

Review of Probability Rules

Cheng Peng

Contents

| | | |
|----------|---|----------|
| 1 | Introduction | 1 |
| 2 | Definitions of Probability | 1 |
| 2.1 | Experiments, Sample Space, and Event | 1 |
| 2.2 | Calculation of Probability | 2 |
| 3 | General Addition Rule | 2 |
| 3.1 | Additive Rule | 3 |
| 3.2 | Special Cases | 4 |
| 4 | Multiplicative Rule and Independence | 5 |
| 4.1 | Conditional Probability | 5 |
| 4.2 | General Multiplication Rule | 6 |
| 4.3 | Checking Independence Using Multiplicative Rule | 6 |
| 4.4 | Bayes' Theorem | 7 |
| 4.5 | Errors in Binary Decision | 8 |
| 5 | Appendix: Calculus Review Topics | 9 |
| 5.1 | Notation, Domain, and Range | 9 |
| 5.2 | Operations of Functions | 9 |
| 5.3 | Composite Functions | 10 |
| 5.4 | Inverse Function | 10 |
| 5.5 | Practice Exercises | 11 |

1 Introduction

This note reviews the basic concepts of probability and important probability rules. At the very end, we briefly review the concepts of functions.

2 Definitions of Probability

In general, probability is a measure of the likelihood of some outcome. We use it not to describe what will happen in one particular event, but rather, what the long-term proportion of that outcome will occur.

2.1 Experiments, Sample Space, and Event

In probability, an **experiment** is any process where the results are uncertain. We call the **sample space**, S , the collection of all possible outcomes. A **probability event** is any collection of outcomes from the experiment.

Example 1: Suppose we have a family with three children, and we consider the sex of those three children.

Solution: If we let B represent a boy and G represent a girl, here is the sample space:

$$S = \{BBB, BBG, BGB, GBB, BGG, GBG, GGB, GGG\}$$

There are thus 8 outcomes in this experiment. One possible event might be:

E = the family has exactly two girls = $\{BGG, GBG, GGB\}$

Unusual Events: Typically, we say that an event with a probability less than 5% is **unusual**, but this isn't a hard cutoff. It depends on the context

Suppose we're planning on making a decision one way unless the probability of a particularly "unusual" event is too high. For example, suppose we're planning a picnic on a nice summer day. If the risk of a rain shower isn't too high, we'll plan on the picnic. In this case, we might set our cutoff at 20% - anything less than that is too unusual (or unlikely) to happen, so we'll risk it.

2.2 Calculation of Probability

There are two primary methods for calculating probabilities

Frequency Approach: The first is to simply look at what has happened in the past and assume the probability is the same as the **relative frequency** of that particular outcome. This is called the empirical probability of that event.

$$P(E) \approx \text{relative frequency of } E = (\text{frequency of } E) / (\text{total number of trials})$$

Example 2: In a statistics class, 38 out of 50 students earned a grade of B or better (denoting this event as an event E). By the above definition, we have $P(E) \approx 38/50 = 0.76$.

Classic Approach: The second primary method for calculating probabilities is the **classical method**. The key to this method is to assume that all outcomes are equally likely.

$$P(E) = (\text{number of ways } E \text{ can occur}) / (\text{total number of possible outcomes}) = N(E) / N(S)$$

Example 3: Let's consider again the probability experiment of rolling two fair six-sided dice. Let the event E = the sum of the two dice is 6. Find $P(E)$

Solution We first list the elements in the sample space.

Since there are five ways to get a sum of 6, $P(E) = 5/36 \approx 0.14$.

3 General Addition Rule

John Venn developed a great way to visualize sets. This visual is usually called the Venn diagram. Because events are sets of outcomes of an experiment, it works well to visualize probability as well.

Here's an example of a Venn diagram showing two disjoint outcomes, E and F .



Figure 1: Figure1. Sample space of the experiment of rolling two dice.

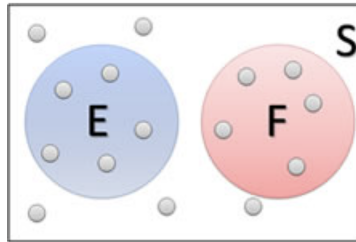


Figure 2: Figure 2. Venn diagram of two events in a sample space.

