ANSWERS TO CHAPTER 2

Probability

2.1 (a) 4/52 or 1/13. (There are 4 aces in a pack of 52 cards.)

(b) 12/52 or 3/13. Again, this is calculated simply by counting the possibilities, 12 court cards (3 in each suit).

(c) 1/2.

(d) $4/52 \times 3/52 \times 2/52 = 3/17{,}576$ (0.017%).

(e) (4/52)3 = 0.000455.

2.2 (a) Proportions unemployed in the various categories may be interpreted as probabilities according to the frequentist approach. Thus, given someone aged 16–19 who is out of work, the probability that they have been unemployed for ≤ eight weeks is 27.2%. Of course, this pattern of unemployment might change over time, altering the probabilities in the future.

(b) (i) True, (ii) false (19% is the probability of an unemployed person aged 16–19 being out of work for over one year, not the other way round), (iii) false (56.2% is the proportion of all currently unemployed (whenever they became unemployed) who are out of work for over one year), (iv) false (8.9% is the proportion of the unemployed, not the economically active, out of work for less than eight weeks), (v) true.

(c) 1,022.7 have been out of work over one year. (This is 19% × 273.4 + 36.8% × 442.5 + … and so on.) Of these, 51.9 (19% of 273.4, all in thousands) are aged 16–19, so the probability is 5.08% (= 51.9/1,022.7).

2.3 (a) 0.25 (three to one against means one win for every three losses, so one win in four races), 0.4, 5/9.

(b) 'Probabilities' are 0.33, 0.4, and 0.5, which sum to 1.23. Therefore, these cannot be real probabilities. The difference leads to a (expected) gain to the bookmaker.

(c) Suppose the true probabilities of winning are proportional to the odds, that is. 0.33/1.23, 0.4/1.23, 0.5/1.23, or 0.268, 0.325, 0.407. If £1 was bet on each horse, then the bookie would expect to pay out $0.268 \times 3 + 0.325 \times 1.5 + 0.407 \times 0.8 = 1.6171$, plus one of the £1 stakes, that is £2.62 in total. He would thus gain 38 pence on every £3 bet, or about 12.7%.

2.4 (a) 8/21 (0.381), 0.667, 0.231.

(b) 0.500 + 0.286 + 0.154 + 0.100 + 0.059 = 1.099.

(c) Basing odds on amounts bet allows the bookie to make a guaranteed profit (no uncertainty). Diverging from this towards the true probabilities (presumably better known by the bookie than the punter) increases the expected gain, but involves increased risk of a loss to the bookmaker.

2.5 A number of factors might help: statistical ones such as the ratio of exports to debt interest, the ratio of GDP to external debt, the public sector deficit and so on, and political factors such as the policy stance of the government. More insight could be gained by looking at the current interest rate on the debt. A high interest rate suggests investors believe there is a greater chance of default, other things equal.

2.6 Factors such as profitability, expected future incomes, size of debt relative to earnings and so on. The credit default swap (CDS) market provides insurance against a company not repaying its bonds, and the premium charged would reflect the market's view of that risk.

2.7 (a) is the more probable, since it encompasses her being active or not active in the feminist movement. Many people get this wrong, which shows how one's preconceptions can mislead. People tend to read part (a) as 'Judy is a bank clerk, not active in the feminist movement'. A simple way of stating this mathematically is $\Pr(B) \geq \Pr(B \text{ and } A)$ where B indicates a bank clerk and A indicates an activist. It has to be true since $\Pr(A) \leq 1$.

2.8 Not very good performance. If one was guessing at random ($P = 0.5$), the probability of getting two or less wrong is $6C0 \times 0.56 + 6C1 \times 0.56 + 6C2 \times 0.56 = 34\%$. Hence, the evidence is reasonably consistent with someone just guessing. If the clinic got only one wrong out of six, that would be more convincing. The chance of guessing correctly at random five out of six times is 11%.

2.9 The advertiser is a trickster and guesses at random. Every correct guess (P = 0.5) nets a fee, every wrong one costs nothing except reimbursing the fee. The trickster would thus keep half the money sent in. You should be wary of such advertisements!

