MATH4406 – HW1

Madeleine Nargar, 41214538

Exercise 1

a) $\emptyset \in \mathcal{F}$

b) A $\in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

c) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_i A_i \in \mathcal{F}$

1. $\Omega = \emptyset^c$
   (a) + (b) $\Rightarrow \emptyset^c \in \mathcal{F}$
   $\Rightarrow \Omega \in \mathcal{F}$

2. By (b), if $A_i \in \mathcal{F}$ then $A_i^c \in \mathcal{F}$. By (c), $A_1^c, A_2^c, \dots \in \mathcal{F}$ gives $\bigcup_i A_i^c \in \mathcal{F}$
   By De Morgan's Law, $\bigcup_i A_i^c = \bigcap_i A_i \in \mathcal{F}$

Exercise 2

1.

$$
\begin{aligned}
1 &= \mathbb{P}(\Omega) \\
&= \mathbb{P}(A \cup A^c) \\
&= \mathbb{P}(A) + \mathbb{P}(A^c) && \text{(by sum rule)} \\
\mathbb{P}(A^c) &= 1 - \mathbb{P}(A)
\end{aligned}
$$

2. $\emptyset = \Omega^c$. Using the result above (Ex 2.1), $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$.

3. Expressing A, B and A $\cup$ B as a union of disjoint subsets:
   $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$ (1)
   $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)$ (2)
   $\mathbb{P}(A \cup B) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) + \mathbb{P}(A^c \cap B)$
   Substituting (1) and (2) gives,
   $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Exercise 4

A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. So for independent A and B,

$$\begin{aligned}
\mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} \\
&= \mathbb{P}(A) \\
\text{Similarly, } \mathbb{P}(A) &= \mathbb{P}(A|B) \\
&= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\
\Rightarrow \mathbb{P}(A \cap B) &= \mathbb{P}(A)\mathbb{P}(B) \quad \text{i.e. A, B independent}
\end{aligned}$$

Exercise 5

$$\begin{aligned}
\mathbb{P}(6|even) &= \frac{\mathbb{P}(6 \cap even)}{\mathbb{P}(even)} \\
&= \frac{\mathbb{P}(6)}{\mathbb{P}(even)} \\
&= \frac{1/6}{1/2} \\
&= 1/3
\end{aligned}$$

Exercise 6

1. This follows directly from the definition of conditional probability, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
2. For $B_1, B_2, \ldots$ forming a partition of the sample space, $A \cap B_i$ are disjoint for all $i \neq j$ so we can use the sum rule, giving: $\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i)$. Applying the product rule proved above, $\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)$.

3.

$$\begin{aligned}
\mathbb{P}(B_i|A) &= \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)} && \text{(using product rule)} \\
&= \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_j \mathbb{P}(A|B_j)\mathbb{P}(B_j)} && \text{(apply law of total probability to A)}
\end{aligned}$$

Exercise 7

Thinking about a comparable problem makes the solution more intuitive. Imagine that instead of three boxes there are a thousand, with only one containing a prize. After you pick a box the host opens all but one of the other boxes. In this case the choice to switch seems obvious, as you're gambling on whether the prize was more likely to be in your original box (a one in a thousand chance), or in any of the other 999 boxes. Given the host will never open the box that actually has the prize in it, they have effectively "combined" 999 one-in-a-thousand chances (of the prize being in one of the boxes you didn't pick) into a single box. So it should be an obvious decision to switch from your original box (1/1000 chance of winning) to the one other unopened box (999/1000 chance of winning).

Exercise 8

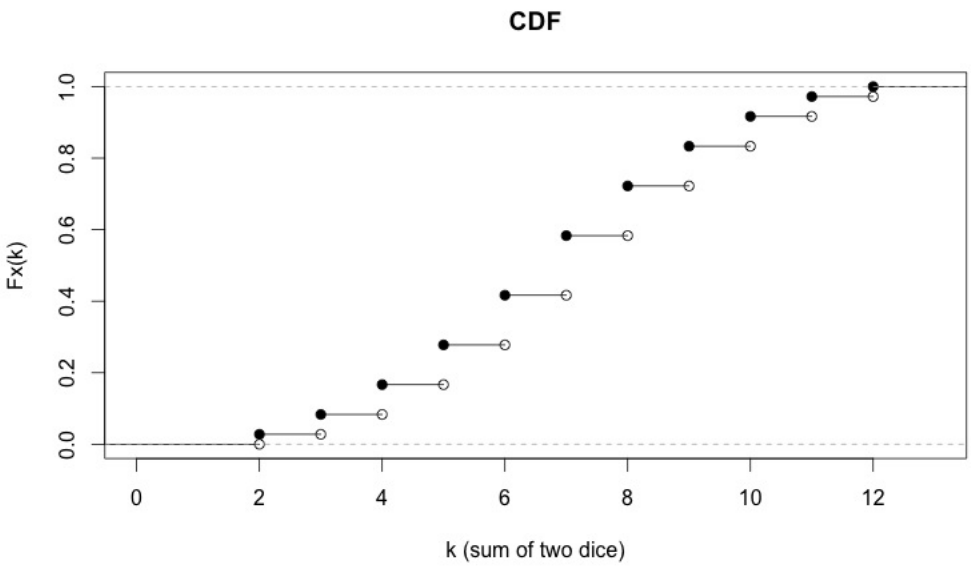
1. $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x])$
   $\lim_{x\to\infty} F_X(x) = \mathbb{P}(X \in \emptyset) = \mathbb{P}(\emptyset) = 0$