3.1 Additive Rule

Let E and F be defined in a sample space S of an experiment. The General Addition Rule is given by

$$P(E \text{ or } F) = P(E) + P(F) - P(E \text{ and } F)$$

The above general additive is depicted in the following Venn diagram

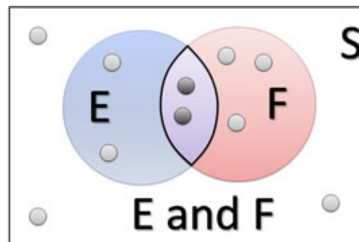


Figure 3: Figure 3. Venn diagram representation of additive rule.

Some facts about the additive rule:

- $0 \leq P(E \text{ or } F) \leq 1$.
- $P(E \text{ or } F) \geq P(E)$ **AND** $P(E \text{ or } F) \geq P(F)$.
- $P(E \text{ and } F) \leq P(E)$ **AND** $P(E \text{ and } F) \leq P(F)$.

Example 4: Consider a standard playing card. Suppose we draw one card at random from the deck and define the following events:

E = the card drawn is an ace

F = the card drawn is a king

Use these definitions to find $P(E \text{ or } F)$.

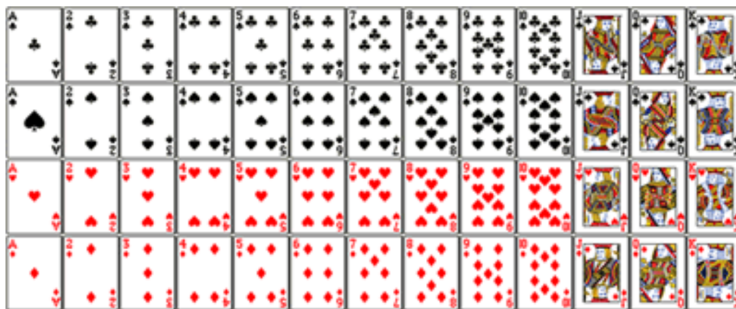


Figure 4: Figure 4. additive rule - playing card example.

Solution: Since E and F have no outcomes in common, we can use the Addition Rule for Disjoint Events:

$$P(E \text{ or } F) = P(E) + P(F) = 4/52 + 4/52 = 8/52 = 2/13.$$

3.2 Special Cases

Case #1: If the two events are disjoint (mutually exclusive), the general additive rule is reduced to

$$P(E \text{ or } F) = P(E) + P(F)$$

Example 5: Suppose we have a family with three children, and we consider the sex of those three children. If we let B represent a boy and G represent a girl, here is the sample space:

$$S = \{BBB, BBG, BGB, GBB, BGG, GBG, GGB, GGG\}$$

let's define the following events:

E = the family has exactly two boys

F = the family has exactly one boy

Describe the event **E or F** and find its probability.

Solution: **E or F** is the event that the family has either one or two boys.

Clearly, both of these events can't occur at the same time, so they are disjoint. The probability of the family having either one or two boys is then:

$$P(E \text{ or } F) = P(E) + P(F) = 3/8 + 3/8 = 6/8 = 3/4.$$

Case #2 - The Complement Rule.

The complement of E , denoted by E^c , is all outcomes in the sample space that are not in E .

So essentially, the complement of E is everything but the outcomes in E . In fact, some texts actually write it as "not E ". The following Venn diagram illustrates the event and its complement.

Since $P(S) = 1$. Clearly, E and E^c are disjoint, so $P(E \text{ or } E^c) = P(E) + P(E^c)$. Combining those two facts, we get the following complementary Rule

$$P(E) + P(E^c) = 1$$

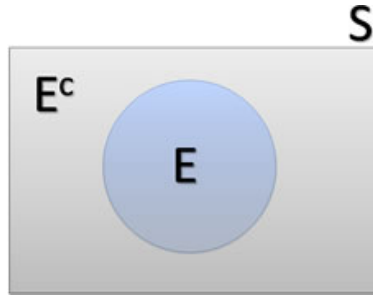


Figure 5: Figure 5. Complementary events.

4 Multiplicative Rule and Independence

Two events E and F are **independent** if the occurrence of event E does not affect the probability of event F.

Disjoint vs. Independent: It is very common for students to confuse the concepts of disjoint (mutually exclusive) events with independent events. Recall from the last section:

Two events are disjoint if they have no outcomes in common. (Also commonly known as mutually exclusive events.)

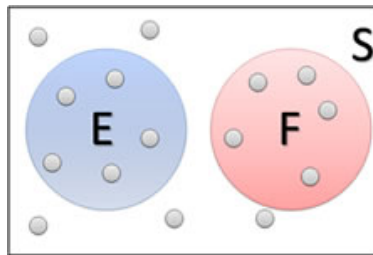


Figure 6: Figure 6. Disjoint and independent events.

