Solutions for Exercise 1

1. Suppose that the balls are labelled S_1, S_2, S_3, G_1, G_2 with an obvious convention for their colours. A suitable sample space is the set of all possible pairs of the 5 balls, with order taken into account. There are 25 such possible pairs. This gives

$$\Omega = \{(S_1, S_1), (S_1, S_2), \ldots, (G_2, G_2)\}$$

They can be put into a table as follows:

	First Ball				
	S_1	S_2	S_3	G_1	G_2
Second Ball					
S_1	○	○	○	○	○
S_2	○	○	○	○	○
S_3	○	○	○	○	○
G_1	○	○	○	○	○
G_2	○	○	○	○	○

Each dot represents an outcome in Ω. We can mark the outcomes in events E_1 with $\square$ and E_2 with $\triangle$.

	First Ball				
	S_1	S_2	S_3	G_1	G_2
Second Ball					
S_1	$\square\triangle$	$\square\triangle$	$\square\triangle$	$\triangle$	$\triangle$
S_2	$\square\triangle$	$\square\triangle$	$\square\triangle$	$\triangle$	$\triangle$
S_3	$\square\triangle$	$\square\triangle$	$\square\triangle$	$\triangle$	$\triangle$
G_1	$\square$	$\square$	$\square$	○	○
G_2	$\square$	$\square$	$\square$	○	○

To find $P(E_1)$, $P(E_2)$ and $P(E_1 \cap E_2)$, since the outcomes in the sample space are equally likely, it is enough to count them.

$$P(E_1) = 15/25 = 3/5.$$
$$P(E_2) = 15/25 = 3/5.$$
$$P(E_1 \cap E_2) = 9/25.$$

Notice that $P(E_1 \cap E_2) = P(E_1)P(E_2)$ so that E_1 and E_2 are *independent* events.

If we sample without replacement, then the outcomes on the diagonal of the table are no longer needed. All the others remain equally likely.

	First Ball				
	S_1	S_2	S_3	G_1	G_2
Second Ball					
S_1		$\square\triangle$	$\square\triangle$	$\triangle$	$\triangle$
S_2	$\square\triangle$		$\square\triangle$	$\triangle$	$\triangle$
S_3	$\square\triangle$	$\square\triangle$		$\triangle$	$\triangle$
G_1	$\square$	$\square$	$\square$		○
G_2	$\square$	$\square$	$\square$	○	

Recounting the outcomes

$$P(E_1) = 12/20 = 3/5.$$

$$P(E_2) = 12/20 = 3/5.$$

$$P(E_1 \cap E_2) = 6/20 = 3/10.$$

This time the events are not independent, but we can clearly see that $P(E_1) = P(E_2)$, which is not otherwise completely transparent for sampling without replacement.

2. The probabilities of not inspecting the two faulty motors are easily seen to be

   (a) $\binom{10}{2}/\binom{12}{2} = (10 \times 9)/(12 \times 11) = 15/22$.

   (b) $(\binom{5}{1}/\binom{6}{1})^2 = 25/36$.

   (c) $(\binom{4}{1}/\binom{6}{1}) \times 1 = 2/3$.

   The second one is the largest.

3. (a) There are 8 equally likely ordered arrangements of heads and tails when 3 coins are tossed. All except for HHH and TTT have an "odd one out". So the probability of having a loser on a given turn is $6/8 = 3/4$.

   (b) The probability of an even number of turns is

   $$\sum_{i=1}^{\infty} P(2i \text{ turns needed})$$

   which is

   $$\sum_{i=1}^{\infty} \left(\frac{1}{4}\right)^{2i-1} \frac{3}{4} = \frac{3}{16} \sum_{i=0}^{\infty} \left(\frac{1}{16}\right)^{i} = (3/16)/(1 - 1/16) = 1/5.$$

   Notice that we need to add here the probabilities for a whole sequence of disjoint events to get the answer.

4. There are a lot of ways of doing this. A throws $n+1$ times and B throws n times. Consider A's first n throws together with B's throws; a total of $2n$ throws, each having 2 possible outcomes, so there is a total of 2^{2n} equally likely outcomes. Now let H_A be the number of heads in A's first n throws and H_B be the number of heads in B's throws. We can partition the sample space into three events;

   E_1 event that $H_A > H_B$ E_2 event that $H_A = H_B$ E_3 event that $H_A < H_B$

   By mutual exclusive and exhaustive properties, the sum of the number of outcomes in each event must be the total number of outcomes in the sample space; $|E_1| + |E_2| + |E_3| = 2^{2n}$. By symmetry we see that $|E_1| = |E_3|$ and so $2|E_1| + |E_2| = 2^{2n}$.

   Now consider A's final throw. This brings the total number of throws to $2n+1$ and hence the size of the sample space to 2^{2n+1}. If event E_1 has occurred then A has won with either a head or a tail on the final throw; this gives $2|E_1|$ outcomes. If E_2 has occurs then A needs to throw a head to win; giving $|E_2|$ outcomes. Thus the total number of ways that A can win is $2|E_1| + |E_2|$. We established above that $2|E_1| + |E_2| = 2^{2n}$. Thus $P(A\text{wins}) = 2^{2n}/2^{2n+1} = 1/2$.

   Another way is to use a few probability generating functions. Suppose we have m coin tosses for A and n for B. It is obvious by inspection that for $r = -n, -n+1, \ldots, m$ the coefficient of t^r in

   $$(1+t)^m (1+1/t)^n / 2^{m+n} = (1+t)^{m+n} / [2^{m+n} t^n]$$

is the probability that there are exactly r more heads for A than for B. Expanding the binomial on the right we evaluate the probability as

$$\binom{m+n}{n+r}/2^{m+n}.$$

This is the probability of $n+r$ heads in $m+n$ tosses. So the probability that A has more heads than B is the same as the probability of more than n heads in $m+n$ tosses of the coin. This is obviously 0.5 if $m = n+1$.

It is easy to obtain directly the probability given above of exactly r more heads for A than for B. The probability of r more heads for A than for B is the sum for $s = \max(0, -r), 1, \ldots, \min(n, m-r)$ of the probabilities that we have $s+r$ heads in m tosses and also s heads from n tosses. The probabilities summed are the same as the probabilities that there are $s+r$ heads in m tosses and also $n-s$ heads from n tosses. The sum of those probabilities is the same as the probability of $n+r$ heads in $m+n$ tosses.

5. It is easiest to use a sample space which has equally likely outcomes, and then to count outcomes to find the required probabilities. Let's number the positions in the row from 1 to 6. A and B are equally likely to occupy any 2 positions that are distinct. We can use a diagram to show the outcome space - each circle shows a possible position for A and B, and all 30 outcomes are equally likely.

Position of B \ Position of A	1	2	3	4	5	6
1		○	○	○	○	○
2	○		○	○	○	○
3	○	○		○	○	○
4	○	○	○		○	○
5	○	○	○	○		○
6	○	○	○	○	○	

Now we can mark each outcome with the number of people between A and B. This gives

Position of B \ Position of A	1	2	3	4	5	6
1		0	1	2	3	4
2	0		0	1	2	3
3	1	0		0	1	2
4	2	1	0		0	1
5	3	2	1	0		0
6	4	3	2	1	0	

Then counting the number of outcomes for each case gives

$$\begin{aligned} P(0) &= 10/30 = 1/3 \\ P(1) &= 8/30 = 4/15 \\ P(2) &= 6/30 = 1/5 \\ P(3) &= 4/30 = 2/15 \\ P(4) &= 2/30 = 1/15. \end{aligned}$$

Now it is easy to generalise to the case of n positions. There are $n^2 - n = n(n-1)$ equally likely positions for A and B, and $2(n-r-1)$ of these have r people between A and B. So

$$P(r \text{ people between } A \text{ and } B) = \frac{2(n-r-1)}{n(n-1)}$$

for $r = 0, 1, 2, \ldots, n-2$.

If the positions for A and B are in a ring, and we look clockwise for the number of people between them, the same diagram will work, but we need to put in different labels for the numbers between.

	Position of A					
	1	2	3	4	5	6
Position of B						
1		4	3	2	1	0
2	0		4	3	2	1
3	1	0		4	3	2
4	2	1	0		4	3
5	3	2	1	0		4
6	4	3	2	1	0	

In this case it is obvious that all numbers of people between 0 and $n-2$ have the same probability. This is also true in the general case of n positions. It is often true that arrangements on a circle have simpler properties than arrangements on a line.

