3. Probability and applications

“We balance probabilities and choose the most likely. It is the scientific use of the imagination.”  Sherlock Holmes, The Hound of the Baskervilles, 1902.

Chapter 1 tells us something about what to do with a data set, or at least how to look at it in a sensible way. Chapter 2 provides an indication of where the data we analyse comes from. In this chapter, and the next, we look at models for the data. The basic approach is based on sampling in which we assume a model for the population, from which we select a sample. The data are the sample.

3.1 Probability: The Basics

Probability (chance, likelihood) has an everyday usage which gives some sort of rough gradation between the two extremes of impossible and certain:

impossible  not likely  maybe  probably  certain
no way  possibly  fair chance  no worries

We wish to make probability numerical — to define a mathematical probability. To do this we need a structure and some rules.

random experiment:
event space, $\Omega$:
outcome, $\omega$:
event, $A$:
realisation:

Example  Tossing a coin three times. (Treating three patients.)

\[ \Omega = \{hhh, hht, hth, thh, htt, tht, tth, ttt\} ; \]
\[ A = \text{“at least two heads”} = \{hhh, hht, hth, thh\} ; \]
\[ \omega_o = \]
**Probability**, $Pr(A)$, is a number assigned to each event $A$, which reflects its (probability, chance, likelihood) of occurrence. This (mathematical) Probability must obey some rules: very reasonable rules, but rules nevertheless.

**Properties of Pr**

1. $0 \leq Pr(A) \leq 1$;
2. $Pr(\emptyset) = 0, Pr(\Omega) = 1$;
3. $Pr(A') = 1 - Pr(A)$;
4. $A \subseteq B \Rightarrow Pr(A) \leq Pr(B)$;
5. $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$.

**Example** If $Pr(A) = 0.45$, then $Pr(A') = 0.55$. If $Pr(B) = 0.35$ and $A$ and $B$ are mutually exclusive; then $Pr(A \cup B) = 0.80$.

**Note:** It is enough to assume the three axioms I: $Pr(A) \geq 0$; II: $Pr(\Omega) = 1$; III: $Pr(A \cup B) = Pr(A) + Pr(B)$ if $A$ and $B$ are mutually exclusive. All the other properties above can be deduced from these axioms.

**Assigning values to Pr**

- symmetry
- long-term relative frequency
- subjective
- model

**Example** What is the value for $Pr(6)$ for this die?

Venn diagram:

**Probability table:**

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$\gamma$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>$A'$</td>
<td>$\beta$</td>
<td>1</td>
</tr>
</tbody>
</table>

**Example** $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$.

*[Similar tables can be made for 3 events, and for 4 events. We will not consider them though.]*
In this probability table, intersections are represented by one square: \( A \cap B, A \cap B', A' \cap B, A' \cap B' \); unions are three squares, i.e. an L-shape: \( A \cup B, A \cup B', A' \cup B, A' \cup B' \).

To complete the probability table, and hence to work out anything involving \( A \) and \( B \), we need three (separate) items of information.

**Example**  \( \Pr(A) = 0.6, \Pr(B) = 0.2, \Pr(\{A \cap B\}) = 0.1 \).

\[
\begin{array}{cc}
0.1 & 0.5 \\
0.1 & 0.3 \\
\end{array}
\]

\( \Pr(A \cap B') = 0.5, \Pr(\{A \cup B\}) = 0.9, \ldots \)

**Example**  \( \Pr(A \cup B) = 0.7, \Pr(\{A \cap B'\}) = 0.2, \Pr(A) = 0.3 \).

\[
\begin{array}{cc}
0.1 & 0.2 \\
0.4 & 0.3 \\
\end{array}
\]

\( \Pr(B) = 0.5, \Pr(\{A \cap B\}) = 0.1, \ldots \)

Probability tables are very useful devices. Their data counterparts are called “contingency tables.”

### 3.2 Conditional probability

\( \Pr(\cdot \mid H) \) denotes the probability of \( \cdot \) given the information \( H \) about the outcome of the random experiment. \( H \) is an event.

