Introduction to probability

Example: Larry Bird



Free throws (FTs) are typically shot in pairs. Below is a “contingency table” of counts summarizing Larry Bird’s first and second FT attempts during the 1980-1 and 1981-2 NBA seasons. The data source is Wardrop (*American Statistician*, 1995)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Second |   |
|   |   | Made | Missed | Total |
| First | Made | 251 | 34 | 285 |
| Missed | 48 | 5 | 53 |
|   | Total | 299 | 39 | 338 |

Interpreting the table:

* 251 first AND second FTs were both made
* 34 first FTs were made AND the second FTs were missed
* 48 first FTs were missed AND the second FTs were made
* 5 first AND second FTs were both missed
* 285 first FTs were made regardless what happened on the second attempt
* 299 second FTs were made regardless what happened on the first attempt
* 338 FT pairs were shot during these seasons

More formally,

* Let A = 1st is made.
* Let B = 2nd is made.

Often, A and B are called “events” because they describe possible outcomes. The opposite of A and B are denoted formally as  (1st FT is missed) and  (2nd FT is missed). These opposites are referred to as “complements”. Note that many books will denote the complement as Ac or A′.

Using this notation, the body of the contingency table becomes

|  |  |  |
| --- | --- | --- |
|   | B |  |
| A | 251 | 34 |
|  | 48 | 5 |

The counts can be transformed into a table of “probabilities” by dividing each numerical cell by 338.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Second |   |
|   |   | Made | Missed | Total |
| First | Made | 0.7426 | 0.1006 | 0.8432 |
| Missed | 0.1420 | 0.0148 | 0.1568 |
|   | Total | 0.8846 | 0.1154 | 1 |

These probabilities can be stated symbolically,

P(1st made) = P(A) = 0.8432

“P(A)” is read as “the probability that A occurs” or more simply “the probability of A”.

Also, P(B) = P(2nd made) = 0.8846, P() = 0.1568, and P() = 0.1154. These probabilities on the margins of the table (total column and row) are often called “marginal probabilities”. Notice that

P(A) + P() = 0.8432 + 0.1568 = 1

Why?

Also, note that P(A) = 1 – P() and P() = 1 – P(A). This result can be very useful!

In the body of the table, we can symbolically represent these probabilities as well. For example,

P(1st made and 2nd made) = P(A ∩ B) = 0.7426

where ∩ symbol is often referred to as “intersect”. Also,

P(1st made and 2nd missed) = P(A ∩ ) = 0.1006

Probabilities in the body of the table are often called “joint probabilities”.

Notice that P(A) = P(1st made)

= P(1st made ∩ 2nd made) + P(1st made ∩ 2nd missed)
= P(A ∩ B) + P(A ∩ )
= 0.7426 + 0.1006

= 0.8432

Thus, P(A∩B) + P(A∩) = P(A).

What is the probability the 1st FT or the 2nd FT is made? Symbolically,

P(1st made or 2nd made) = P(A ∪ B)

where ∪ symbol is often referred to as “union”. There are a few different ways to figure this out.

1.

|  |  |  |
| --- | --- | --- |
|  |  | Second |
|   |   | Made | Missed |
| First | Made | 0.7426 | 0.1006 |
| Missed | 0.1420 | 0.0148 |

Add the probabilities in yellow.

2. P(A∪B) = P(A) + P(B) – P(A∩B)
 = **0.8432** + **0.8846** – **0.7426** = 0.9852

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Second |   |
|   |   | Made | Missed | Total |
| First | Made | **0.7426** | 0.1006 | **0.8432** |
| Missed | 0.1420 | 0.0148 | 0.1568 |
|   | Total | **0.8846** | 0.1154 | 1 |

Why does P(A∩B) need to be subtracted out?

Note that P(A∪B) = P(A) + P(B) – P(A∩B) is sometimes referred to as the “probability of the union” or the “addition rule”. Also, note that P(A∩B) = P(A) + P(B) – P(A∪B).

3. P(A∪B) = 1 –  using the complement

What is ? Look at the contingency table from 1. above. It is the cell that is not colored in! Thus,  = P(∩). This result is also referred to as coming from one of “De Morgan’s laws”.

Then P(A∪B) = 1 – 

= 1 – P(∩)

= 1 – 0.0148 = 0.9852

Two events A and B are “mutually exclusive” events if
P(A ∩ B) = 0.

Question: Are A and  mutually exclusive?

Conditional probability

Conditional probability – The probability an event happens conditioned on another event happening.

Consider two events A and B. The probability that A occurs given that B occurred is called a conditional probability. It is denoted by P(A|B). This is read as “the probability of A GIVEN B”.

This probability can be found from

,

provided P(B) ≠ 0.