6. The urn contains 8 distinct disks.

	A	B	C	D
red	$\circ$	$\circ$		
green	$\circ$	$\circ$		
blue	$\circ$	$\circ$	$\circ$	$\circ$

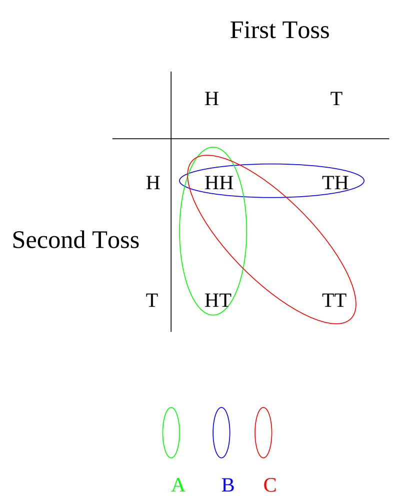
By inspection we can see;

(a) $P(A) = 3/8$, $P(\text{red}) = 1/4$ and $P(A \cap \text{red}) = 1/8$ hence not independent,

(b) $P(A \cup D) = 1/2$, $P(\text{red}) = 1/4$ and $P((A \cup D) \cap \text{red}) = 1/8$ hence independent,

(c) $P(A \cup \text{blue}) = 3/4$, $P(\text{red}) = 1/4$ and $P((A \cup \text{blue}) \cap \text{red}) = 1/8$ hence not independent,

7. The probabilities can be read off from the diagram:



8. (a) We can use a sample space of eight equally likely outcomes if we take into account birth order. Using M and F for boys and girls, we can write

$$\Omega = \{MMM, MMF, MFM, FMM, MFF, FMF, MFF, FFF\}$$

Then $A = \{MMF, MFM, FMM, MFF, FMF, MFF\}$,
$B = \{MMM, MMF, MFM, FMM\}$ and $A \cap B = \{MMF, MFM, FMM\}$. Counting the outcomes in the events gives

$$\begin{aligned} P(A) &= 6/8 = 3/4 \\ P(B) &= 4/8 = 1/2 \\ P(A \cap B) &= 3/8 \end{aligned}$$

Obviously $P(A \cap B) = P(A)P(B)$, so $A \perp B$. It is hard to see this independence intuitively. One needs to verify it to be sure.

(b) For four children families there are 16 equally likely family outcomes. Just 2 of these have all the children of the same gender, so $P(A) = 14/16 = 7/8$, and there is $\binom{4}{0} = 1$ family with all boys and $\binom{4}{1} = 4$ families with 1 girl. All the other families have more than 1 girl, so $P(B) = 5/16$. There are 4 families with children of both genders and no more than 1 girl, so $P(AB) = 4/16 = 1/4$. In this case there is not independence between A and B because

$$P(A \cap B) = \frac{1}{4} \neq \frac{7}{8} \times \frac{5}{16} = P(A)P(B).$$

9. (a) Since $A \cap B$ and $A \cap B^c$ are disjoint,

$$A = (A \cap B) \cup (A \cap B^c)$$

implies

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

or

$$P(A \cap B^c) = P(A) - P(A \cap B)$$

Since $A \perp B$, $P(A \cap B) = P(A)P(B)$, and so the last equation implies

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A)P(B) \\ &= P(A)[1 - P(B)] \\ &= P(A)P(B^c). \end{aligned}$$

So $A \perp B^c$.

(b) If

$$P(A|B) < P(A)$$

then since $P(A|B) = P(A \cap B)/P(B)$, it follows that

$$\frac{P(A \cap B)}{P(B)} < P(A)$$

and so that

$$P(A \cap B) < P(A)P(B).$$

Since $P(A) > 0$ we can divide both sides by $P(A)$ to get

$$\frac{P(A \cap B)}{P(A)} < P(B)$$

which is

$$P(B|A) < P(B).$$

Notice that there is an intuitive interpretation of this result. It says that if A makes B less probable, then B makes A less probable.

(c) This is a false statement. It says that A and B are conditionally independent given event C. To show that this does not imply in general that A and B are independent, you must find an explicit counter-example. One such is found by choosing $A = B = C$ and $P(A) < 1$. Then $P(A|C) = P(B|C) = P(A \cap B)|C) = 1$, and so the conditional independence trivially holds true. However, $P(A \cap B) = P(A)$ and $P(A)P(B) = P(A)^2$, so there is not independence between A and B, for that implies $P(A) = P(A)^2$, which is false as $P(A) \neq 0$ or 1.

(d) The statement in the question says that B has the same probability if A occurs as if A does not occur. One would expect that this should imply A and B independent, and it does!

$$P(B|A) = P(B|A^c)$$

implies

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

Adding the numerators and denominators, this implies

$$\frac{P(A \cap B)}{P(A)} = \frac{P(A \cap B) + P(A^c \cap B)}{P(A) + P(A^c)}$$

which is

$$\frac{P(A \cap B)}{P(A)} = \frac{P(B)}{1}$$

Cross-multiplying gives

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

which shows that $A \perp B$.

10. Let A_i be the event that ith face does not appear. We will use the formulae for unions

$$P(\bigcup_{i=1}^{6} A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \ldots - P(A_1 \cap \ldots \cap A_6). \quad (1)$$

If we throw the die n times then

$P(A_1) = (5/6)^n$ $\quad$ $P(A_1 \cap A_2) = (4/6)^n$ $\quad$ $P(A_1 \cap A_2 \cap A_3) = (3/6)^n$

$P(A_1 \cap \ldots \cap A_4) = (2/6)^n$ $\quad$ $P(A_1 \cap \ldots \cap A_5) = (1/6)^n$ $\quad$ $P(A_1 \cap \ldots \cap A_6) = 0$

Let E be event each face appears at least once. Using (1) and exploiting symmetry yields;

$$\begin{aligned} P(E) &= 1 - P(A_1 \cup \ldots \cup A_6) \\ &= 1 - 6(5/6)^n + 15(2/3)^n - 20(1/2)^n + 15(1/3)^n - 6(1/6)^n. \end{aligned}$$

11. For X and Y mutually exclusive $P(X \cup Y) = P(X) + P(Y)$. For general events A and B;

(a) $A \cup A^c = \Omega \Rightarrow P(A \cup A^c) = 1$.
A and A^c mutually exclusive $\Rightarrow P(A \cup A^c) = P(A) + P(A^c)$.
Thus $P(A) + P(A^c) = 1$ and so $P(A^c) = 1 - P(A)$.

(b) $B = (A \cap B) \cup (A^c \cap B) \Rightarrow P(B) = P((A \cap B) \cup (A^c \cap B))$
$A \cap B$ and $A^c \cap B$ mutually exclusive $\Rightarrow P((A \cap B) \cup (A^c \cap B)) = P(A \cap B) + P(A^c \cap B)$
Thus $P(A \cap B) + P(A^c \cap B) = P(B)$ and so $P(A^c \cap B) = P(B) - P(A \cap B)$.

(c) $A \cup B = A \cup (A^c \cap B) \Rightarrow P(A \cup B) = P(A \cup (A^c \cap B))$.
A and $(A^c \cap B)$ mutually exclusive $\Rightarrow P(A \cup (A^c \cap B)) = P(A) + P(A^c \cap B)$.
Thus $P(A \cup B) = P(A) + P(A^c \cap B) = P(A) + P(B) - P(A \cap B)$ by part 11b.

(d) $(A \cup B) \cap (A^c \cup B^c) = (A \cap B^c) \cup (A^c \cap B) \Rightarrow P((A \cup B) \cap (A^c \cup B^c)) = P((A \cap B^c) \cup (A^c \cap B))$.
$(A \cap B^c)$ and $(A^c \cap B)$ are mutually exclusive so $P((A \cap B^c) \cup (A^c \cap B)) = P(A \cap B^c) + P(A^c \cap B)$.
Thus $P((A \cup B) \cap (A^c \cup B^c)) = P(A \cap B^c) + P(A^c \cap B) = P(A) + P(B) - 2P(A \cap B)$ by 11b.

We could interpret the event $(A \cup B) \cap (A^c \cup B^c)$ as the event that either A or B happens but not both.