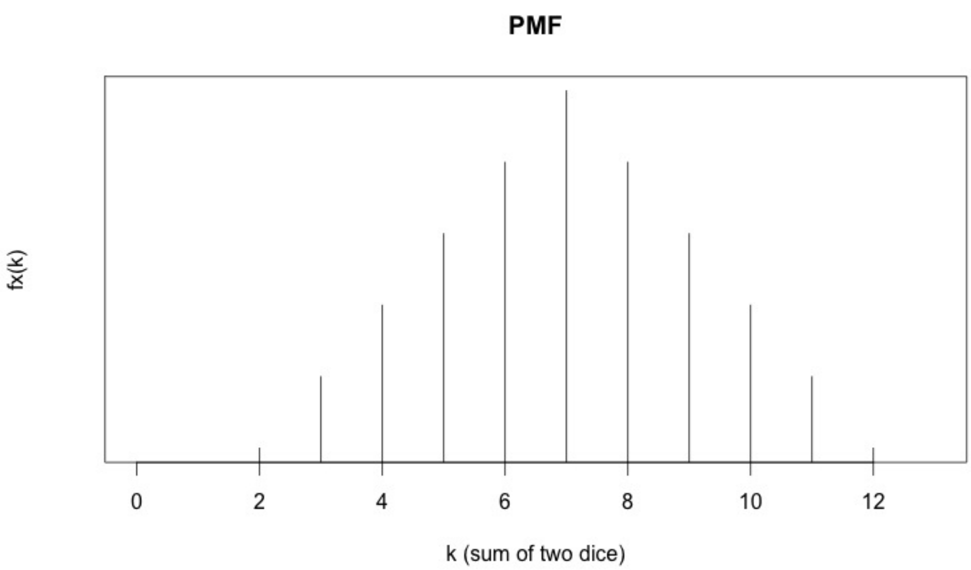
2. $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x])$
   $\lim_{x\to\infty} F_X(x) = \mathbb{P}(X \in (-\infty, \infty)) = \mathbb{P}(\Omega) = 1$

3. $F_X(x)$ is non-decreasing if for all $a < b,\ F_X(a) \leq F_X(b)$.

$$\begin{aligned}
F_X(b) &= \mathbb{P}(X \leq b) \\
&= \mathbb{P}(X \leq a) + \mathbb{P}(a < X \leq b) && \text{(by sum rule)} \\
&= F_X(a) + \mathbb{P}(a < X \leq b) \\
\Rightarrow F_X(b) &\geq F_X(a) && \text{(as } \mathbb{P}(x) \geq 0\ \forall x)
\end{aligned}$$

Exercise 9, 10





Heights of the bars in the PMF plot gives the size of the jumps between points in the CDF.

Exercise 11

Using the probabilities from above:

x	≤ 1	2	3	4	5	6	7	8	9	10	11	12	> 12
$p_X(x)$	0	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	0

$$\begin{aligned}
\mathbb{E}[X] &= \sum_k k p_X(k) \\
&= 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} + 8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36} + 11\frac{2}{36} + 12\frac{1}{36} \\
&= 7
\end{aligned}$$

Exercise 12

$$\begin{aligned}
\sum_{k=0}^{\infty} \bar{F}_k(k) &= \sum_{k=0}^{\infty} \mathbb{P}(X > k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) \\
&= \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} p_X(i) \\
&= \sum_{i=1}^{\infty} \sum_{k=1}^{i} p_X(i) \\
&= \sum_{i=1}^{\infty} i p_X(i) \\
&= \sum_{i=-\infty}^{\infty} i p_X(i) \qquad \text{as non-negative, } p_X(i) = 0 \ \forall i < 0 \\
&= \mathbb{E}[X]
\end{aligned}$$

Exercise 13

Using $\mathbb{E}[h(X)] = \sum_k h(k) p_X(k)$

1. Setting $h(x) = cX$,
   $\mathbb{E}[cX] = \sum_k ckp_X(k) = c\sum_k kp_X(k) = c\mathbb{E}[X]$

2.

