PART VII: Probability and statistics

Chapter 39: Probability

39.1. A sample space is the set of all possible outcomes of a random experiment. We may be able to associate several sample spaces with an experiment in each case.
(a) Assuming that the coins are spun and placed in sequence in a row. A single spin of each coin will give a sequence of 5 heads(H) or tails(T)). A typical outcome could be $\{H, T, H, T, T\}$ Since each coin could be H or T the number of possible outcomes is $2 \times 2 \times 2 \times 2 \times 2 = 2^5 = 32$ in this sample space.

Another sample space could be the the number of heads which occur when 5 coins are spun. The coins could show 0 heads, 1 head, 2 heads and so on. The sample space is therefore

$$S = \{0, 1, 2, 3, 4, 5\},$$

which has 6 elements.

(b) Assume that the sample space is the sum of the faces shown. Any sum between 3 and 18 is possible. Hence the sample space has 16 possible outcomes.

(c) Each outcome of the coin can be combined with the outcomes of the die. A typical outcome could be $\{H, 5\}$. Hence the number of outcomes is $2 \times 6 = 12$.

(d) A standard dartboard is divided into 20 sectors numbered $1, 2, \ldots, 20$, each sector has areas which the values are double and tripled. There are also inner and outer bulls. Assuming that a dart always hits the board there are $20 \times 3 + 2 = 62$ scoring areas on the board where the dart can score. We might be interested in the probability that the dart lands in one of these parts of the dartboard.

A second space could be the score which occurs when the dart lands on the board. Taking account of the doubles and triples, and the inner bull which scores 50 and the outer bull which scores 25, the sample space is

$$\begin{aligned} S &= \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, \\ & \quad 25, 26, 27, 28, 30, 32, 33, 34, 36, 38, 39, 40, 42, 45, 48, 50, 51, 54, 57, 60\}. \end{aligned}$$

There are 43 possible outcomes.

39.2. The sample space for two dice has 36 possible outcomes which are listed in Example 39.2. They are equally likely. Of the 36 outcomes, 6 score 7. Hence the probability of a score of 7 is 6/36=1/6.

At least one 5 appears in 11 possible outcomes. Hence the probability that no 5 appears is

$$\frac{36 - 11}{36} = \frac{25}{36}.$$

In the list in Example 39.2, a score of 7 or less appears in 21 outcomes. Hence the probability of a score of 7 or less is 21/36=7/12.

39.3. The set of the scores of two dice can be expressed as

$$S = \{(i, j)|i, j = 1, 2, 3, 4, 5, 6\}$$

(see also Example 39.2). S has 36 elements. We can write A and B as

$$A = \{(i, j)|i + j = 5\}, \quad B = \{(i, j)|i = 4 \text{ or } j = 4 \text{ or both} = 4\}.$$

As lists:

$$A = \{(1,4),(2,3),(3,2),(4,1)\},$$

$$B = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(1,4),(2,4),(3,4),(5,4),(6,4)\}.$$

The elements in the union and intersection of A and B are

$$\begin{aligned} A \cup B &= \{(2,3),(3,2),(4,1),(4,2),(4,3),(4,4),(4,5), \\ &\quad (4,6),(1,4),(2,4),(3,4),(5,4),(6,4)\} \end{aligned}$$

$$A \cap B = \{(1,4),(4,1)\}.$$

39.4. (a) Only B occurs is the set

$$B \backslash (A \cup C).$$

(b) As in (a) only A occurs is $A \backslash (B \cup C)$ and only C occurs is $C \backslash (A \cup B)$. Therefore exactly one of A, B or C occurs is the union of (a) with only A and only C, that is,

$$(B \backslash (A \cup C)) \cup (A \backslash (B \cup C)) \cup (C \backslash (A \cup B)).$$

39.5. The number of elements in $A \cup B$ is the sum of the number of elements in the intersection of A and B plus the number of elements in A but not in B and the number of elements in B but not in A. The intersection of A and B is $A \cap B$. $\overline{A}$ is the complement of A, that is, the set of elements not in A, so that $\overline{A} \cap B$ is the set of elements in B but not in A. Similarly, $\overline{B} \cap A$ is the set of elements in A but not in B. Therefore

$$\mathrm{n(A \cup B) = n(A \cap B) + n(\overline{A} \cap B) + n(\overline{B} \cap A).}$$

The sets A and B are given by

$$A = \{(i,j) | i + j = 6\}, \quad B = \{(i,j) | i = j\}.$$

The lists of elements in A and B are

$$A = \{(1,5),(2,4),(3,3),(4,2),(5,1)\},$$

$$B = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}.$$

Also

$$A \cap B = \{(3,3)\},$$

$$\overline{A} \cap B = \{(1,1),(2,2),(4,4),(5,5),(6,6)\},$$

$$A \cap \overline{B} = \{(1,5),(2,4),(4,2),(5,1)\}.$$

Therefore

$$\begin{aligned} \mathrm{n(A \cup B)} &= \mathrm{n(A \cap B) + n(\overline{A} \cap B) + n(\overline{B} \cap A)} \\ &= 1 + 5 + 4 = 10 \end{aligned}$$

39.6. A is the event that an ace is drawn, B is the event that a heart is drawn, and C is the event that a black card is drawn.
(a) $A \cap B$ is the event that the ace of hearts is drawn.
(b) $A \cap C$ is the event that the ace of clubs or spades is drawn.
(c) $A \cup B$ is the event that any ace or any heart is drawn.
(d) $A \cup B \cup C$ is the event that any ace or any heart or any black card is drawn.
(e) $A \backslash B$ is the event that the ace of clubs or the ace of diamonds or the ace of spades is drawn.
(f) $\overline{A} \backslash \overline{B}$ is the event that any heart other than the ace is drawn.
(g) $\overline{A} \backslash \overline{C}$ is the event a black card but not an ace is drawn.

(h) $(A \cap B) \cup C$ is the event that the ace of hearts or any black card is drawn.
(i) $(A \cap B) \cup (A \cap \overline{C})$ is the event an ace is drawn (the same as A).

39.7. (a) There are 4 kings in a pack of 52 cards. Hence

$$P(\text{king}) = \frac{4}{52} = \frac{1}{13}.$$

(b) As in (a) the probability that the first card is a king is $\frac{1}{13}$. Given that one king has been drawn there will be 3 kings in the remaining 51 cards. Hence

$$P(2\text{ kings}) = \frac{1}{13}\frac{3}{51} = \frac{1}{221}.$$

(c) If the second card is not a king it must be one of 48 remaining cards and the probability that this occurs is $\frac{48}{51}$. Similarly the probability that the third card is not a king is $\frac{47}{50}$ and so on. Hence

$$P(\text{king,not king,not king,king}) = \frac{1}{13}\frac{48}{51}\frac{47}{50}\frac{3}{49} = \frac{1128}{270725} = 0.004166\ldots.$$

39.8. This is an example of selection with replacement: no cards are drawn or removed from the pack at each cut unlike the preceding problem. The probability of one ace being shown is $\frac{1}{13}$. Therefore the probability of two aces being shown in two consecutive cuts is

$$P(\text{two aces}) = \frac{1}{13}\frac{1}{13} = \frac{1}{169}.$$

39.9. From Sections 1.17 and 39.4, the permutation ${}_nP_r$ is defined by

$${}_nP_r = \frac{n!}{(n-r)!}.$$

(a)
$${}_5P_3 = \frac{5!}{2!} = 60.$$

(b)
$${}_{10}P_4 = \frac{10!}{6!} = 10.9.8.7 = 5040.$$

(c)
$${}_7P_7 = \frac{7!}{0!} = 5040,$$

(0! is defined to be equal to 1: see eqn (1.38c)).

