All of Statistics

My proposed solutions to the book:

All of Statistics - A Concise Course in Statistical Inference.

I really needed a proper refresher on statistics.

Let me know if you find any mistakes! I am sure there are plenty. :(

Contents

1 Chapter 3 - Expectation

Exercises

3.1

Define X as the wealth after n games. The probability of winning and losing is the same for each outcome so $p = 1/2$.

$$\mathbb{E}[X] = \frac{1}{2} \cdot 2c + \frac{1}{2} \cdot \left(\frac{1}{2}\right) c = c + \frac{c}{4} = \frac{5}{4}c$$

We expect to have $5/4 \cdot c$ after n games. We can also verify this result with a simulation in **R**.

```
> games = sample(c(2, 0.5), size = 1000000, replace = TRUE)
> mean(games)
[1] 1.250973
> 5/4
[1] 1.25
```

R

3.2

Claim. $\mathrm{Var}(X) = 0$ if and only if $\mathbb{P}(X = c) = 1$ for some constant c.

Proof.
$\Rightarrow$) Set $c = \mu$ and assume $\mathrm{Var}(X) = 0$, which means that

$$\mathbb{E}[(X - \mu)^2] = 0 \quad \Longrightarrow \quad \int (x - \mu)^2 dF(x) = 0.$$

This can only be 0 when $x = \mu = c$ for the entire domain of X. Hence $\mathbb{P}(X = c) = 1$.

$\Leftarrow$) Assume $\mathbb{P}(X = c) = 1$. When calculating the expectation:

$$\mu = \mathbb{E}[X] = \int c dF(x) = c$$

When calculating the variance:

$$\mathrm{Var}(X) = \int (x - \mu)^2 dF(x) = \int (c - c)^2 dF(x) = 0,$$

since $x = c$ for all x in the domain of X. $\square$

3.3

Let $X_1, \ldots, X_n \sim U(0,1)$ and define $Y = \max(X_1, \ldots, X_n)$. We will calculate $\mathbb{E}[Y]$. It is not stated in the exercise, but we will assume that the X_i are independent. Finding the CDF for Y.

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leqslant y) \\
&= \mathbb{P}(\max(X_1, \ldots, X_n) \leqslant y) \\
&= \mathbb{P}(X_1 \leqslant y) \cap \ldots \cap \mathbb{P}(X_n \leqslant y) \\
&= \mathbb{P}(X_1 \leqslant y)\mathbb{P}(X_2 \leqslant y) \cdots \mathbb{P}(X_n \leqslant y) && \text{(Independence)} \\
&= (F_X(y))^n
\end{aligned}$$

Differentiating to get the PDF for Y.

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}(F_X(y))^n = n(F_X(y))^{n-1} f_X(y)$$

Since X_i are uniformly distributed, we know that $F_X(y) = y$ and $f_X(y) = 1$, so:

$$f_Y(y) = ny^{n-1}$$

Now we can calculate the expectation of Y.

$$\mathbb{E}[Y] = \int_0^1 y \cdot ny^{n-1} dy = n \int_0^1 y^n dy = n\Big[\frac{y^{n+1}}{n+1}\Big]_0^1 = \frac{n}{n+1}$$

Confirming this result with a numeric simulation in **R**.

```
> # 3.3
> N = 1000000
> U1 = runif(N)
> U2 = runif(N)
> U3 = runif(N)
> U4 = runif(N)
> U5 = runif(N)
> U6 = runif(N)
> U7 = runif(N)
> U8 = runif(N)
> U9 = runif(N)
> U10 = runif(N)
> Y = pmax(U1, U2, U3, U4, U5,
+          U6, U7, U8, U9, U10)
> mean(Y)
[1] 0.9091151
> # Theoretical Result
> 10/11
[1] 0.9090909
```

R

As we can see, the theoretical result is very close to the simulated result for $n = 10$.

3.4 - Random Walk
A particle starts in the origin and jumps left, a step of -1, with probability p and jumps right, a step of 1, with probability $1-p$. The expected location will be:

$$\mathbb{E}[X] = (-1)p + (1)(1-p) = -p + 1 - p = 1 - 2p$$

To calculate the variance, we start by finding the second moment:

$$\mathbb{E}[X^2] = (-1)^2 p + (1)^2(1-p) = p + 1 - p = 1$$

So the variance is:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1 - (1-2p)^2 = 1 - (1 - 4p + 4p^2) = 4p - 4p^2$$

3.5
Tossing a fair coin until we get H. Finding the expected number of tosses. The reasoning is as follows. We get H on the first toss with probability 1/2, first H on the second toss with probability $1/2^2 = 1/4$ and so on. The pattern becomes as follows, for the first 7 cases:

Tosses	**Outcome**	**Probability**
1	$\{H\}$	1/2
2	$\{TH\}$	1/4
3	$\{TTH\}$	1/8
4	$\{TTTH\}$	1/16
5	$\{TTTTH\}$	1/32
6	$\{TTTTTH\}$	1/64
7	$\{TTTTTTH\}$	1/128

Define T to be the number of tosses to get H.

$$\begin{aligned}\mathbb{E}[T] &= (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{4}\right) + \ldots + (k)\left(\frac{1}{2^k}\right) + \ldots \\ &= \sum_{k=1}^{\infty} \frac{k}{2^k} \\ &= 2\end{aligned}$$

Not delving in to the mathematics of the infinite sum, but it can be shown that this sum becomes 2 which will be the expected number of tosses to get a H. Here is a numeric approximation in **R**.

```
> sumApprox = 0
> for (k in 1:1000) {
+   sumApprox = sumApprox + k/2^k
+ }
> sumApprox
[1] 2
```