Set $Z(\omega) = X(\omega) + Y(\omega)$. Let an event $A_k = \{\omega \in \Omega : X(\omega) = x, Y(\omega) = y, Z(\omega) = z = x + y\}$

$$\begin{aligned}\mathbb{E}[X+Y] = \mathbb{E}[Z] &= \sum_z z\mathbb{P}(Z(\omega) = z : X(\omega) \in A, Y(\omega \in B)) \\ &= \sum_z (X+Y)\mathbb{P}(Z(\omega) = z : X(\omega) = x, Y(\omega = y)) \\ &= \sum_x X\mathbb{P}(X(\omega) = x) + \sum_y Y\mathbb{P}(Y(\omega) = y) \\ &= \mathbb{E}[X] + \mathbb{E}[Y]\end{aligned}$$

Exercise 14

$$Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b \qquad \text{(by ex 13.1, 13.2, with } \mathbb{E}(c) = c)$$

$$\begin{aligned}Var(aX+b) &= \mathbb{E}((aX + b - (a\mathbb{E}(X) + b))^2) \\ &= \mathbb{E}(a^2(X - \mathbb{E}(X))^2) \\ &= a^2\mathbb{E}((X - \mathbb{E}(X))^2) \\ &= a^2\mathbb{V}ar(X)\end{aligned}$$

Exercise 15

For non-negative Z, $\mathbb{E}(Z)$ is non-negative (as $\mathbb{P}(A) \geq 0$ for any A).
So $Var(X) = \mathbb{E}((X - \mathbb{E}X)^2) = \sum_k (X - \mathbb{E}X)^2\ p_X(k)$ can only be zero if $X(k) = \mathbb{E}X$ for all k (otherwise we'd need $p_X(k) = 0$ for all k, violating $p_X(\Omega) = 1$). This implies X is a constant, which we will call k_0
Using $k_0 = X = \mathbb{E}X = \sum_{k\in\{k_0\}} k_0\ p_X(k) = k_0\ p_X(k_0)$
$\Rightarrow p_X(k_0) = 1$

Exercise 16

For the binary sequences $B_n(\omega) = (b_1, b_2, \ldots, b_n)$ with $b_i \in \{0,1\}$, $b_i = \mathbf{I}\{b_i = 1\}$. So setting X as the number of successes in B, $X(\omega) = \sum_{i=1}^n \mathbf{I}\{b_i = 1\}$. b_i are independent events, each with $\mathbb{P}(\{b_i = 1\}) = p$. For a particular sequence $B_n(\omega)$ consisting of $\{b_i = 1\}$ for a particular k of the events, with $\{b_i = 0\}$ for the remaining $(n-k)$ events,

$\mathbb{P}(\{B_n(\omega)\}) \ = \ \mathbb{P}(b_1)\mathbb{P}(b_2)\ldots\mathbb{P}(b_n) \ = \ p^k(1-p)^{n-k}$
$\mathbb{P}(X(\omega)=k) = \sum_A \mathbb{P}(B_n(\omega))$ with $A = \{B_n(\omega) : b_i = 1 \text{ for k of i} \in (0,n)\}$
$= \sum_A p^k(1-p)^{n-k}$
$= \binom{n}{k} p^k(1-p)^{n-k}$ as $|A| = \binom{n}{k}$

Exercise 17

According to the binomial theorem, $(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} p^i(1-p)^{n-i}$.
Setting $a = p,\ b = 1-p$,
$\sum_{i=0}^{n} \mathbb{P}(X=i) = \sum_{k=0}^{n} \binom{n}{i} p^i(1-p)^{n-i} = (p+(1-p))^n = 1^n = 1.$

Exercise 18

1.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=-\infty}^{\infty} k\mathbb{P}(X=k) \\
&= \sum_{k=0}^{n} k\binom{n}{k} p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!} p^k(1-p)^{n-k} \qquad \text{(first term zero)} \\
&= \sum_{k=1}^{n} np\frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k} \\
&= \sum_{k'=0}^{n-1} np\frac{(n-1)!}{k'!(n-1-k')!} p^{k'}(1-p)^{(n-1)-k'} \qquad \text{(replacing k with k'=k-1, then n-k=n-1-k')} \\
&= np\sum_{k'=0}^{n-1} \binom{n-1}{k'} p^{k'}(1-p)^{(n-1)-k'} \\
&= np(p-(1-p)^{n-1} \qquad \text{(as above, using binomial theorem with a=p, b=1-p)} \\
&= np
\end{aligned}$$