2.10 Pr(winning) = (1/6)6 = 0.000021433 (about 2 in 100,000). Your expected winnings are therefore $0.000021433 \times 40\,000 - 1 = -£0.143$. With 400 punters during the fair, the probability of the car not being won is therefore $(1 - 0.000021433)400 = 0.991463$; so, the probability of it being won is 0.008537. The expected loss is therefore $40{,}000 \times 0.008537 = £340$ approximately. This is less than the premium, but the premium also has to pay for someone independent to monitor the game, otherwise the organisers could cheat and claim the car had been won.

2.11 (a) $E(\text{winnings}) = 0.520 \times £1\text{billion} + (1 - 0.520) \times -£100 = £853.67$.

(b) Despite the positive expected value, most would not play because of their aversion to risk. Would you? The size of the prize is also not credible – would they actually pay up?

2.12 (a) Accident probabilities: four-engine plane: $P = \Pr(3 \text{ or } 4 \text{ engines fail}) = 4C3 \times 0.0013 \times 0.999 + 0.0014 = 0.000000004001$; two-engine plane: $P = \Pr(2 \text{ engines fail}) = 0.0012 = 0.000001$. Hence, a four-engine plane is 'safer' by a factor of 250, on this basis.

(b) However, the probabilities of not crashing are 0.999999996 and 0.999999 respectively. Thus, the former is 0.00009% safer, not a lot.

(c) The crucial assumption is independence. Since the engines share many features (e.g. fuel supply), this is probably not accurate. Dependence would tend to reduce the differences between the types of aircraft. Within the limit, if engines all worked or failed together, both types would effectively be single-engine planes.

Flying regulations dictate that aircraft with fewer engines are required to follow a flight path that remains closer to airports. Modern aircraft are much more reliable so these regulations have been relaxed over the years.

2.13 (a), (b) and (d) are independent, though legend says that rain on St Swithin's Day means rain for the next 40 days, so (d) is arguable according to legend.

2.14 (a) is likely to be independent. The others are not. IBM and Dell operate in the similar markets so their profits are likely to be correlated. In football, it is possible for teams to have a winning (or losing) 'streak' (the better the team is though, the less it should suffer from this phenomenon). No claims bonuses in car insurance reflect the dependence – if you have an accident, you reveal the information that you're more likely to have another.

2.15 (a) There are 15 ways where a 4–2 score could be arrived at, of which this is one. Hence, the probability is 1/15.

(b) Six of the routes through the tree diagram involve a 2–2 score at some stage, so the probability is 6/15.

2.16 (a) 97.03% (= 0.993).

(b) 99.9996% Pr(first three get it right) + Pr(any two get it right) + Pr(any one plus the fourth computer get it right) = $0.993 + 0.992 \times 0.01 \times 3C2 + 0.99 \times 0.012 \times 3C1 \times 0.99$).

(c) 0.014.

(d) $1 - 0.999996 - 0.014 = 0.00000397$.

2.17 Pr(guessing all six) = $6/50 \times 5/49 \times ... \times 1/45 = 1/15{,}890{,}700$.

Pr(six from 10 guesses) = $10/50 \times 9/49 \times ... \times 5/45 = 151{,}200/11{,}441{,}304{,}000$. This is exactly 210 times the first answer, so there is no discount for bulk gambling.

2.18 (a) $6/49 \times 5/48 \times 4/47 ... \times 1/44 = 1/13{,}983{,}816 = 0.000000072$. Note this is $1/49C6$, the number of ways of choosing 6 numbers from 49.

(b) $6/49 \times ... \times 2/45 \times 43/44 \times 6C5 \times 1/43 = 1/2{,}330{,}636 = 0.000000429$.

(c) Third prize: $6/49 \times ... \times 2/45 \times 43/44 \times 6C5 = 1/54{,}200 = 0.000018450$; fourth prize: $6/49 \times ... \times 3/46 \times 43/45 \times 42/44 \times 6C4 = 1/1{,}032 = 0.00096862$; fifth prize: $6/49 \times ... \times 4/47 \times 43/46 \times 42/45 \times 41/44 \times 6C3 = 1/57 = 0.017650404$.

(d) Summing the probabilities gives 0.018637974 or about a 1 in 54 chance of winning. Note that the events are mutually exclusive; you cannot win both the first and the second prize with the same ticket, for example, so the addition rule is valid.

(e) The second prize is six times more probable than the jackpot, yet the prize is only one-twentieth. Relatively, more is put into the jackpot since this is what tends to attract customers. This is true for all other prizes except the fifth. This pot too was 'over-weighted', presumably to get a lot of (small) winners and increase the attractiveness of the lottery.