R

3.6 Theorem - The Rule of the Lazy Statistician
Proving the following result for the discrete case. Let $Y = r(X)$, then

$$\mathbb{E}[Y] = \mathbb{E}[r(X)] = \sum_x r(x) f_X(x)$$

Proof. Set $\mathcal{X}, \mathcal{Y}$ to be the set of all values for X and Y. We have the functions $r : \mathcal{X} \to \mathcal{Y}$ and $r^{-1} : \mathcal{Y} \to \mathcal{X}$. We will define $s := r^{-1}$ for convenience. By definition of the expectation:

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{y \in \mathcal{Y}} y \cdot f_Y(y) = \sum_{y \in \mathcal{Y}} y \cdot \mathbb{P}(Y = y) \\
&= \sum_{y \in \mathcal{Y}} y \sum_{x \in s(y)} \mathbb{P}(X = x) \\
&= \sum_{y \in \mathcal{Y}} \sum_{x \in s(y)} y \mathbb{P}(X = x) \\
&= \sum_{x \in \mathcal{X}} r(x) \mathbb{P}(X = x) \\
&= \sum_{x \in \mathcal{X}} r(x) f_X(x)
\end{aligned}$$

□

A simplified case to make the proof easier a bit easier to understand. We define the variables X and $Y = r(X) = X^2$, and use the following distributions.

x	$\mathbb{P}(X = x)$
-1	$1/4$
0	$1/4$
1	$1/2$

$Y = r(X) = X^2$

y	$\mathbb{P}(Y = y)$
0	$1/4$
1	$3/4$

Then the calculations above become:

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{y \in \mathcal{Y}} y \cdot f_Y(y) = (0)(1/4) + (1)(3/4) \\
&= \sum_{y \in \mathcal{Y}} y \sum_{x \in s(y)} \mathbb{P}(X = x) = (0)(1/4) + (1)\big[1/4 + 1/2\big] \\
&= \sum_{y \in \mathcal{Y}} \sum_{x \in s(y)} y \mathbb{P}(X = x) = (0)(1/4) + (1)(1/4) + (1)(1/2) \\
&= \sum_{x \in \mathcal{X}} r(x) \mathbb{P}(X = x) = (0)^2(1/4) + (-1)^2(1/4) + (1)^2(1/2) \\
&= \sum_{x \in \mathcal{X}} r(x) f_X(x)
\end{aligned}$$

In both cases, the expectation becomes $3/4$.

3.7
$X \sim F$ is continuous and we suppose that $\mathbb{P}(X > 0) = 1$ and that $\mathbb{E}[X]$ exists. Show that $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)dx$.

PROOF. By definition of the expectation, and since the domain of X must be $(0, \infty)$:

$$\mathbb{E}[X] = \int_0^\infty x \cdot f_X(x)dx$$

Using integration by parts, and setting $u = x$ and $v' = f_X$, which gives $u' = 1$ and $v = F_X$:

$$\int uv' = uv - \int u'v,$$

which gives us:

$$\begin{aligned}
\mathbb{E}[X] = \int_0^\infty x \cdot f_X(x)dx &= xF_X(x)\Big|_0^\infty - \int_0^\infty F_X(x)dx \\
&= xF_X(x)\Big|_0^\infty - \int_0^\infty \mathbb{P}(X \leqslant x)dx \\
&= xF_X(x)\Big|_0^\infty - \int_0^\infty 1 - \mathbb{P}(X > x)dx \\
&= xF_X(x)\Big|_0^\infty - x\Big|_0^\infty + \int_0^\infty \mathbb{P}(X > x)dx \\
&= -\big(\lim_{x\to\infty} x(1 - F_X(x)\big) + \int_0^\infty \mathbb{P}(X > x)dx \\
&= \int_0^\infty \mathbb{P}(X > x)dx
\end{aligned}$$

which proves the result. In the second to last step, we applied the hint regarding the limit. □

3.8 - Theorem 3.17
Let $X_1, \ldots, X_n$ be IID and let $\mu = \mathbb{E}[X_i]$, $\sigma^2 = \mathrm{Var}(X_i)$. Then:

$$\mathbb{E}[\overline{X}] = \mu, \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}, \quad \mathbb{E}[S^2] = \sigma^2.$$

PROOF. Start with the expectation of the sample mean:

$$\mathbb{E}[\overline{X}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^n \mu = \frac{n \cdot \mu}{n} = \mu.$$

Next, the variance.

$$\mathrm{Var}[\overline{X}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}[X_i] = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Finally, the expectation of the sample variance. This calculation is a lot more involved.

$$\mathbb{E}[S^2] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right] \quad \Longrightarrow \quad (n-1)\mathbb{E}[S^2] = \mathbb{E}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right]$$

We will use that $\overline{X} = (1/n)\Sigma X_i$ means that $\Sigma X_i = n\overline{X}$. Running through the calculations:

$$\begin{aligned}
(n-1)\mathbb{E}[S^2] &= \mathbb{E}\left[\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i\overline{X} + \sum_{i=1}^{n}\overline{X}^2)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] - \mathbb{E}\left[2\sum_{i=1}^{n} X_i\overline{X}\right] + \mathbb{E}\left[\sum_{i=1}^{n}\overline{X}^2)\right] \\
&= \sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[2\overline{X}\sum_{i=1}^{n} X_i\right] + \mathbb{E}\left[\overline{X}^2\sum_{i=1}^{n} 1)\right] && (\overline{X} \text{ indp. of } i) \\
&= n\mathbb{E}\left[X_i^2\right] - 2n\mathbb{E}\left[\overline{X}^2\right] + n\mathbb{E}\left[\overline{X}^2\right] && (\Sigma X_i = n\overline{X}) \\
&= n\mathbb{E}\left[X_i^2\right] - n\mathbb{E}\left[\overline{X}^2\right]
\end{aligned}$$

