Solutions to additional questions for Exercise 1.

1. The number of different orderings of n balls equals $n!$. However, by the same token, r_i balls of colour i can be reshuffled in $r_i!$ equivalent ways, $i = 1, \ldots, k$. Therefore, the total number of different ways of putting n balls in a row, of which r_i are of colour i is

$$\binom{n}{r_1, \ldots, r_k} = \frac{n!}{r_1! \cdots r_k!}.$$

This is called a multinomial coefficient and it multiplies the $x_1^{r_1} \cdots x_k^{r_k}$ term in an expansion of $(x_1 + \cdots + x_k)^n$.

2. We draw balls without replacement. There are $\binom{6}{2}$ ways of drawing 2 balls out of 6. We can draw two white balls in $\binom{2}{2}$ ways and two black balls in $\binom{4}{2}$ ways, therefore the probability of drawing two balls of the same colour from an urn with 2 white and 4 black balls is equal to

$$\frac{\binom{2}{2} + \binom{4}{2}}{\binom{6}{2}} = \frac{7}{15}.$$

Similarly, we can draw two balls of different colour in $\binom{2}{1}\binom{4}{1}$, therefore the interesting probability is equal to

$$\frac{\binom{2}{1}\binom{4}{1}}{\binom{6}{2}} = \frac{8}{15}.$$

In sum, we see that it is more probable to draw 2 balls of different colour.

3. Consider first placing 10 boys and 10 girls in a line, so that no two boys or two girls stand next to each other. This will give the same result as placing them at the round table, as we can choose one particular seat at the table (for convenience) and then placing boys and girls from a line to the table (clockwise or anticlockwise) starting from this chosen seat. The two probabilities will be the same, as the two ways of arranging 20 boys and girls are equivalent.

In total there are 20! ways of arranging 20 people in a row. If no two boys or two girls are supposed to stand next to each other, after first girl there should be a boy followed by a second girl and so on; there are 10!10! such arrangements. But we could as well start from a boy, followed by a girl and so on, which gives another 10!10! possibilities.

To sum up, the probability that no two boys or two girls sit next to each other at the round table is

$$\frac{2 * 10!10!}{20!} \approx 1.08 * 10^{-5}.$$

4. $\{H_i\}$ form a partition of Ω, which means that $H_i \in \Sigma$, $\bigcup H_i = \Omega$, $H_i \cap H_j = \emptyset$ for $i \neq j$; we assume additionally that $P(H_i) > 0$. I assume that the partition is countable, for finite partitions the proof is almost the same. Take $A \in \Sigma$, then

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P(A \cap \bigcup H_i) = P(\bigcup A \cap H_i) \\ &= \sum P(A \cap H_i) = \sum P(A|H_i)P(H_i), \end{aligned}$$

where the last equality follows from the definition of conditional probability and the second-next one from the countable additivity assumption on P. Thus, we proved the Total Probability Law.

To prove the Bayes' Law note that from the definition of the conditional probability (we assume additionally that $P(A) > 0$) we have two factorizations:

$$\begin{aligned} P(A \cap H_i) &= P(A|H_i)P(H_i), \\ P(A \cap H_i) &= P(H_i|A)P(A). \end{aligned}$$

Equating them and rearranging we obtain

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{P(A)} = \frac{P(A|H_i)P(H_i)}{\sum P(A|H_j)P(H_j)},$$

which is the Bayes' Law. The last equality follows from the Total Probability Law; notice however, that $P(A)$ in the denominator above is merely a normalizing factor, therefore the Bayes' Law is sometimes expressed (using proportionality symbol $\propto$) as

$$P(H_i|A) \propto P(A|H_i)P(H_i).$$

5. By $A \geq B$ given C we understand that $P(A|C) \geq P(B|C)$, for $P(C) > 0$. $A \geq B$ similarly means that $P(A) \geq P(B)$. Now assume that $0 < P(C) < 1$, from $A \geq B$ given C and $A \geq B$ given C^c, by elementary probabilistic calculations we obtain he Sure Thing Principle (STP)

$$\begin{aligned} P(A) &= P(A \cap (C \cup C^c)) = P(A \cap C) + P(A \cap C^c) \\ &= P(A|C)P(C) + P(A|C^c)P(C^c) \\ &\geq P(B|C)P(C) + P(B|C^c)P(C^c) = P(B). \end{aligned}$$

