Econ 5023: Statistics for Decision Making

Univariate Statistics (I): Brief History and Basic Probability Theory

Le Wang

2018-09-24

Road Map

1. Brief History
2. Basic Probability Theory
   - Philosophical Debate about the Interpretations of Probability
   - Basic Definitions and Their implementations in R
   - A Naive way to Assign Probability
   - Counting techniques and Implementation in R
   - Three Axioms for Probility and Their Implications
3. Common Fallacies or Issues in Practice

Reading

1. Chapter 5 in Lind et al. (basics on pages 131-145; counting on pages 160-166)
2. Penn State Stat 414: Probability Theory and Mathematical Statistics Lessons 2 and 3.
3. Introduction to Probability: Chapter 2 Counting

A bit of History:

Question: Could I have thought that?

A bit of History:

Clockwork Universe $\Longrightarrow$ **Error View** $\Longrightarrow$ **Statistics View**

Frequentist vs Bayesian View

Frequentist: probability represents the *limit* of relative frequency, defined as the ratio between the number of times the event occurs and the number of trials (defined below), in **repeated trials** under **the same conditions**.

Bayesian: probability is a measure of one's **subjective belief** about the likelihood of an event occurring based on whatever information is available.

Note that: all bolded words represent the sources of controversies for each view.

Frequentist vs Bayesian View

Frequentist: probability represents the *limit* of relative frequency, defined as the ratio between the number of times the event occurs and the number of trials (defined below), in **repeated trials** under **the same conditions**.

1. The probability of winning a game is 50%. **Issue:** Every game is different, not the same conditions.

2. The probability of Trump winning the 2016 US. presidential election is xx%. **Issue:** The 2016 election occurred only once. What do you mean by **repeated trials**?

Frequentist vs Bayesian View

Bayesian: probability is a measure of one's **subjective belief** about the likelihood of an event occurring based on whatever information is available.

Issue if scientists have identical sets of empirical evidence, they should arrive at the same conclusion rahther than reporting different probabilities of the same event!

Under the Bayesian framework, probability simply becomes a tool to describe one's belief system. (Bayesian Response: everyone's subjective, we should explicitly recognize the role of such subjectivity.)

Frequentist vs Bayesian View

Regardless of the controversy about its **interpretation**, probability was established as a mathematical theory by Andrey Kolmogorov.

Both views adopt the mathematical theory and the disagreement is about **interpretation** and is NOT **mathematical**.

Basic Definitions and Three Axioms

The definitions of probability requires the following concepts: $(\Omega, \mathcal{F}, P)$

1. **experiment**: an action or a set of actions (or a process) that produce **stochastic** or **uncertain** events of interest.
2. **outcome**: a particular result of an experiemnt
3. **sample space**: a set of all possible outcomes of the experiment, typically denoted by Ω
4. **event**: a subset of the sample space

Loosely speaking, a **set** is a collection of distinct objects.

Basic Definitions and Axioms (Example)

Experiment	Roll a Die
Outcome	e.g., 1
Sample space	1,2,3,4,5,6

Sample Space in R

```r
library("prob")
tosscoin(1)
```

```
##   toss1
## 1     H
## 2     T
```

```r
tosscoin(3)
```

```
##   toss1 toss2 toss3
## 1     H     H     H
## 2     T     H     H
## 3     H     T     H
## 4     T     T     H
## 5     H     H     T
## 6     T     H     T
## 7     H     T     T
## 8     T     T     T
```

```r
rolldie(1)
```

```
##   X1
## 1  1
## 2  2
## 3  3
## 4  4
## 5  5
## 6  6
```

```r
head(cards(), n = 15)
```

```
##    rank    suit
## 1     2    Club
## 2     3    Club
## 3     4    Club
## 4     5    Club
## 5     6    Club
## 6     7    Club
## 7     8    Club
## 8     9    Club
## 9    10    Club
## 10    J    Club
## 11    Q    Club
## 12    K    Club
## 13    A    Club
## 14    2 Diamond
## 15    3 Diamond
```

Sample Space and Events

Experiment	Roll a Die
Outcome	e.g., 1
Sample space	1,2,3,4,5,6
Event	A number greater than 4, including 5 and 6 (a subset)
Event	A number has three letters, (one, two, six)