12. (a) Partition B_1 vicious $P(B_1) = 0.1$
B_2 not vicious $P(B_2) = 0.9$
If R is the event that a randomly chosen worm is red, we can see from the questions that $P(R|B_1) = 4/5$ and $P(R|B_2) = 1/4$.
Law of total probability to get probability worm is red;

$$\begin{aligned} P(R) &= P(R|B_1)P(B_1) + P(R|B_2)P(B_2) \\ &= 4/5 \times 0.1 + 1/4 \times 0.9 = 0.305. \end{aligned}$$

Bayes' theorem to get probability that a worm is vicious given that it is red;

$$\begin{aligned} P(B_1|R) &= P(R|B_1)P(B)/P(R) \\ &= 4/50 \times 200/61 = 16/61 = 0.262. \end{aligned}$$

(b) Define events R_1 - worm red; R_2 - parent red; R_3 - grandparent red. From part 12a $P(R_1) = 0.305 = P(R_2) = P(R_3)$. From question $P(R_1|R_2) = 3/4$, $P(R_1|R_3) = 2/3$ and $P(R_1|R_2 \cap R_3) = 4/5$.

i. Probability red and parent red is given by
$P(R_1 \cap R_2) = P(R_1|R_2)P(R_2) = 3/4 \times 0.305 = 0.22875$.

ii. Probability red or parent red or grandparent red given by
$P(R_1 \cup R_2 \cup R_3) = P(R_1) + P(R_2) + P(R_3) - P(R_1 \cap R_2) - P(R_2 \cap R_3) - P(R_1 \cap R_3) + P(R_1 \cap R_2 \cap R_3)$.
Elements of this equation:
$P(R_1 \cap R_2) = 0.22875$
$P(R_2 \cap R_3) = 0.22875$ (parent to grandparent $\sim$ child to parent)
$P(R_1 \cap R_3) = P(R_1|R_3)P(R_3) = 0.305 \times 2/3 = 0.20333$
$P(R_1 \cap R_2 \cap R_3) = P(R_1|R_2 \cap R_3)P(R_2 \cap R_3) = 4/5 \times 0.22875 = 0.183$
$\Rightarrow P(R_1 \cup R_2 \cup R_3) = 0.4371$.

13. This question just uses the law of total probability.

(a) When we transfer just one disk;

Partition: R_1 transfer red disk, $P(R_1) = 4/9$
G_1 transfer green disk, $P(G_1) = 5/9$

Let R_2 be the event that the disk we draw from the second urn is red then

$$\begin{aligned} P(R_2) &= P(R_2|R_1)P(R_1) + P(R_2|G_1)P(G_1) \\ &= 6/10 \times 4/9 + 5/10 \times 5/9 = 49/90 = 0.544. \end{aligned}$$

(b) Transferring two disks;

Partition: RR transfer 2 red disks, $P(RR) = 4/9 \times 3/8 = 1/6$
RG transfer 1 red and 1 green disk, $P(RG) = 2 \times 4/9 \times 5/8 = 5/9$
GG transfer 2 green disks, $P(GG) = 5/9 \times 4/8 = 5/18$

$$\begin{aligned} P(R_2) &= P(R_2|RR)P(RR) + P(R_2|RG)P(RG) + P(R_2|GG)P(GG) \\ &= 7/11 \times 1/6 + 6/11 \times 5/9 + 5/11 \times 5/18 = 53/99 = 0.535 \end{aligned}$$

Solutions for Exercise 2

1. (a)

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} \frac{1}{\pi(1+y^2)} dy \\ &= \left[\frac{1}{\pi} \arctan y\right]_{-\infty}^{x} \\ &= \frac{1}{\pi} \arctan x - \frac{1}{\pi}\frac{-\pi}{2} \\ &= \frac{1}{\pi} \arctan x + \frac{1}{2}. \end{aligned}$$

(b)

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} \frac{e^{-y}}{(1+e^{-y})^2} dy \\ &= \left[\frac{1}{1+e^{-y}}\right]_{-\infty}^{x} \\ &= \frac{1}{1+e^{-x}}. \end{aligned}$$

(c) For $x \geq 0$,

$$\begin{aligned} F_X(x) &= \int_{0}^{x} \frac{a-1}{(1+y)^a} dy \\ &= \left[-\frac{1}{(1+y)^{a-1}}\right]_{0}^{x} \\ &= 1 - \frac{1}{(1+x)^{a-1}}. \end{aligned}$$

For $x < 0$ it is obvious that $F_X(x) = 0$, so we could write the result in full as

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - \frac{1}{(1+x)^{a-1}} & x \geq 0. \end{cases}$$

(d) For $x \geq 0$,

$$\begin{aligned} F_X(x) &= \int_{0}^{x} c\tau y^{\tau-1} e^{-cy^{\tau}} dy \\ &= \left[-e^{-cy^{\tau}}\right]_{0}^{x} \\ &= 1 - e^{-cx^{\tau}}. \end{aligned}$$

For $x < 0$ it is obvious that $F_X(x) = 0$, so we could write the result in full as

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-cx^{\tau}} & x \geq 0. \end{cases}$$

2. (a) We can find the s^{th} moment about the origin, and use this result to get the mean and variance.

$$\begin{aligned}
E[X^s] = m_s &= \int_0^\infty x^s f_X(x)dx = \int_0^\infty x^s \frac{1}{(r-1)!} e^{-kx} x^{r-1} k^r dx \\
&= \int_0^\infty \frac{1}{(r-1)!} e^{-kx} x^{s+r-1} k^r dx \\
&= \frac{(s+r-1)!}{(r-1)!} \frac{1}{k^s} \int_0^\infty \frac{1}{(s+r-1)!} e^{-kx} x^{(s+r)-1} k^{s+r} dx \\
&= \frac{(s+r-1)!}{(r-1)!} \frac{1}{k^s},
\end{aligned}$$

since the integrand is a Gamma density function with shape parameter $s+r$ and scale parameter k. So

$$m_s = \frac{(s+r-1)!}{(r-1)!k^s} = \frac{r(r+1)...(r+s-1)}{k^s}$$

which is sometimes written as

$$\frac{(r+s-1)^{(s)}}{k^s}.$$

Note that this result holds also for non-integer r, using Gamma function results.

Using the result,

$$EX = m_1 = \frac{r}{k}$$

and

$$EX^2 = m_2 = \frac{r(r+1)}{k^2}.$$

The variance is

$$\begin{aligned}
m_2 - (m_1)^2 &= \frac{r(r+1)}{k^2} - \frac{r^2}{k^2} \\
&= \frac{r}{k^2}.
\end{aligned}$$

Both mean and variance increase with r increasing and decrease with k increasing.

(b) The easiest direct way here is to find the r^{th} *factorial* moment, $\mu_{(r)} = EX^{(r)} = EX(X-1)\ldots(X-r+1)$. This works out very simply. Then we can convert to the mean and the variance. The critical property that makes $\mu_{(r)}$ work out easily is that for x not an integer from 1 to $r-1$

$$\frac{x^{(r)}}{x!} = \frac{x^{(r)}}{x^{(x)}} = \frac{1}{(x-r)^{(x-r)}} = \frac{1}{(x-r)!}.$$

$$\begin{aligned}
EX^{(r)} &= \sum_{x=0}^{\infty} x^{(r)} e^{-\lambda} \lambda^x / x! \\
&= \sum_{x=r}^{\infty} \frac{x^{(r)}}{x!} e^{-\lambda} \lambda^x \\
&= \sum_{x=r}^{\infty} \frac{1}{(x-r)!} e^{-\lambda} \lambda^x \\
&= \lambda^r \sum_{x=r}^{\infty} \frac{1}{(x-r)!} e^{-\lambda} \lambda^{x-r} \\
&= \lambda^r \sum_{y=0}^{\infty} \frac{1}{y!} e^{-\lambda} \lambda^y \\
&= \lambda^r.
\end{aligned}$$

The last step follows because we are adding together all the probabilities for a Poisson distribution with parameter λ.

Now it is easy to get the mean and the variance.

$$EX = EX^{(1)} = \lambda$$

and since $X(X-1) + X = X^2$,

$$m_2 = EX^{(2)} + EX = \mu_{(2)} + \mu_{(1)} = \lambda^2 + \lambda$$

and so $\mathrm{Var}X = m_2 - (m_1)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

(c) We can use the factorial moments for the negative binomial distribution too. The important identity in this case is that for x not an integer from 1 to $r-1$

$$\begin{aligned}
x^{(r)}\binom{a+x-1}{x} &= x^{(r)}\frac{(a+x-1)^{(x)}}{x!} \\
&= \frac{(a+x-1)^{(x)}}{(x-r)!} = \frac{(a+x-1)^{(x)}}{(x-r)!} \\
&= \frac{(a+x-1)^{(x-r)}(a+r-1)^{(r)}}{(x-r)!} \\
&= (a+r-1)^{(r)}\binom{a+x-1}{x-r} = (a+r-1)^{(r)}\binom{(a+r)+(x-r)-1}{x-r}
\end{aligned}$$

$$\begin{aligned}
EX^{(r)} &= \sum_{x=0}^{\infty} x^{(r)}\binom{a+x-1}{x}\left[\frac{b}{1+b}\right]^a\left[\frac{1}{1+b}\right]^x \\
&= \sum_{x=r}^{\infty} x^{(r)}\binom{a+x-1}{x}\left[\frac{b}{1+b}\right]^a\left[\frac{1}{1+b}\right]^x \\
&= \sum_{x=r}^{\infty}(a+r-1)^{(r)}\binom{(a+r)+(x-r)-1}{x-r}\left[\frac{b}{1+b}\right]^a\left[\frac{1}{1+b}\right]^x \\
&= (a+r-1)^{(r)}\left[\frac{1+b}{b}\right]^r\left[\frac{1}{1+b}\right]^r\sum_{x=r}^{\infty}\binom{(a+r)+(x-r)-1}{x-r}\left[\frac{b}{1+b}\right]^{a+r}\left[\frac{1}{1+b}\right]^{x-r} \\
&= \frac{(a+r-1)^{(r)}}{b^r}\sum_{y=0}^{\infty}\binom{(a+r)+y-1}{x-r}\left[\frac{b}{1+b}\right]^{a+r}\left[\frac{1}{1+b}\right]^y \\
&= \frac{(a+r-1)^{(r)}}{b^r} = \frac{a(a+1)(a+2)\dots(a+r-1)}{b^r}.
\end{aligned}$$

