Statistics For Librarians: Probability and Odds

Caitlin Bakker, MLIS, AHIP-D[a]

[a]Discovery Technologies Librarian, Dr. John Archer Library, University of Regina, Regina, Saskatchewan, Canada, https://www.orcid.org/0000-0003-4154-8382,
caitlin.bakker@uregina.ca

Cite as: Bakker C. Statistics For Librarians: Probability and Odds. Hypothesis. 2025;37(1). doi:10.18060/28737

Probability refers to how likely it is that an event would occur. It is fundamental to statistics, as it tells us the likelihood of particular outcomes and allows us to calculate risk, make predictions, and measure our level of certainty.

Probabilities are always between 0 and 1, where 0 would indicate that an event is impossible while 1 would indicate that it always happens. Probabilities can be expressed as fractions, decimals, or percentages. While we may not formally calculate probabilities regularly, most people use probabilities on a daily basis, whether that's looking at a weather forecast, trying to predict the stock market, or making decisions about medical treatment. Understanding probability is crucial, both to research and to daily life.

There are different types of probability and different ways of calculating probability depending on circumstances. This column will attempt to introduce some of these concepts. Readers looking for a more in-depth resource may wish to consult Kunin et al.'s *Seeing Theory*[1], Grinstead & Snell's *Introduction to Probability*[2,] or Orloff and Kamrin's course, *Introduction to Probability and Statistics*[3].

Types of Probability

Theoretical probability can be used when no experiment has occurred, and no observations have been made. It is probability that's based on available data and logic and is essentially reflecting what is expected. For example, we can predict the probability of flipping a coin and it landing on heads, or rolling a die and it landing on a specific number. The formula for theoretical probability is:

$$P(A) = \frac{\textit{The number of ways A can happen}}{\textit{The total number of outcomes}}$$

In the above formula, A refers to our event of interest, like a coin being flipped and landing on heads. While theoretical probability can be useful, it also has limitations. Theoretical probability assumes somewhat idealized conditions, in which all outcomes have the same probability of occurring. In real life, this is often not the case. For example, the outcomes in medical trials may not all be equally likely. In these cases, we could not find the probability of an outcome without having further information or making assumptions. Scientists may be able to use their expertise, combined with information about the specific disease and treatment, to make assumptions about probability in those cases.

Unlike theoretical probability, where we calculate the probability based on what is expected to happen, in **empirical probability** we instead focus on what actually happened. The formula for empirical probability is:

$$P(A) = \frac{\textit{The number of times A occurred}}{\textit{The number of observed occurrences}}$$

Let's imagine that we have a library that has hosted a Blind Date with a Book event. After the event, the library reviews data gathered from its gate counter to determine patron foot traffic and circulation statistics to see how many of the books circulated. They found that:

1. 1000 patrons visited the library
2. 165 patrons selected one item from "Blind Date with a Book"

3. 23 patrons selected two items from “Blind Date with a Book”
4. 12 patrons selected three or more items from “Blind Date with a Book”

If we wanted to calculate the probability that a person visiting the library would also participate in the Blind Date with a Book event, we could do so as follows:

$$P(participation) = \frac{(165 + 23 + 12)}{1000} = \frac{200}{1000} = \frac{1}{5}(or\ 0.2\ or\ 20\%)$$

We can calculate that the probability a randomly selected patron would be a participant in the Blind Date with a Book event is 20%. Having this sort of data can help the library plan for future events by providing a baseline for assessment and planning.

It’s important to note that, if the probability of participation is 20%, then the probability of non-participation is 80%. This is due to the **complement rule**, which states that the sum of the probabilities of all possible outcomes must equal 1 (or 100%).

While empirical probability does not make the same assumptions that theoretical probability does, it has its own limitations, particularly when working with small data sets. For example, if we only observed five patrons entering the library, three of whom participated in the Blind Date with a Book event, you might conclude that the probability of participation is 60%.

As additional data is gathered, empirical probability will approach theoretical probability. This is known as the **law of large numbers**, which states that if you repeat an experiment a large number of times, the result obtained should be close to the expected value. One of the most famous examples of this was conducted by mathematician John Kerrich, who flipped a coin 10,000 times, recording whether it landed on heads or tails each time[4]. The first three flips landed on tails. At that point, one would erroneously calculate that the probability of landing on tails was 100%. However, by the end of the 10,000 flips, Kerrich had recorded 5,067 heads and 4,933 tails. The theoretical probability of landing on tails (49.33%) closely mirrors the theoretical probability of landing on heads (50%).

