Math Excursions for Liberal Arts

Ben Weng

Math Excursions for Liberal Arts

Pangyen (Ben) Weng, PhD
Minneapolis, Minnesota, USA
Email: drweng.net@gmail.com
URL: http://drweng.net

Chapter 2

Probability in Your Life

2.1 What is Probability

A probability is a number used to measure how likely something would happen. This number is between 0 and 1.

- The higher probability for something, the more likely it would happen. For example, something with 0.5 probability is more likely to happen than something with 0.2 probability.
- Some people prefer to use percentages. In that way, the probability is between 0% and 100%.
- 0 (0%) means impossible to happen and 1 (100%) means absolutely happening.

Frequency As Probability

People get a sense of *how likely* something happens by observing how *frequently* it happens. This is done by *looking at a large data*, or *conducting a large experiment or survey*. For example, during the 2019–2020 NBA Season, LA Lakers player LeBron James made 239 free throws out of 343 attempts. The frequency is $239/343 = 69.7\%$, which can be viewed as the probability for him to make a free throw at his next attempt.

As for using large experiments, assume that a factory wants to know how good a production line is, so it randomly tests 1000 items produced on the line: 998 of them are good and 2 of them are defective. Then the probability for the production line to produce a good item is $998/1000 = 99.8\%$.

Why does it have to be large enough? Because when the observation is small, what you see may not be representative of the general likelihood for something to happen. This is the underlying idea of *the Law of Large Numbers*, which says that the data needs to be large enough for the observed frequency to reflect the probability in reality.

Example: M&M's Chocolates

In a big bag of M&M's chocolates, 24% blue, 20% orange, 16% green, 14% yellow, 13% red and 13% brown. If you pick one chocolate randomly from the bag, the probability of getting an orange one is 24% because if you pick a lot of them, about 24% of them would be orange.

Question. Can we say that the probability of each color is $1/6$ because there are 6 colors? Why or why not?

Random Choices

When someone makes a random choice, it means the person gives each possible options the same chance (or probability). For example, if Ben has 5 ties and he picks his ties randomly, each tie has a 20% probability to be chosen. The total probability of 100% is *evenly* divided by the 5 ties. Here are some more examples.

1. 72 employees, 30 men and 42 women, enter a random drawing for an iPad. Each person has $1/72$ chance to win. The men collectively have $30/72$ chance to win and the women have $42/72$ chance to win.

2. The Powerball is an American lottery. There are $292,201,338$ different combinations and only 1 winning combination. If the lottery is random (or fair), the probability for a ticket to have the winning combination is $1/292,201,338$, or

approximately 1 out of three hundred million.

Not All Choices Are Random

In fact, most choices in life are not random, where the outcomes are NOT equally likely. Thus we cannot give each outcome equal probability. For example, a professional football team has 3 quarterbacks, but the starting quarterback always gets to start the game (probability 100%). Another example: You are taking the Math for Liberal Arts course, which you may pass or fail. This doesn't mean that you only have 50% chance of passing it.

Too Good (or Bad) to Be True

The probability is small for something that rarely happens. When it happens, people may express disbelief: "*No way! That's too lucky (or too bad)!*"

It's possible that sometimes someone does get very lucky or unlucky, but an unlikely event can also be the result of unfairness, like cheating. The 1980 Pennsylvania Lottery Scandal is a good example: It was too good to be true, and it turned out to be the result of cheating.

Example: Workplace Discrimination

A company hires 20 new employees *with the same job description*, and 3 of them are black.

The manager gives the 3 worst shifts to the three black employees while claiming that the assignment is *randomly*, and the three black employees are just unlucky. Does this sound reasonable?

Mathematics can calculate and show that there are $1,142$ different ways to assign the three worst shifts to 20 employees, so the probability for the three black employees to randomly get these shifts is $1/1142 = 0.00088 < 0.1\%$. This probability is extremely small, making it highly unlikely to happen. This result suggests that the assignment is not random but intentional. In this case, it is a strong indication that there is racial discrimination in the work assignment.

Note that this analysis only works against the *randomness assumption*. For example, a company may have over 100 employees but the bathrooms are always cleaned by the two custodians. That doesn't mean the company discriminates against the custodians. It only means that the cleaning job is NOT randomly assigned to all employees.

Discussions

1. **120% Sure?** Sometimes people can go above 100% when comparing quantities, like *this year he is making 120% of last year's salary*. But can you be more than

100% sure about something? Explain your answer.

2. **100% Free Throw Shooter.** In the 2019/2020 NBA Season, Chicago Bull's Mark Strus made 1 free throw out of 1 attempt, which means his free throw percentage is $1/1 = 100\%$. Does this mean that he has 100% probability of making his next free throw? What's your take on this?