Moving on, let me present my view on the Simpson's Paradox. Compute a fraction of positive results in each subcase defined by patient's sex and the drug administered, as well as total; we obtain the following result

	drug1	drug2
male	93.1%	90%
female	72.4%	68.8%
total	77.6%	85.14%

Denoting by $d1+$ and $d2+$ positive result after taking drug 1 or 2, M F if the patient is male or female, it seems that the STP is violated. $93.1\% = P(d1+|M) > P(d2+|M) = 90\%$ and $72.4\% = P(d1+|F) > P(d2+|F) = 68.8\%$ and yet $77.6\% = P(d1+) < P(d2+) = 85.14\%$!

How come? First note that from algebraical point of view there is nothing paradoxical in the above table if we move back from percentages to ratios; we may find many possible numbers a, b, c, d, A, B, C and D such that

$$a/b < A/B \quad \text{and} \quad c/d < C/D,$$

$$\text{and yet} \quad (a + c)/(b + d) > (A + C)/(B + D).$$

What is more, percentages hide some potentially useful information. In the table below I present the full information given in the formulation of the paradox. There is strikingly more females then males taking drug 1 (265 to 87), and the reverse holds for drug 2; therefore, we suspect some sort of selectivity-bias. Moreover, it may be that these two combinations, *i.e.* $(M, d2)$ and $(F, d1)$, somehow dominate the sample, and as for these cases drug 2 gives better results (90% of recoveries compared to 72.4%) than drug 1, the reversal of inequalities (in the whole sample) takes place.

	drug1	drug2
male	81/87	243/270
female	192/265	55/80
total	273/352	298/350

What is the solution to the Simpson's paradox? Note that in fact we should consider three events: R representing recovery (positive result), S representing sex (simply M or $F = M^c$), and drug administered D ($d1$ or $d2 = d1^c$). Note further that the cells in the table above (combinations (S, D)) represent probabilities conditional on two events, *e.g.* $81/87 = P(R|M, d1)$, and what we are in fact interested is the probability $P(R|d1)$ and $P(R|d2)$. From the Total Probability Law applied to $P(\cdot|D)$ we have

$$P(R|d1) = P(R|d1, M)P(M|d1) + P(R|d1, F)P(F|d1), \quad \text{and} \tag{1}$$

$$P(R|d2) = P(R|d2, M)P(M|d2) + P(R|d2, F)P(F|d2). \tag{2}$$

Now, the point is that although the probabilities $P(R|d1, S)$ are higher in the first line than $P(R|d2, S)$ in the second one (suggesting that $P(R|d1)$ should be higher than $P(R|d2)$), the skewed weights $P(S|D)$ representing sample-selection bias, cause the reversal of probabilities (so that $P(R|d1)$ is in fact lower than $P(R|d2)$). If the experiment was randomized (so that the weights equalled 0.5, and there is no problem of selectivity-bias), then indeed we would get $P(R|d1) > P(R|d2)$ just as it happens in the sub-groups. As by using percentages we ignore formulas (1) and (2), and thus assume implicitly random experiment setting, we are surprised by the fact that in total drug 2 gives better results.

6. a) As both Y_1 and Y_2 can take values in $\{0, 1\}$, their arithmetic mean takes values in $\{0, 0.5, 1\}$.

b) Note that $2\bar{Y} \sim binomial(2, 0.5)$. Denote weights of this binomial distribution function by $p_2(n)$, $n = 0, 1, 2$. Then the distribution of $\bar{Y}$ is given by the sequence of pairs $\{(n/2, p_2(n)),\ n = 0, 1, 2\}$.

c) If N Bernoulli-distributed random variables are given and we are interested in the distribution of their mean, we may generalize the above argument to obtain $\{(n/N, p_N(n)),\ n = 0, 1, \ldots, N\}$.