From the above figure, we can see very clearly that if event E occurs (that is, the outcome - gray circle- is in event E), it cannot possibly be in event F. So E and F are mutually exclusive.

4.1 Conditional Probability

The notation $P(F|E)$ is read **the probability of F given E** and represents the probability that event F occurs, given that event E has already occurred. Mathematically, the probability of F giving E is given by

$$P(F|E) = \frac{P(E \text{ and } F)}{P(E)}$$

Example 6: The following Venn diagram illustrates conditional probability.

Based on the above Venn diagram, find $P(E|F)$.

Solution: $P(E|F)$ is the probability that event E occurs given that event F has already occurred. Let's change the image a bit. Since F has occurred, we can focus on just those **outcomes** in F. And then of those, we want the probability that the **outcome** is in E. Since there are 2 of those in E, and 5 total,

$$P(E|F) = \frac{P(E \text{ and } F)}{P(F)} = 2/5.$$

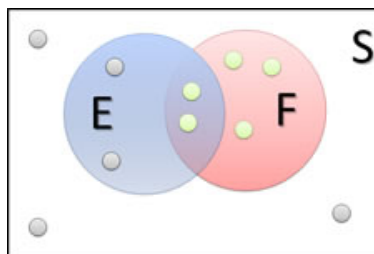


Figure 7: Figure 7. Conditional probability.

4.2 General Multiplication Rule

The probability that two events E and F both occur can be obtained from the definition of conditional probability in the following form

$$P(E \text{ and } F) = P(E) \times P(F|E)$$

The general multiplicative rule essentially expresses the joint probability as the product of a conditional probability and the probability of the conditioning event.

Example 7. Let's try a new probability experiment. This time, consider a bag of marbles, containing 10 red, 20 blue, and 15 green marbles. Suppose that two marbles are drawn without replacement. (The first marble is not put back in the bag before drawing the second.)

What is the probability that both marbles drawn are red?

Solution: Let's define a couple of events:

E = first marble is red

F = second marble is red

We want $P(E \text{ and } F)$. Using the General Multiplication Rule, we see

$$P(E \text{ and } F) = P(E) \times P(F|E) = (10/45) \times (9/44) \approx 0.0455.$$

4.3 Checking Independence Using Multiplicative Rule

Two events E and F are independent if event E 's occurrence does not affect event F 's probability.

Looking at this in terms of conditional probability, if the occurrence of E doesn't affect the probability of F , then $P(F|E) = P(F)$. This is a good way to test for independence. In fact, we can redefine independence using this concept.

Two events E and F are independent if $P(F|E) = P(F)$.

Example 8: Consider a survey given to 52 students in an introductory statistics class, with the following responses to the statement **I enjoy statistics**.

| | Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree | Total |
|--------|----------------|-------|---------|----------|-------------------|-------|
| Male | 6 | 10 | 3 | 0 | 0 | 19 |
| Female | 8 | 14 | 7 | 4 | 0 | 33 |
| Total | 14 | 24 | 10 | 4 | 0 | 52 |

Suppose a student is selected at random from those surveyed and we define the events E and F as follows:

E = student selected is female

F = student enjoys math

Are events E and F independent?

Solution: To answer this, we'll need to see if $P(F|E) = P(F)$.

$$P(F) = 38/52 \approx 0.7308$$

$$P(F|E) = 22/33 \approx 0.6667$$

Since $P(F) \neq P(F|E)$, events E and F are dependent.

4.4 Bayes' Theorem

Bayes' Theorem (also called Bayes Rule) is a way of finding a (posterior) probability when we know certain other (prior) probabilities.

The formula is:

$$P(A|B) = P(A) \times P(B|A)$$

Bayes Rule tells us how often A happens given that B happens - This conditional probability is denoted by $P(A|B)$.

When we need to know in order to calculate $P(A|B)$?

1. how often B happens given that A happens, denoted by $P(B|A)$.
2. how likely A is on its own, denoted by $P(A)$ or how likely B is on its own, denoted by $P(B)$

Example 9: According to a research of publicly-traded companies, 60% of the companies that increased their share price by **more than** 5% in the last three years replaced their CEOs during the period.

At the same time, only 35% of the companies that did not increase their share price by **more than** 5% in the same period replaced their CEOs. Knowing that the probability that the stock prices grow by **more than** 5% is 4%, find the probability that the shares of a company that fires its CEO will increase by **more than** 5%.