Sample Space and Events (Graphical Illustration)

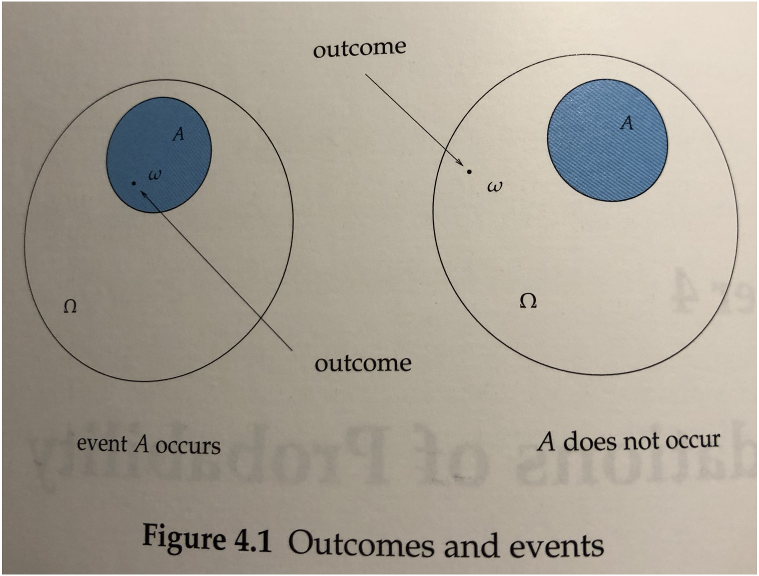


Figure 1

Subset or Events in R

```r
S <- cards()
head(S, n = 20)
```

```
##    rank    suit
## 1     2    Club
## 2     3    Club
## 3     4    Club
## 4     5    Club
## 5     6    Club
## 6     7    Club
## 7     8    Club
## 8     9    Club
## 9    10    Club
## 10    J    Club
## 11    Q    Club
## 12    K    Club
## 13    A    Club
## 14    2 Diamond
## 15    3 Diamond
## 16    4 Diamond
## 17    5 Diamond
## 18    6 Diamond
## 19    7 Diamond
## 20    8 Diamond
```

```r
subset(S, suit == "Heart")
```

```
##    rank  suit
## 27    2 Heart
## 28    3 Heart
## 29    4 Heart
## 30    5 Heart
## 31    6 Heart
## 32    7 Heart
## 33    8 Heart
## 34    9 Heart
## 35   10 Heart
## 36    J Heart
## 37    Q Heart
## 38    K Heart
## 39    A Heart
```

```r
subset(S, rank %in% 7:9)
```

```
##    rank    suit
## 6     7    Club
## 7     8    Club
## 8     9    Club
## 19    7 Diamond
## 20    8 Diamond
## 21    9 Diamond
## 32    7   Heart
## 33    8   Heart
## 34    9   Heart
## 45    7   Spade
## 46    8   Spade
## 47    9   Spade
```

```r
head(rolldie(3), n = 4)
```

```
##   X1 X2 X3
## 1  1  1  1
## 2  2  1  1
## 3  3  1  1
## 4  4  1  1
```

```r
subset(rolldie(3), X1 + X2 + X3 > 16)
```

```
##     X1 X2 X3
## 180  6  6  5
## 210  6  5  6
## 215  5  6  6
## 216  6  6  6
```

Other Set operations in R

You can use

1. **Union**: `union(A, B)`
2. **Intersection**: `intersect(A, B)`
3. **Set Difference**: `setdiff(A, B)`

A Naive Way to Assign Probability (or the Classical Approach)