2. If X is the number of successes ($\{x_i = 1\}$) in n Bernoulli trials, then $Z = n - X =$ number of failures ($\{x_i = 0\} \equiv \{z_i = 1\}$) in n Bernoulli trials. We can think of Z as

binomially distributed with $\mathbb{P}(z_i = 1) = \mathbb{P}(x_i = 0) = 1 - p$.
$\Rightarrow \mathbb{P}(Z = k) = \binom{n}{k}(1-p)^k p^{n-k}$

Exercise 19

This is binomially distributed with $X = \#$ of correct answers, $n = 20,\ p = 0.25$.

$$\begin{aligned}
\mathbb{P}(X \geq 10) &= \sum_{k=10}^{20} \mathbb{P}(X = k) \\
&= \sum_{k=10}^{20} \binom{20}{k} 0.25^k 0.75^{20-k} \\
&= 0.013864 \qquad \text{looks like passing by guessing is much harder than I thought!}
\end{aligned}$$

Exercise 20

For the event $X = k$ (that the k^{th} Bernoulli trial is the first success) to occur there must be k-1 failures followed by one success. For independent trials b_i, each with probability of success p:

$\mathbb{P}(X = k) = \mathbb{P}(\{b_1 = 0\})\mathbb{P}(\{b_1 = 0\}) \ldots \mathbb{P}(\{b_{k-1} = 0\})\mathbb{P}(\{b_k = 1\}) = (1-p)^{k-1}p$,
for $k = 1, 2, \ldots$

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathbb{P}(X = k) &= \sum_{k=1}^{\infty} (1-p)^{k-1} p \\
&= \sum_{k'=0}^{\infty} (1-p)^{k'} p \\
&= \frac{p}{1-(1-p)} \qquad \text{(infinite geometric series, } a = p, r = 1-p, |r| < 1\text{,converges to } \frac{a}{1-r}) \\
&= 1
\end{aligned}$$

Exercise 21

Here we want 20 successes (no flat tire) in 20 trials. Assuming the chances of a flat tire on consecutive rides are independent (presumably they aren't really...), we have X=# successes in 20 trials, p=0.99, k=20, n=20. $\mathbb{P}(X = 20) = \binom{20}{20} 0.99^{20} 0.01^0 = 0.8179$. I used binomial distribution for number of puncture-free rides, but would be same using geometric distribution for number of rides taken without a "success" (with success = flat tire with probability 0.01).

Exercise 22

If X is the number of trials until a success, then Y = X - 1 is the number of failures until a success. So based on the distribution and support of X:

$\mathbb{P}(Y = k) = \mathbb{P}(X = k + 1) = (1 - p)^k p, \qquad k = 0, 1, \ldots$

Exercise 23

Let $B_k^i = (b_1, b_2 \ldots, b_k)$ be a sequence of k Bernoulli trials each with probability of success p. For a given B_k^i consisting of m successes and $k-m$ failures, $\Pr(B_k) = (1-p)^{k-m} p^m$. If we let X be the number of Bernoulli trials until m successes, then $\Pr(X = k) = \sum_I \Pr(\{B_k^i\})$, where I is the set of all B_k^i consisting of $m - 1$ successes and $k - m$ failures (in any order) then ending in a success. $|I| = \binom{k-1}{m-1}$. As $\{B_k^i\}$ are equilikely,
$\Pr(X = k) = \binom{k-1}{m-1} (1 - p)^{k-m} p^m \qquad \text{for } k = m, m + 1, \ldots$

As above, if X is the number of trials until m successes, then let Y be the number of failures until m successes. Y = X - m, giving:
$\mathbb{P}(Y = k) = \mathbb{P}(X = k + m) = \binom{k+m-1}{m-1} (1 - p)^k p^m, \qquad k = 0, 1, \ldots$

Exercise 24

If X is the number of gold fish caught in n trials, then

$$\begin{aligned}\Pr(X = k) &= \frac{\text{\# outcomes with X=k}}{|\Omega|} \\ &= \frac{\text{number of ways to catch } k \text{ gold fish and } n - k \text{ brown fish}}{\text{number of ways to catch } n \text{ fish}}\end{aligned}$$

We have N gold fish and M brown fish, so $N+M$ total fish, so:

$$|\Omega| = \binom{N+M}{n}$$

$$\#\text{ ways to catch } k \text{ gold fish from } N = \binom{N}{k}$$

$$\#\text{ ways to catch } (n-k) \text{ brown fish from } M = \binom{M}{n-k}$$

$$\Rightarrow \qquad \Pr(X=k) = \frac{\binom{N}{k}\binom{M}{n-k}}{\binom{N+M}{n}}, \qquad k \in \{0,1,\ldots,n\}$$