By dividing both sides of the equality by n, we have shown:

$$\frac{n-1}{n}\mathbb{E}[S^2] = \mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[\overline{X}^2\right] \tag{3.8.1}$$

We have the second moment for X. From the definition of the variance:

$$\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - \mu^2 \quad \Longrightarrow \quad \mathbb{E}[X_i^2] = \mathrm{Var}(X_i) + \mu^2 = \sigma^2 + \mu^2$$

From the previous results:

$$\mathrm{Var}(\overline{X}) = \mathbb{E}[\overline{X}^2] - \mathbb{E}[\overline{X}]^2 \quad \Longrightarrow \quad \mathbb{E}[\overline{X}^2] = \mathrm{Var}(\overline{X}) + \mathbb{E}[\overline{X}]^2 = \frac{\sigma^2}{n} + \mu^2$$

Replacing each of these into (3.8.1) and isolating $\mathbb{E}[S^2]$.

$$\begin{aligned}
\mathbb{E}[S^2] &= \frac{n}{n-1}\left(\sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2\right)\right) \\
&= \frac{n}{n-1}\left(\sigma^2 - \frac{\sigma^2}{n}\right) \\
&= \frac{n}{n-1}\left(\frac{n-1}{n}\right)\sigma^2 \\
&= \sigma^2
\end{aligned}$$

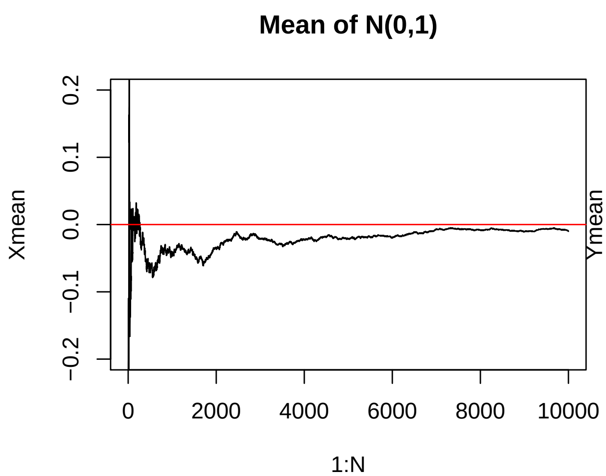
and we have finally shown the final result. $\square$

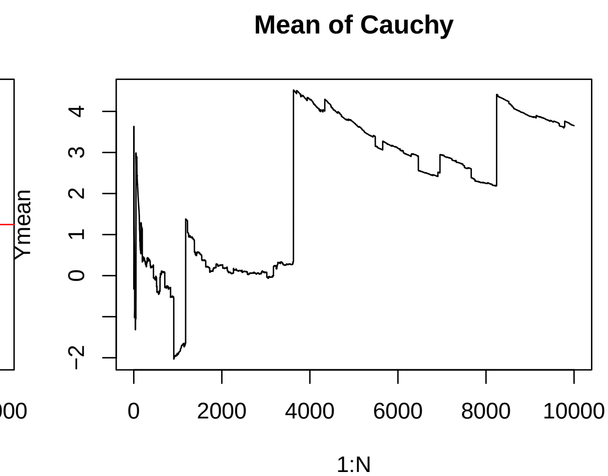
3.9

Computer experiment; studying the effects of the mean over 10.000 simulations from the standard normal distribution and the Cauchy distribution.

As we can see the simulation of the normal distribution quickly stabilizes to a value of around 0 which is the expected value. The Cauchy distribution is a very heavy-tailed distribution and the expectation does not exist. Therefore, there is no mean it can stablize to and it ends up behaving erraticaly even after 10.000 simulations. The jumps that can be seen are extreme outliers that are added to the data, which are relatively common in the Cauchy distribution.

Results:






```
# Setting the number of simulations
N = 10000

# Simulating N standard normal values
X = rnorm(N)
Xmean = cumsum(X)/1:N
plot(1:N, Xmean, type="l", ylim = c(-0.2, 0.2),
main = "Mean of N(0,1)")
abline(h = 0, col="red")

# Simulating N Cauchy values with loc = 0, scale = 1
Y = rcauchy(N)
Ymean = cumsum(Y)/1:N
plot(1:N, Ymean, type="l",
main = "Mean of Cauchy")
```

R

3.10
Let $X \sim N(0,1)$ and define $Y = e^X$. Find $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$.

One possible way to solve this is defining $r(X) = e^X$ and evaluating:

$$\mathbb{E}[Y] = \mathbb{E}[r(X)] = \int_{\mathbb{R}} e^x f_X(x)dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(x - \frac{1}{2}x^2\right) dx.$$

Completing the square in the exponential term:

$$x - \frac{1}{2}x^2 = -\frac{1}{2}\left(x^2 - 2x\right) = -\frac{1}{2}\left(x^2 - 2x + 1 - 1\right) = -\frac{1}{2}\left(x-1\right)^2 + \frac{1}{2}$$

We can rewrite the integral:

$$\mathbb{E}[Y] = e^{1/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}(x-1)^2\right) dx$$

We will integrate this by substitution:

$$u = x - 1 \quad \Longrightarrow \quad du/dx = 1 \quad \Longrightarrow \quad du = dx$$

$$\mathbb{E}[Y] = e^{1/2} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}u^2\right) du}_{=1} = e^{1/2} \approx 1.6487$$

Confirming with numerical simulation.

```
> N = 10000000; X = rnorm(N); Y = exp(X)
> mean(Y)
[1] 1.648564
> exp(0.5)
[1] 1.648721
> var(Y)
[1] 4.676382
> exp(2) - exp(1)
[1] 4.670774
```

We can calculate the variance by first finding the second moment: $\mathbb{E}[Y^2]$.

$$\mathbb{E}[Y^2] = \mathbb{E}[r(X)^2] = \int_{\mathbb{R}} e^{2x} f_X(x)dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(2x - \frac{1}{2}x^2\right) dx.$$

Just as above, we complete the square and get: $-(1/2)(x-2)^2 + 2$ and move e^2 outside the integral. Then we use integration by substitution: $u = x - 2$ which gives $du = dx$. We end up with a similar integral:

$$\mathbb{E}[Y^2] = e^2 \underbrace{\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}u^2\right) du}_{=1} = e^2$$

Now we can calculate the variance of Y.