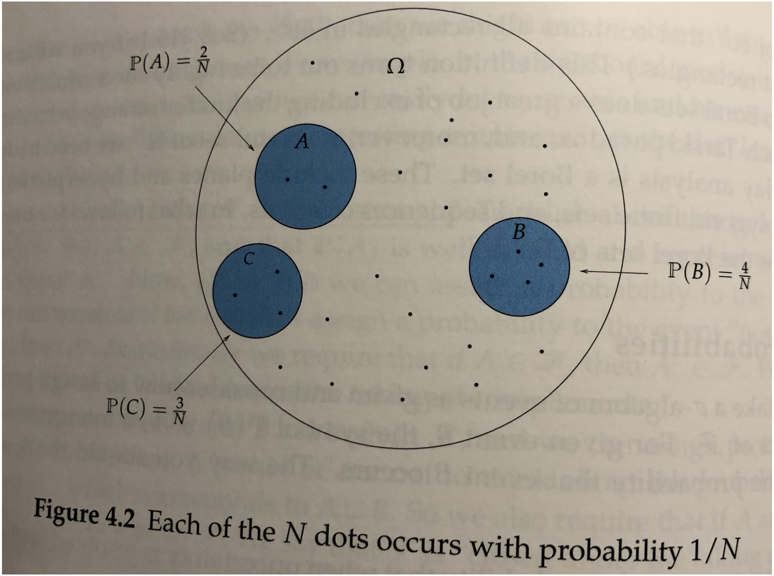
As long as the outcomes in the sample space are equally likely, the probability of event A is

$$P(A) = \frac{|A|}{|S|}$$

where $|A| :=$ the number of elements in set A. $|S|$ is the number of elements in set S, the sample space.

```r
A <- subset(rolldie(3), X1 + X2 + X3 > 16)
S <- rolldie(3)
Pr.A <- nrow(A)/nrow(S)
Pr.A
```

```
## [1] 0.01851852
```



$\mathbb{P}(A) = \frac{2}{N}$
Ω
A
B
C
$\mathbb{P}(B) = \frac{4}{N}$
$\mathbb{P}(C) = \frac{3}{N}$
Figure 4.2 Each of the N dots occurs with probability $1/N$

Figure 2

Naive Probability, Sample Space, and Counting

As you can see, the naive definition of probability depends on the number of elements in the sample space. It is therefore important to be able to count this number (especially when built-in commands such as `rolldie()` or `tosscoin()` are not available).

Counting techniques are ways to achieve this goal.

Notation: factorial

"n factorial" is written as $n\,!$, defined as $n \times (n-1) \times (n-2) \cdots \times 1$.

For example

"4 factorial" $(4\,! = 4 \times 3 \times 2 \times 1)$

```
factorial(4)
```

```
## [1] 24
```

	without Replacement	with Replacement
Ordered	(1) $\frac{N\,!}{(N-K)\,!} = N \times (N-1) \times \ldots (N-K+1)$	(3)$N^K = N \times N \times \cdots \times N$
Unordered	(2) $\binom{N}{K} = \frac{N\,!}{(N-K)\,!K\,!}$	(4) $\binom{N-1+K}{K}$

In the book: (1): Permutation Formula (2): Combination Formula (Get rid of duplications) (4): A result of Multiplication Formula

Examples

Ordered with replacement: Flip a coint 7 times. How many possible outcomes? $N = 2, K = 7$

$$2^7$$

	without Replacement	with Replacement
Ordered	(1) $\frac{N!}{(N-K)!} =$ $N \times (N-1) \times \ldots (N-K+1)$	(3)$N^K = N \times N \times \cdots \times N$
Unordered	(2) $\binom{N}{K} = \frac{N!}{(N-K)!K!}$	(4) $\binom{N-1+K}{K}$

Examples

Unordered with replacement: You rent 5 movies to watch over the span of two nights and decide to watch 2 on the first night. How many possible choices do you have? $N = 5, K = 2$

$$\binom{5}{2} = \frac{5\,!}{(5-2)\,!2\,!} = 10$$

	without Replacement	with Replacement
Ordered	(1) $\frac{N\,!}{(N-K)\,!} =$ $N \times (N-1) \times \ldots (N-K+1)$	(3)$N^K = N \times N \times \cdots \times N$
Unordered	(2) $\binom{N}{K} = \frac{N\,!}{(N-K)\,!K\,!}$	(4) $\binom{N-1+K}{K}$

Counting in R

It is tedious to remember these formulae and calculate with them. We could do it in `R`

```r
choose(5, 2)
```

```
## [1] 10
```

In fact, there is a command called `nsamp()` for all four cases

```
library("prob")
```

```
nsamp(n = 3, k = 2, replace = TRUE, ordered = TRUE)
```

```
## [1] 9
```

```
nsamp(n = 3, k = 2, replace = FALSE, ordered = TRUE)
```

```
## [1] 6
```

```
nsamp(n = 3, k = 2, replace = TRUE, ordered = FALSE)
```

```
## [1] 6
```

```
nsamp(n = 3, k = 2, replace = FALSE, ordered = FALSE)
```

```
## [1] 3
```

Limitations for the Classical Approach

One is the conceptual challenge when $|S|$ (the number of possible outcomes) goes to infinity.