We would like to modify the probability in the light of this additional information:  \( \Pr(\cdot) \rightarrow \Pr(\cdot \mid H) \).

**Example** (tossing three fair coins) \( A = “at least two heads” \) and \( H = “first toss is a head” \). \( \Pr(A) = \frac{1}{2}, \) but \( \Pr(A \mid H) > \frac{1}{2} \).

How should this modification be done?

Given that \( H \) has occurred, any outcome in \( H' \) is impossible. Now any possible outcome is in \( H \).

\[
\begin{array}{cc}
\Pr(\cdot \cap H) & 0 \\
\Pr(\cdot \cap H') & 0 \\
\end{array}
\]

(the possible universe is reduced to \( H \)).

We make the modified probabilities proportional to what they were, i.e. \( \Pr(\omega \mid H) = k \Pr(\omega) \), but rescale them so that they sum to 1. This means that \( k = 1 / \Pr(H) \) and hence \( \Pr(A \mid H) = \frac{\Pr(A \cap H)}{\Pr(H)} \).

Note: this is an axiom, not a theorem; i.e. it’s an assumption, it can’t be proved.

When we consider \( \Pr(\cdot \mid H) \), we act as if the universe is \( H \).

Really, all probability is conditional: \( \Pr(\cdot) = \Pr(\cdot \mid \Omega) \).

**Example** (tossing three fair coins)

\[
\Pr(A \mid H) = \frac{3/8}{1/2} = \frac{3}{4} \quad \text{OR} \quad \frac{1}{4} + \frac{1}{4} + 0 + \frac{1}{4} \quad \text{OR} \quad \Pr(\geq 1/2 \text{ heads}) = \frac{3}{4}
\]

**Example**  \( \Pr(A) = 0.4, \Pr(B \mid A) = 0.8, \Pr(B \mid A') = 0.3 \)

**Example**  \( \Pr(A) = 0.3, \Pr(B) = 0.4, \Pr(A \mid B) = 0.6 \)

**Example** (exposure and disease)

\( \Pr(E) = 0.3, \Pr(D \mid E') = 0.001, \Pr(D \mid E) = 0.011 \)

Note: \( \Pr(A \mid B) \) is not the same as \( \Pr(B \mid A) \): they are in different universes!

Note: \( \Pr(A \mid B) \) is not equal to \( 1 - \Pr(A \mid B') \): different universes again.
Multiplication rule

\[ \Pr(A \cap B) = \Pr(A) \Pr(B \mid A) = \Pr(B) \Pr(A \mid B) \]

Dividing through by \( \Pr(A) \Pr(B) \) gives:

\[ \frac{\Pr(A \cap B)}{\Pr(A) \Pr(B)} = \frac{\Pr(B \mid A)}{\Pr(B)} = \frac{\Pr(A \mid B)}{\Pr(A)} \quad [= k, \text{say}] \]

if \( k > 1 \), then:
\( \Pr(B \mid A) > \Pr(B) \) and \( \Pr(A \mid B) > \Pr(A) \) and \( \Pr(A \cap B) > \Pr(A) \Pr(B) \);
and we say that \( A \) and \( B \) are positively related, since each increases the chance of the other occurring.

if \( k < 1 \), the \( A \) and \( B \) are negatively related: each decreases the chance of the other occurring.

Note: if \( A \) and \( B \) are positively related, then \( A \) and \( B' \) are negatively related.

It follows that

\[ \Pr(A \mid B) \gtrsim \Pr(A) \gtrsim \Pr(A \mid B') \]

with > for positively related events, and < for negatively related events.

Example (exposure and disease)

If the exposure \( E \) is positively related to disease \( D \), then
\[ \Pr(D \mid E) > \Pr(D) > \Pr(D \mid E') \]

\[ [0.011] \quad [0.004] \quad [0.001] \quad \text{—— from example above.} \]

The relative risk of a disease with respect to an exposure \( E \) is given by

\[ \text{relative risk, } RR = \frac{\Pr(D \mid E)}{\Pr(D \mid E')} \quad [11] \quad \text{in the example.} \]

The relative risk is the ratio of the probability of the disease given the exposure and the probability of disease given non-exposure.