Dependence

While we may want to calculate the probability of an individual event occurring, there are also times when we will need to calculate the probability of multiple events occurring. To understand how to calculate these probabilities, we need to know how the events relate to each other.

Independent events are those in which the probability of one event does not change the probability of the other. For example, we know that the probability that a library patron would participate in the Blind Date with a Book event is 20%. Based on previously gathered data, we also know that the probability that a library patron would access WiFi is 95%. These two events are independent, since accessing WiFi (or not) has no impact on participating in Blind Date with a Book (or not). When calculating the probability for independent events, we use the **multiplication rule**:

$$P(A\ and\ B) = P(A)\ x\ P(B)$$

In this case, if we want to know the probability that a patron would both access the WiFi and participate in the Blind Date with a Book event, we can do so as follows:

$$P(\textit{WiFi and participation}) = P(\textit{WiFi})\ x\ P(\textit{participation}) = \frac{95}{100} x \frac{1}{5} = \frac{95}{500} = 0.19$$

Ultimately, we find that the probability of the patron participating in Blind Date with a Book and accessing the WiFi is 19%.

Dependent events are those that are affected by the outcome of other events. For example, at the Blind Date with a Book event, the first patron comes into the event and selects a biography. If that item isn't replaced with another biography, the original probabilities change.

Type of Book	Original Probability	After the First Patron
Mystery	10/25 = 0.4 = 40%	10/24 = 0.42 = 42%
Science Fiction	10/25 = 0.4 = 40%	10/24 = 0.42 = 42%
Biography	5/25 = 0.2 = 20%	4/24 = 0.17 = 17%

The previous event of a biography being selected impacts the probability of the other items being selected, making these dependent events.

Some dependent events are also what are called mutually exclusive events. **Mutually exclusive events** are events that cannot happen at the same time. For example, a patron takes out a single book as part of the Blind Date with a Book event and we want to know the probability that they would pick either a science fiction book or a biography. These are mutually exclusive events, since one book (presumably) couldn't be both. With independent events, we calculate the probability that both happened (you can think of these as "AND" events), while with mutually exclusive events, we can calculate the probability that either happened ("OR" events). We can do this using the **addition rule**:

$$P(A\ or\ B) = P(A) + P(B)$$

In this case, if we want to calculate the probability that a randomly selected patron would choose either a science fiction novel or a biography, we could do so as follows:

$$P(\textit{science fiction or biography}) = \frac{10}{25} + \frac{5}{25} = \frac{15}{25} = 0.6$$

We find that there is a 60% chance that a randomly selected patron would choose either a science fiction novel or a biography.

Odds

Odds and probability are closely related, but not synonymous. They are both ways of expressing the likelihood of an event. We know that the probability of a patron participating in the Blind Date with a Book event is 20%. But, if you were to say that the odds of a patron participating were 1/5, that would be incorrect.

While empirical probability is the likelihood of an event occurring compared to the possible outcomes, odds are the probability of an event occurring dividing by the probability of it not occurring. We can calculate the odds for an event using the probability of that event:

$$Odds(A) = \frac{P(A)}{1 - P(A)}$$

Unlike probability, odds aren't restricted to a 0 to 1 scale. Because odds can be greater than 1, they are typically presented as fractions or decimals, rather than percentages. If the probability of participation is 20%, we could calculate the odds as follows:

$$Odds(participation) = \frac{0.2}{1 - 0.2} = \frac{0.2}{0.8} = \frac{1}{4}$$

While it may strike you as strange that odds and probability are different, you can think of them as two different ways of expressing likelihood. With probability, we can say that there is a 20% chance that a patron would have participated in the event. With odds, we can say that for every patron that would have participated, four would not have. While both are communicating likelihood, they allow us to emphasize different things and can ultimately be used in different, more advanced calculations.

References

1. Kunin D, Guo J, Devlin TD, Xiang D. Seeing theory. [Cited 2025 Jan 11]. Available from: http://seeingtheory.io

2. Grinstead CM, Snell JL. Introduction to probability [Internet]. American Mathematical Society; 1997. [Cited 2025 Jan 11]. Available from: https://open.umn.edu/opentextbooks/textbooks/21

3. Orloff J, Kamrin JF. Introduction to probability and statistics. MIT OpenCourseWare; 2022. [Cited 2025 Jan 11]. Available from: https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/pages/syllabus/

4. Kerrich JE. *An experimental introduction to the theory of probability*. Einar Munksgaard; 1946.