Notes:

*  implies that P(A∩B) = P(A|B)×P(B); this is often referred to as the multiplicative rule
* Another conditional probability could also be stated as P(B|A) = P(A∩B)/P(A)

Where does the formula  come from?

* Suppose the event B occurs and it had a particular probability (P(B)) of occurring. This now limits the possibility of what other events occur.
* To determine the probability that A occurs, we must examine P(A∩B) since B occurs.
* To find the probability that A occurs given the B occurred, we use P(A∩B)/P(B). This gives us the probability of A occurring out of all possibilities where B occurred.

Example: Larry Bird

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Second |   |
|   |   | Made | Missed | Total |
| First | Made | 0.7426 | 0.1006 | 0.8432 |
| Missed | 0.1420 | 0.0148 | 0.1568 |
|   | Total | 0.8846 | 0.1154 | 1 |





Notice how we are limiting ourselves by looking at the 1st missed row ONLY by conditioning on 1st missed. We then adjust the joint probabilities in the row to account for the probability of being in that row is only 0.1568.

Written in terms of

|  |  |  |
| --- | --- | --- |
|   | B |  |
| A |  |  |
|  |  |  |

P(B|) = P(∩B)/P() = 0.1420/0.1568 = 0.9057

Therefore it is still very likely that Larry Bird will make the second free throw even if the first one is missed.

Question for basketball fans: Why would this probability be important to know?

Verify on your own that P(2nd made | 1st made) = 0.8807.

Example: The Showcase Showdown on the Price is Right



On the game show, The Price is Right, three contestants are given an opportunity to spin the big wheel. The big wheel has monetary values of 5, 10, …, 100 cents on it. The contestant that is closest to a dollar (100 cents) in one or a combination of two consecutive spins, without going over, wins the game. If there is a tie, the tied players are given one additional spin with the player having the highest number in that spin winning.

See <https://youtu.be/hp1a0BHmyc0> for an example.

Coe and Butterworth (*American Statistician*, 1995) compute conditional win probabilities for the first person spinning the big wheel. The probabilities are shown in the table below.

| **First Spin****(i)** | **P(win | spin once & 1st spin=i)** | **P(win | spin twice & 1st spin=i)** |
| --- | --- | --- |
| **5** | .00034 | .20595 |
| **10** | .00121 | .20589 |
| **15** | .00285 | .20574 |
| **20** | .00540 | .20547 |
| **25** | .00906 | .20502 |
| **30** | .01415 | .20431 |
| **35** | .02101 | .20326 |
| **40** | .03009 | .20176 |
| **45** | .04190 | .19966 |
| **50** | .05704 | .19681 |
| **55** | .08346 | .19264 |
| **60** | .11829 | .18672 |
| **65** | .16319 | .17856 |
| **70** | .21563 | .16778 |
| **75** | .28416 | .15357 |
| **80** | .36818 | .13517 |
| **85** | .46990 | .11167 |
| **90** | .59169 | .08209 |
| **95** | .73606 | .04528 |
| **100** | .90567 | .00000 |

For example, P(win | spin once & 1st spin=5 cents) = 0.00034

What is the optimal strategy the first person should follow in deciding whether or not to spin twice?

What if there are more than two events?

Consider the events of A1, A2, …, Ak. Then

* P(A1∩A2) = P(A1)×P(A2|A1)
* P(A1∩A2∩A3) = P(A1)×P(A2∩A3|A1)

 = P(A1)×P(A2|A1)×P(A3|A2∩A1)

Why is P(A2∩A3|A1) = P(A2|A1)×P(A3|A2∩A1)?

Remember that P(A2∩A3) = P(A2)×P(A3|A2)

In general, P(A1∩A2∩A3∩…∩Ak)
= P(A1) × P(A2|A1) × P(A3|A1∩A2) × … ×
 P(Ak|A1∩A2∩...∩Ak-1)

Independence – Events A and B are independent if P(A|B) = P(A) or equivalently P(B|A) = P(B)

In words, this means the probability of event A is not affected by event B and vice versa.

As a result of the conditional probability equation,
P(A∩B) = P(A)×P(B) also means independence. Why?

There is another way to check for independence in a contingency table structure. Find all of the conditional probabilities given the row events (could also do for columns too) so that you have a table of conditional probabilities:

|  |  |  |
| --- | --- | --- |
|   | B |  |
| A | P(A|B) | P(A|) |
|  | P(|B) | P(|) |

If there is equality across the columns, e.g., P(A|B) =
P(|B), there is independence! Why?

What is P(A|B) under independence and what is P(|B) under independence?

One could also rewrite the table as

|  |  |  |
| --- | --- | --- |
|   | B |  |
| A | P(B|A) | P(|A) |
|  | P(B|) | P(|) |

If there is equality across the rows, e.g., P(B|A) =
P(B|), there is independence!