(note some k in this set may have zero probability, i.e. in cases where $k > N$ or $n-k > M$)

Exercise 26

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} && (0^{th} \text{ term} = 0) \\
&= e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= e^{-\lambda} \lambda \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!} \\
&= e^{-\lambda} \lambda e^{\lambda} && (e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}) \\
&= \lambda
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{k}{(k-1)!} \lambda^k \\
&= e^{-\lambda} \left(\sum_{k=2}^{\infty} \frac{(k-1)\lambda^k}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \right) \\
&= e^{-\lambda} \left(\lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) \\
&= e^{-\lambda} \left(\lambda^2 \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} + \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \right) \\
&= e^{-\lambda}(\lambda^2 e^{\lambda} + \lambda e^{\lambda}) \\
&= \lambda^2 + \lambda
\end{aligned}$$

$$\begin{aligned}
Var(X) &= \mathbb{E}[X]^2 - \mathbb{E}[X^2] \\
&= \lambda - (\lambda^2 + \lambda) && = \lambda^2
\end{aligned}$$

Exercise 27

$$\begin{aligned}
\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} &= \lambda^k\frac{n!}{k!(n-k)!}\frac{1}{n^k}\left(\frac{1-\lambda}{n}\right)^{-k}\left(\frac{1-\lambda}{n}\right)^n \\
&= \frac{\lambda^k}{k!}\left(\frac{1-\lambda}{n}\right)^n\frac{n!}{(n-k)!}\left(\frac{n}{(n-\lambda)n}\right)^k \\
&= \frac{\lambda^k}{k!}\left(\frac{1-\lambda}{n}\right)^n\frac{n(n-1)(n-2)\dots(n-k+1)}{(n-\lambda)^k} \\
\lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} &= \lim_{n\to\infty}\frac{\lambda^k}{k!}\left(\frac{1-\lambda}{n}\right)^n\frac{n}{(n-\lambda)}\frac{n-1}{(n-\lambda)}\cdots\frac{n-k+1}{(n-\lambda)} \\
&= \frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}$$

Exercise 28

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty}k\frac{1}{k(k+1)} \\
&= \sum_{k=0}^{\infty}k\frac{1}{k(k+1)} \qquad \text{(as 0th term = 0)} \\
&= \sum_{k=0}^{\infty}\frac{1}{k+1} = \sum_{k'=1}^{\infty}\frac{1}{k'}
\end{aligned}$$

This is a p-series with p=1, so diverges to infinity. Thus the mean is infinite.

Exercise 29

For any partition L of the sample space $(l_1, l_2, \dots)$,
$p_X(k) = \sum_{l\in L}(\Pr(X = k|Y = l)\Pr(Y = j) = \sum_{l\in L}\Pr(\{X = k\}\cap\{Y = l\}) = \sum_{l\in L}p_{X,Y}(k,l)$ (using sum rule as for a given x, the events $\{X = x, Y = l_1\}, \{X = x, Y = l_2\}, \dots$ are disjoint).

Exercise 30

$$
\begin{aligned}
\mathbb{C}ov(X,Y) &= \mathbb{E}((X-\mathbb{E}X)(Y-\mathbb{E}Y)) \\
&= \mathbb{E}(XY - X\mathbb{E}Y - Y\mathbb{E}X + \mathbb{E}X\mathbb{E}Y) \\
&= \mathbb{E}(XY) - \mathbb{E}(X\mathbb{E}Y) - \mathbb{E}(Y\mathbb{E}X) + \mathbb{E}(\mathbb{E}X\mathbb{E}Y)) \quad \text{(by linearity of expectation)} \\
&= \mathbb{E}(XY) - \mathbb{E}Y\mathbb{E}X - \mathbb{E}X\mathbb{E}Y + \mathbb{E}X\mathbb{E}Y) \\
&= \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y
\end{aligned}
$$

Exercise 31

As X and Y are independent, $p_{XY}(kl) = p_{X,Y}(k,l) = p_X(k)p_Y(l)$

$$
\begin{aligned}
\mathbb{E}(XY) &= \sum_k \sum_l k\ l\ p_{X,Y}(k,l) \\
&= \sum_k \sum_l k\ l\ p_X(k)p_Y(l) \\
&= \sum_k k\ p_X(k) \sum_l l\ p_Y(l) \\
&= \mathbb{E}X\mathbb{E}Y
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{C}ov(X,Y) &= \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y \\
&= \mathbb{E}X\mathbb{E}Y - \mathbb{E}X\mathbb{E}Y \qquad \text{(for independent X,Y)} \\
&= 0
\end{aligned}
$$

