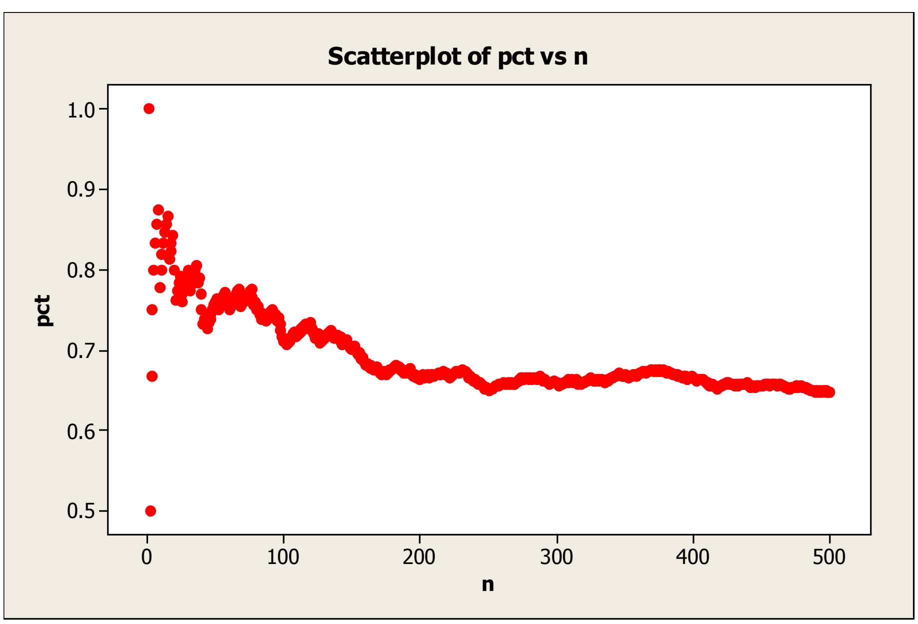
Chapter 1 – Probability

Statisticians help us make intelligent *decisions* (statistical inference) and *model* complex phenomena, and probability is the "language of statistics". Note that the *scientific method* is indeed an "endless cycle"

The *outcome space* (denoted S) is a listing of all outcomes of an experiment; some examples are on p.3. An *event* (e.g., denoted A) is a collection of such outcomes so $A \subset S$. Important definitions related to the *null set*, *subsets*, *unions*, *intersections*, and *complements* are given on p.3, and the pictures on p.4 are helpful to visualize these. Events are *mutually exclusive* if they are pairwise disjoint, and they are *exhaustive* if their union is the whole space S. Probability is defined on p.4 as the limit of a relative frequency.

Example 1.1-7 on pp.4-5 illustrates probability, and we perform our own simulation with the following program and graph.

```
data one;
  sumy=0;
  do nn=1 to 500;
    sumy=sumy+ranbin(1234321,1,(31031/46656));
    pct=sumy/nn;
    output;
  end;
```



Here, a die is thrown 6 times (labeled trials k = 1, 2 … 6) and a "Success" occurs at least one of {trial k results in a roll of k} occurs. The probability of this happening is $1 - (5/6)^6 = 31031/46656 = 0.665$, and this process is repeated n = 500 times. Notice in the above graph (as on p.5), the relative frequencies approach this limit – although not uniformly. Probability is a long-run phenomenon. Example 1.1-8 provides another illustration.

Probability is rigorously defined on p.6: it is a real-valued set function P that assigns to each set A the number P(A) and is such that

(a) $P(A) \geq 0$

(b) $P(S) = 1$

(c) For (at most countably infinite) pairwise disjoint events A_1, A_2, A_3 … (so that all $A_k \cap A_m = 0$), we have

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$$

Several results follow including the following six theorems.

Theorem 1.1-1. For all events A, $P(A') = 1 - P(A)$, where A' is the complement of A.

Theorem 1.1-2. The probability of the null set is zero: $P(\varnothing) = 0$.

Theorem 1.1-3. For all events A and B such that $A \subset B$, $P(A) \leq P(B)$.