Then we quickly get

$$EX = \mu_{(1)} = \frac{a}{b}$$

and

$$\begin{aligned}
\mathrm{Var}X = \mu_{(2)} + \mu_{(1)} - (\mu_{(1)})^2 &= \frac{a(a+1)}{b^2} + \frac{a}{b} - \frac{a^2}{b^2} \\
&= \frac{a(a+1)+ab-a^2}{b^2} = \frac{a(1+b)}{b^2}.
\end{aligned}$$

(d) We can directly integrate for m_r, writing the integral as a beta function after a transformation. It is easier to work with $Y = X + 1$, noting that $EX = EY - 1$ and $\text{Var}Y = \text{Var}X$.

$$\begin{aligned} EY^r &= \int_0^\infty (x+1)^r \frac{a-1}{(1+x)^a} \\ &= \frac{a-1}{a-r-1} \int_0^\infty \frac{a-r-1}{(1+x)^{a-r}} \\ &= \frac{a-1}{a-r-1}, \end{aligned}$$

provided that $a - r > 1$ (or else the integral is not defined). So

$$EY = \frac{a-1}{a-2} \;\Rightarrow\; EX = \frac{1}{a-2}$$

for $a > 2$, and provided $a > 3$,

$$\begin{aligned} \text{Var}X = \text{Var}Y = EY^2 - (EY)^2 &= \frac{a-1}{a-3} - \left[\frac{a-1}{a-2}\right]^2 \\ &= \frac{(a-1)[(a-2)^2 - (a-1)(a-3)]}{(a-2)^2(a-3)} \\ &= \frac{a-1}{(a-2)^2(a-3)} \end{aligned}$$

3. We will do this by showing that the cumulant generating function is $K_X(t) = \mu t + \sigma^2 t^2/2$. Once we have shown this the definition of cumulants will give us our result. First evaluate the moment generating function.

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \int_{-\infty}^\infty e^{tx} f_X(x) dx \\ &= \int_{-\infty}^\infty e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx \\ &= \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2 - 2\sigma^2 t]/(2\sigma^2)} dx \\ &= e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-[x-(\mu+\sigma^2 t)]^2/(2\sigma^2)} dx \\ &= e^{\mu t + \sigma^2 t^2/2} \end{aligned}$$

Taking logs now gives us the cumulant generating function and hence our result.

4. The description of Y says that $Y = |X - a|$, so

$$\begin{aligned} F_Y(y) &= P[Y \le y] \\ &= P[a - y \le X \le a + y] = P[a - y < X \le a + y] \\ &= \begin{cases} F(a+y) - F(a-y) & y \ge 0 \\ 0 & y < 0 \end{cases} \end{aligned}$$

The density function $f_Y(y)$ of Y is obtained by differentiating with respect to y, so

$$f_Y(y) = \begin{cases} f(a+y) + f(a-y) & y \ge 0 \\ 0 & y < 0. \end{cases}$$

In the case that $X \sim N(\mu, \sigma^2)$ and $a = \mu$, the density function of Y is

$$f_Y(y) = \begin{cases} \frac{2}{\sigma\sqrt{2\pi}} \exp\left[-\frac{y^2}{2\sigma^2}\right] & y \ge 0 \\ 0 & y < 0 \end{cases}$$

This is sometimes called a *half-normal distribution*. It is important for residual plots.

5. It is probably better to use a finite upper limit before integrating by parts, because some standard results for integration by parts do not allow an infinite range of integration. The question implies, however, that the mean is defined. We can use a limiting operation to get rid of the infinite range.

$$\begin{aligned}
\int_0^\infty [1 - F_X(x)]dx &= \lim_{a\to\infty} \int_0^a [1 - F_X(x)]dx \\
&= \lim_{a\to\infty} \left[x[1 - F_X(x)]\Big|_0^a - \int_0^a (-f_X(x))x dx \right] \\
&= \lim_{a\to\infty} a[1 - F_X(a)] + \lim_{a\to\infty} \int_0^a f_X(x) x dx \\
&= \lim_{a\to\infty} a[1 - F_X(a)] + EX.
\end{aligned}$$

The result is proved if we can show that $\lim_{a\to\infty} a[1-F_X(a)] = 0$. This is not true in general (check the Pareto distributions), but *is* true provided that the mean is defined. We can see this through a simple inequality that sandwiches $a[1 - F_X(a)]$ between 0 and a quantity tending to 0. Since

$$EX = \int_0^\infty x f_x(x)dx = \lim_{a\to\infty} \int_0^a x f_x(x)dx$$

it follows that $\lim_{a\to\infty} \int_a^\infty x f_x(x)dx = 0$. Now, for $a > 0$

$$\int_a^\infty x f_x(x)dx \geq \int_a^\infty a f_x(x)dx = a[1 - F_X(a)]$$

so we see that $\lim_{a\to\infty} a[1 - F_X(a)] = 0$.

There is a similar result for random variables on the whole line

$$EX = -\int_{-\infty}^0 F_X(x)dx + \int_0^\infty [1 - F_X(x)]dx.$$

The result is also true for discrete distributions. For instance, if X is a non-negative integer-valued random variable,

$$\begin{aligned}
EX = \int_0^\infty [1 - F_X(x)]dx &= \sum_{i=0}^\infty \int_i^{i+1} [1 - F_X(x)]dx \\
&= \sum_{i=0}^\infty \int_i^{i+1} [1 - F_X(i)]dx \\
&= \sum_{i=0}^\infty [1 - F_X(i)] = \sum_{i=0}^\infty \sum_{j=i+1}^\infty P[X = j].
\end{aligned}$$

This is easily seen directly.

6. We have

$$f(y) = \begin{cases} 0 & y < 0 \\ y f_X(y)/\mu & y \geq 0 \end{cases}$$

so it is obvious that since $\mu > 0$, $f(y) \geq 0$. The only other property to check to verify that $f(y)$ is a density function is that it integrates to 1.

$$\int_0^\infty \frac{y f_X(y)}{\mu} dy = \frac{\mu}{\mu} = 1.$$

Now, if Y is the random variable with density function $f(y)$ we must have $\mathrm{Var}Y \geq 0$ which is $EY^2 \geq (EY)^2$. Assuming that all the moments are defined, it is obvious that $E[Y^r] = E[X^{r+1}]/\mu$. So, from the inequality for Y,

$$E[X^3]/\mu \geq [E[X^2]/\mu]^2.$$

which is the result in the question. Many similar results are available.

7. We must check

$$P[X > x + y \mid X > x] = P[X > y].$$

This can be written in terms of the survival function of X, because for $y > 0$

$$\begin{aligned} 1 - F_X(y) = P[X > y] &= P[X > x + y \mid X > x] \\ &= \frac{P[X > x + y \cap X > x]}{P[X > x]} \\ &= \frac{P[X > x + y]}{P[X > x]} \\ &= \frac{1 - F_X(x + y)}{1 - F_X(x)}. \end{aligned}$$

If X has an exponential distribution of rate λ, then the survival function is $1 - F_X(x) = e^{-\lambda x}$. The no memory property is verified by noting that

$$e^{-\lambda y} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}}.$$

If X has a geometric distribution, then the survival function is $1 - F_X(x) = q^x$. The no memory property is verified by noting that

$$q^y = \frac{q^{x+y}}{q^x}.$$

The 'no memory' property is saying that 'old is as good as new'. If we think in terms of lifetimes, it says that you are equally likely to survive for y more years whatever your current age x may be. This is unrealistic for humans for widely different ages x, but may work as a base model in other applications.