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = e^2 - (e^{1/2})^2 = e^2 - e^1,$$

and as we can see in the simulation above, this corresponds to the simulated variance.

3.11

Simulating stocks. For each day Y_i can be either -1 or 1, each with probability 1/2. We have $X_n = \sum_{i=1}^n Y_i$ which is the cumulative value of the stock.

(a) Calculating the expectation and variance of X_n. First we calculate it for Y_i.

$$\mathbb{E}[Y_i] = (-1)(1/2) + (1)(1/2) = -1/2 + 1/2 = 0$$
$$\mathbb{E}[Y_i^2] = (-1)^2(1/2) + (1)^2(1/2) = 1/2 + 1/2 = 1$$
$$\text{Var}(Y_i) = \mathbb{E}[Y_i^2] - \mathbb{E}[Y_i]^2 = 1 - 0 = 1$$

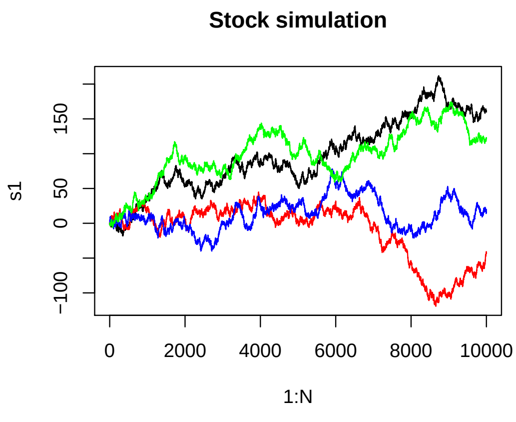
Now for X_n.

$$\mathbb{E}[X_n] = \mathbb{E}\left[\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n \mathbb{E}[Y_i] = 0$$
$$\text{Var}(X_n) = \text{Var}\left[\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n \text{Var}[Y_i] = \sum_{i=1}^n 1 = n$$

(b) Simulating four stocks.

```
simStock <- function(N) {
    dailyMovements = sample(c(-1, 1), size=N, replace = TRUE)
    totalMovements = cumsum(dailyMovements)
    return(totalMovements)
}
N = 10000
s1 = simStock(N);s2 = simStock(N)
s3 = simStock(N);s4 = simStock(N)
plot(1:N, s1, type="l",
    ylim=c(-119, 212), # Adjust according to simulation results
    main = "Stock simulation")
lines(1:N, s2, type="l", col="red")
lines(1:N, s3, type="l", col="blue")
lines(1:N, s4, type="l", col="green")
```

R



The simulations have mean 0, so they should vary around the x-axis. Since the variance is n, it is dependent on the time. The longer the simulations last, the more they will vary which is what we see.

The standard deviation in this example will be the square root of n, which is about 100, and that is also what we see (and it is expected that they will not be exactly on the standard deviation).

3.12
Deriving the general expression for the expectation and variance for a whole bunch of probability distributions.

- **Bernoulli** with parameter p. Possible values: $\mathcal{X} = \{0, 1\}$.

$$f(x) = p^x(1-p)^{x-1}$$

Expectation.

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot f(x) = (0) + (1)p^1 = p$$

Second moment.

$$\mathbb{E}[X^2] = \sum_{x^2 \in \mathcal{X}} x^2 \cdot f(x) = (0) + (1)^2 p^1 = p$$

Variance.

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1-p)$$

- **Poisson** with parameter λ. Possible values: $\mathcal{X} = \mathbb{N} \cup \{0\}$.

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

Expectation. First term disappears. Cancel x against the factorial. Switch to $k = x - 1$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in \mathcal{X}} x f(x) = \sum_{x=0}^{\infty} x \cdot e^{-\lambda}\frac{\lambda^x}{x!} \\
&= 0 + \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}$$

Second moment. We will split the sum into two Poisson sums.

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{x^2 \in \mathcal{X}} x^2 \cdot f(x) = \sum_{x=0}^{\infty} x^2 \cdot e^{-\lambda}\frac{\lambda^x}{x!} \\
&= 0 + \lambda e^{-\lambda} \sum_{x=1}^{\infty} x \cdot \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} (x - 1 + 1) \cdot \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \left[(x-1)\frac{\lambda^{x-1}}{(x-1)!} + \frac{\lambda^{x-1}}{(x-1)!}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \lambda e^{-\lambda} \left(\sum_{x=2}^{\infty} (x-1) \frac{\lambda^{x-1}}{(x-1)!} + \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right) \\
&= \lambda e^{-\lambda} \left(\lambda \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right) \\
&= \lambda e^{-\lambda} \left(\lambda \sum_{j=0}^{\infty} \frac{\lambda^{j}}{j!} + \sum_{k=0}^{\infty} \frac{\lambda^{k}}{k!} \right) \\
&= \lambda e^{-\lambda} \left(\lambda e^{\lambda} + e^{\lambda} \right) \\
&= \lambda^2 + \lambda
\end{aligned}
$$

The remaining calculation for the variance is easy.

$$
\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda
$$

- **Uniform** on (a, b). Possible values: $\mathcal{X} = (a, b)$.

$$
f(x) = \frac{1}{b-a}
$$

Expectation.

