Introduction to Probability

Lecture 3

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.2     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.1.0
-- Conflicts ----------------------------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

1 Outline

- Views of Probability
- Elementary Properties of Probability
- Calculating Probabilities
- Bayes' Theorem
- Screening Tests

2 Views of Probability

2.1 Objective Probability

Probability was thought of by statisticians and mathematicians only as an objective phenomenon derived from objective processes.

Note

There are two categories of objective probability:

- classical/a priori probability
- frequentist/a posteriori probability

2.2 Classical Probability

Probabilities for random events related to games of chance can be calculated by the processes of abstract reasoning. Therefore, it is not necessary for these events to happen to compute these probabilities.

Classical/A Priori Probability

Probability is defined as follows: If an event can occur in N mutually exclusive and equally likely ways, and if m of these possess a trait E, the probability of the occurrence of E is equal to m/N.

$$P(E) = m/N$$

Note

As an example, the probability of getting a three after tossing a **fair** six-sided die can be calculated by assuming that each of the six sides is equally likely to be observed.
The number of outcomes that include a three is 1 out of a total of 6 mutually exclusive outcomes.
Hence, the probability is 1/6.

2.3 Frequentist Probability

The frequentist approach to probability depends on the repeatability of some process and the ability to count the number of repetitions, as well as the number of times that some event of interest occurs. The probability describes the long-run relative frequency of occurrences from the process.

Frequentist/A Posteriori Probability

Probability is defined as follows: If some process is repeated a large number of times, n, and if some resulting event with the characteristic E occurs m times, the relative frequency of occurrence of E, m/n, will be *approximately equal* to the probability of E.

$$P(E) = m/n$$

Important

Note that m/n is only an estimate of the probability, and that n should be large to have a better estimate of $P(E)$.
Example: Monty-Hall Problem

2.4 Subjective Probability

This view holds that probability measures the confidence that a particular individual has in the truth of a particular proposition.

Note

The most famous example of subjective probability is *Bayesian*, which is a mathematically formal method to combine experimental data and expert information in producing probabilities.

3 Properties of Probability

3.1 Helpful Definitions

Sample Space

A sample space is a list of all possible outcomes that might be observed.

Event

An event is any collection of outcomes in the sample space.

Mutually Exclusive

If two events E_i and E_j contain no elements in common, these two events are said to be disjoint or mutually exclusive.

3.2 Axioms of Probability

The properties of probability was mathematically formalized by Russian mathematician A.N. Kolgomorov by defining the following axioms:

1. For each event E_i, the probability of E_i is non-negative.
2. The sum of the probabilities of the mutually exclusive events is equal to 1.
3. Consider any two mutually exclusive events, E_i and E_j. The probability of occurrence of either E_i or E_j is equal to the sum of their individual probabilities.

3.3 Calculating Marginal Probabilities

We can calculate the probability of occurrence for events using the sample point method.

Sample Point Method

Under the assumption that every outcome in the sample space is equally likely to occur, the probability that some event E_i occurs is defined as the number of outcomes corresponding to E_i divided by the total number of outcomes in the sample space.
In frequency distributions, this corresponds to the relative frequency of a specific outcome.

Important

Probabilities computed using this method are often referred to as **marginal probability**. Marginal probabilities are unconditional and are specific to an event.

3.4 Example

Out of 1,325 university students, 532 of them reported to consume caffeine through coffee. What is the marginal probability of selecting a university student who consumes coffee?

$$P(E) = \frac{532}{1325}$$

The resulting probability is 0.402.

3.5 Example 2

Consider the `AMSSurvey` data in the package `carData`. The data set includes the counts of new PhDs in the mathematical sciences for 2008-09 and 2011-12 categorized by type of institution, gender, and US citizenship status. You can learn more about the data set by typing `?carData::AMSsurvey` after installing the `car` package.

3.5.1 Question + Data

What is the probability of selecting a participant of the survey at random who is not a US citizen in 2008? Use the `count` variable.

```r
library(carData)
AMSsurvey
```

```
    type    sex citizen count count11
1  I(Pu)   Male      US   132     148
2  I(Pu) Female      US    35      40
3  I(Pr)   Male      US    87      63
4  I(Pr) Female      US    20      22
5     II   Male      US    96     161
6     II Female      US    47      53
7    III   Male      US    47      71
8    III Female      US    32      28
9     IV   Male      US    71      89
10    IV Female      US    54      55
11    Va   Male      US    34      42
12    Va Female      US    14      21
13 I(Pu)   Male  Non-US   130     136
14 I(Pu) Female  Non-US    29      32
15 I(Pr)   Male  Non-US    79      82
16 I(Pr) Female  Non-US    25      26
17    II   Male  Non-US    89     116
18    II Female  Non-US    50      56
19   III   Male  Non-US    53      61
20   III Female  Non-US    39      30
21    IV   Male  Non-US   122     153
22    IV Female  Non-US   105     115
23    Va   Male  Non-US    28      27
24    Va Female  Non-US    12      17
```

3.5.2 Answer

The total counts can be calculated using the `sum` function. Or simple addition.

```
sum(AMSsurvey$count)
```

```
[1] 1430
```

We can also count the total number of non-US citizens in the data set.

```
130+29+79+25+89+50+53+39+122+105+28+12
```

```
[1] 761
```

Thus, the probability can be calculated as 761/ 1430 = 0.532