8. We are given for $\lambda \geq 0, 0 < \rho < 1$,

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \rho \exp(-\lambda(1-\rho)x) & x \geq 0 \end{cases}$$

which is defined as 0 for all negative x, so $F(-\infty) = 0$. Also,

$$F(\infty) = lim_{x\to\infty} F(x) = lim_{x\to\infty} 1 - \rho \exp(-\lambda(1-\rho)x) = 1.$$

All we need to check now is that $F(x)$ is non-decreasing in x, and that it is right continuous. It is clearly non-decreasing and continuous for $x < 0$, being defined as 0 there. Similarly, $1 - \rho \exp(-\lambda(1-\rho)x)$ is continuous and strictly increasing in x for $x > 0$. At $x = 0$, $F(x) = 1 - \rho$, which is the limit as $x \downarrow 0$ of $1 - \rho \exp(-\lambda(1-\rho)x)$. This shows the continuity from the right at $x = 0$. For x just less than 0, we have $F(0-) = 0 < 1 - \rho = F(0)$, so $F(x)$ increases at $x = 0$. That completes checking all the properties. The distribution has a non-zero probability of a waiting time 0, and a continuous distribution for waiting times greater than 0.

One could also simply write the given function as a mixture of two distribution functions to immediately verify that it is itself a distribution function. This can be done with

$$F(x) = (1-\rho)H_0(x) + \rho \begin{cases} (1 - \exp(-\lambda(1-\rho)x)) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

which shows the distribution as a mixture with mixing probabilities $1-\rho$ and ρ of a distribution giving probability 1 to $x = 0$ and an exponential distribution with rate $\lambda(1-\rho)$.

9. First notice that the failure rate is

$$\lambda(x) = -\frac{d \ln \bar{F}(x))}{dx}$$

(a) For the Weibull distribution, $\bar{F}(x) = e^{-cx^\tau}$, so

$$\lambda(x) = -\frac{d(-cx^\tau)}{dx} = c\tau x^{\tau-1}.$$

For the Pareto distribution, $\bar{F}(x) = 1/(1+x)^{a-1}$, so

$$\lambda(x) = -\frac{d(-(a-1)\ln(1+x))}{dx} = (a-1)/(1+x).$$

Neither of these gives a very flexible shape for the failure rate if one wants to use it for modelling.

(b) We are asked to show that

$$\lambda(x) \leq \lambda(x+y)$$

when $y \geq 0$.

We are told that $\bar{F}(x+y)/\bar{F}(x)$ and so $\ln \bar{F}(x+y) - \ln \bar{F}(x)$ has a non-positive derivative with respect to x. Differentiating with respect to x,

$$-\lambda(x+y) + \lambda(x) \leq 0$$

which is the result required.

Solutions for Exercise 3

1. We are asked to find
$$M_X(t) = E\left[e^{tX}\right]$$
for $X \sim N(\mu, \sigma^2)$. We may write $X = \mu + \sigma Z$ where $Z \sim N(0,1)$, so that
$$\begin{aligned} M_X(t) &= E\left[e^{t(\mu+\sigma Z)}\right] = e^{\mu t} E\left[e^{\sigma t Z}\right] \\ &= e^{\mu t} M_Z(\sigma t). \end{aligned}$$
So one need only find the MGF for a standard normal random variable.
$$\begin{aligned} M_Z(t) &= \int_{-\infty}^{\infty} e^{zt} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[(z-t)^2 - t^2]} dz \\ &= e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz \\ &= e^{\frac{1}{2}t^2}. \end{aligned}$$
The last integral is the total probability for a $N(t,1)$ random variable, and so is equal to 1. The step from line 1 to line 2 follows from the simple algebraic identity
$$-\frac{1}{2}[(z-t)^2 - t^2] = -\frac{1}{2}z^2 + zt.$$
The MGF for the general normal distribution is
$$M_X(t) = \exp(\mu t + \sigma^2 t^2/2).$$
The mean of Y is
$$E[e^X] = M_X(1) = \exp(\mu + \sigma^2/2).$$
Also,
$$E[Y^2] = M_X(2) = \exp(2\mu + 4\sigma^2/2).$$
The variance of Y is therefore
$$\exp(2\mu + 4\sigma^2/2) - \exp(2\mu + 2\sigma^2/2) = \exp(2(\mu + \sigma^2/2))(\exp(\sigma^2) - 1).$$

2. The Markov inequality says that if X is a non-negative random variable, then for $\lambda > 0$,
$$P[X \geq \lambda] \leq \frac{E[X]}{\lambda}.$$
The proof is fairly easy.
$$\begin{aligned} P[X \geq \lambda] &= E[I_{[\lambda,\infty)}(X)] \\ &\leq E[\frac{X}{\lambda} I_{[\lambda,\infty)}(X)] \\ &= \frac{1}{\lambda} E[X I_{[\lambda,\infty)}(X)] \\ &\leq \frac{1}{\lambda} E[X]. \end{aligned}$$

If $t > 0$, then the events $Y \geq \gamma$ and $e^{tY} \geq e^{t\gamma}$ are the same. We can take, in the Markov inequality, $X = e^{tY}$ where $t > 0$, to obtain

$$\begin{aligned} P[Y \geq \gamma] = P[e^{tY} \geq e^{t\gamma}] &\leq \frac{E[e^{tY}]}{e^{t\gamma}} \\ &= \frac{M_Y(t)}{e^{t\gamma}}. \end{aligned}$$

Notice that this is true for every $t > 0$, so we can choose that value of t that gives the sharpest upper bound.

If $Z \sim N(0, 1)$, then $M_Z(t) = e^{t^2/2}$, so the previous inequality gives

$$P[Z \geq \gamma] \leq \frac{e^{t^2/2}}{e^{\gamma t}}$$

and choosing $t = \gamma$ gives, for $\gamma > 0$

$$P[Z \geq \gamma] \leq e^{-\gamma^2/2}.$$

The inequality is also obviously true for $\gamma = 0$. One can show that taking $t = \gamma$ gives the sharpest result.

3. (a) We know that

$$M_X(t) = \sum_{j=0}^{\infty} m_j \frac{t^j}{j!}$$

This gives

$$K_X(t) = \log M_X(t) = \log \left[\sum_{j=0}^{\infty} m_j \frac{t^j}{j!} \right].$$

the right hand side is

$$\log \left[1 + m_1 \frac{t}{1!} + m_2 \frac{t^2}{2!} + m_3 \frac{t^3}{3!} + \dots \right]$$

Using the Taylor series expansion $\log(1 + x) = x - x^2/2 + x^3/3 - \dots$, we get

$$\begin{aligned} K_X(t) = &\left[m_1 \frac{t}{1!} + m_2 \frac{t^2}{2!} + m_3 \frac{t^3}{3!} + \dots \right] \\ &- \left[m_1 \frac{t}{1!} + m_2 \frac{t^2}{2!} + m_3 \frac{t^3}{3!} + \dots \right]^2 \Big/ 2 \\ &+ \left[m_1 \frac{t}{1!} + m_2 \frac{t^2}{2!} + m_3 \frac{t^3}{3!} + \dots \right]^3 \Big/ 3 + \dots \end{aligned}$$

Retaining terms up to degree t^3 gives

$$\begin{aligned} K_X(t) = &\left[m_1 t/1! + m_2 t^2/2! + m_3 t^3/3! + \dots \right] \\ &- \left[(m_1)^2 t^2 + 2m_1 m_2 t^3/2 + \dots \right]/2 \\ &+ \left[(m_1)^3 t^3 \right]/3 + \dots. \end{aligned}$$

This gives

$$\begin{aligned} K_X(t) = &\, m_1 \frac{t}{1!} + \left[m_2 - (m_1)^2 \right] t^2/2! \\ &+ \left[m_3 - 3m_1 m_2 + 2(m_1)^3 \right] t^3/3! + \dots. \end{aligned}$$

So, identifying coefficients of $t^r/r!$ on each side for $r = 1, 2, 3$

$$\begin{aligned}
\kappa_1 &= m_1 \\
\kappa_2 &= m_2 - (m_1)^2 = \mu_2 \\
\kappa_3 &= m_3 - 3m_1m_2 + 2(m_1)^3 = \mu_3.
\end{aligned}$$

(b) IF $Y = X - m_1$, then by definition, $m_r^{(Y)} = \mu_r^{(X)} = \mu_r$ for $r = 2, 3, \ldots$. We can also see that

$$\begin{aligned}
M_Y(t) &= E[e^{Yt}] \\
&= E[e^{t(X-m_1)}] \\
&= e^{-tm_1} M_X(t).
\end{aligned}$$

It follows that

$$\begin{aligned}
K_Y(t) &= -tm_1 + K_X(t) \\
&= -\kappa_1 t + K_X(t).
\end{aligned}$$

It follows that **all** the cumulants of Y and X must be the same except for the first. In other words, the cumulants of order greater than 1 **do not depend on the location** of the distribution. We can get $\kappa_2, \kappa_3, \kappa_4, \kappa_5$ from the moments of Y, which are just the central moments for X.

$$K_Y(t) = \log[1 + \mu_2 t^2/2! + \mu_3 t^3/3! + \mu_4 t^4/4! + \mu_5 t^5/5! + \ldots]$$