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{\mathcal{X}} x \cdot f(x) = \frac{1}{b-a} \int_a^b x dx \\
&= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \left(\frac{b^2 - a^2}{2} \right) \\
&= \frac{1}{b-a} \left(\frac{(b+a)(b-a)}{2} \right) \\
&= \frac{b+a}{2}
\end{aligned}
$$

Second moment.

$$
\begin{aligned}
\mathbb{E}[X^2] &= \int_{\mathcal{X}} x^2 \cdot f(x) = \frac{1}{b-a} \int_a^b x^2 dx \\
&= \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{1}{b-a} \left(\frac{b^3 - a^3}{3} \right) \\
&= \frac{1}{b-a} \left(\frac{(b-a)(b^2 + ab + a^2)}{3} \right) \\
&= \frac{b^2 + ab + a^2}{3}
\end{aligned}
$$

Variance.

$$
\begin{aligned}
\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 &= \frac{b^2 + ab + a^2}{3} - \frac{(b+a)^2}{4} \\
&= \frac{4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2}{12} = \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}
\end{aligned}
$$

• **Exponential** with parameter β. Possible values: $\mathcal{X} = \{x \in \mathbb{R} : x \geqslant 0\}$.

$$f(x) = \frac{1}{\beta}\exp(-x/\beta)$$

Expectation.

$$\mathbb{E}[X] = \int_{\mathcal{X}} x \cdot f(x) = \int_0^\infty x \cdot \frac{1}{\beta} e^{-x/\beta} dx$$

We will integrate this with the substitution $u = x/\beta$, so $du/dx = 1/\beta$ which gives us $dx = \beta du$. Note that the antiderivative of ue^{-u} is $-ue^{-u} - e^{-u}$.

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x \cdot \frac{1}{\beta} e^{-x/\beta} dx = \beta \int_0^\infty ue^{-u} du \\
&= \beta\Big[-ue^{-u} - e^{-u}\Big]_0^\infty \\
&= \beta\Big(\lim_{u\to\infty}(-ue^{-u} - e^{-u}) - (-1)\Big) \\
&= \beta
\end{aligned}$$

Skipping the evaluation of the limit. The first term can be shown to converge to 0 by using L'Hôpital's rule. The second term in the limit obviously becomes 0.
Second moment. Will derive the expression for the expectation in the calculations.

$$\mathbb{E}[X^2] = \int_{\mathcal{X}} x^2 \cdot f(x) = \int_0^\infty x^2 \cdot \frac{1}{\beta} e^{-x/\beta} dx$$

Using integration by parts: $u = x^2$ so $u' = 2x$. Setting $v' = (1/\beta)e^{-x/\beta}$ which yields $v = -e^{-x/\beta}$. From the integration-by-parts formula:

$$\int uv' = uv - \int u'v$$

$$\begin{aligned}
\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot \frac{1}{\beta} e^{-x/\beta} dx &= \Big[-x^2 e^{-x/\beta}\Big]_0^\infty - \int_0^\infty -2xe^{-x/\beta} dx \\
&= 0 + \int_0^\infty 2xe^{-x/\beta} dx \\
&= 2\beta \int_0^\infty x \cdot \frac{1}{\beta} e^{-x/\beta} dx \\
&= 2\beta\mathbb{E}[X] \\
&= 2\beta^2
\end{aligned}$$

The limit can be shown to converge to 0 by applying L'Hôpital's rule twice. We use a clever trick and substitute in the expectation and get the result. Finally, we calculate the variance.

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 2\beta^2 - \beta^2 = \beta^2$$

• **Gamma** with parameters α, β. Possible values: $\mathcal{X} = \{x \in \mathbb{R} : x > 0\}$.

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta)$$

We will use the definition of the Gamma function, and the 'Gamma difference' chaining property:

$$\Gamma(\alpha+1) = \int_0^\infty t^\alpha e^{-t} dt, \quad \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

In step 3, we make the substitution $t = x/\beta$ which makes $x = \beta t$. $dt/dx = 1/\beta$, so $dx = \beta dt$.

$$\begin{aligned}
\mathbb{E}[X] = \int_{\mathcal{X}} x \cdot f(x) &= \int_0^\infty x \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta) dx \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha} \exp(-x/\beta) dx \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} (\beta t)^{\alpha} \exp(-t)\beta dt \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} \beta^\alpha t^{\alpha} \exp(-t)\beta dt \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} \beta^{\alpha+1} t^{\alpha} \exp(-t) dt \\
&= \frac{\beta^{\alpha+1}}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty t^{\alpha} \exp(-t) dt \\
&= \frac{\beta}{\Gamma(\alpha)} \int_0^\infty t^{\alpha} \exp(-t) dt \\
&= \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} \\
&= \frac{\beta\alpha\Gamma(\alpha)}{\Gamma(\alpha)} \\
&= \alpha\beta
\end{aligned}$$

For the Gamma distribution, we will not calculate the second moment. Instead we will go straight to the variance calculation and calculate it directly. We will use the same substitution and use the definition of the Gamma function again.

$$\begin{aligned}
\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 &= \int_0^\infty x^2 \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta) dx - \alpha^2\beta^2 \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha+1} \exp(-x/\beta) dx - \alpha^2\beta^2 \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty (\beta t)^{\alpha+1} \exp(-t)\beta dt - \alpha^2\beta^2 \\
&= \frac{\beta^{\alpha+2}}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty t^{\alpha+1} \exp(-t) dt - \alpha^2\beta^2
\end{aligned}$$

Using the definition of the Gamma function.