3. **What's riskier?** According to a study,

   - The probability of dying in a plane crash is $1/205,552$,
   - The probability of dying while cycling is $1/4,050$, and
   - The probability of dying in a car accident is $= 1/102$.

   Based on these numbers,

   a) Which one do you THINK is riskier?
   b) Which one do you FEEL is riskier?
   c) If they are different, is there an explanation?

4. **50-50 Right?** During an interview, NBA legend Shaquille "Shaq" O'Neal was asked to predict the outcome of

an upcoming basketball game. Shaq famously responded, "Well, you either win or lose... I guess it's 50-50." What's your take on this?

5. **Randomness and Equal Opportunity.** The Equal Employment Opportunity Act of 1972 address employment discrimination against African Americans and other minorities. What is your take on equal opportunity? Do you think there is a connection between equal opportunity and the random choices discussed in this math lesson? Please explain.

2.2 *The Conditional Probability*

Probability with Context

Probability can be considered in a particular context. When it is within a context, we call the probability a *conditional probability* and the context a *condition*. For example, the president's overall approval rating is a *general probability*, and the president's approval rating *in Minnesota* is a conditional probability, with the condition being the people of Minnesota. Another example: The probability of anyone being a Facebook user is a general probability, and the chance for any teenager to be a Facebook user is a conditional probability with the condition being teenagers.

Example: Nursing Students

Assume that at University X, the student population can be broken down to 6 groups. These percentages are of the entire student population, with total percentage 100%.

	Nursing	other
Male	1%	35%
Female	7%	47%
LGBTQ	1%	9%

1. If we randomly choose *a female student*, what is the probability that she is in nursing? First, the female students are $7\% + 47\% = 54\%$ of the entire university and the female nursing students are 7% of the entire university. So among all the female students, the conditional probability for a student to be in nursing is $(7\%)/(54\%) = 13\%$.

	Nursing	other
Female	$\mathbf{7\%}$	$\mathbf{47\%}$

2. If we choose among all the nursing students, what's the probability that the student is LGBTQ? The condition is nursing major, which is $(1\%) + (7\%) + (1\%) = 9\%$ of all, while 1%

of all is LGBTQ in Nursing. So the conditional probability of getting an LGBTQ student among Nursing students is $(1\%)/(9\%) = 11\%$.

	Nursing
Male	**1%**
Female	**7%**
LGBTQ	**1%**

Conditional Probability vs General Probability
When comparing the conditional probability with the general probability of an event, if the conditional probability is higher than the general probability, the condition is said to be *favorable* for the event. If the conditional probability is lower than the general probability, the condition is *unfavorable* for the event. And if the conditional probability the same as the general probability, the event is called *independent* of the condition.

In baseball we call someone a *clutch hitter* if (s)he hits better than usual when there is a scoring opportunity. So, the scoring opportunity is a favorable condition for this player. It doesn't mean this player hits at 50% or higher under the condition, just that (s)he does better than usual.

Keeping one's cholesterol levels in the normal range can lower the risk of stroke. So, for the probability of getting a stroke, keeping normal cholesterol levels is an unfavorable condition because it *reduces* the probability. (It is a favorable and healthy decision for the person, though.)

Back to the example of nursing students at University X:

	Nursing	other
Male	1%	35%
Female	7%	47%
LGBTQ	1%	9%

The general probability for a student to be in nursing is $(1\%)+(7\%)+(1\%) = 9\%$, and the conditional probability for a female student to be in nursing is 13%, which is higher than the general probability. So the being female is a favorable condition to being in nursing. In common language, we say that *you are more likely to find a nursing student among the female students.*

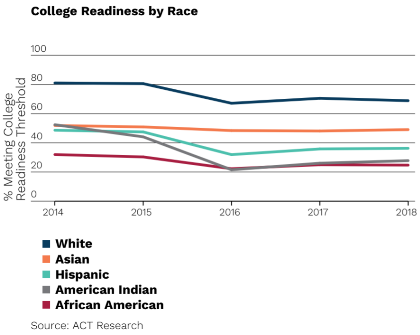
Conditional Probability and Stereotypes

According to https://nces.ed.gov, in 2015-2016, of all the US graduating college students, 18% are in the STEM fields (Science, Technology, Engineering and Math), while of all the US graduating Asian college students, 33% are in the STEM fields. So, an Asian student is more likely to have a STEM degrees *than college students in general.* On the other hand, this conditional probability is still under 50%, so we shouldn't assume the stereotype of Asian students being in STEM field because less than half of them are.