Example (Hypertension)

An adult is considered as hypertensive if their diastolic blood pressure (DBP) \( \geq 95 \). A baby is considered as hypertensive if its DBP \( \geq 80 \). From extensive hospital records, the probability that a mother is hypertensive is 0.1, that a baby is hypertensive is 0.2, and that both mother and baby are hypertensive is 0.05. If a mother is hypertensive, what is the probability that her baby is hypertensive?

Example (Breast cancer)

Consider a study that examine the risk factors for breast cancer among women participating in a national Health and Nutrition Examination Survey. In this study, in a sample of 4540 women who gave birth to their first child before the age of 25, 65 developed breast cancer. Of the 1628 women who first gave birth at age 25 or older, 31 were diagnosed with breast cancer. Give an estimate of the relative risk of developing breast cancer, given that the woman’s age at first birth is 25 or older.
3.3 Law of Total Probability & Bayes’ Theorem

These results generally apply in the context of:

```
mutually exclusive and exhaustive “causes” \( A_1, A_2, \ldots, A_k \)
of some “result” \( H \)
```

where we know the probability of the possible “causes”, i.e. \( \Pr(A_1), \Pr(A_2), \ldots, \Pr(A_k) \); and the probability of the “result” given each of the “causes”, i.e. \( \Pr(H \mid A_1), \Pr(H \mid A_2), \ldots, \Pr(H \mid A_k) \).

Law of Total Probability gives \( \Pr(H) \); Bayes’ theorem gives \( \Pr(A_j \mid H) \).

The formulae will be given later. But first we’ll learn how to work these out. Then you don’t need the formulae!

Standard applications are:

```
“causes” = exposure \( \rightarrow \) “result” = disease;
“causes” = disease; \( \rightarrow \) “result” = test result.
```

**Example** (alcohol and headache)

<table>
<thead>
<tr>
<th></th>
<th>( H )</th>
<th>( H' )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>0.3</td>
<td>(0.01)</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>0.5</td>
<td>(0.1)</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>0.2</td>
<td>(0.9)</td>
</tr>
</tbody>
</table>

LTP: \( \Pr(H) = 0.233 \); BT: \( \Pr(A_3 \mid H) = 0.773 \).

**Example** (exposure and disease) \( A_1 = E, A_2 = E', H = D \).

\[
\begin{array}{ccc}
| & D & D' \\
| E & 0.0033 & 0.3 (0.011) \\
| E' & 0.0007 & 0.7 (0.001) \\
\end{array}
\]

LTP: \( \Pr(D) = 0.004 \); BT: \( \Pr(E \mid D) = 0.825 \).

We already knew how to do this! See the earlier example.

**Example** (Ophthalmology)

We are planning a 5-year study of cataracts in a population of 5000 people 60 years of age and older. We know from census data that 45% of this population are ages 60-64, 28% are ages 65-69, 20% are ages 70-74, and 7% are age 75 or older. We also know from Framingham Eye Study that 2.4%, 4.6%, 8.8% and 15.3% of the people in those respective age groups will develop cataracts over the next 5 years. What percentage of our population will develop cataracts over the next 5 years, and how many people does this percentage represent?

**Example** (diagnostic test for disease \( D \))

Let \( P \) denote a Positive result to the test, i.e. indicating that the disease is present; then \( P' \) denotes a Negative result, indicating that the disease is not present.

\[
\begin{array}{ccc}
| & P & P' \\
| D & 0.01 (0.95) \\
| D' & 0.99 (0.1) \\
\end{array}
\]

LTP: \( \Pr(P) = 0.1085 \); BT: \( \Pr(D \mid P) = 0.0876 \).

Thus, this is not all that clever a test (although \( \Pr(D) = 0.01 \) has been increased to \( \Pr(D \mid P) = 0.09 \)).