Exercise 32

1. Consider the following:

$p_{X,Y}(x,y)$		x = 1	2	3	$p_Y(y)$
y	1	1/6	1/6	2/6	2/3
	2	0	1/6	1/6	1/3
$p_X(x)$		1/6	2/6	3/6	

Here $\mathbb{E}X = \frac{14}{6},\ \mathbb{E}Y = \frac{8}{6},\ \mathbb{E}XY = \frac{19}{6}$
$\mathbb{C}ov(X,Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y = \frac{19}{6} - \frac{14}{6}\frac{8}{6} \neq 0.$

2. Consider the following:

$p_{X,Y}(x,y)$		x			
		1	2	3	$p_Y(y)$
y	1	1/3	0	1/3	2/3
	2	0	1/3	0	1/3
$p_X(x)$		1/3	1/3	1/3	

Here $\mathbb{E}X = 2,\ \mathbb{E}Y = \frac{4}{3},\ \mathbb{E}XY = \frac{8}{3}$, so $\mathbb{C}ov(X,Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y = \frac{8}{3} - 2\frac{4}{3} = 0$. However it can be seen that $p_{X,Y} \neq p_X p_Y$, so X and Y aren't independent.

Exercise 33

By conditioning on Y we are just choosing a subset the sample space, so instead of being interested in events satisfying $(x \in K, y \in L)$, we now only want events with $(x \in K, y = l)$. $p_{X,Y}(k,l) \geq 0$ for all k,l (as the joint pmf was is valid pmf), which means $p_{X,Y}(.,l) \geq 0$ for any fixed l. Thus for $p_Y(l) > 0,\ p_{X|Y=l}(.,l) = \frac{p_{X,Y}(.,l)}{p_Y(l)} \geq 0$.
$\sum_k p_{X,Y}(k,l) = p_Y(l) \Rightarrow \sum_k p_{X|Y=l}(k,l) = p_Y(l)/p_Y(l) = 1$, as required.
We also want the probability of a union of disjoint events to be equal to the sum of their probabilities. This property must be satisfied by the joint pmf (as it is a valid pmf), and as the conditional pmf just gives the probabilities of events in a subset of the joint pmf (rescaled by $p_Y(y = l)$), the conditional pmf will also satisfy this property.

Exercise 34

If X and Y are independent random variables then $p_{X,Y}(k,l) = p_X(k)p_Y(l)$.

$$\begin{aligned} p_{X,Y}(.,l) &= \frac{p_{X,Y}(.,l)}{p_Y(l)} \\ &= \frac{p_X(.)p_Y(l)}{p_Y(l)} \\ &= p_X(.) \end{aligned}$$

Exercise 35

For the first example ($\mathbb{C}ov(X,Y) \neq 0$):

$$p_{X|Y=1} = \begin{cases} 1/4 & x = 1,2 \\ 1/2 & x = 3 \end{cases} \qquad p_{X|Y=2} = \begin{cases} 0 & x = 1, \\ 1/2 & x = 2,3 \end{cases}$$

$$p_{Y|X=1} = \begin{cases} 1, & y = 1 \\ 0, & y \neq 1 \end{cases} \qquad p_{Y|X=2} = 1/2,\ y = 1, 2 \qquad p_{Y|X=3} = \begin{cases} 2/3, & y = 1 \\ 1/3, & y = 2 \end{cases}$$

Exercise 36

1.

$$\begin{aligned} \Pr(X > k) &= 1 - \sum_{n=1}^{k} \Pr(X = n) \\ &= 1 - \sum_{n=1}^{k} (1-p)^{n-1} p \\ &= 1 - p \sum_{n'=0}^{k-1} (1-p)^{n'} \\ &= 1 - p \frac{1-(1-p)^k}{1-(1-p)} \qquad\qquad \text{for} (1-p) \neq 1 \\ &= (1-p)^k \end{aligned}$$

As $X > s + t$ occuring implies $X > s$ has also occured,

$$\begin{aligned} \Pr(X > s + t | X > t) &= \frac{\Pr(X > s + t, X > t)}{\Pr(X > t)} \\ &= \frac{\Pr(X > s + t)}{\Pr(X > t)} \\ &= \frac{(1-p)^{s+t}}{(1-p)^t} \\ &= (1-p)^s \\ &= \Pr(X > s) \end{aligned}$$

2. As the geometric distribution is produced by a sequence of Bernoulli trials which are iid, it makes sense that the probability of t more failures occurring given s failures have already occurred should be the same as the probability of the first t trials being failures - i.e. what's already happened doesn't influence future Bernoulli trials.