Once again expanding the right hand side we get

$$\begin{aligned}
K_Y(t) = {} & [\mu_2 t^2/2! + \mu_3 t^3/3! + \mu_4 t^4/4! + \mu_5 t^5/5! + \ldots] \\
& - [\mu_2 t^2/2! + \mu_3 t^3/3! + \mu_4 t^4/4! + \mu_5 t^5/5! + \ldots]^2/2 + \ldots
\end{aligned}$$

Keeping terms to the fifth power of t,

$$\begin{aligned}
K_Y(t) = {} & [\mu_2 t^2/2! + \mu_3 t^3/3! + \mu_4 t^4/4! + \mu_5 t^5/5! + \ldots] \\
& - [\mu_2^2 t^4/4 + 2\mu_2\mu_3 t^5/12 + \ldots]/2 + \ldots
\end{aligned}$$

which gives

$$\begin{aligned}
K_Y(t) = {} & \mu_2 t^2/2! + \mu_3 t^3/3! + (\mu_4 - 3\mu_2^2)t^4/4! \\
& + (\mu_5 - 10\mu_3\mu_2)t^5/5! + \ldots.
\end{aligned}$$

Picking out the coefficients of $t^r/r!$ on each side

$$\begin{aligned}
\kappa_2^{(Y)} &= \mu_2 \\
\kappa_3^{(Y)} &= \mu_3 \\
\kappa_4^{(Y)} &= \mu_4 - 3\mu_2^2 \\
\kappa_5^{(Y)} &= \mu_5 - 10\mu_3\mu_2.
\end{aligned}$$

(c) If $Z = Y/\sigma$, then

$$K_Z(t) = \log M_Z(t) = \log E[e^{Zt}] = \log E[e^{Yt/\sigma}] = K_Y(t/\sigma).$$

It follows that $\kappa_r^{(Z)} = \kappa_r^{(Y)}/\sigma^r$, and so

$$\begin{aligned}
\kappa_2^{(Z)} &= 1 \\
\kappa_3^{(Z)} &= \mu_3/\mu_2^{3/2} \\
\kappa_4^{(Z)} &= (\mu_4 - 3\mu_2^2)/\mu_2^2 \\
\kappa_5^{(Z)} &= (\mu_5 - 10\mu_3\mu_2)/\mu_2^{5/2}.
\end{aligned}$$

(d) For the normal distribution, $K_X(t) = \mu t + \sigma^2 t^2/2$, so that

$$\begin{aligned}
\kappa_1 &= \mu \\
\kappa_2 &= \sigma^2 \\
\kappa_r &= 0 \text{ for } r > 2.
\end{aligned}$$

So

$$\begin{aligned}
m_1 &= \kappa_1 = \mu \\
m_2 &= \kappa_2 + (m_1)^2 = \sigma^2 + \mu^2 \\
m_3 &= \kappa_3 + 3m_2m_1 - 2(m_1)^3 = 3\sigma^2\mu + \mu^3.
\end{aligned}$$

4. We shall require that $-1 < t < 1$ for some parts of the following derivation:

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} e^{xt}\frac{e^{-|x|}}{2}dx \\
&= \int_{-\infty}^{0} e^{xt}\frac{e^{x}}{2}dx + \int_{0}^{\infty} e^{xt}\frac{e^{-x}}{2}dx \\
&= \int_{-\infty}^{0} \frac{e^{x(1+t)}}{2}dx + \int_{0}^{\infty} \frac{e^{-x(1-t)}}{2}dx \\
&= \frac{e^{x(1+t)}}{2(1+t)}\Big|_{-\infty}^{0} + \frac{-e^{-x(1-t)}}{2(1-t)}\Big|_{0}^{\infty} \\
&= \frac{1}{2(1+t)} + \frac{1}{2(1-t)} \\
&= \frac{1}{1-t^2}.
\end{aligned}$$

$$\begin{aligned}
K_X(t) &= -\log(1-t^2) = t^2 + (t^2)^2/2 + (t^2)^3/3 + \ldots \\
&= 0t + 2\frac{t^2}{2!} + 0\frac{t^3}{3!} + 12\frac{t^4}{4!} + \ldots.
\end{aligned}$$

So,

$$\begin{aligned}
\kappa_1 &= 0 \\
\kappa_2 &= 2 \\
\kappa_3 &= 0 \\
\kappa_4 &= 12.
\end{aligned}$$

5. (a) This can't work, because if it were true

$$F(\infty, \infty) = H(\infty) + G(\infty) = 1 + 1 = 2.$$

This is inconsistent with $F(\infty, \infty)$ being a probability.

(b) If we consider the joint distribution of independent random variables X with distribution function $H(x)$ and Y with distribution function $G(y)$, then

$$\begin{aligned} F(x, y) &= P[X \leq x, Y \leq y] \\ &= P[X \leq x]P[Y \leq y] \\ &= H(x)G(y). \end{aligned}$$

This verifies that $H(x)G(y)$ is a joint distribution function.

(c) This one does not work because

$$F(-\infty, \infty) = \max[H(-\infty), G(\infty)] = \max[0, 1] = 1.$$

But this should be 0.

(d) As the question suggests, consider the singular joint distribution for which X and Y always take exactly the same value, and for which the marginal distribution function of X is $H(x)$. A typical value for (X, Y) is (x, x). The support of this distribution in the XY plane is the line $y = x$. We have

$$\begin{aligned} F(x, y) &= P[X \leq x, Y \leq y] \\ &= P[X \leq x, X \leq y] \\ &= \min[H(x), H(y)]. \end{aligned}$$

So we have shown that this is a distribution function, though for a singular distribution.

A more difficult exercise would be to show that $F(x, y) = \min[H(x), G(y)]$ is a joint distribution function.

Solutions for Exercise 4

1.

$$\begin{aligned}
P(K=k) &= \sum_{r=k}^{\infty} P[K=k|R=r]P(R=r) \\
&= \sum_{r=k}^{\infty} \binom{r}{k} p^k (1-p)^{r-k} e^{-\lambda}\lambda^r/r! \\
&= \frac{e^{-\lambda}\lambda^k p^k}{k!} \sum_{r=k}^{\infty} \frac{[\lambda(1-p)]^{r-k}}{(r-k)!} \\
&= \frac{e^{-\lambda}\lambda^k p^k}{k!} e^{\lambda(1-p)} \\
&= e^{-\lambda p}\frac{(\lambda p)^k}{k!}.
\end{aligned}$$

So K has a Poisson distribution with mean λp, which is what one might expect for a random thinning of random events.

2. Suppose that the density function for Λ is, for $\lambda > 0$,

$$f_\Lambda(\lambda) = e^{-k\lambda} k^r \lambda^{r-1}/\Gamma(r),$$

where k and r are positive. Then the marginal distribution of X is given by

$$\begin{aligned}
P(X=x) &= \int_0^\infty P[X=x|\Lambda=\lambda] f_\Lambda(\lambda) d\lambda \\
&= \int_0^\infty e^{-\lambda}\frac{\lambda^x}{x!} e^{-k\lambda}\frac{k^r\lambda^{r-1}}{\Gamma(r)} d\lambda \\
&= \frac{\Gamma(r+x)k^r}{(1+k)^{x+r}x!\Gamma(r)} \int_0^\infty e^{-\lambda(1+k)} \frac{\lambda^{r+x-1}(1+k)^{r+x}}{\Gamma(r+x)} d\lambda \\
&= \frac{\Gamma(r+x)k^r}{(1+k)^{x+r}x!\Gamma(r)} \\
&= \binom{r+x-1}{x} \left[\frac{1}{1+k}\right]^x \left[\frac{k}{1+k}\right]^r.
\end{aligned}$$

This is a natural way to produce a negative binomial distribution for which there is no requirement that r be integer.

3. The marginal density function for X is, for $x > 0$,

$$\begin{aligned}
f_X(x) &= \int_0^\infty f_{X|\Theta}(x|\Theta=\theta) f_\Theta(\theta) d\theta \\
&= \int_0^\infty \theta e^{-x\theta} e^{-\lambda\theta} \frac{\lambda^\alpha \theta^{\alpha-1}}{\Gamma(\alpha)} d\theta \\
&= \frac{\lambda^\alpha\Gamma(\alpha+1)}{(x+\lambda)^{(\alpha+1)}\Gamma(\alpha)} \int_0^\infty e^{-\theta(x+\lambda)} \frac{\theta^{(\alpha+1)-1}(x+\lambda)^{(\alpha+1)}}{\Gamma(\alpha+1)} d\theta \\
&= \frac{\lambda^\alpha\Gamma(\alpha+1)}{(x+\lambda)^{(\alpha+1)}\Gamma(\alpha)} \\
&= \frac{\alpha}{\lambda\left(1+\frac{x}{\lambda}\right)^{\alpha+1}}.
\end{aligned}$$

This is a Pareto density function (with scale parameter λ).