(f) There would be a danger of guaranteeing the jackpot – you might get 20 winners and hence a huge payout. With the fifth prize, this is much less likely so the prize is guaranteed.

(g) The expected and actual numbers of winners were:

	Expected	**Actual**	**Ratio**
First	3.5	7	2.0
Second	21	39	1.9
Third	904	2,139	2.4
Fourth	47,462	76,731	1.6
Fifth	864,870	1,071,084	1.2
Total	913,261	11,50,000	

There were about 25% more winners than expected, probably because of the preponderance of low winning numbers in the draw (3, 5, 14, 22, 30, 44 and bonus – 10) and since many people used birth dates to select their numbers. This may account for the large number of third-prize winners (with five correct numbers) – five of the six numbers were 31 or below. The largest deviations from expectations (as one would expect) are for the first, second and third prizes. The number of fifth prize winners is very close to expected, given the large number of tickets sold.

2.19

	Prior	**Likelihood**	**Prior × likelihood**	**Posterior**
Fair coin	0.5	0.25	0.125	0.2
Two heads	0.5	1.00	0.500	0.8
			0.625	

2.20 Pr(A|+ test) = $\dfrac{0.99\times 0.01}{0.99\times 0.01+0.01\times 0.99}=0.5$. Half of all positive tests will be false ones. This is likely to spread undue alarm.

2.21 (a) Write the initial probability of guilt as Pr(G) = ½. The probability the witness says the defendant is guilty, given they are guilty, is Pr(W|G) = p. Using Bayes' theorem, the probability of guilt, given the witness's statement, Pr(G|W), is

$$\frac{\Pr(\mathrm{W}\mid\mathrm{G})\times\Pr(\mathrm{G})}{\Pr(\mathrm{W}\mid\mathrm{G})\times\Pr(\mathrm{G})+\Pr(\mathrm{W}\mid\mathrm{not}-\mathrm{G})\times\Pr(\mathrm{not}-\mathrm{G})}=\frac{p\times 0.5}{p\times 0.5+(1-p)\times 0.5}=p.$$

(b) Again using Bayes' theorem, and writing Pr(2W|G) for the probability that both witnesses claim the defendant is guilty and so on, we obtain Pr(G|2W) as

$$\frac{\Pr(2\mathrm{W}\mid\mathrm{G})\times\Pr(\mathrm{G})}{\Pr(2\mathrm{W}\mid\mathrm{G})\times\Pr(\mathrm{G})+\Pr(2\mathrm{W}\mid\mathrm{not}-\mathrm{G})\times\Pr(\mathrm{not}-\mathrm{G})}=\frac{p^2\times 0.5}{p^2\times 0.5+(1-p)^2\times 0.5}=$$

$$\frac{p^2}{p^2+(1-p)^2}.$$

(c) If $p < 0.5$, then the value in part (b) is less than the value in (a). The agreement of the second witness *reduces* the probability that the defendant is guilty. Intuitively, this seems unlikely. The fallacy is that they can lie in many different ways, so Bayes' theorem is not applicable here.

2.22 The probability of the assailant having red hair is, via Bayes' theorem, $(0.8 \times 0.1)/(0.8 \times 0.1 + 0.2 \times 0.9) = 0.31$. You probably guessed a much higher value, as would most jurors. This could be a contentious issue especially if applied to racial minorities, rather than the neutral issue of hair colour. The use of statistical evidence in court is sometimes a contentious issue as many people wrongly interpret the data.

2.23 (a) Expected values are 142, 148.75 and 146, respectively. Hence, B is chosen.

(b) The minima are 100, 130 and 110; so, B has the greatest minimum. The maxima are 180, 170 and 200; so, C is chosen.

(c) The regret table is

	low	middle	high	max
A	30	5	20	30
B	0	0	30	30
C	20	15	0	20

So, C has the minimax regret figure.

(d) The EV assuming perfect information is 157.75, against an EV of 148.75 for project B; so, the value of information is 9.

2.24 (a) EVs are 316, 365 and 260; hence, medium is preferred.

(b) The minima are 300, 270, 50, so small is preferred.

The maxima are 330, 420, 600, so large is chosen.

(c) The maximum regrets are 270, 180 and 250, and the min of these is associated with medium.

(d) The EV with perfect information is 410 against an EV of 365 for the medium factory; so, the value is 45.

