 – *P*(none defective)

= 1 – *P*(1^{st} not defective) · *P*(2^{nd} not defective | 1st not defective)

$$= 1 - \frac{4970}{5000} \cdot \frac{4969}{4999}$$

$$\approx 1 - .988$$

$$\approx .012$$

With or Without Replacement:

In examples 6*a* and 7, we assumed the computer chips were drawn from the population *without replacement*. That makes sense since a defective chip should not be returned to the box. Also, whether or not it was defective, we would not want to pick it right back up and test the chip again.

Some problems will set up a procedure where you have replacement. Once a card has been drawn from a deck, or a marble picked from a bag of colored marbles, or a song played from a playlist, it is entirely possible for that same card or marble or song to be selected again. This is called an experiment *with replacement*. Here we see an example.

expl 8: You listen to an album which has 12 songs. You like 4 of the songs. Answer the following questions.

a.) If your music player plays two songs at random, what is the probability that you will like both of them? Assume the first song *cannot* be played again. (This is *without replacement*.)

b.) If your music player plays two songs at random, what is the probability that you will like both of them? Assume the first song *can* be played again. (This is *with replacement*.)

c.) If your music player plays two songs at random, what is the probability that you will like neither of them? Assume the first song *can* be played again. (This is *with replacement*.)

d.) If your music player plays two songs at random, what is the probability that you will like exactly one song? Assume the first song *can* be played again. (This is *with replacement*.)

What is the relationship between the events in parts *b*, *c*, and *d*?