Theorem 1.1-4. For all events A, we have that $P(A) \leq 1$.

Theorem 1.1-5. For all events A and B, we have that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Theorem 1.1-6. For all events A, B and C, we have that

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

To illustrate Theorem 1.1-5, work through Exercise 1.1-3 (a) – (c) on p. 10. Other illustrations are Examples 1.1-10 and 1.1-11.

The *equally likely* idea on p.9 is very intuitive and demonstrated in Exercise 1.1-9 on p. 11.

Enumeration Methods (Section 1.2, pp. 11-17)

When considering enumerating different combinations, the multiplication principle is easily demonstrated using *probability trees* (although these are sometimes impractical). Here, 'n factorial' is defined as $n! = n(n-1)\ldots(2)(1)$ for $n \geq 1$ and as $0! = 1$.

If we have n different objects, then each of the n! arrangements in a row of the n objects is called a permutation of the n objects. For example, the 4! = 24 permutations of $\{A, B, C, D\}$ are

ABCD	**ABDC**	**ACBD**	**ACDB**	**ADBC**	**ADCB**
BACD	**BADC**	**BCAD**	**BCDA**	**BDAC**	**BDCA**
CABD	**CADB**	**CBAD**	**CBDA**	**CDAB**	**CDBA**
DABC	**DACB**	**DBAC**	**DBCA**	**DCAB**	**DCBA**

If there are only $r \leq n$ positions, then the number of possible ordered arrangements (called a permutation of n objects taken r at a time) is

$$_nP_r = n(n-1)(n-2)\ \ldots\ (n-r+1) = \frac{n!}{(n-r)!}$$

So, in the above example with n=4, if there are only r = 2 positions, then the number of possible ordered arrangements is 24/2 = 12; in the above list, these twelve arrangements are

AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC

When order is not relevant – for example, in the above list if AB and BA are really the same occurrence – then we need to consider combinations. From Definition 1.2-6 on p.14, each of the $_nC_r$

unordered subsets is called a combination of n objects taken r at a time; here

$$_n C_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

In the above illustration, there are $_4C_2 = 6$ combinations of n = 4 objects taken r = 2 at a time – these are AB, AC, AD, BC, BD, CD. As discussed on p.15, the numbers $_nC_r$ are called binomial coefficients and come from the expansion of $(a+b)^n$ in powers of 'a' and 'b'. Note also Definitions 1.2-3 (ordered sample of size r), 1.2-4 (sampling with replacement), 1.2-5 (sampling without replacement), and 1.2-6 (combination of n objects taken r at a time).

Work through the 13 examples in this section and Ex. 1.2-7 and 1.2-9.

Conditional Probability (Section 1.3, pp. 18-25)

Consider Example 1.3-1 (p.19) regarding tulips.

	Early (E)	Late (L)	Total
Red (R)	5	8	13
Yellow (Y)	3	4	7
Total	8	12	20

Note that this table is conveying that the probability that a randomly selected tulip of this type is both Red and blooms Early is $P(R \cap E) = 5/20 = 25\%$, that one such tulip is Red is $P(R) = 13/20 = 65\%$, that one such tulip blooms Early is $P(E) = 8/20 = 40\%$, and that one such tulip is Red given that it blooms Early is $P(R/E) = 5/8 = 62.5\%$. This latter probability is called a *conditional probability*, and the following rule holds in general (provided $P(B) \neq 0$):

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Example 1.3-3 and Figure 1.3-1 (p.20) help to visualize conditional probability. Also, on p.21, it is shown that conditional probabilities follow the axioms of a probability function, and the proofs follow directly from results for unconditional probabilities.

A direct consequence of the above definition is the multiplication rule:

$$P(A \cap B) = P(A) \times P(B/A) = P(B) \times P(A/B)$$

Also, by extension,

$$P(A \cap B \cap C) = P(A) \times P(B/A) \times P(C/A \cap B)$$

(and so on).