(d)
$${}_7P_1 = \frac{7!}{6!} = 7.$$

39.10. This is a permutation problem. The first letter can be chosen in 5 ways, the second in 4 ways, and so on. The number of different words which can be made up is

$${}_5P_3 = \frac{5!}{2!} = 5.4.3 = 60.$$

39.11. (a) Since the first digit cannot be zero, there are 9 possible numbers which can be chosen for the first digit. Since zero can chosen for any other position and numbers are selected without replacement there are 9 possible digits in the second position, 8 digits for the third and so on. Hence the number of possible distinct 5 digit numbers is

$$9.9.8.7.6 = 27216.$$

(b) With any number of repetitions the number of distinct 5 digit numbers is

$$9.10.10.10.10 = 90000.$$

(c) The final digit must be either 0 or 5 and so can be chosen in 2 ways. Since there is no replacement the first number can be chosen in 8 ways, the second in 7 ways, and so on. Hence the number of 5 digit numbers is

$$8.7.6.5.2 = 3360.$$

39.12. From Sections 1.17 and 39.4 (see also eqn (1.38c)), the combination ${}_nC_r$ is defined as

$${}_nC_r = \frac{n!}{(n-r)!r!}.$$

(a) $${}_7C_3 = \frac{7!}{4!3!} = 35.$$

(b) $${}_{99}C_{96} = \frac{99!}{3!96!} = 156849.$$

(c) $${}_{11}C_5 = \frac{11!}{6!5!} = 462.$$

39.13. Using the definition of ${}_nC_r$ (see previous problem),

$$\begin{aligned} {}_{n-1}C_r + {}_{n-1}C_{r-1} &= \frac{(n-1)!}{(n-1-r)!r!} + \frac{(n-1)!}{(n-r)!(r-1)!} \\ &= \frac{(n-1)!}{r!(n-r)!}[(n-r)+r] \\ &= \frac{n!}{r!(n-r)!} = {}_nC_r \end{aligned}$$

39.14. (a) Reverse the identity. Then, by the binomial theorem (1.44a) with $x = 1$,

$$\begin{aligned} 2^n = (1+1)^n &= 1 + {}_nC_1 + {}_nC_2 + \cdots + {}_nC_{n-1} + {}_nC_n \\ &= \sum_{r=0}^{n} {}_nC_r \end{aligned}$$

(b) By the binomial theorem with $x = 3$,

$$\begin{aligned} 4^n = (1+3)^n &= 1 + {}_nC_1 3 + {}_nC_2 3^2 + \cdots + {}_nC_{n-1} 3^{n-1} + {}_nC_n 3^n \\ &= \sum_{r=0}^{n} {}_nC_r 3^r \end{aligned}$$

39.15. This is a combination problem: order in the hands is immaterial. The number of ways in which 4 cards can be chosen from 52 is

$${}_{52}C_4 = \frac{52!}{48!4!} = \frac{52.51.50.49}{1.2.3.4} = 270725.$$

4 card hands from the same suit are taken from 13 cards and there are 4 suits. Hence the number of hands with cards from the same suit is

$$4 \times {}_{13}C_4 = \frac{4}{13!}9!4! = 2860.$$

The probability that a random hand of 4 cards contains cards from the same suit is, by counting, equal to the ratio of the two previous results, namely

$$\frac{4 \times {}_{13}C_4}{{}_{52}C_4} = \frac{2860}{270725} = 0.0105\ldots.$$

39.16. The number a_n of different n card hands which can be dealt from 52 cards is given by

$$a_n = {}_{52}C_n = \frac{52!}{(52-n)!n!}.$$

The number b_n of different n card hands of the same suit which can be dealt from 52 cards is given by

$$b_n = 4 \times {}_{13}C_n = \frac{4 \times 13!}{(13-n)!n!}.$$

The probability p_n that b_n occurs is given by

$$p_n = P(4 \text{ cards of same suit}) = \frac{b_n}{a_n} = \frac{4 \times {}_{13}C_n}{{}_{52}C_n} = \frac{4 \times 13!(52-n)!}{(13-n)!52!}.$$

The probabilities p_n have been computed for $n = 1, 2, 3, 4, 5$: they are shown in the table below.

n	1	2	3	4	5
p_n	1	0.235	0.517×10^{-1}	0.106×10^{-1}	0.198×10^{-2}

39.17. The box contains 22 balls of which 7 are red(r), 9 are white(w) and 6 are black(b). firstly, imagine that all the balls are individually distinguished in some way. The total number of 4 ball selections from the box is

$$N = {}_{22}C_4 = \frac{22!}{18!4!} = \frac{22.21.20.19}{1.2.3.4} = 7315.$$

the balls of one colour are indistinguishable. Noe consider how to take account of the fact that all
(a) 3 balls can be chosen from 7 red balls in ${}_7C_3$ ways, and then the white ball can be chosen in 9 ways. Therefore

$$P(3r + 1w) = \frac{9 \times {}_7C_3}{N} = \frac{315}{7315} = \frac{9}{209} = 0.043.$$

(b) 4 balls can be chosen from 7 red balls in ${}_7C_4 = 35$ ways. Hence

$$P(4r) = \frac{{}_7C_4}{N} = \frac{35}{7315} = \frac{1}{209} = 0.0048.$$

(c) The 4 balls are either all red, all white or all black. Therefore

$$\begin{aligned} P(4r \text{ or } 4w \text{ or } 4b) &= P(4r) + P(4w) + P(4b) = \frac{1}{N}({}_7C_4 + {}_9C_4 + {}_6C_4) \\ &= \frac{35 + 126 + 15}{7315} = \frac{176}{7315} \\ &= \frac{16}{665} = 0.024 \end{aligned}$$

(d) If a selection contains a least one ball of each of the three colours, then we can count instead the number of selections which contain not more than two colours and subtract this from the total number of possible selections. The number of different selections which contain no reds is ${}_{15}C_4$, the number which contain no whites is ${}_{13}C_4$ and the number which contain no blacks is ${}_{16}C_4$. Hence the probability that any selection contains at least one of each colour is

$$\begin{aligned}
1 - P(\text{not more than two colours}) &= \frac{{}_{15}C_4 + {}_{13}C_4 + {}_{16}C_4}{N} \\
&= \frac{1}{7315}\left[\frac{15!}{11!5!} + \frac{13!}{9!4!} + \frac{16!}{11!4!}\right] \\
&= \frac{683}{1463} = 0.467
\end{aligned}$$