Solution: Before finding the probabilities, we must first define the notation of the probabilities.

First of all, we define two events A and B in the following:

A = Stock price increases by 5%

B = CEO was replaced.

With the definitions of the above two events, we need to find the following probabilities in order to use the Bayes Theorem.

- $P(A)$ – the probability that the stock price **increases by 5%**, that is
- $P(B)$ – the probability that the CEO is replaced
- $P(A|B)$ – the probability of the stock price **increases by 5% given that** the CEO has been replaced
- $P(B|A)$ – the probability of the CEO replacement **given that** the stock price has **increased by 5%**.

We next find the above probabilities based on the given conditions:

The first sentence: 60% of the companies that increased their share price by **more than** 5% [A] in the last three years replaced their CEOs during the period [B] . This means $P(B|A) = 0.6$.

The second sentence: only 35% of the companies that did not increase their share price by **more than** 5% in the same period replaced their CEOs. That is, $P(B) = 0.35$.

The third sentence: Knowing that the probability that the stock prices grow by **more than** 5% is 4%. That is, $P(A) = 0.04$.

The Question: find the probability that the shares of a company that fires its CEO will increase by **more than 5%** is equivalent to $P(A|B) = ?$

Using the Bayes rule, we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) \times P(A)}{P(B)} = \frac{0.6 \times 0.04}{0.35} \approx 0.06857143.$$

Example 9: Allergy or Not? - Hunter says she is itchy. There is a test for Allergy to Cats, but this test is not always right:

- For people that really do have the allergy, the test says **Yes** 80% of the time
- For people that do not have the allergy, the test says **Yes** 10% of the time (**false positive**)
- If 1% of the population has the allergy, and Hunter's test says **Yes**, what are the chances that Hunter really has the allergy?

We want to know the chance of having the allergy when the test says **Yes**, written $P(\text{Allergy}|\text{Yes})$

Solution: Let's get our formula:

$$P(\text{Allergy}|\text{Yes}) = \frac{P(\text{Allergy})P(\text{Yes}|\text{Allergy})}{P(\text{Yes})}$$

Note that we are given that $P(\text{Allergy}) = 0.01$

$P(\text{Yes}|\text{Allergy})$ is the probability of the test saying **Yes** for people with allergy = 80%. That is, $P(\text{Yes}|\text{Allergy}) = 0.8$

We need to find $P(\text{Yes})$ is the probability of the test saying **Yes** (to anyone) = ??%

We can calculate it by adding up those with, and those without the allergy:

- 1% have the allergy, and the test says **Yes** to 80% of them
- 99% do not have the allergy and the test says **Yes** to 10% of them

Then $P(\text{Yes}) = 1\% \times 80\% + 99\% \times 10\% = 10.7\%$, which means that about 10.7% of the population will get a **Yes** result. Therefore,

$$P(\text{Allergy}|\text{Yes}) = 1\% \times 80\% / 10.7\% = 7.48\%$$

That is, $P(\text{Allergy}|\text{Yes}) \approx 7\%$.

4.5 Errors in Binary Decision

A binary decision is a choice between two alternatives, for instance between taking some specific action or not taking it. For example, graduate admission decision: Accept or decline (assuming no waiting-list); physician's diagnostic decision: diseased or disease-free, hire manager's decision on a job applicant: offer the job or do not offer the job, etc.

In practice applications, all decisions are based on certain evidence defined based on certain criteria.

One of the famous uses for the Bayes Theorem is **False Positives and False Negatives**.

We use PCR test result of COVID-19 as an example to explain

False Positives = A person's PCR is positive but was not infected with COVID-19.

False Negatives = A person's PCR is negative but was infected with COVID-19.

Acknowledgment

'The content of this note is based on <https://faculty.elgin.edu/dkernler/statistics/ch05/>

5 Appendix: Calculus Review Topics

This appendix reviews the basics of single variable functions.

5.1 Notation, Domain, and Range

- **Notation:** Functions can also be written in the form of $f(x)$, pronounced **f of x**. When someone says **y is a function of x**, it means that the value of y is determined by the value of x. Here, y is the dependent variable and x is the independent variable. $f(x)$ is just the shortened form of **function of x**.