3. The discrete uniform distribution is not memory-less. e.g. For discrete uniform

random variable with $p_X(x) = 1/L, x \in \{a, a+1, \ldots, a+L\}$:

$$\begin{aligned}
\Pr(X > s+t | X > t) &= \frac{\Pr(X > s+t)}{\Pr(X > t)} \\
&= \frac{(t-L)L}{(s+t-L)L} \\
&= \frac{t-L}{s+t-L} \neq \frac{1}{sL} = \Pr(X > s)
\end{aligned}$$

Exercise 38

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[h(X)|Y]] &= \sum_y \sum_x h(x)\ p_{X|Y=y}(x,y)\ p_Y(y) \\
&= \sum_y \sum_x h(x) \frac{p_{X,Y}(x,y)}{p_Y(y)} p_Y(y) \\
&= \sum_y \sum_x h(x) p_{X,Y}(x,y) \\
&= \sum_x h(x) p_X(x) \\
&= \mathbb{E}[h(X)]
\end{aligned}$$

Exercise 40

1. $\int_{-\infty}^{\infty} f_X(x)dx = \Pr(\Omega) = 1$
2. $F_x(k) = \Pr(X \leq k) = \int_{-\infty}^{k} f_X(x)dx$. The CDF of a continuous random variable is continuous everywhere (set $f_X(x) = 0$ for any x outside support).
3. Set $K = \frac{1}{\int \tilde{f}(x)dx}$. This will then satisfy the three axioms of probability.

Exercise 41

$\mathbb{E}X = \int x \frac{1}{b-a} dx = \frac{1}{b-a}\frac{1}{2}(b^2 - a^2) = \frac{1}{b-a}\frac{1}{2}(a+b)(b-a) = \frac{a+b}{2}$. This is half the length of the interval, which is what we'd expect the mean to be if uniformly distributed over a length.

$$
\begin{aligned}
(\mathbb{E}X)^2 &= \frac{(a+b)^2}{4} = \frac{a^2+2ab+b^2}{4} \\
\mathbb{E}(X^2) &= \frac{1}{b-a}\int x^2 dx = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3} \\
Var(X) &= \mathbb{E}(X^2) - (\mathbb{E})^2 \\
&= \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4} \\
&= \frac{a^2+b^2-2ab}{12} = \frac{(b-a)^2}{12}
\end{aligned}
$$

Exercise 42

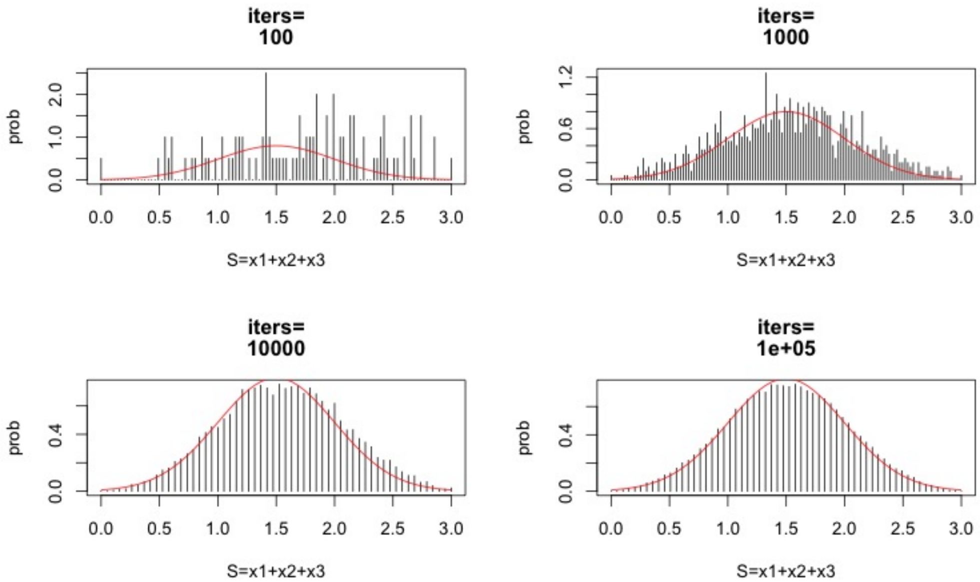
$$
f_X(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \le x \le b \\ 0 & x > b \end{cases} \Rightarrow F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}
$$