39.18. (See Example 39.11.) Let

$A_1 = \{\text{event: component made by } M_1\}$
$A_2 = \{\text{event: component made by } M_2\}$
$B = \{\text{event: component not faulty}\}$

We are given the following probabilities:

$$P(A_1) = 0.7, \quad P(A_2) = 0.3, \quad P(B|A_1) = 0.89, \quad P(B|A_2) = 0.83.$$

(a) This is simply $P(B|A_1) = 0.89$.
(b) The events A_1 and A_2 are mutually exclusive so that, by the law of total probability (39.9)

$$\begin{aligned}
P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) \\
&= 0.89 \times 0.7 + 0.83 \times 0.3 = 0.872
\end{aligned}$$

(c) The event that a component is faulty is $\overline{B}$. We require the value of the conditional probability $P(A_1|\overline{B})$. By Bayes' theorem (39.13),

$$\begin{aligned}
P(A_2|\overline{B}) &= \frac{P(\overline{B}|A_2)P(A_2)}{P(\overline{B})} = \frac{[1 - P(B|A_2)]P(A_2)}{1 - P(B)} \\
&= \frac{(1 - 0.83)0.3}{1 - 0.872} \text{ (by (b) above)} \\
&= 0.40.
\end{aligned}$$

39.19. Let

$A_i = \{\text{event: component made by } M_i\}, (i = 1, 2, 3)$
$B = \{\text{event: component not faulty}\}$

We are given the following probabilities:

$$P(A_1) = 0.45, \quad P(A_2) = 0.30, \quad P(A_3) = 0.25,$$

$$P(B|A_1) = 0.87, \quad P(B|A_2) = 0.84, \quad P(B|A_3) = 0.91.$$

(a) The events A_1, A_2 and A_3 are mutually exclusive. Hence by the law of total probability (39.9)

$$\begin{aligned}
P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \\
&= 0.87 \times 0.45 + 0.84 \times 0.30 + 0.91 \times 0.25 \\
&= 0.871
\end{aligned}$$

(b) We require the conditional probability $P(A_2|\overline{B})$. By Bayes' theorem (39.13)

$$\begin{aligned}
P(A_2|\overline{B}) &= \frac{P(\overline{B}|A_2)P(A_2)}{P(\overline{B})} = \frac{[1 - P(B|A_2)]P(A_2)}{1 - P(B)} \\
&= \frac{(1 - 0.84)0.30}{1 - 0.87} \text{ (by (a) above)} \\
&= 0.37
\end{aligned}$$

(c) We require the conditional probability $P(A_1 \cup A_2|\overline{B})$. Since A_1 and A_2 are independent,

$$\begin{aligned} P(A_1 \cup A_2|\overline{B}) &= P(A_1|\overline{B}) + P(A_2|\overline{B}) \\ &= \frac{[1 - P(B|A_1)]P(A_1)}{1 - P(B)} + \frac{[1 - P(B|A_2)]P(A_2)}{1 - P(B)} \text{ (as in (b))} \\ &= \frac{(1 - 0.87)0.45}{1 - 0.87} + 0.37 \\ &= 0.45 + 0.37 = 0.82. \end{aligned}$$

39.20. The parallel and series probabilities are given in Example 39.12. The components with failure probabilities p_1, p_2 and p_3 are in parallel and the probability of failure is $p = p_1p_2p_3$. Similarly the probability of failure of the parallel components with failure probabilities r_1 and r_2 is $r = r_1r_2$. The components with probabilities p, q and r are in series. Hence by the probability is (see Example 39.12)

$$\begin{aligned} f &= (p + q + r) - (qr + rp + pq) + pqr \\ &= (p_1p_2p_3 + q + r_1r_2) - (qr_1r_2 + r_1r_2p_1p_2p_3 + p_1p_2p_3q) + p_1p_2p_3qr_1r_2 \end{aligned}$$

If the probability of failure in all components is the same, namely 0.98, then $f = 0.999953$.

39.21. In a batch of microprocessors, 5 are defective.
(a) The probability that one microprocessor chosen at random is defective is

$$P(1 \text{ defective}) = \frac{5}{100} = \frac{1}{20}.$$

(b) The first microprocessor will be chosen from 100 and the second from 4 defectives in 99. Hence

$$P(2 \text{ defective}) = \frac{5}{100}\frac{4}{99} = \frac{1}{495}.$$

(c) After 1 defective has been chosen, the second defective will be chosen from 4 in 99. Hence

$$P(\text{second defective given first defective}) = \frac{4}{99}.$$

39.22. The order in which the numbers are chosen is immaterial. Hence the number of ways in which 6 numbers can be chosen from 49 is

$${}_{49}C_6 = \frac{49!}{43!6!} = 13983816.$$

Hence the probability p_6 of obtaining the 6 numbers is

$$p_6 = P(6 \text{ correct}) = \frac{1}{13983816}.$$

For 5 correct numbers, the 5 numbers must be chosen from the 6 which can occur in ${}_6C_5$ but the remaining number must not be one of the 6: this can be chosen in ${}_{43}C_1$ ways. Hence the probability that 5 correct numbers are obtained is

$$p_5 = P(5 \text{ correct}) = \frac{{}_6C_5\,{}_{43}C_1}{{}_{49}C_6} = \frac{43}{2330636} = 0.000018.$$

Similarly the remaining probabilities are

$$p_4 = P(4 \text{ correct}) = \frac{{}_6C_4\,{}_{43}C_2}{{}_{49}C_6} = \frac{645}{665896} = 0.000969,$$

$$p_3 = P(3 \text{ correct}) = \frac{{}_6C_3\,{}_{43}C_3}{{}_{49}C_6} = \frac{8815}{499422} = 0.017650.$$

The probability of winning with one lottery ticket is

$$p = p_3 + p_4 + p_5 + p_6 = 0.018637,$$

which is a probability of about 1 in 53.7 of winning.