$$
\begin{aligned}
&= \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)} - \alpha^2\beta^2 \\
&= \frac{\beta^2\Gamma(\alpha+2) - \alpha^2\beta^2\Gamma(\alpha)}{\Gamma(\alpha)} \\
&= \frac{\beta^2\big(\Gamma(\alpha+2) - \alpha^2\Gamma(\alpha)\big)}{\Gamma(\alpha)} \\
&= \frac{\beta^2\big((\alpha+1)\Gamma(\alpha+1) - \alpha^2\Gamma(\alpha)\big)}{\Gamma(\alpha)} \\
&= \frac{\beta^2\big((\alpha+1)\alpha\Gamma(\alpha) - \alpha^2\Gamma(\alpha)\big)}{\Gamma(\alpha)} \\
&= \frac{\beta^2\Gamma(\alpha)\big((\alpha+1)\alpha - \alpha^2\big)}{\Gamma(\alpha)} \\
&= \beta^2\big((\alpha+1)\alpha - \alpha^2\big) \\
&= \beta^2\big(\alpha^2\alpha - \alpha^2\big) \\
&= \alpha\beta^2
\end{aligned}
$$

And the calculation for the Gamma distribution has been completed.

- **Beta** with parameter α, β. Possible values: $\mathcal{X} = (0, 1)$.

$$
f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}
$$

In the following, we will use the following definition:

$$
\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt
$$

Expectation.

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{\mathcal{X}} x \cdot f(x) = \int_0^1 x \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1}dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} \\
&= \frac{\alpha}{\alpha+\beta} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}
$$

Variance. Just like for the Gamma distribution, we calculate the second moment directly in the variance calculation. Will use a lot of the same tricks as when calculating the expectation.

$$
\begin{aligned}
\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 &= \int_0^1 x^2 \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+1}(1-x)^{\beta-1} dx - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+1}(1-x)^{\beta-1} dx - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \qquad (3.12.1) \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\
&= \frac{\alpha^2+\alpha}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{(\alpha^2+\alpha)(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= \frac{\alpha^3+\alpha^2\beta+\alpha^2+\alpha\beta-\alpha^3-\alpha^2\beta-\alpha^2)}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= \frac{\cancel{\alpha^3}+\cancel{\alpha^2\beta}+\cancel{\alpha^2}+\alpha\beta\cancel{-\alpha^3}\cancel{-\alpha^2\beta}\cancel{-\alpha^2})}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}
$$

In step (3.12.1) we used the difference/chaining property of the Gamma function.

$$\Gamma(\alpha+2) = (\alpha+1)\Gamma(\alpha+1) = (\alpha+1)\alpha\Gamma(\alpha)$$

and,

$$\Gamma(\alpha+\beta+2) = (\alpha+\beta+1)\Gamma(\alpha+\beta+1) = (\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)$$

This concludes the last calculation. I found this exercise to be a little tedious, to be honest...

3.13

A variable X is generated by assuming a value in $U(0,1)$ if a coin toss is 0 (H), and X is in $U(3,4)$ if the coin toss is 1 (T). The coin is fair, so each of these has a probability of 1/2.

(a) Calculating the mean of X. This will be a conditional expectation, and we define Y to be the coin toss. The marginal distributions might depend on the outcome, but when we calculate them:

$$f_{X|Y}(x|Y=0) = \frac{1}{1-0} = 1, \quad f_{X|Y}(x|Y=1) = \frac{1}{4-3} = 1,$$

we find that it will be $f_{X|Y}(x|y) = 1$ either way.

Now we can calculate the expected value for specific outcomes.

$$\mathbb{E}[X|Y=0] = \int_{\mathcal{X}} x \cdot f_{X|Y}(x|y) = \int_0^1 x dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1-0}{2} = \frac{1}{2}$$

$$\mathbb{E}[X|Y=1] = \int_{\mathcal{X}} x \cdot f_{X|Y}(x|y) = \int_3^4 x dx = \left[\frac{x^2}{2}\right]_3^4 = \frac{16-9}{2} = \frac{7}{2}$$

Using that $p_Y(y)$ is known, we can calculate the expected value of X:

$$\begin{aligned}
\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] &= \sum_{y\in\{0,1\}} \mathbb{E}[X|Y=y]\mathbb{P}(Y=y) \\
&= \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) + \left(\frac{7}{2}\right)\left(\frac{1}{2}\right) \\
&= \frac{1}{4} + \frac{7}{4} = \frac{8}{4} = 2
\end{aligned}$$

(b) Calculating the second moment.

$$\mathbb{E}[X^2|Y=0] = \int_{\mathcal{X}} x^2 \cdot f_{X|Y}(x|y) = \int_0^1 x^2 dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1-0}{3} = \frac{1}{3}$$

$$\mathbb{E}[X^2|Y=1] = \int_{\mathcal{X}} x^2 \cdot f_{X|Y}(x|y) = \int_3^4 x^2 dx = \left[\frac{x^3}{3}\right]_3^4 = \frac{4^3-3^3}{3} = \frac{37}{3}$$

$$\begin{aligned}
\mathbb{E}[X^2] = \mathbb{E}[\mathbb{E}[X^2|Y]] &= \sum_{y\in\{0,1\}} \mathbb{E}[X^2|Y=y]\mathbb{P}(Y=y) \\
&= \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{37}{3}\right)\left(\frac{1}{2}\right) \\
&= \frac{1}{6} + \frac{37}{6} = \frac{38}{6} \\
&= \frac{19}{3}
\end{aligned}$$

Calculating the variance:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{19}{3} - 4 = \frac{7}{3} \approx 2.333$$

And finally the standard deviation:

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{7}{3}} \approx 1.5275$$

To verify these results, we did a numeric simulation.

```
N = 10000; Y = sample(c(0, 1), size=N, replace = TRUE); X = rep(0, N)
for (i in 1:N) {
     if (Y[i] == 1) X[i] = runif(1)
     else X[i] = runif(1, min=3, max=4)
}
> mean(X)
[1] 2.01424
> var(X)
[1] 2.332699
> sd(X)
[1] 1.527318
```

R

3.14
Claim: Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be random variables and $a_1, \ldots, a_m$ and $b_1, \ldots, b_n$ be constants. Then:

$$\text{Cov}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_i Y_i\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \text{Cov}(X_i, Y_j).$$

Proof. This can be verified with a direct proof. We will do a simplified example with $m = n = 2$, which is not a formal proof, but can easily be extended to one. We will repeatedly use:

$$\text{Cov}(A, B) = \mathbb{E}[AB] - \mathbb{E}[A]\mathbb{E}[B].$$

Starting with the covariance, we apply the identity above. We have, for $m = n = 2$:

$$\begin{aligned}\text{Cov}(a_1X_1 + a_2X_2, b_1Y_1 + b_2Y_2) &= \mathbb{E}[(a_1X_1 + a_2X_2)(b_1Y_1 + b_2Y_2)] - \mathbb{E}[a_1X_1 + a_2X_2]\mathbb{E}[b_1Y_1 + b_2Y_2] \\ &= S_1 - S_2\end{aligned}$$

Multiplying and rewriting the first term:

$$\begin{aligned}S_1 = \mathbb{E}[(a_1X_1 + a_2X_2)(b_1Y_1 + b_2Y_2)] &= \mathbb{E}[a_1b_1X_1Y_1 + a_1b_2X_1Y_2 + a_2b_1X_2Y_1 + a_2b_2X_2Y_2] \\ &= a_1b_1\mathbb{E}[X_1Y_1] + a_1b_2\mathbb{E}[X_1Y_2] + a_2b_1\mathbb{E}[X_2Y_1] + a_2b_2\mathbb{E}[X_2Y_2]\end{aligned}$$

The second term:

$$\begin{aligned}S_2 = \mathbb{E}[a_1X_1 + a_2X_2]\mathbb{E}[b_1Y_1 + b_2Y_2] &= (a_1\mathbb{E}[X_1] + a_2\mathbb{E}[X_2])\,(b_1\mathbb{E}[Y_1] + b_2\mathbb{E}[Y_2]) \\ &= a_1b_1\mathbb{E}[X_1]\mathbb{E}[Y_1] + a_1b_2\mathbb{E}[X_1]\mathbb{E}[Y_2] + a_2b_1\mathbb{E}[X_2]\mathbb{E}[Y_1] + a_2b_2\mathbb{E}[X_2]\mathbb{E}[Y_2]\end{aligned}$$

Subtracting:

$$\begin{aligned}S_1 - S_2 &= a_1b_1\mathbb{E}[X_1Y_1] + a_1b_2\mathbb{E}[X_1Y_2] + a_2b_1\mathbb{E}[X_2Y_1] + a_2b_2\mathbb{E}[X_2Y_2] \\ &\quad - a_1b_1\mathbb{E}[X_1]\mathbb{E}[Y_1] - a_1b_2\mathbb{E}[X_1]\mathbb{E}[Y_2] - a_2b_1\mathbb{E}[X_2]\mathbb{E}[Y_1] - a_2b_2\mathbb{E}[X_2]\mathbb{E}[Y_2] \\ &= a_1b_1(\mathbb{E}[X_1Y_1] - \mathbb{E}[X_1]\mathbb{E}[Y_1]) + a_1b_2(\mathbb{E}[X_1Y_2] - \mathbb{E}[X_1]\mathbb{E}[Y_2]) \\ &\quad + a_2b_1(\mathbb{E}[X_2Y_1] - \mathbb{E}[X_2]\mathbb{E}[Y_1]) + a_2b_2(\mathbb{E}[X_2Y_2] - \mathbb{E}[X_2]\mathbb{E}[Y_2]) \\ &= a_1b_1\text{Cov}(X_1, Y_1) + a_1b_2\text{Cov}(X_1, Y_2) + a_2b_1\text{Cov}(X_2, Y_1) + a_2b_2\text{Cov}(X_2, Y_2) \\ &= \sum_{i=1}^{2}\sum_{i=j}^{2} a_ib_j\text{Cov}(X_i, Y_j)\end{aligned}$$

□

3.15
Defining the joint distribution:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{3}(x+y) & 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

We will calculate $\mathrm{Var}(2X - 3Y + 8)$. Constants disappear in the variance, and by Theorem 2.30:

$$\mathrm{Var}(2X - 3Y + 8) = 4\mathrm{Var}(X) + 9\mathrm{Var}(Y) - 12\mathrm{Cov}(X,Y)$$

Finding the marginal distributions:

$$f_X(x) = \int_0^2 \frac{1}{3}x + \frac{1}{3}y dy = \left[\frac{1}{3}xy + \frac{1}{6}y^2\right]_0^2 = \frac{2x}{3} + \frac{2}{3}$$

$$f_Y(y) = \int_0^1 \frac{1}{3}x + \frac{1}{3}y dx = \left[\frac{1}{6}x^2 + \frac{1}{3}xy\right]_0^1 = \frac{y}{3} + \frac{1}{6}$$

Calculating the expectation of X.

$$\mathbb{E}[X] = \int_0^1 x\left(\frac{2x}{3} + \frac{2}{3}\right) dx = \int_0^1 \left(\frac{2x^2}{3} + \frac{2x}{3}\right) dx = \left[\frac{2x^3}{9} + \frac{2x^2}{6}\right]_0^1 = \frac{5}{9}$$