Example 10: A function can be denoted in different forms. The following are the two forms of the same function.

$$y = x^2 + 2 \text{ and } f(x) = x^2 + 2$$

- **Domain:** The domain of a function is all the possible values of the independent variable, x , for which y is defined.
- **Range:** The range of a function is all the possible values of the dependent variable y .

Example 11: The given function $f(x) = x^2$ has a domain of all real numbers (x can be anything) and a range that is greater than or equal to zero.

5.2 Operations of Functions

We can define new functions based on the existing functions using basic arithmetic operations.

Example 12 Let $f(x) = x^2 + 1$ and $g(x) = \sqrt{x+1}$. We can define the following new functions.

- $h_1(x) = f(x) + g(x) = x^2 + \sqrt{x+1} + 1$.
- $h_2(x) = f(x) - g(x) = x^2 + 1 - \sqrt{x+1}$.
- $h_3(x) = f(x) \times g(x) = (x^2 + 1) \times \sqrt{x+1}$.
- $h_4(x) = f(x)/g(x) = (x^2 + 1)/\sqrt{x+1}$.

5.3 Composite Functions

A **composite function** is a function created when one function is used as the input value for another function. Essentially, the output of the inner function (the function used as the input value) becomes the input of the outer function (the resulting value).

- **Notation:** For the functions $f(x)$ and $g(x)$, when $g(x)$ is used as the **input** of $f(x)$, the composite function is written as:

$$(f \circ g)(x)$$

The \circ symbol denotes a composite function - it looks similar to the multiplication symbol, \cdot , but does not mean the same thing. $(f \circ g)(x)$ is the same thing as $f(g(x))$.

$(f \circ g)(x)$ is not the same thing as $(g \circ f)(x)$. $(g \circ f)(x)$ is the same thing as $g(f(x))$, which will often be different than $f(g(x))$.

You can use composite functions to check if two functions are inverses of each other because they will follow the rule:

$$(f \circ g)(x) = (g \circ f)(x) = x$$

We can find the composite of two functions by replacing every x in the outer function with the equation for the inner function (the input).

Example 13: Given: $f(x) = 4x^2 + 3$ and $g(x) = 2x + 1$.

$$(f \circ g)(x) = f(g(x)) = 4(2x + 1)^2 + 3 = 4(4x^2 + 4x + 1) + 3 = 16x^2 + 16x + 7$$

$$(g \circ f)(x) = g(f(x)) = 2(4x^2 + 3) + 1 = 8x^2 + 6 + 1$$

5.4 Inverse Function

Inverse functions are a way to “undo” a function. In the original function, plugging in x gives back y , but in the inverse function, plugging in y (as the input) gives back x (as the output). If a function were to contain the point $(3,5)$, its inverse would contain the point $(5,3)$.

If the original function is $f(x)$, then its inverse $f^{-1}(x)$ is **not the same as** $1/f(x)$.

To find the inverse of a function, we need to do the opposite of what the original function does to x . To be more specific, we can use the following steps:

1. In the original equation, replace $f(x)$ with y ;
2. Replace every x in the original equation with a y and every y in the original equation with an x ;
3. Solve for y ;
4. Change y to $f^{-1}(x)$;

Example 14: Find the inverse function of $f(x) = 5\sqrt{x+4} - 3$.

Solution. We follow the above steps to find the inverse function.

1. In the original equation, replace $f(x)$ with y :

$$f(x) = 5\sqrt{x+4} - 3$$

to

$$y = 5\sqrt{x+4} - 3$$

2. Replace every x in the original equation with a y and every y in the original equation with an x

Note: It is much easier to find the inverse of functions that have only one x term. For functions that have more than one x term, you will need to solve for y by moving all y terms to one side of the equation and factoring out a y .

Example

$$y = 5\sqrt{x+4} - 3$$

$$x = 5\sqrt{y+4} - 3$$

3. Solve for y :

$$(1) x = 5\sqrt{y+4} - 3$$

$$(2) x + 3 = 5\sqrt{y+4}$$

$$(3) \frac{x+3}{5} = \sqrt{y+4}$$

$$(4) \left(\frac{x+3}{5}\right)^2 = y+4$$

$$(5) y = \left(\frac{x+3}{5}\right)^2 - 4$$

4. Change y to $f^{-1}(x)$:

$$f^{-1}(x) = \left(\frac{x+3}{5}\right)^2 - 4$$

5.5 Practice Exercises

The following link <https://github.com/pengdsci/WCUSTA504/raw/main/topic01/FunctionsWorksheet.pdf> gives a list of exercises that reflect the concepts reviewed in this appendix.