The probability of obtaining 5 correctly from the 6 is p_5 and the probability of correctly obtaining the bonus ball is 1/42. Hence

$$q = P(5 \text{ correct and bonus ball}) = \frac{p_5}{42} = \frac{43}{97886712}.$$

39.23. The probability of obtaining 1 head and $(n-1)$ tails from the n players is

$$_nC_{n-1}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{n-1},$$

since the order of the $(n-1)$ tails is immaterial. The probability of obtaining 1 tail and $(n-1)$ heads is the same. Hence the probability that the game ends at a given play is

$$2_nC_{n-1}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{n-1} = \frac{n}{2^{n-1}}.$$

The probability that it does not end is

$$1 - \frac{n}{2^{n-1}}.$$

The probability that the game finishes at the ith play is the probability that is continues for $r = 1, 2, \ldots, i-1$ and ends at the next play. Hence

$$P(\text{game ends at } i\text{th}) = \frac{n}{2^{n-1}}\left(1 - \frac{n}{2^{n-1}}\right)^{i-1}.$$

This is a geometric distribution (see Section 40.5 in the following chapter). The mean number of plays is

$$\begin{aligned}\mu &= \sum_{i=1}^{\infty}\frac{ni}{2^{n-1}}\left(1-\frac{n}{2^{n-1}}\right)^{i-1}\\ &= \frac{n}{2^{n-1}}\sum_{i=1}^{\infty} i\left(1-\frac{n}{2^{n-1}}\right)^{i-1}\end{aligned}$$

To find the sum of this series consider the series

$$S = \sum_{i=1}^{\infty} ix^i.$$

Multiply both sides by x and subtract from the series for S so that

$$(1-x)S = \sum_{i=1}^{\infty} x^{i-1} = \frac{1}{1-x},$$

the latter series being geometric. Therefore $S = 1/(1-x)^2$ so that

$$\mu = \frac{n}{2^{n-1}}\frac{1}{[1-(1-\frac{n}{2^{n-1}})]^2} = \frac{2^{n-1}}{n}.$$

(*Comment*: the last part of this problem would be more appropriate in Chapter 40.)

Random variables and probability distributions

40.1. Let H denote heads and T tails. The sample space of possible outcomes is

$$S = \{(HHH), (THH), (HTH), (HHT), (TTH), (THT), (HTT), (TTT)\}.$$

If X is the random variable of the number of heads (the 'events' being considered) define a sample space S_X is given by

$$S_X = \{0, 1, 2, 3\}.$$

Since $P(H) = 0.45$ and $P(T) = 0.55$, it follows from S that the probability distribution of X is

x_i	0	1	2	3
p_i	$0.55^3 =$ 0.166	$0.55^2 \times 0.45 \times 3 =$ 0.408	$0.55 \times 0.45^2 \times 3 =$ 0.334	$0.45^3 =$ 0.91

The probability that $X \geq 1$ is

$$P(X \geq 1) = p_1 + p_2 + p_3 = 0.833.$$

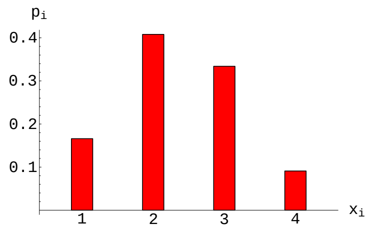

Figure 1: Problem 40.1

40.2. Obviously $p_j > 0$ and

$$\begin{aligned}
\sum_1^\infty \frac{1}{3}\left(\frac{1}{2^j} + \frac{1}{2^{j-1}}\right) &= \frac{1}{3}\left(\sum_{j=1}^\infty \frac{1}{2^j} + \sum_{j=1}^\infty \frac{1}{2^{j-1}}\right) \\
&= \frac{1}{3}\left(\sum_{j=1}^\infty \frac{1}{2^j} + 2\sum_{j=1}^\infty \frac{1}{2^j}\right) \\
&= \sum_{j=1}^\infty \frac{1}{2^j} \\
&= \frac{\frac{1}{2}}{1-\frac{1}{2}} = 1,
\end{aligned}$$

the series being geometric (see Section 1.15). Hence the sequence $\{p_j\}$ is a possible probability distribution. The probability that the random variable $X \geq 6$ is given by

$$\begin{aligned}
P(X \geq 6) &= \sum_{j=6}^\infty p_j = \frac{1}{3}\sum_{j=6}^\infty \left(\frac{1}{2^j} + \frac{1}{2^{j-1}}\right) = \frac{1}{3}\left(\frac{1}{2^5} + \frac{2}{2^5}\right) \\
&= \frac{1}{2^5} = \frac{1}{32}
\end{aligned}$$

40.3. (see Section 40.3.) The probability of i successes in n trials is the binomial distribution ${}_nC_i p^i(1-p)^{n-i}$ where p is the probability of success in any trial. In this problem $n = 12$, $p = \frac{1}{3}$ and $i = 0, 1, 2, \ldots, 12$. The computed probabilities are to 3 decimal places:

i	0	1	2	3	4	5	6
p_i	0.008	0.462	0.127	0.212	0.238	0.191	0.111

i	7	8	9	10	11	12
p_i	0.048	0.015	0.003	0.000	0.000	0.000

From Section 40.4, the mean and standard deviation of the binomial distribution are p and $\sqrt{np(1-p)}$ respectively. Hence

$$\text{mean} = 12 \times \frac{1}{3} = 4,$$

$$\text{standard deviation} = \sqrt{12 \times \frac{1}{3} \times \frac{2}{3}} = 1.633.$$

40.4. The probability density function and the corresponding cumulative density function of the

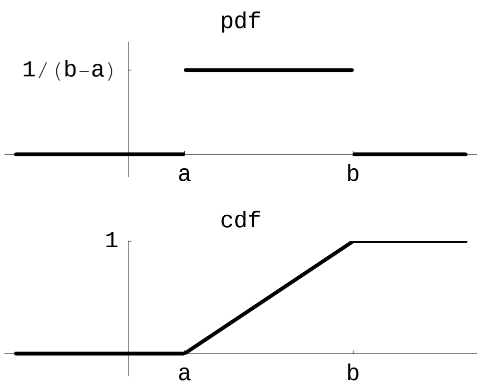

Figure 2: Problem 40.4: pdf and cdf for the uniform distribution on (a, b).

uniform distribution

$$f(x) = \begin{cases} 1/(b-a) & a < x < b \\ 0 & \text{elsewhere} \end{cases}$$

are shown in Figure 2. The mean is given by (see (40.12))

$$\begin{aligned} \mu &= E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x = \int_a^b \frac{x\mathrm{d}x}{b-a} = \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b \\ &= \frac{1}{2}(b+a). \end{aligned}$$

The variance of the uniform distribution is

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) \mathrm{d}x = \int_a^b (x-\mu)^2 \frac{1}{b-a} \mathrm{d}x \\ &= \frac{1}{b-a}\left[\frac{1}{3}(x-\mu)^3\right]_a^b = \frac{1}{3(b-a)}\left[(b-\mu)^3 - (a-\mu)^3\right] \\ &= \frac{1}{3(b-a)}\left[\frac{1}{8}(b-a)^3 - \frac{1}{8}(a-b)^3\right] \\ &= \frac{1}{12}(b-a)^2 \end{aligned}$$