4.

$$\begin{aligned}\text{Cov}[X,Y] &= E[XY] - E[X]E[Y] \\ &= E[XE[Y|X]] - E[X]E[E[Y|X]] \\ &= E[X(1+X)/2] - E[X]E[(1+X)/2] \\ &= E[X]/2 + E[X^2]/2 - E[X]/2 - (E[X])^2/2 \\ &= [\text{Var}X]/2.\end{aligned}$$

For $0 < x < 1$,

$$f_X(x) = k(1-x)$$

for some constant k. One can easily see that $k = 2$, so for $0 < x < 1$

$$f_X(x) = 2(1-x).$$

The mean of X is $1/3$, and

$$E[X^2] = \int_0^1 x^2 2(1-x)dx = 2/3 - 1/2 = 1/6.$$

The variance of X is $1/6 - 1/9 = 1/18$, and the $\text{Cov}(X,Y) = 1/36$.

Notice that $E[Y|X]$, which is often called the regression of Y on X, is linear in X, since $E[Y|X] = (1+X)/2$. The regression slope coefficient is $1/2$. When the regression is linear we know that the slope coefficient is

$$\frac{\text{Cov}(X,Y)}{\text{Var}X}$$

in concordance with the results used above.

5. For $0 < x < 1$ the support for the conditional distribution of Y given $X = x$ is the strip going from $(x,0)$ to (x,x). Outside this strip the conditional density is 0. To get the marginal density of X, we must integrate along strips at each different value of x. For $0 < x < 1$,

$$\begin{aligned}f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy &= \int_0^x 8xydy \\ = 8x\int_0^x ydy &= 4x^3.\end{aligned}$$

Obviously, the marginal density function for X at values x outside the interval $(0,1)$ must be 0. The conditional density function of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

which for $0 < x < 1$ and $0 < y < x$ is

$$f_{Y|X}(y|x) = \frac{8xy}{4x^3} = \frac{2y}{x^2}.$$

The conditional mean and variance are now easy to find. For $0 < x < 1$,

$$\begin{aligned} E[Y^r|X = x] &= \int_0^x y^r \frac{2y}{x^2} dy \\ &= \int_0^x \frac{2y^{r+1}}{x^2} dy \\ &= \frac{2x^{r+2}}{(r+2)x^2} \\ &= \frac{2x^r}{r+2}. \end{aligned}$$

For other values x, the conditional mean values are not defined. We get $E[Y|X = x] = 2x/3$, $E[Y^2|X = x] = 2x^2/4 = x^2/2$, $\text{Var}[Y|X = x] = x^2/2 - 4x^2/9 = x^2/18$.

$E[XY|X = x] = xE[Y|X = x] = 2x^2/3$, so

$$\begin{aligned} E[XY] &= E[E[XY|X]] = E[2X^2/3] \\ &= \int_0^1 \frac{2x^2}{3} 4x^3 dx \\ &= \int_0^1 \frac{8x^5}{3} dx \\ &= 4/9. \end{aligned}$$

To get the covariance of X, Y we need also $E[X] = \int_0^1 x4x^3 dx = 4/5$, and $E[Y] = E[E[Y|X]] = E[2X/3] = 8/15$. The covariance of X, Y is

$$\frac{4}{9} - \frac{4}{5}\frac{8}{15} = 4/225.$$

Notice that $E[Y|X]$, which is often called the regression of Y on X, is linear in X, since $E[Y|X] = 2X/3$. The regression slope coefficient is $2/3$. When the regression is linear we know that the slope coefficient is

$$\frac{\text{Cov}(X, Y)}{\text{Var}X}$$

so we could say from this that $\text{Cov}(X, Y) = \frac{2}{3}\text{Var}X$, as may be checked directly.

6. It is always true that if $X \perp Y$ then $\text{Cov}(X, Y) = 0$ and so the correlation is 0. It is necessary to prove that for this family of distributions the reverse implication is also true.

   If $a = 0$ then the joint distribution is uniform over the unit square, and so $X \perp Y$. It is enough therefore to show that when $\text{Cov}(X, Y) = 0$, $a = 0$.

   The marginal density function of X for $0 < x < 1$ is

$$f_X(x) = \int_0^1 [1 - a(1 - 2x)(1 - 2y)]dy = 1.$$

So X, and by symmetry Y have $U(0,1)$ distributions. It follows that $EX = EY = 1/2$.

$$\begin{aligned}
E(XY) &= \int_0^1 \int_0^1 xy[1 - a(1-2x)(1-2y)]dydx \\
&= \frac{x^2}{2}\Big|_0^1 \frac{y^2}{2}\Big|_0^1 - a\int_0^1 x(1-2x)dx \int_0^1 y(1-2y)dy \\
&= \frac{1}{4} - a\left[\frac{x^2}{2} - \frac{2x^3}{3}\right]\Big|_0^1 \left[\frac{y^2}{2} - \frac{2y^3}{3}\right]\Big|_0^1 \\
&= 1/4 - a[1/2 - 2/3]^2 \\
&= 1/4 - a/36.
\end{aligned}$$

So $\text{Cov}(X,Y) = 1/4 - a/36 - (1/2)^2 = -a/36$. If the the correlation of X, Y is 0, then so is the covariance and so $a = 0$. This implies that $X \perp Y$.

Note. This is another bivariate family with regressions that are linear. It is easy to see that

$$E[Y|X] = 1/2 - a(1-2x)/6$$

so that the slope coefficient is a/3. We can verify directly that

$$\text{Cov}(X,Y) = \frac{a}{3}\text{Var}X = \frac{a}{3}\frac{1}{12}.$$

7. Consider exponential distributions $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$. If the means are equal the $\lambda = \mu$. Let $Z = X + Y$ then

$$\begin{aligned}
f_Z(z) &= \textstyle\int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx \\
&= \textstyle\int_0^z \lambda e^{-\lambda x}\lambda e^{-\lambda(z-x)} \\
&= \lambda^2 \textstyle\int_0^z e^{-\lambda z}dx \\
&= \lambda^2 z e^{-\lambda z} \text{ for } z \geq 0
\end{aligned}$$

For the general case

$$\begin{aligned}
f_Z(z) &= \textstyle\int_0^z \lambda e^{-\lambda x}\mu e^{-\mu(z-x)}dx \\
&= \lambda\mu e^{-\mu z}\textstyle\int_0^z e^{-x(\lambda-\mu)}dx \\
&= \lambda\mu e^{-\mu z}[-e^{-x(\lambda-\mu)}/(\lambda-\mu)]_0^z \\
&= [\lambda\mu/(\lambda-\mu)](e^{-\mu z} - e^{-\lambda z}) \text{ for } z \geq 0
\end{aligned}$$

8. We will interpret $\sum_{i=1}^N X_i$ as 0 for $N = 0$, and $\prod_{i=1}^N$ operations as giving 1 when $N = 0$. Then

$$\begin{aligned}
M_Y(t) &= E[e^{Yt}] \\
&= E[E[e^{Yt}|N]] = E[E[e^{t\sum_{i=1}^N X_i}|N]] \\
&= E[\prod_{i=1}^N E[e^{tX_i}]] \\
&= E[\prod_{i=1}^N M_X(t)] \\
&= E[M_X(t)^N] = E[e^{N \ln M_X(t)}] \\
&= E[e^{NK_X(t)}] = M_N(K_X(t)).
\end{aligned}$$

Taking logs on both sides gives the result, though the form above may often be more useful.