Axioms for Probability

Let's put aside the interpretation of probability, and focus on what mathematically is generally accepted as probability

Axioms for Probability

The **probability axioms** are giveny by the following three rules

1. The probability of any event A is non-negative
$$P(A) \geq 0$$
2. The probability that one of the outcomes in the sample space occurs is 1:
$$P(\Omega) = 1$$
3. (*Countable Additivity*) if events $A_1, A_2, A_3, \ldots$ are **pairwise disjoint**, then
$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Empty Sets

It is useful to note that $\emptyset$ is a subset of the sample space, representing **impossible events**. It contains no elements and formally defined as follows

$$\emptyset := \{x : x \neq x\}$$

Also, disjoint events are those who do not overlap. Mathmetically, A and B are mutually exclusive or disjoint if

$$A \cap B = \emptyset$$

Some Tricky things about Empty Sets

Actually, empty set is a subset of any subsets, and two empty sets are disjoint

$$\emptyset \subset A, \quad \emptyset \cap \emptyset = \emptyset$$

The following sets are considered to be empty sets

$$(a, a] \quad \text{or} \quad [b, b)$$

Countable additivity implies the finite additivity as follows

(*Addition Rule*) If events A and B are mutually exclusive, then

$$P(A \text{ or } B) = P(A) + P(B)$$

Note Two events/outcomes are called **mutually exclusive** or **disjoint** if they cannot both happen.

Examples of Addition Rule

$$P(A \text{ or } B) = P(A) + P(B)$$

Considering no third-party candidate running for the presidential election,

Pr[Either Trump or Hilarly Wins] = ?

Pr[Either Trump or Hilarly Wins] =
Pr[Trump Wins] + Pr[Hilarly Wins] ?

Visualizing Probability

Question: How will you visualize the concept of probability?

There are many ways to do so! Just find something that ACTUALLy helps you.

Visualizing Probability

Think about probability visually as an area. The larger the area, the higher the probability

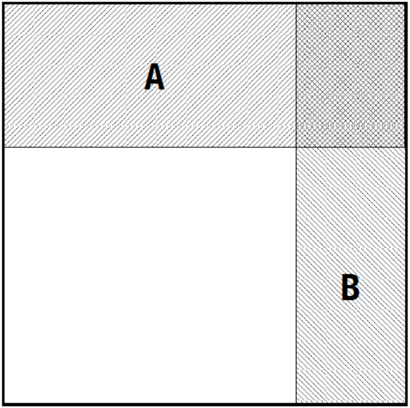


Figure 3

Several Implications from the Axioms (1)

Complement of a set: We can also be more concise when representing the collection of all outcomes in the sample space that do not belong to a set, say, A

$$A^c$$

$$\Pr[A^c] = 1 - \Pr[A]$$

Example:

$\Pr[\text{Trump Does not Win}] = 1 - \Pr[\text{Trump Wins}]$

Several Implications from the Axioms (2)

The notation $\{A \text{ and } B^c\}$ translates to "all outcomes of A that do not belong to B".

You can imagine any event decomposed into two mutually exclusive events

A consists of the part of A that belongs to B ($\{A \text{ and } B\}$) and the part that does NOT belong to B ($\{A \text{ and } B^c\}$).