Example: Larry Bird

What does independence mean in this example?

 = 0.9057

P(2nd made) = 0.8846

Dependence exists - but notice how close they are.

Notes:

* Only one conditional probability needs to be checked.
* Typically, one would consider the 338 free throws here a sample from the population of all Larry Bird’s free throw attempts. This would be especially desirable if Larry Bird still was playing basketball professionally. Questions about whether this is a representative sample would need to be addressed. Assuming it was a representative sample, one may be interested in drawing an inference from the sample to the population all free throws. A chi-square hypothesis test for independence could be conducted using the data. The result is there is not sufficient evidence to prove dependency. We will discuss later how to perform this test. Also, my Categorical Data Analysis course shows how to do this test as well.

Independence is a VERY important concept to understand and we will be using this frequently in the future.

Bayes’ formula

This formula is useful to find a conditional probability, like P(A|B), if other conditional probabilities, like P(B|A), are available. When there are two events A and B, the Bayes’ formula is



Where does this formula come from? Remember that we can write the conditional probability as



Now,  to obtain the numerator. Also,



to obtain the denominator. Look back at the contingency table for the Larry Bird example if you do not understand why .

When there are more than two events, such as A1, …, Ak and B1, …, Bm, Bayes’ formula is



Bayes’ formula is used in MANY places! In particular, there is a whole way of approaching statistics, called Bayesian analysis, that relies on this type of formula. In our class, we are going to focus on specific applications of it. One of these involves the sensitivity and specificity of a diagnostic test.

Example: HIV testing

Diagnostic tests are used to determine if a person has a disease or not. These tests are not always correct. The makers of the tests try to make them very “accurate” in detecting a disease. However, this form of accuracy comes at a cost in terms of incorrectly saying that some people have the disease when they do not really have it.

Suppose a clinical trial is being conducted on a new HIV test. The test measures a number of different variables related to the presence of HIV. Using the observed results for a patient, the test decides if a person is HIV positive or not. Below are the possible outcomes:

|  |  |  |
| --- | --- | --- |
|  |  | HIV test results |
|  |  | Negative | Positive |
| HIV actual | No | Correct=True Negative | Error=False positive |
| Yes | Error=False Negative | Correct=True positive |

The test is correct when a person with HIV actually tests positive. Similarly, the test is correct when a person without HIV actually test as negative. There is the possibility the test could be incorrect. This happens when someone has HIV and the test says the person is negative. Also, this happens when someone does not have HIV and the test says the person is positive. Obviously, it is important to control the probabilities of making these errors.

Statisticians, epidemiologists, physicans, … are interested in particular probabilities associated with the contingency table above:

* Sensitivity = P(Test is Positive | Actual is Yes)

This is the probability a person tests positive, given the person actually has HIV.

* Specificity = P(Test is Negative | Actual is No)

This is the probability a person tests negative, given the person does not actually have HIV.

* Positive predictive value = P(Actual is Yes | Test is Positive)

This is the probability a person actually has HIV, given the person tests positive.

* Negative predictive value = P(Actual is No | Test is Negative)

This is the probability a person actually does not have HIV, given the person tests negative.

All of the above values are widely used to measure the accuracy of any infectious disease test. Specific information regarding the sensitivity and specificity of HIV tests is summarized in Branson et al. (2014, <http://stacks.cdc.gov/view/cdc/23447>). This reference provides observed sensitivities of 96.3% to 100% and observed specificities ranging from 99.03% to 100% based on over 30 other research articles.

Suppose an HIV test had a sensitivity of 0.993 and the specificity is 0.9999. What is the probability of making these errors:

* P(Test is Negative | Actual is Yes) = 0.007
* P(Test is Positive | Actual is No) = 0.0001

Hint: P(A|B) = 1-P(|B).

Recent HIV prevalence estimates by the CDC give a value of 0.004.

This means P(Actual is Yes) = 0.004 or 40,000 people out of 10,000,000 people have HIV.

What is the positive predictive value?

P(Actual is Yes | Test is Positive)



What is the negative predictive value?

P(Actual is No | Test is No)





Questions:

* + 1. Suppose your doctor tells you that your HIV test is positive. What should be your first question to your doctor?
		2. What is the probability that a negative test result was actually wrong?

Not all infectious disease tests are this accurate. One can examine many different scenarios by changing the sensitivity, specificity, and overall disease prevalence. For example, suppose there was a very rare disease with a prevalence of 0.000025.

P(Actual is Yes | Test is Positive)



P(Actual is No | Test is No)



Go back and answer the same questions as given above.

For more on determining sensitivity and specificity, see my lectures notes from a “multivariate statistical analysis” course that I have taught in the past. Specifically, see the discussion on receiver and operating characteristic (ROC) curves.