The example has a geometric distribution for N.

$$\begin{aligned}
M_N(t) &= \sum_{n=1}^{\infty} e^{tn} p^{n-1}(1-p) \\
&= e^t(1-p)\sum_{n=1}^{\infty}(pe^t)^{n-1} \\
&= \frac{e^t(1-p)}{1-pe^t}.
\end{aligned}$$

X_i has a Gamma$(\alpha, 2)$ distribution with MGF

$$M_X(t) = \frac{\alpha^2}{(\alpha-t)^2}.$$

Using the result we obtained above,

$$\begin{aligned}
M_Y(t) = M_N(K_X(t)) &= \frac{e^{K_X(t)}(1-p)}{1-pe^{K_X(t)}} \\
&= \frac{M_X(t)(1-p)}{1-pM_X(t)} \\
&= \frac{(1-p)\frac{\alpha^2}{(\alpha-t)^2}}{1-p\frac{\alpha^2}{(\alpha-t)^2}} \\
&= \frac{(1-p)\alpha^2}{(\alpha-t)^2-p\alpha^2} \\
&= \frac{(1-p)\alpha^2}{[(\alpha-t)-\sqrt{p}\alpha][(\alpha-t)+\sqrt{p}\alpha]} \\
&= \frac{(1-p)\alpha^2}{[\alpha(1-\sqrt{p})-t][\alpha(1+\sqrt{p})-t]} \\
&= \frac{(1-\sqrt{p})(1+\sqrt{p})\alpha}{2\sqrt{p}}\frac{1}{\alpha(1-\sqrt{p})-t} - \frac{(1-\sqrt{p})(1+\sqrt{p})\alpha}{2\sqrt{p}}\frac{1}{\alpha(1+\sqrt{p})-t} \\
&= \frac{1+\sqrt{p}}{2\sqrt{p}}\frac{\alpha(1-\sqrt{p})}{\alpha(1-\sqrt{p})-t} - \frac{1-\sqrt{p}}{2\sqrt{p}}\frac{\alpha(1+\sqrt{p})}{\alpha(1+\sqrt{p})-t}
\end{aligned}$$

We can identify the moment generating function as a mixture of two exponential distributions with scale parameters $\alpha(1-\sqrt{p})$ and $\alpha(1+\sqrt{p})$ and mixing weights $\frac{1+\sqrt{p}}{2\sqrt{p}}$ and $-\frac{1-\sqrt{p}}{2\sqrt{p}}$. One of the weights is negative. The corresponding density function is

$$\begin{aligned}
f_Y(y) &= \frac{1+\sqrt{p}}{2\sqrt{p}}\alpha(1-\sqrt{p})e^{-\alpha(1-\sqrt{p})y} - \frac{1-\sqrt{p}}{2\sqrt{p}}\alpha(1+\sqrt{p})e^{-\alpha(1+\sqrt{p})y} \\
&= \frac{\alpha(1-p)}{2\sqrt{p}}\left[e^{-\alpha(1-\sqrt{p})y} - e^{-\alpha(1+\sqrt{p})y}\right].
\end{aligned}$$

9. Notice that because the variances of X, Y are both 1, $\kappa_{11} = \rho$. We need to work out

$$\mathrm{Cov}(X^2, Y^2) = E[X^2Y^2] - E[X^2]E[Y^2] = \mu_{22} - \mu_{20}\mu_{02}$$

because we have zero means. So

$$\begin{aligned}
\mathrm{Cov}(X^2, Y^2) &= (\kappa_{22} + \kappa_{20}\kappa_{02} + 2\kappa_{11}^2) - \kappa_{20}\kappa_{02} \\
&= \kappa_{22} + 2\kappa_{11}^2 = 2\kappa_{11}^2.
\end{aligned}$$

Also

$$\begin{aligned}\text{Var}(X^2) &= E[X^4] - [E(X^2)]^2 = \mu_{40} - \mu_{20}^2 \\ &= [\kappa_{40} + 3\kappa_{20}^2] - \kappa_{20}^2 \\ &= \kappa_{40} + 2\kappa_{20}^2 = 0 + 2 \times 1^2 = 2.\end{aligned}$$

The variance of Y^2 is the same as the variance of X^2, so the correlation coefficient is

$$\frac{2\kappa_{11}^2}{2} = \rho^2.$$

This is less than ρ unless $\rho^2 = 1$: squaring reduces the strength of the linear association.

10. Once you know X you also know X^2, so the joint distribution of $X, Y = X^2$ is a degenerate distribution with all its probability in the (X, Y) plane concentrated on the curve $Y = X^2$. Nevertheless, we can use the usual definitions even for a case like this.

$$M_{X,X^2}(s,t) = E[e^{sX+tX^2}] = \int_{-\infty}^{\infty} e^{sx+tx^2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Now we can make the integrand like a normal density function.

$$sx + tx^2 - x^2/2 = -x^2(1-2t)/2 + sx = -(1-2t)(x - s/(1-2t))^2/2 + s^2/[2(1-2t)]$$

So provided that $1 - 2t > 0$, we manufacture a normal density with mean $\frac{s}{1-2t}$ and variance $\frac{1}{1-2t}$.

$$\begin{aligned}M_{X,X^2}(s,t) &= \frac{e^{s^2/[2(1-2t)]}}{\sqrt{1-2t}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi/(1-2t)}} e^{-(1-2t)(x-s/(1-2t))^2/2} dx \\ &= \frac{e^{s^2/[2(1-2t)]}}{\sqrt{1-2t}}.\end{aligned}$$

Putting $t = 0$ gives $e^{s^2/2}$, which is the mgf for X, and putting $s = 0$ gives $\frac{1}{\sqrt{1-2t}}$ which is the $\chi^2_{(1)}$ mgf for X^2.

11. Suppose that $W = XY$ then

$$M_W(u) = E(M_{W|X}(u|X)) = E(E(e^{uXY}|X)) = E(M_Y(uX))$$

Since Y is a standard normal. $M_Y(t) = e^{t^2/2}$. Thus

$$M_W(u) = E(e^{u^2X^2/2}) = M_{X^2}(u^2/2)$$

But X is also standard normal so X^2 is χ_1^2 and $M_{X^2}(t) = (1-2t)^{-1/2}$. Thus

$$M_W(u) = \frac{1}{\sqrt{1-u^2}}$$

This MGF does not correspond to any 'standard' distributions. However, if we take $Z = UV + XY$. Then by independence the MGF of the sum is the product of the MGFs so

$$M_Z(u) = \frac{1}{1-u^2}$$

which is the MGF of a Laplace distribution.

12. The transformation here is $u = x/y$ and $v = x + y$ so the inverse transformation is $x = uv/(1+u)$ and $y = v/(1+u)$. So

$$\begin{array}{ll} \frac{\partial x}{\partial u} = v/(1+u)^2 & \frac{\partial x}{\partial v} = u/(1+u) \\ \frac{\partial y}{\partial u} = -v/(1+u)^2 & \frac{\partial y}{\partial v} = 1/(1+u) \end{array}$$

and the Jacobian of the transformation is $v/(1+u)^2$. The joint density is

$$f_{U,V}(u,v) = f_{X,Y}(x(u,v), y(u,v))v/(1+u)^2 = \lambda^2 v e^{-\lambda v}/(1+u)^2 \text{ for } u, v > 0$$

13. In this case we will work out the Jacobian for the transformation and then invert. Here $x = r\cos\theta$ and $y = r\sin\theta$ so

$$\begin{array}{ll} \frac{\partial x}{\partial r} = \cos\theta & \frac{\partial x}{\partial \theta} = -r\sin\theta \\ \frac{\partial y}{\partial r} = \sin\theta & \frac{\partial y}{\partial \theta} = r\cos\theta \end{array}$$

So the Jacobian is just r. Thus

$$f_{X,Y}(x,y) = f_{R,\Theta}(r(x,y), \theta(x,y))/r$$

But from the question R and Θ are independent and Θ is uniformly distributed over $[0, 2\pi]$. Thus

$$f_{X,Y}(x,y) = f_R(r)/2\pi r = \frac{1}{2\pi\sqrt{x^2+y^2}} f_R(\sqrt{x^2+y^2})$$

This is a bivariate density that has circular contours of equal density.

14. (a) Let's use the result for the mean from the solutions for Exercise 4.

$$E[X] = -\int_{-\infty}^{0} F_X(z)dz + \int_0^{\infty} [1 - F_X(z)]dz$$

$$E[Y] = -\int_{-\infty}^{0} F_Y(z)dz + \int_0^{\infty} [1 - F_Y(z)]dz$$

Taking these two together,

$$E[Y] - E[X] = \int_{-\infty}^{0} [F_X(z) - F_Y(z)dz + \int_0^{\infty} [F_X(z) - F_Y(z)]dz$$

so that since the integrands are positive for all z

$$E[Y] - E[X] > 0.$$

(b) This is a very weak result, so easy to prove.

$$F_X(z) > F_Y(z)$$

so

$$F_{X,Y}(z,\infty) > F_{X,Y}(\infty, z)$$
$$F_{X,Y}(z,\infty) - F_{X,Y}(z,z) > F_{X,Y}(\infty,z) - F_{X,Y}(z,z).$$

So

$$F_{X,Y}(z,\infty) - F_{X,Y}(z,z) > 0$$

and since $P(Y > X) \geq F_{X,Y}(z,\infty) - F_{X,Y}(z,z)$, the result follows.

(c) Here we can just calculate directly:

$$\begin{aligned}
P(Y > X) &= \int_{-\infty}^{\infty} [\int_{x}^{\infty} f_X(x) f_Y(y) dy] dx \\
&= \int_{-\infty}^{\infty} [1 - F_Y(x)] f_X(x) dx \\
&> \int_{-\infty}^{\infty} [1 - F_X(x)] f_X(x) dx \\
&= -[1 - F_X(x)]^2/2 \Big|_{-\infty}^{\infty} \\
&= 1/2.
\end{aligned}$$