Several Implications from the Axioms (2)

Law of Total Probability

$$\begin{aligned}\Pr[A] = & \quad \Pr[\{A \text{ and } B\} \text{ or } \{A \text{ and } B^c\}] \\ = & \quad \Pr[\{A \text{ and } B\}] + \Pr[\{A \text{ and } B^c\}]\end{aligned}$$

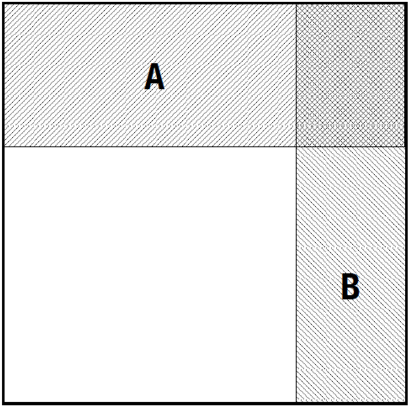


Figure 4

Several Implications from the Axioms (2b)

We can apply this law of total probability further to two events: What is

$$\Pr[A \text{ or } B]$$

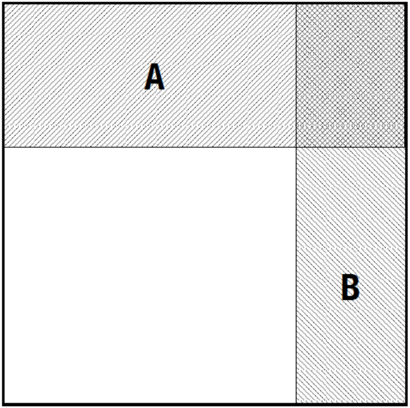


Figure 5

Several Implications from the Axioms (2b)

We can further decompose it into three mutually exclusive events

1. A and B^c (the part of A that does not belong to B)
2. A and B (the part of A that does not belong to B)
3. B and A^c (the part of B that does not belong to A)

$$\begin{aligned}\Pr[A \text{ or } B] = \quad & \Pr[\{A \text{ and } B^c\}] + \Pr[\{B \text{ and } A^c\}] \\ & + \Pr[\{A \text{ and } B\}]\end{aligned}$$

Several Implications from the Axioms (3): General Addition Rule

$$\Pr[A \text{ or } B] = \Pr[A] + \Pr[B] - \Pr[\{A \text{ and } B\}]$$

Example:

$$\begin{aligned}\Pr[\text{Either Trump or Hilarly loses}] = \quad & \Pr[\text{Trump loses}] + \Pr[\text{Hilarly loses}] \\ & - \Pr[\text{Trump and Hilarly loses}]\end{aligned}$$

Question: What does it mean if $\Pr[\{A \text{ and } B\} = 0$

Examples (more complicated case)

Question: What is the probability of the Yellow Region or Event Y (assuming that you have information on the probability related to A and B)?

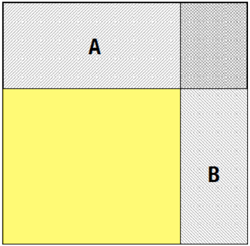


Figure 6

$Y =$ Not (A or B)

Examples (more complicated case)

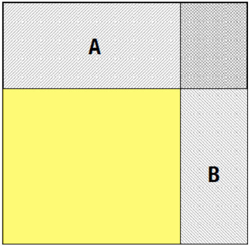


Figure 7

$Y = \text{Not } (A \text{ or } B)$

$$\begin{aligned} \Pr[Y] &= 1 - \Pr[(A \text{ or } B)] \\ &= 1 - (\Pr[A] + \Pr[B] - \Pr[\{A \text{ and } B\}]) \end{aligned}$$

Examples (more complicated case)

Florida Voters Data

```
##           gender
## race                f           m
##   asian    0.009107868 0.010095468
##   black    0.074399210 0.056622408
##   hispanic 0.073082410 0.057719741
##   native   0.001865467 0.001316800
##   other    0.017337869 0.016679469
##   white    0.360035115 0.321738176
```

In your homework, you will see an example in which you can't calculate this by hand.

Examples (more complicated case): Implementation in R

```r
sum(FL.dist[, 1])
```

```
## [1] 0.5358279
```

We obtain 0.5358279.

$$
\begin{aligned}
\Pr[\text{female}] = \; & \Pr[\text{female and asian}] + \Pr[\text{female and black}] \\
& + \Pr[\text{female and hispanic}] + \Pr[\text{female and native}] \\
& + \Pr[\text{female and other}] + \Pr[\text{female and white}]
\end{aligned}
$$

Examples (more complicated case)

$$\Pr[A] = \sum_{i=1}^{N} \Pr[A \text{ and } B_i]$$

We can also use `colSums()` or `rowSums()` in `R`.

```
colSums(FL.dist)
```

```
##         f         m
## 0.5358279 0.4641721
```

In your homework, you will be asked to try the other command as well.