Exercise 44

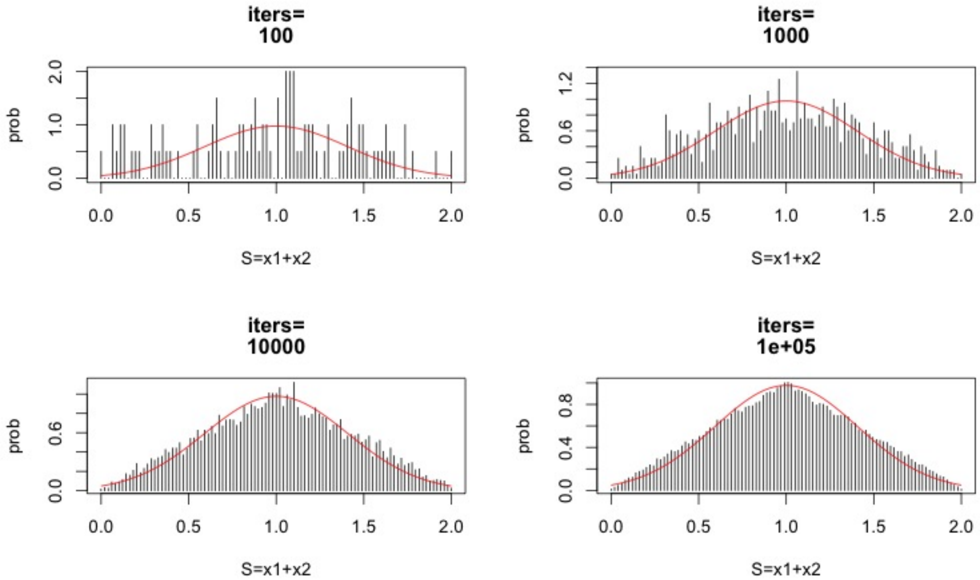
```
Console ~/
> for (i in seq(-10,10,by=5)){
+   for (j in 1:5){
+     gauss <- function(x,v=j,m=i) {
+     (1/(v*sqrt(2*pi)))*exp((-(x-m)^2)*(1/(2*v*v))) }
+     print(integrate(gauss,-Inf,Inf))
+ }}
1 with absolute error < 6e-07
1 with absolute error < 4.9e-06
1 with absolute error < 4.6e-06
1 with absolute error < 2.8e-07
1 with absolute error < 5.5e-06
1 with absolute error < 2e-06
1 with absolute error < 1.6e-06
1 with absolute error < 4.5e-05
1 with absolute error < 1.7e-09
1 with absolute error < 8.9e-09
1 with absolute error < 9.4e-05
1 with absolute error < 1.5e-06
1 with absolute error < 9.4e-05
1 with absolute error < 6.5e-07
1 with absolute error < 2.6e-07
1 with absolute error < 2e-06
```

Exercise 50

Plotting a histogram (of densities not counts) of $S_n = X1 + X2 + X3$, against a normal distribution (with mean 3/2, variance 3/12 (using $var(S) = 3 * var(X_i)$ for independent X_i, mean(S)=3*mean(X_i):



The Central Limit Theorem says that for sufficiently large n, the distribution of S_n, where S_n is the sum of n iid random variables, is approximately normal. This can be see in the plots above - the normal distribution is a good approximation for distribution of the sum of three uniformly distributed random variables. Similarly for S = X1+X2, plotting against a normal distribution (with mean 1, variance 1/6 (using var(S)=2*var(Xi) for independent Xi, mean(S)=2*mean(Xi). The approximation here is less good.



For n=3:

```
par(mfrow = c(2,2))
for (iters in c(100,1000,10000,100000)) {
  S <- array(dim=iters)
  for (i in 1:iters) {
    S[i] <- sum(runif(3)) }
  histn <- hist(S,breaks=100,plot=FALSE)
  rng <- seq(0,3,by=3/(length(histn$counts)-1))
  plot(rng,histn$density,type="h",ylab="prob",xlab="S=x1+x2+x3",main=c("iters=",iters))
  points(rng,dnorm(rng,mean=1.5,sd=1/sqrt(4)),type="l",col="red")
  #plotting against pdf of normal distn with mean=1.5, var=sqrt (3 * 1/12)
}
```

Exercise 51

```
B <- array(0, dim=10000)
for (i in 1:10000) {
    if (runif(1)<=0.25) #set to success if unif random number <= .25
    {B[i] <- 1    }
}
print(c("sample mean=",mean(B),"sample var=",var(B),"theoretical mean = p = 0.25","theoretical var = p*(1-p)",0.25*.75))
```

These sample values are close to the theoretical mean and variance.

```
[1] "sample mean="               "0.2491"                    "sample var="               "0.187067896789679"
[5] "theoretical mean = p = 0.25" "theoretical var = p*(1-p)" "0.1875"
```