The standard deviation is therefore

$$\sigma = \frac{1}{2\sqrt{3}}(b-a).$$

40.5. From Section 40.5 the mean μ of the geometric distribution $p_i = (1-p)^{i-1}p$ is $\mu = 1/p$. Let $q = 1-p$. The variance of the geometric distribution is given by

$$\begin{aligned}\sigma^2 &= E(X^2) - \mu^2 = p\sum_{i=1}^{\infty} i^2 q^{i-1} - \mu^2 = p(1 + 2^2 q + 3^2 q^2 + \cdots) - \mu^2 \\ &= p\frac{\mathrm{d}}{\mathrm{d}q}(q + 2q^2 + 3q^3 + \cdots) - \mu^2\end{aligned}$$

Let

$$S = q + 2q^2 + 3q^3 + \cdots.$$

Then

$$S - qS = q + q^2 + q^3 + \cdots = \frac{q}{1-q},$$

which is the sum of a geometric series. Hence

$$S = \frac{q}{(1-q)^2}.$$

Therefore

$$\begin{aligned}\sigma^2 &= p\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{q}{(1-q)^2}\right] - \mu^2 = p\left[\frac{1}{(1-q)^2} + \frac{2q}{(1-q)^3}\right] - \mu^2 \\ &= \frac{1}{p} + \frac{2}{p^2} - \frac{2}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}\end{aligned}$$

40.6. Let $p = 0.012$ be the probability that a particular component is faulty. Let p_i be the probability that the first faulty component occurs for the ith component. Then,

$$p_i = (1-p)^{i-1}p,$$

since the first $i-1$ components must be not faulty (with probability $1-p$). This is a geometric distribution. From Section 40.5 and Problem 40.5, the mean of the geometric distribution is given by $\mu 1/p$ but this includes the faulty component. If it is excluded the average number of component until the first faulty component is

$$\frac{1}{p} - 1 = \frac{1-p}{p} = 82.33.$$

The variance is given by (it is not affected by the the exclusion or not of the first faulty component)

$$\sigma^2 = \sum_{i=1}^{\infty} i^2 p_i - \mu^2 = \frac{1-p}{p^2}.$$

The standard deviation is therefore

$$\sigma = \frac{\sqrt{1-p}}{p} = 82.83.$$

40.7. This requires the geometric distribution with $p = \frac{1}{2}$. For the geometric distribution (see (40.9))

$$p_i = (1-p)^{i-1}p.$$

For 8 heads before the first tail, put $i = 9$. Then

$$p_9 = \left(\frac{1}{2}\right)^8 \frac{1}{2} = \frac{1}{2^9}.$$

40.8. We must have $X = i$ if success occurs at the ith trial and $r - 1$ successes have occurred previous to that. The order in which the $r - 1$ successes happen in the $i - 1$ trials is not material, so that

$$p_i = {}_{i-1}C_{r-1} p^r (1-p)^{i-r}, \quad (i = r, r-1, \ldots),$$

which is the negative binomial distribution. The sum of the probabilities which start at $i = r$ (it is possible that the first r trials could all be successes) is

$$\begin{aligned}
\sum_{i=r}^{\infty} p_i &= p^r \sum_{i=r}^{\infty} {}_{i-1}C_{r-1} p^r (1-p)^{i-r} \\
&= p^r \left[{}_{r-1}C_{r-1} + {}_{r}C_{r-1}(1-p) + {}_{r+1}C_{r-1}(1-p)^2 + \cdots \right] \\
&= p^r \left[r + r(1-p) + \frac{r(r+1)}{2!}(1-p)^2 + \cdots \right] \\
&= rp^r \frac{1}{[1-(1-p)]^{r+1}} \text{ (by the binomial expansion)} \\
&= \frac{r}{p}
\end{aligned}$$

The mean value of X is given by

$$\begin{aligned}
E(X) &= \sum_{i=r}^{\infty} i p_i = p^r \sum_{i=r}^{\infty} i_{i-1}C_{r-1} p^r (1-p)^{i-r} \\
&= p^r \left[r_{r-1}C_{r-1} + (r+1)_r C_{r-1}(1-p) + (r+2)_{r+1}C_{r-1}(1-p)^2 + \cdots \right] \\
&= rp^r \left[1 + (r+1)(1-p) + \frac{(r+1)(r+2)}{2!}(1-p)^2 + \cdots \right] \\
&= \frac{rp^r}{[1-(1-p)]^{r+1}} = \frac{r}{p}
\end{aligned}$$

The expected value of X^2 is given by (replacing $1 - p$ by q)

$$\begin{aligned}
E(X^2) &= \sum_{i=r}^{\infty} i^2 p_i = p^r \sum_{i=r}^{\infty} i^2{}_{i-1}C_{r-1} p^r (1-p)^{i-r} \\
&= p^r \left[r^2{}_{r-1}C_{r-1} + (r+1)^2{}_r C_{r-1}(1-p) + (r+2)^2{}_{r+1}C_{r-1}(1-p)^2 + \cdots \right] \\
&= rp^r \left[r + (r+1)^2 q + \frac{(r+1)(r+2)^2}{2!} q^2 + \cdots \right] \\
&= \frac{rp^r}{q^{r-1}} \frac{\partial}{\partial q} \left[q^r + (r+1)q^{r+1} + \frac{(r+1)(r+2)}{2!} + \cdots \right] \\
&= \frac{rp^r}{q^{r-1}} \frac{\partial}{\partial q} \left[\frac{q^r}{(1-q)^{r+1}} \right] = \frac{rp^r}{q^{r-1}} \left[\frac{rq^{r-1}}{(1-q)^{r+1}} + \frac{(r+1)q^r}{(1-q)^{r+2}} \right] \\
&= \frac{r(r+1-p)}{p^2}.
\end{aligned}$$

Finally, the variance is given by

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{r(r+1-p)}{p^2} - \frac{r^2}{p^2} = \frac{r(1-p)}{p^2}.$$

40.9. This is the problem as described at the beginning of Section 40.5. We are interested in the number of bottles filled until the first failure. If X is this random variable (including the first failure) and p is the probability that an individual bottle fails the weight test, then

$$P(X = i) = (1-p)^i p,$$

which is a geometric distribution. The expected value until the first breakdown is $E(X) = 1/p$. On average this is 1503 bottles. Hence $E(X) = 1503 = 1/p$ so that the probability of an individual bottle failing is $p = 1/1503 = 0.000665\ldots$.

40.10. The random variable X has the exponential distribution

$$f(t) = \begin{cases} 1.5\mathrm{e}^{-1.5t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

With $\alpha = 1.5$,
(a) $P(0 < X < 1) = \int_0^1 \alpha \mathrm{e}^{-\alpha t}\mathrm{d}t = [-\mathrm{e}^{-\alpha t}]_0^1 = 1 - \mathrm{e}^{-\alpha} = 0.777$;

(b) $P(X < 0) = 0$;

(c) $P(X \geq 1) = \int_1^\infty \alpha \mathrm{e}^{-\alpha t}\mathrm{d}t = [-\mathrm{e}^{-\alpha t}]_1^\infty = \mathrm{e}^{-\alpha} = 0.223$;

(d) $P(X \leq 1) = 1 - \mathrm{e}^{-\alpha} = 0.777$, by (a) since $f(t) = 0$ for $t < 0$;

(d) $P(X < 2)$ or $P(X < 1) = \int_0^2 \alpha \mathrm{e}^{-\alpha t}\mathrm{d}t = 1 - \mathrm{e}^{-2\alpha} = 0.950$.