Most of the positive results come from individuals without the disease.
Probability diagram representation:

<table>
<thead>
<tr>
<th></th>
<th>$H$</th>
<th>$\bar{H}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>$\Pr(A_1) \Pr(H \mid A_1)$</td>
<td>$\cdots$</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$\Pr(A_2) \Pr(H \mid A_2)$</td>
<td>$\cdots$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\cdots$</td>
</tr>
<tr>
<td>$A_k$</td>
<td>$\Pr(A_k) \Pr(H \mid A_k)$</td>
<td>$\cdots$</td>
</tr>
<tr>
<td></td>
<td>$\Pr(H)$</td>
<td>$\cdots$</td>
</tr>
</tbody>
</table>

Law of Total Probability: \[ \Pr(H) = \sum_{i=1}^{k} \Pr(A_i) \Pr(H \mid A_i). \]

Bayes’ Theorem: \[ \Pr(A_j \mid H) = \frac{\Pr(A_j) \Pr(H \mid A_j)}{\sum_{i=1}^{k} \Pr(A_i) \Pr(H \mid A_i)}. \]

3.4 Diagnostic Testing

The diagnostic testing scenario is very important in medicine. There are a bunch of names for many of the probabilities and conditional probabilities that you need to know about.

- “false negative” = $D \cap P'$
- “false positive” = $D' \cap P$
- Prevalence = $\Pr(D)$

<table>
<thead>
<tr>
<th></th>
<th>$P$</th>
<th>$P'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>$\checkmark$</td>
<td>$\times$</td>
</tr>
<tr>
<td>$D'$</td>
<td>$\times$</td>
<td>$\checkmark$</td>
</tr>
</tbody>
</table>

ppv = $\Pr(D \mid P)$ npv = $\Pr(D' \mid P')$

- The **sensitivity** (sn) of a test is the probability that the test is positive given that the person has the disease: $\text{sn} = \Pr(P \mid D)$.
- The **specificity** (sp) of a test is the probability that the test is negative given that the person does not have the disease: $\text{sp} = \Pr(P' \mid D')$.
- The **positive predictive value** (ppv) of the test is the probability that a person has the disease, given the test is positive: $\text{ppv} = \Pr(D \mid P)$.
- The **negative predictive value** (npv) of the test is the probability that a person does not have the disease, given that the test is negative: $\text{npv} = \Pr(D' \mid P')$.

*Note that all these conditional probabilities are concerned with “getting it right”, given $D$, given $D'$, given $P$ and given $P'$.***
A false negative is when the test is negative, but the person has the disease, i.e. \( fn = D \cap P' \).

A false positive is when the test is positive, but the person does not have the disease, i.e. \( fp = D' \cap P \).

The prevalence of a disease is the probability that a randomly chosen person from a population has the disease.

Note: the incidence of a disease is the probability that an individual with no prior disease will develop the disease over some specified time period.

Example A government (USA) study showed that in 1974 in a sample of 144,380 people 17 years of age or older, there were 22,626 that were hypertensive. So, the prevalence of hypertension for people 17 years or older is \( \frac{22,626}{144,380} = 15.7\% \).

Example (diagnostic test)
Consider a diagnostic test with sensitivity 99% and specificity 99% applied to a population with disease prevalence 5%. Find the positive predictive value for this test.

\[
\begin{array}{ccc}
D & P & P' \\
\hline
D & 0.0495 & 0.0005 & 0.05 \\
D' & 0.0095 & 0.9405 & 0.95 \\
0.0590 & 0.9410 & 1 \\
\end{array}
\]

\( (sn=0.99) \)
\( (sp=0.95) \)

Thus \( ppv = \Pr(D \mid P) = \frac{0.0495}{0.0590} = 0.839 \).

Example (hypertension)
Suppose 84% of hypertensives and 23% of normotensives are classified as hypertensive by an automated blood-pressure machine. What is the positive predictive value and negative predictive value negative of the machine, assuming that 20% of the adult population is hypertensive?