To illustrate, in the above if there are 20 tulips in a box and we randomly select two of them *<u>without replacement</u>*, then the probability that they are both red is:

$$\frac{13}{20} \times \frac{12}{19} = 41.05\%$$

Note that (a) this is slightly less than the answer we would obtain 'with replacement': $(0.65)^2 = 42.25\%$, and (b) another way to get this answer using the results of the previous section is via $\binom{13}{2} \Big/ \binom{20}{2}$.

Good illustrations are Examples 1.3-6 and 1.3-7 on p.23 and Examples 1.3-11 and 1.3-12 on pp.24-25. Example 1.3-10 on p.24 is easily solved using probabilities trees; we'll return to this example after we consider Bayes Theorem in Section 1.5 below.

<u>Independent Events (Section 1.4, pp. 27-31)</u>

This is analogous to sampling with replacement (or sampling from an infinitely large population). Events A and B are independent if the occurrence of one of them does not affect the probability of the

occurrence of the other, i.e., if $P(A/B) = P(A)$; note that $P(A/B) = P(A)$ implies that $P(B/A) = P(B)$. A direct consequence of this (and an alternate definition of independence) is that events A and B are independent if $P(A \cap B) = P(A) \times P(B)$. Note that by Theorem 1.4-1, independence of A and B carries over to A' and B'.

As Example 1.4-4 on p.29, care needs to be exercised with three events: events A, B and C are independent if they are pairwise independent and if $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$.

To illustrate, note that we used the multiplication rule for independent events in our simulation example on p.1 of these notes (Example 1.1-7); see Example 1.4-6 on p.30. For another illustration, in Example 1.4-8,

- **the probability of exactly one defect detected is $P(1) = 0.001376$**
- **the probability of exactly two defects detected is $P(2) = 0.067224$**
- **the probability of exactly three defects detected is $P(3) = 0.931392$**

We can find the probability of at least one defect by adding these three numbers or (better yet) by the calculation:

$$1 - (0.01)\times(0.02)\times(0.04) = 0.999992$$

Example 1.4-9 on p.31 is interesting and is hinting at the Binomial probability distribution introduced and illustrated on p.59ff.

Bayes's Theorem (Section 1.5, pp. 33-36)

In Example 1.5-1, we consider three bowls (B_1, B_2, and B_3) with Red (R) and White (W) chips in them:

2R 4W	1R 2W	5R 4W
B_1 (1/3)	B_2 (1/6)	B_3 (1/2)

Also, the probabilities that we select the respective bowls are above in RED; these probabilities are called the *prior probabilities* for the respective bowls.

Let's find the probability that a randomly selected chip is RED (with no knowledge of which bowl it came from). This is best-done using probability trees, and turns out to equal 4/9.

Next, imagine that we were told that a randomly chosen chip is RED, and we are asked to calculate the conditional probabilities associated with each of the three bowls. We do this using the conditional probability rules of Section 1.3. For example,

$$P(B_1/R) = \frac{P(B_1)\times P(R/B_1)}{P(B_1)\times P(R/B_1)+P(B_2)\times P(R/B_2)+P(B_3)\times P(R/B_3)}$$

The denominator in this calculation is just the P(R) = 4/9 that we found above. Then, we get $P(B_1/R)$ = ¼ = 0.25, $P(B_2/R)$ = 1/8 = 0.125, and $P(B_3/R)$ = 5/8 = 0.625; these are called the *posterior probabilities* (i.e., given that the chip is RED). Note that they have changed from the prior probabilities given above. The probability associated with the third bowl (B_3) has increased since it has more RED chips.

In general, if the events B_1, B_2 … B_m constitute a partition of Ω – so that they are pairwise disjoint and their union is Ω – and we are given the prior probabilities $P(B_k)$ for all k and if we know for event A all conditional probabilities $P(A/B_1)$, then the posterior probabilities can be calculated using Bayes's theorem:

$$P(B_k/A) = \frac{P(B_k)\times P(A/B_k)}{\sum_{j=1}^{m} P(B_j)\times P(A/B_j)}$$

Applications of Bayes's theorem in medicine are widespread as in Example 1.5-3 on pp.35-36. The *sensitivity* of this test is 84% and the *specificity* of the test is 81%.