40.11. The distribution is exponential with density function

$$f(t) = \begin{cases} 1.5\mathrm{e}^{-1.5t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

where t is the time between calls in minutes. The mean of the exponential distribution is $1/\alpha$. Therefore, since the mean time between calls is 20 minutes, $\alpha = 1/20 = 0.05$.
(a) The probability that there are no calls in a one-hour interval is

$$P(X > 60) = \int_{60}^\infty \alpha \mathrm{e}^{-\alpha t}\mathrm{d}t = [-\alpha \mathrm{e}^{-\alpha t}]_{60}^\infty = \mathrm{e}^{-3} = 0.050.$$

(b) The probability that at least one call within a 15-minute interval is

$$P(X < 15) = \int_0^{15} \alpha \mathrm{e}^{-\alpha t}\mathrm{d}t = [-\mathrm{e}^{-\alpha t}]_0^{15} = 1 - \mathrm{e}^{-15/20} = 0.528.$$

40.12. For

$$P(X = n) = \frac{\mathrm{e}^{-\lambda}\lambda^n}{n!}, \quad (n = 0, 1, 2, \ldots),$$

the mean is given by

$$\begin{aligned} E(X) &= \sum_{n=1}^\infty n\frac{\mathrm{e}^{-\lambda}\lambda^n}{n!} = \mathrm{e}^{-\lambda}\sum_{n=1}^\infty \frac{\lambda^n}{(n-1)!} \\ &= \mathrm{e}^{-\lambda}\lambda\left(1 + \lambda + \frac{\lambda^2}{2!} + \cdots\right) \\ &= \mathrm{e}^{-\lambda}\lambda\mathrm{e}^{\lambda} = \lambda. \end{aligned}$$

The variance is given by

$$\sigma^2 = E(X^2) - E(X)^2 = \sum_{n=1} \frac{\mathrm{e}^{-\lambda}\lambda^n}{n!} - \lambda^2$$

$$
\begin{aligned}
&= e^{-\lambda}\left(\lambda + 2\lambda^2 + \frac{3\lambda^3}{2!} + \cdots\right) - \lambda^2 \\
&= e^{-\lambda}\lambda\frac{d}{d\lambda}\left(\lambda + \lambda^2 + \frac{\lambda^3}{2!} + \cdots\right) - \lambda^2 \\
&= e^{-\lambda}\lambda\frac{d}{d\lambda}(\lambda e^{\lambda}) - \lambda^2 = e^{-\lambda}(\lambda e^{\lambda} + \lambda^2 e^{\lambda}) - \lambda^2 \\
&= \lambda
\end{aligned}
$$

The probability that 5 or more hits occur in the time interval is

$$
\begin{aligned}
P(X \geq 5) &= \sum_{n=5}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} \\
&= e^{-\lambda}\left(e^{\lambda} - 1 - \lambda - \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} - \frac{\lambda^4}{4!}\right)
\end{aligned}
$$

40.13. The standardized normal distribution has the probability density function

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

The numerical values can be calculated from the cumulative distribution table in Appendix H(b).

(a)
$$P(Z \geq 0.8) = \frac{1}{\sqrt{2\pi}} \int_{0.8}^{\infty} e^{-\frac{1}{2}z^2} dz = 0.212.$$

(b)
$$P(Z \leq 0.7) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0.7} e^{-\frac{1}{2}z^2} dz = 0.758.$$

(c)
$$P(-0.5 \leq Z \leq 0.8) = \frac{1}{\sqrt{2\pi}} \int_{-0.5}^{0.8} e^{-\frac{1}{2}z^2} dz = 0.480.$$

40.14. The density function is

$$f(t) = \begin{cases} 0.1 & 33 < t < 43 \\ 0 & \text{elsewhere} \end{cases}$$

Refer back to Problem 40.4 and put $a = 33$ and $b = 43$. The mean and variance of the operation are respectively

$$\mu = \frac{1}{2}(b + a) = \frac{1}{2}(43 + 33) = 38,$$

$$\sigma^2 = \frac{1}{12}(b - a)^2 = \frac{1}{12}(43 - 33)^2 = \frac{100}{12} = 8.33.$$

The probability that an operation takes longer than 40 secs is

$$P(X \leq 40) = \int_{40}^{43} 0.1 dt = 0.1 \times 3 = 0.3.$$

Therefore about a third of the operations take longer than 40secs.

40.15. The function

$$f(t) = \begin{cases} A(a^2 - t^2) & -a \leq t \leq a \\ 0 & \text{elsewhere} \end{cases}$$

is a probability density function if $f(t) \geq 0$ and

$$\int_{-\infty}^{\infty} f(t)\mathrm{d}t = 1.$$

Hence

$$\int_{-\infty}^{\infty} f(t)\mathrm{d}t = \int_{-a}^{a} A(a^2 - t^2)\mathrm{d}t = A\left[a^2 t - \frac{1}{3}t^3\right]_{-a}^{a} = \frac{4a^3}{3} = 1,$$

if $A = 3/(4a^3)$. Therefore

$$f(t) = \begin{cases} \frac{3(a^2-t^2)}{4a^3} & -a \leq t \leq a \\ 0 & \text{elsewhere} \end{cases}$$

The variance is given by

$$\sigma^2 = \frac{3}{4a^3}\int_{-a}^{a} t^2(a^2 - t^2)\mathrm{d}t = \frac{3}{4a^3}\left[\frac{a^2t^3}{3} - \frac{t^5}{5}\right]_{-a}^{a} = \frac{a^2}{5}.$$

The standard deviation is given by $\sigma = a/\sqrt{5}$ which is 1 if $a = \sqrt{5}$.

40.16. The engines are modelled by $N(1200, \sigma^2)$. The normalized standard random variable is

$$Z = \frac{X - 1200}{\sigma},$$

where X is the random variable of the time to failure. When $X = 1000$, then $Z = -200/\sigma$. We require σ such that

$$P\left(Z \geq \frac{-200}{\sigma}\right) = \int_{-200/\sigma}^{\infty} \frac{\mathrm{e}^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\mathrm{d}z = 0.95.$$

From the table in Appendix H(b), it follows that σ should satisfy

$$\frac{200}{\sigma} = 1.645, \text{ or } \sigma = \frac{200}{1.645} = 121.6.$$

40.17. (see Section 40.8.) The exponential distribution has the probability density function

$$f(t) = \begin{cases} \alpha\mathrm{e}^{-\alpha t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

Let X be the random variable of the time to failure. The mean of the exponential distribution is $E(X) = 1/\alpha = 500$ hours. Hence $\alpha = 1/500 = 0.002$.