Note that you need more two dimensions for `colSums()` to work.

Test your knowledge of Probability

Which of the following outcomes is more likely?

1. OU will lose one game
2. OU will lose one game but still make it to the Big XII championship.

Test your knowledge of Probability ("Linda"" Problem)

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Which is more probable?

1. Linda is a bank teller.
2. Linda is a bank teller and is active in the feminist movement.

Test your knowledge of Probability (“Linda”" Problem)

The **conjunction fallacy** (also known as the **Linda problem**) is a formal fallacy that occurs when it is assumed that **specific conditions** are more probable than a **single general one**.

Further Example Tversky, Amos; Kahneman, Daniel (October 1983). “Extension versus intuitive reasoning: The conjunction fallacy in probability judgment”. Psychological Review.

Policy experts were asked to rate the probability that the Soviet Union would invade Poland, and the United States would break off diplomatic relations, all in the following year. They rated it on average as having a 4% probability of occurring.

Another group of experts was asked to rate the probability simply that the United States would break off relations with the Soviet Union in the following year. They gave it an average probability of only 1%.

Test your knowledge of Probability (“Linda”" Problem)

$$\Pr[A \text{ and } B] \leq \Pr[A]$$

or

$$\Pr[A \text{ and } B] \leq \Pr[B]$$

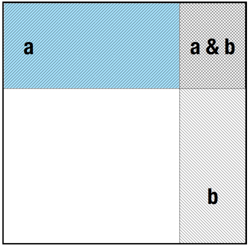


Figure 8

Probability, Gambling, and Dutch Book

In a gambling setting, odds are given as **odds against**. For example, the "odds against A" is defined as

$$\frac{\Pr[\text{Not-}A]}{\Pr[A]} = \frac{1 - \Pr[A]}{\Pr[A]}$$

Winner of 2017 Kentucky Derby

Figure 9

Probability, Gambling, and Dutch Book

If $A =$ "Always Dreaming wins the derby" with $\Pr[A] = .20$, the corresponding odds (against A) would be

$$\begin{aligned}
\frac{\Pr[\text{Not-}A]}{\Pr[A]} &= \frac{1 - \Pr[A]}{\Pr[A]} \\
&= \frac{1 - .2}{.2} \\
&= \frac{.8}{.2} \\
&= 4 : 1
\end{aligned}$$

For bookies to make money, the stated odds are typically not the actual probabilities (or even their best guess of them).

Probability, Gambling, and Dutch Book

Now, suppose your friend wants to bet on the Blazers-Lakers game (so that many of you may know nothing about the game). He sets his odds by looking at the winning percentage from the past season ().

1. Portland Trail Blazers: 41-41 (.5)
2. LA Lakers: 26-56 (.317)

Let's say he gives oods of $1 : 1$ and $7 : 3$.

Question: How will you bet?

Probability, Gambling, and Dutch Book

Let's say he gives oods of 1 : 1 and 7 : 3.

Question: How will you bet? 40 on Blazer and 30 on Laker

Event	Blazer Bet	Laker Bet	Total Profit
Blazer Wins	40	-30	10
Laker Wins	-40	70	30

Probability, Gambling, and Dutch Book

In gambling, a **Dutch book** or lock is a set of odds and bets which guarantees a profit, regardless of the outcome of the gamble. It is associated with probabilities implied by the odds not being coherent.

In economics, a Dutch book usually refers to a sequence of trades that would leave one party strictly worse off and another strictly better off. Typical assumptions in consumer choice theory rule out the possibility that anyone can be Dutch-booked.

Question: What rule did your friend violate?

Probability, Gambling, and Dutch Book

Question: What rule did your friend violate?

1. The probability of Blazer winning: $\frac{1}{1+1} = .5$
2. The probability of Laker winning: $\frac{3}{7+3} = .3$