Example (cervical cancer)
Cervical cancer is a disease for which the chance of containment is high given that it is detected early. The Pap smear is a widely accepted screening procedure that can detect a cancer that is as yet asymptotic. An on-site proficiency test assessed the competency of technicians who scan Pap smear slides for abnormalities.

Overall, 16.25% of the tests performed on women with cancer resulted in false negative outcomes. And 18.64% of women who do not have cancer showed positive results. The prevalence of the disease is 8.3 per 100,000.

(a) Find the sensitivity, specificity, predictive value positive and predictive value negative of the test.

(b) If 1,000,000 women are screened by this test, how many positive tests do you expect? How many of them do you expect to be false positive?
3.5 Independence

Events $A$ and $B$ can be positively or negatively related according as:

$$
\Pr(A \mid B) \gtrless \Pr(A) \gtrless \Pr(A \mid B')
$$

The intermediate case, when they are all equal, i.e. the “no relationship” case is the case of independence. $A$ and $B$ are independent if $B$ has no effect on the probability of $A$ occurring ... and vice versa: i.e.

$$
\Pr(A \mid B) = \Pr(A) = \Pr(A \mid B') \quad \text{and} \quad \Pr(B \mid A) = \Pr(B) = \Pr(B \mid A').
$$

This means that:

$$
\Pr(A \cap B) = \Pr(A) \Pr(B),
$$

which is often taken as the ‘rule’ for independence.

Independent events and Mutually exclusive events are entirely different things.

Example

$A$ and $B$ are mutually exclusive events such that $\Pr(A) = \Pr(B) = 0.4$. Then $\Pr(A \cup B) = 0.4 + 0.4 = 0.8$.

$C$ and $D$ are independent events such that $\Pr(C) = \Pr(D) = 0.4$. Then $\Pr(C \cup D) = 0.4 + 0.4 - 0.4 \times 0.4 = 0.64$.

This multiplication rule extends to $n$ independent events:

If $A_1, A_2, \ldots, A_n$ are independent, then

$$
\Pr(A_1 \cap A_2 \cap \cdots \cap A_n) = \Pr(A_1) \Pr(A_2) \cdots \Pr(A_n),
$$

but the converse is not true.

Similarly

$$
\Pr(A_1 \cup \cdots \cup A_n) = 1 - \Pr(A_1' \cap \cdots \cap A_n') = \Pr(A_1') \cdots \Pr(A_n'),
$$

i.e. Pr(“at least one”) = 1 – Pr(“none”).

Example

Find the probability of at least one six in six rolls of a fair die.

$$
\Pr(A) = 1 - \Pr(A') = 1 - \left(\frac{5}{6}\right)^6 = 0.665.
$$

Find the probability that at least one individual in a sample of 100 has disease $D$ when the prevalence of the disease is 1%.

$$
\Pr(A) = 1 - \Pr(A') = 1 - \left(\frac{99}{100}\right)^{100} = 0.634.
$$

A commonly used probability model is that of “independent trials” (commonly called Bernoulli trials) in which each trial results in one of two outcomes, designated “success” or “failure”, with probabilities $p$ and $q$, where $p + q = 1$.

Simple examples of independent trials are coin-tossing and die-rolling; but the “independent trials” model can be applied quite generally with:

- trial = any (independently) repeatable random experiment;
- success = $A$, any nominated event for the random experiment.

Example A risky heart operation is such that the probability of a patient dying as a result of the surgery is 0.01. If 100 such operations are performed at the hospital in a year, find the probability that at least one of these patients dies as a result of surgery.

Assuming the operations are independent and each has probability of “success”.$This just emphasises the fact that “success” is just a name for some event: it clearly doesn’t have to be something good!

It soon becomes clear that the model is too simple — but it is nevertheless a useful place to start the modelling process.

“You know my method. It is founded upon the observation of trifles.”
Sherlock Holmes, The Boscomb Valley Mystery, 1892.