The probability that a bulb is still functioning after 640 hours is

$$P(X \geq 640) = \int_{640}^{\infty} \alpha\mathrm{e}^{-\alpha t}\mathrm{d}t = [-\mathrm{e}^{-\alpha t}]_{640}^{\infty} = \mathrm{e}^{-640/500} = 0.278.$$

Two bulbs will have failed after 640 hours with probability 0.278 and therefore two bulbs will still be working with probability $(1 - 0.278)^2$. The two failing bulbs can be chosen in ${}_4C_2$ ways. Hence the probability that two bulbs are still working is

$${}_4C_2 \times (0.278)^2 \times (0.722)^2 = 6 \times (0.278)^2 \times (0.722)^2 = 0.242.$$

Chapter 41: Descriptive statistics

41.1. (a) The set of data is

$$\{10, 11, 11, 15, 17, 20, 25, 25, 27, 30, 38, 42, 47\},$$

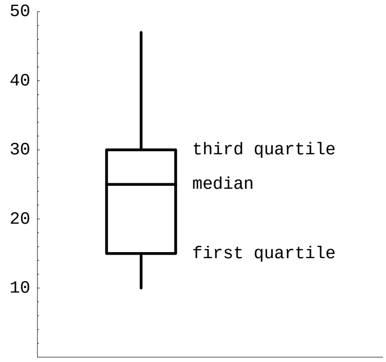

Figure 3: Problem 41.1(a)

which contains 13 observations. The mean is 24.5 and the median is 25. The first quartile is the median of the first 7 observations, that is, 15, and the third quartile is the median of the last 7 observations, that is, 30. The boxplot is shown in Figure 3.
(b) The set of data is

$$\{5, 12, 15, 16, 20, 29, 29, 32, 39, 44\},$$

which contains 10 observations. The mean is 24.1 and the median 24.5. The first quartile is 15 and the third quartile is 32. The boxplot is shown in Figure 4.

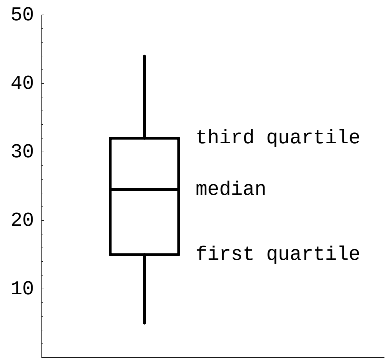

Figure 4: Problem 41.1(b)

41.2. The marks in an examination with 4 papers are are shown in the table. The means and medians are

$$\text{mean(paper 1)} = 54.65, \quad \text{median(paper 1)} = 60.5,$$
$$\text{mean(paper 2)} = 54.60, \quad \text{median(paper 1)} = 52.5,$$
$$\text{mean(paper 3} = 51.65, \quad \text{median(paper 1)} = 49.5,$$
$$\text{mean(paper 4} = 50.95, \quad \text{median(paper 1)} = 52.5,$$

Paper 1	{24,27,27,30,40,42,48,55,58,60, 61,63,64,66,66,68,69,72,78,85},
Paper 2	{30,35,36,38,39,40,44,45,48,51, 54,58,61,64,65,65,69,70,81,90},
Paper 3	{26,29,30,35,36,37,46,48,49,49, 50,54,56,61,69m70,71,71,72,74}
Paper 4	{10,20,22,34,41,44,45,45,45,50, 55,55,55,56,64,65,66,70,85,91}

The boxplots for the four papers can be compared in Figure 4.

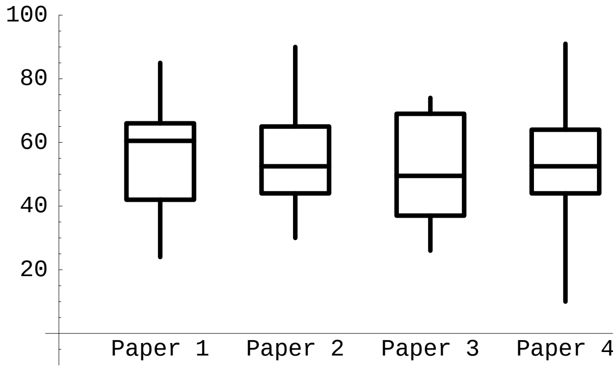

Figure 5: Problem 41.2

41.3. The samples of weights of 10 packets are

$$\{X_i\} = \{25.1, 25.3, 25.0, 25.7, 25.3, 25.2, 25.1, 25.5, 25.7, 25.1\}.$$

The various measures of the data are:

sample mean $= \overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 25.3$;
mode = most often X_i in sample = 25.1;
sample variance $= S^2 = \frac{1}{9}\sum_{i=1}^{10}(X_i - \overline{X})^2 = 0.0644$;
standard deviation $= S = 0.254$;

41.4. The set of data is reproduced in the table, and its histogram over 10 intervals and the frequency polygon are shown in Figure 6.

length interval	frequency of pipes	length interval	frequency of pipes
$9.5 \leq x < 9.6$	1	$10 \leq x < 10.1$	21
$9.6 \leq x < 9.7$	4	$10.1 \leq x < 10.2$	15
$9.7 \leq x < 9.8$	5	$10.2 \leq x < 10.3$	11
$9.8 \leq x < 9.9$	12	$10.3 \leq x < 10..4$	5
$9.9 \leq x < 10.0$	20	$10.4 \leq x < 10.5$	2

(b) The numbers of pipe lengths in each intervals of length 0.2m are listed in the second table. The corresponding bar chart and frequency polygon are shown in Figure 7.

length interval	frequency of pipes	length interval	frequency of pipes
$9.5 \leq x < 9.7$	5	$10.1 \leq x < 10.3$	26
$9.7 \leq x < 9.9$	17	$10.3 \leq x < 10.5$	7
$9.9 \leq x < 10.1$	41		

41.5. (See Section 41.1.) 127 observations over 36 intervals would average about 3 or 4 observations per interval or bin which is too small for interpretation. The general working rule is about $\sqrt{n}$ bins for n observations which would reduce the number of intervals to about 11.