Example: Achievement Gap in Minnesota

Reference: A Statewide Crisis: Minnesota's Education Achievement Gaps by Federal Reserve Bank of Minneapolis

Minnesota constantly ranks among the highest-performing states when it comes to education quality. However, according to the article, there are persistent, significant gaps in the achievement of certain groups of Minnesota students as compared to others. Take college readiness for example, the probability of college readiness is different under different conditions (of racial identity). This means *college readiness is NOT independent of race.*



Discussions

1. **Bias in real life.** The bias towards a person or a group can sometimes be captured by a favorable/unfavorable conditional probability. Give an example and explain what happens in it.

2. **Favorable, independent or unfavorable conditions.** Find an example of each of the following. Be specific about the event and the condition.

   a) A conditional probability with a favorable condition.

   b) A conditional probability with an unfavorable condition.

   c) A conditional probability with an independent condition.

   Sample. *The event is "contracting lung cancer" and the condition is "having second-hand smoke". The condition is in favor of the event because those who have second-hand smoke has a high probability of getting lung cancer than the overall population, according to American Cancer Society.*

3. **A lot?** Someone said to you that *a lot of tall people are good at playing basketball.*

Does this mean that a really large number of tall people are good at basketball, or that being tall is favorable for playing basketball well? Give your take on this, and then find a similar example.

4. **College Readiness by Race.** Read the whole article cited in the lesson about the racial gap in college readiness. State and comment on one evidence from this report. Please refer to the page number.

2.3 COVID-19: Should Everyone Get Tested?

Disclaimer. This is a mathematical model that makes more sense at the beginning of the COVID-19 pandemic with the assumption that there is a small percentage of infected individuals and a good tracking system to control the spread. This model may or may not apply to current day United States or any other country.

COVID-19 in New York

The State of New York was hit the hardest by the pandemic at the beginning of USA's COVID-19 outbreak. On 3/22/2020, the state recorded $15,885$ total number cases, with $15,276$ active cases, 209 total death and $5,440$ daily new cases. At the time, the state (and the entire nation) lacked the capacity to test all people with COVID-like symptoms. While the nation worked on getting enough testing, some people suggested that it would be great if the government would test *everyone*. Is that a good idea? Let's analyze this with mathematics.

Background Information

In March 2020, the population of New York State was approximately 20 million, the COVID-19 infection rate was approximately 2%, and the COVID-19 testing had a 6% false-positive rate and a 1% false-negative rate.

Under these assumptions, 2% of the people in the State of New York had COVID-19, or $(20\,\text{million}) \times (2\%) = 400,000$ people. On the other hand, 98% did not have COVID-19, or $(20\,\text{million}) \times (98\%) = 19,600,000$ people.

Analysis

Suppose all 20 million residents took the test.

1. Of the $19,600,000$ people without COVID-19, 6% would get false-positives, which would be $(19,600,000) \times (6\%) = 1,176,000$ people. 94% of them would get true-negative results, which would be $(19,600,000) \times (94\%) = 18,424,000$ people.

2. Of the $400,000$ people with COVID-19, 1% would get false-negatives: $(400,000) \times (1\%) = 4,000$ people. 99% would get true-positive results:$(400,000) \times (99\%) = 396,000$ people.

Sort this information into this table:

	Testing (+)	Testing (-)
Not having COVID	$1,176,000$ (false-positive)	$18,424,000$
Having COVID	$396,000$	$4,000$ (false-negative)

Looking at all the positive testing results: In total, there would be $1,176,000 + 396,000 = 1,572,000$ of them, but the majority of them ($1,176,000$) were false-positives. If these folks went to emergency rooms or made a few panic calls, it would put unnecessary stress on the medical resources and social stability. And the $396,000$ people with COVID-19 were at the minority, which makes it hard for government to identify and follow up with them.

Remarks.

1. The big problem of universal testing is the large number of false-positive cases that dilute the real COVID-19 cases. This problem can be reduced by prioritizing testing on those who have symptoms or were exposed to confirmed cases.

2. The false-negatives tell infected people that they are okay, but their number is small enough to re-test.

3. When the infection rate is higher, there will be fewer uninfected people, which results in fewer false-positive cases. But there will be more infected people and therefore more false-negative cases.

4. This analysis shows that, even when a medical testing is fairly accurate, if someone gets a positive testing result for an unfavorable disease, they should always test again to be sure.

Discussion

In the article about COVID testing, we identified the main problem as the large number of false-positives.

1. Is there a circumstance when we won't get a large number of false-positives?

2. Does this mean that we should never try to test *everyone*?

What are your thoughts?

2.4 Video: Should Computers Run the World?

Watch this documentary on YouTube: Should Computers Run the World? - with Hannah Fry

Discussions

1. **Sensitive vs specific.** What is your take on the video's comments on the two elements of algorithm: being sensitive and being specific?

2. **Summary of the video.** State and explain at least two takeaways you get from the video.