2.25 The probability of no common birthday is $365/365 \times 364/365 \times 363/365 \times ... \times 341/365 = 0.43$. (This is obtained by noting that the probability of the second person having a birthday on a different day from the first is 364/365, the probability that the third has a birthday different from the first two is 363/365, etc.) Hence, the probability of at least one birthday in common is 0.57, or greater than one-half. Most people underestimate this probability by a large amount. (This result could form the basis of a useful source of income at parties.)

2.26 We start by calculating the odds for the one-against-one gunfights:

A v. B: If A goes first, it's all over. Pr(A wins) = 1, Pr(B) = 0.

If B goes first, then Pr(B) = 0.75, Pr(A) = 0.25 (fairly obvious).

A v. C: If A goes first, Pr(A) = 1, Pr(C) = 0.

If C goes first, Pr(C) = 0.5, Pr(A) = 0.5.

B v. C: More tricky. If B goes first: This is the sum of an infinite series (since there is a small but positive probability that neither ever kills the other).

$\Pr(B) = 0.75 + \{0.25 \times 0.5 \times 0.75\} + \{0.25 \times 0.5 \times 0.25 \times 0.5 \times 0.75\} + \ldots$ (i.e. the probability of B winning with the first, second, third, ... shot). This comes out as Pr(B) = 6/7, Pr(C) = 1/7.

C goes first: Pr(C) = 4/7, Pr(B) = 3/7 by a similar calculation.

Now in the first stage, if A starts, he shoots B (this is his optimal strategy, since he has a better chance against C than he would against B). There is then a 'Shoot-off' between A and C, with C shooting first. Hence, Pr(A wins the tournament|A starts) = 0.5, Pr(C|A) = 0.5. B has no chance.

C starts: He shoots in the air! This is his optimal strategy, surprisingly. This lets A in and we're back to the situation above. Hence, Pr(C|C) = 0.5, Pr(A|C) = 0.5. If he shot at A and hit him, then he'd have only a 1/7 chance against B. In a one-to-one shoot out with A, with C going first, he has a 50:50 chance. Better to deliberately miss and let A pick off B.

B starts: His best strategy is to shoot at A. If he hits, (P = 3/4) he has a 3/7 chance in the shootout against C. If he misses, he's had it. So Pr(B|B) = 9/28. In the case of a shootout between A and C (which occurs with probability 0.25), Pr(A|B) = 1/8 and Pr(C|B) = 1/8.

To find the ultimate probabilities, Pr(A wins) = Pr(A|A starts) × Pr(A starts) + Pr(A|B starts) × Pr(B starts) + Pr(A|C starts) × Pr(C starts) = 1/2 × 1/3 + 1/8 × 1/3 + 1/2 × 1/3 = 0.375.

Similarly, Pr(B) = 0.107 and Pr(C) = 0.518.

Note that C (the worst shot) has the best chance of winning.

Well done if you got this one right.

2.27 (a) Choose low confidence. The expected score is $0.6 \times 1 + 0.4 \times 0 = 0.6$. For medium confidence the score would be $0.6 \times 2 + 0.4 \times -2 = -0.4$, and for high confidence $0.6 \times 3 + 0.4 \times -6 = -0.6$.

(b) Let p be the probability desired. We require E(score|medium) = $p \times 2 + (1 - p) \times -2 > p$ = E(score|low) – the payoff to medium confidence should be greater than low confidence. Hence, $4p - 2 > p \Rightarrow p > 2/3$. Similarly, we also require $p \times 2 + (1 - p) \times -2 > p \times 3 + (1 - p) \times -6$ (the payoff to medium has to exceed the payoff to high). This implies $p < 4/5$. Hence $2/3 < p < 4/5$.

(c) Given $p = 0.85$, the expected scores are $p = 0.85$ (low), $2p - 2(1 - p) = 1.4$ (medium) and $3p - 6(1 - p) = 1.65$. Hence, the expected loss would be $1.65 - 1.4 = 0.25$ if opting for medium confidence and $1.65 - 0.85 = 0.8$ if opting for low confidence.

2.28 (a) 5 out of 20.

(b) 8 + 3 (out of the remaining 12 questions) = 11.

(c) −1/3 (hence, the expected payoff to guessing is $0.25 \times 1 + 0.25 \times (-1/3) + 0.25 \times (-1/3) + 0.25 \times (-1/3) = 0$).