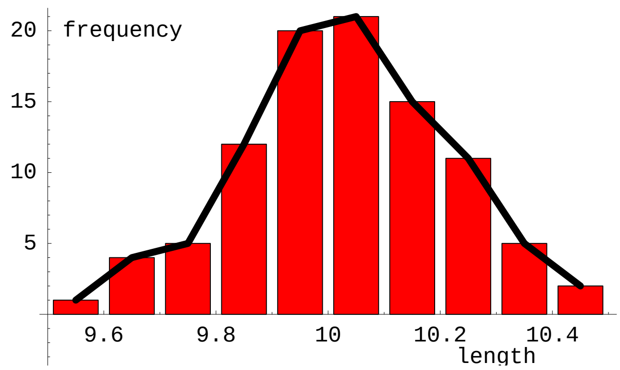

Figure 6: Problem 41.4(a)

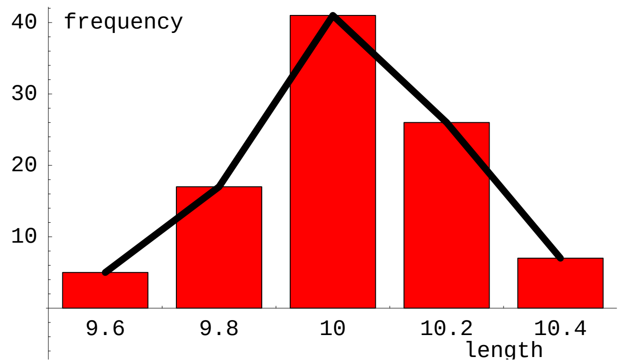

Figure 7: Problem 41.4(b)

41.6. (See Section 41.3 and Problem 40.4 for the uniform distribution.) The uniform distribution is

$$f(x) = \begin{cases} 1, & 1 \leq x \leq 2; \\ 0, & \text{elsewhere} \end{cases}$$

Let the sample of values be $\{X_i\}$ for $i = 1, 2, \ldots 35$. The sample mean $\overline{X}$ is simply the mean of the sample, namely

$$\overline{X} = \frac{1}{35} \sum_{i=1}^{35} X_i.$$

The expected value of the sample mean is the mean of the random variable X, which has a uniform distribution. The mean of the uniform distribution is

$$\mu = \int_1^2 x \mathrm{d}x = \left[\frac{1}{2}x^2\right]_1^2 = \frac{3}{2}.$$

Hence $E(\overline{X}) = \frac{3}{2}$. The variance of the sample mean is (see (41.4))

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{1}{12 \times 35} = \frac{1}{420}$$

for the uniform distribution. From (41.5) an estimator of the sample variance is

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{34} (X_i - \overline{X})^2.$$

41.7. The sample values are the 7 numbers

$$\{X_i\} = \{9.71, 10.26, 9.80, 9.85, 9.99, 10.10, 9.79\}.$$

the sample mean is

$$\overline{X} = \frac{1}{7}\sum_{i=1}^{7} = \frac{1}{7}(9.71 + 10.26 + 9.80 + 9.85 + 9.99 + 10.10 + 9.79) = 9.93.$$

An estimator for the sample variance is given by (41.5), namely,

$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum_{i=1}^{7}(X_i - \overline{X})^2 \\ &= \frac{1}{6}[(9.71-9.93)^2 + (10.26-9.93)^2 + (9.80-9.93)^2 + (9.85-9.93)^2 \\ &\quad +(9.99-9.93)^2 + (10.10-9.93)^2 + (9.79-9.93)^2] \\ &= 0.039 \end{aligned}$$

41.8. (See Example 41.1.) We require k_1 and k_2 so that

$$P(1460 < T < 1540) = \frac{1}{\sqrt{2\pi}}\int_{k_1}^{k_2} \mathrm{e}^{-\frac{1}{2}x^2}\mathrm{d}x. \quad \text{(i)}$$

The mean and variance of X, the random variable that a 1 appears at a single throw of the die, are

$$E(X) = \frac{1}{6}, \quad \mathrm{Var}(X) = \frac{5}{36}.$$

By the central limit theorem

$$P\left(k_1 \le \frac{T - \frac{1}{6}9000}{\frac{\sqrt{5}}{6}\sqrt{9000}} \le k_2\right) = \frac{1}{\sqrt{2\pi}}\int_{k_1}^{k_2} \mathrm{e}^{-\frac{1}{2}x^2}.$$

Comparison with (i) gives

$$\frac{\sqrt{5}}{6}\sqrt{9000}k_2 + \frac{1}{6}9000 = 1540,$$

$$\frac{\sqrt{5}}{6}\sqrt{9000}k_1 + \frac{1}{6}9000 = 1460.$$

Therefore

$$k_2 = \frac{6(1540-1500)}{\sqrt{5}\sqrt{9000}} = 1.131,$$

$$k_1 = \frac{6(1460-1500)}{\sqrt{5}\sqrt{9000}} = -1.131.$$

41.9. The data of the fuel consumption versus car weight is reproduced in the table and displayed in Figure 8.

Vehicle	weight w_i(kg)	fuel consumption (x_i) (km l^{-1})
A	2100	4.96
B	1350	9.10
C	1008	12.04
D	1323	7.68
E	710	15.15
F	1215	10.98
G	1436	7.75
H	1561	8.25
I	2120	4.85
J	1975	4.64
K	1535	5.56

The regression line which is the line of least squares fit to the data has the equation $c = aw + b$, where a and b are given by the linear equations (see (41.7))

$$a \sum_{i=1}^{11} w_i^2 + b \sum_{w_i}^{11} w_i = \sum_{=1}^{11} w_i c_i,$$

$$a \sum_{i=1}^{11} w_i + 11b = \sum_{i=1}^{11} c_i.$$

The solutions of these equations are $a = -0.00707$ and $b = 18.76$, so that the regression line estimator is

$$\hat{c} = -0.00707w + 18.76.$$

The regression line is also shown in Figure 8. It is given that an unbiased estimator in linear

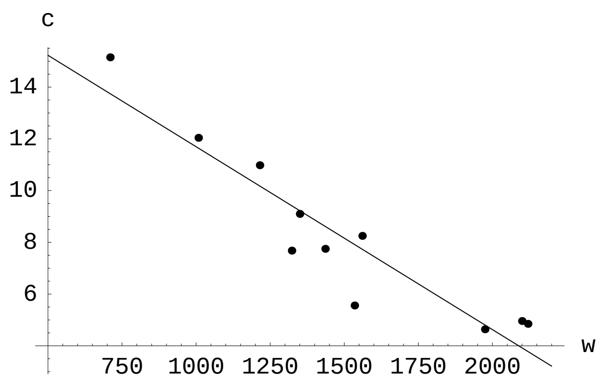

Figure 8: Problem 41.9

regression is

$$S^2 = \sum_{i=1}^{n} \frac{(c_i - \hat{c}_i)^2}{n-2},$$

where $\hat{c}_i = aw_i + b$. The numbers $\hat{c}_i$ are given in the table.

	w_i	$\hat{c}_i$		w_i	$\hat{c}_i$
A	2100	3.92	G	1436	8.61
B	1350	9.22	H	1561	7.73
C	1008	11.64	I	2120	3.78
D	1323	9.41	J	1975	4.80
E	710	13.74	K	1535	7.91
F	1215	10.18			

It follows that an estimate for the variance is $S^2 = 1.56$.

The outlier is vehicle K. If it is deleted from the list, then the regression line estimator has the equation

$$\hat{c} = -0.00700w + 18.90.$$

If K is deleted from the table above the estimator for the variance becomes $S^2 = 1.3$, which is a significant change.