Second moment:

$$\mathbb{E}[X^2] = \int_0^1 x^2\left(\frac{2x}{3} + \frac{2}{3}\right) dx = \int_0^1 \left(\frac{2x^3}{3} + \frac{2x^2}{3}\right) dx = \left[\frac{2x^4}{12} + \frac{2x^3}{9}\right]_0^1 = \frac{7}{18}$$

Variance of X:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{7}{18} - \frac{25}{81} = \frac{13}{162}$$

Calculating the expectation of Y.

$$\mathbb{E}[Y] = \int_0^2 y\left(\frac{y}{3} + \frac{1}{6}\right) dy = \int_0^2 \left(\frac{y^2}{3} + \frac{y}{6}\right) dx = \left[\frac{y^3}{9} + \frac{y^2}{12}\right]_0^2 = \frac{11}{9}$$

Second moment:

$$\mathbb{E}[Y^2] = \int_0^2 y^2\left(\frac{y}{3} + \frac{1}{6}\right) dy = \int_0^2 \left(\frac{y^3}{3} + \frac{y^2}{6}\right) dx = \left[\frac{y^4}{12} + \frac{y^3}{18}\right]_0^2 = \frac{16}{9}$$

Variance of Y:

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \frac{16}{9} - \frac{121}{81} = \frac{23}{81}$$

From the definition of the covariance - these calculations are pretty large, so skipping the details:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\
&= \int_0^2\int_0^1 \left(x-\frac{5}{9}\right)\left(y-\frac{11}{9}\right)\left(\frac{x}{3}+\frac{y}{3}\right)dxdy \\
&= \frac{1}{486}\int_0^2 -9y^2+20y-11dy \\
&= -\frac{1}{81}
\end{aligned}$$

Putting it all together:

$$\begin{aligned}
\mathrm{Var}(2X+3Y-8) &= 4\mathrm{Var}(X)+9\mathrm{Var}(Y)-12\mathrm{Cov}(X,Y) \\
&= 4\left(\frac{13}{162}\right)+9\left(\frac{23}{81}\right)-12\left(-\frac{1}{81}\right) \\
&= \frac{52}{162}+\frac{207}{81}+\frac{12}{81} \;=\; \frac{26}{81}+\frac{207}{81}+\frac{12}{81} \\
&= \frac{245}{81}
\end{aligned}$$

3.16
Let $r(x)$ and $s(y)$ be functions of x and y. Then:

$$\mathbb{E}[r(X)s(Y)|X] = r(X)\mathbb{E}[s(Y)|X], \qquad \mathbb{E}[r(X)|X] = r(X)$$

PROOF. By definition of the conditional expectation in the continuous case.

$$\begin{aligned}
\mathbb{E}[r(X)s(Y)|X=x] &= \int_{\mathcal{Y}} r(x)s(y)f_{Y|X}(y)dy \\
&= r(x)\int_{\mathcal{Y}} s(y)f_{Y|X}(y)dy \\
&= r(X)\mathbb{E}[s(Y)|X]
\end{aligned}$$

For the special case, we can set $s(Y)=1$.

$$\begin{aligned}
\mathbb{E}[r(X)|X=x] &= \int_{\mathcal{Y}} r(x)(1)f_{Y|X}(y)dy \\
&= r(x)\int_{\mathcal{Y}} (1)f_{Y|X}(y)dy \\
&= r(X)\mathbb{E}[(1)|X] \\
&= r(X)
\end{aligned}$$

which proves the statement. It is similar in the discrete case. $\square$

3.17
Proving that

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X]).$$

PROOF. Will disregard the hints.

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathbb{E}[(Y - \mathbb{E}[Y])^2] \\
&= \mathbb{E}\big[((Y - \mathbb{E}[Y|X]) + (\mathbb{E}[Y|X] - \mathbb{E}[Y]))^2\big] \\
&= \mathbb{E}\Big[(Y - \mathbb{E}[Y|X])^2 + 2(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - \mathbb{E}[Y]) + (b_X - \mathbb{E}[Y])^2\Big] \\
&= \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] + 2\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - \mathbb{E}[Y])] + \mathbb{E}[(\mathbb{E}[Y|X] - \mathbb{E}[Y])^2] \\
&= E_1 + 2E_2 + E_3
\end{aligned}$$

Consider each of these in turn. Using the definition of the conditional variance, with $\mu(x) = \mathbb{E}[Y|X]$.

$$E_1 = \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] = \mathbb{E}\big[\mathbb{E}[(Y - \mathbb{E}[Y|X])^2|X]\big] = \mathbb{E}\Big[\int_{\mathcal{Y}} (y - \mu(x))^2 f(y|x) dy\Big] = \mathbb{E}[\mathrm{Var}(Y|X)]$$

For E_3, we define $M = \mathbb{E}[Y|X]$, and $\mu_M = \mathbb{E}[M] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$. On this form, we see that we can use the definition of variance.

$$E_3 = \mathbb{E}[(\mathbb{E}[Y|X] - \mathbb{E}[Y])^2] = \mathbb{E}[(M - \mu_M)^2] = \mathrm{Var}(M) = \mathrm{Var}(\mathbb{E}[Y|X])$$

Finally, we consider E_2. Using iterated expectation and conditioning on X, then $\mathbb{E}[Y|X]$ and $\mathbb{E}[Y]$ become constants wrt. expectation. Define $C := \mathbb{E}[Y|X] - \mathbb{E}[Y]$ in order to clarify.

$$\begin{aligned}
\mathbb{E}[E_2] &= \mathbb{E}\Big(\mathbb{E}\big[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - \mathbb{E}[Y])\big|X\big]\Big) \\
&= \mathbb{E}\Big(\mathbb{E}\big[(Y - \mathbb{E}[Y|X])C\big|X\big]\Big) \\
&= \mathbb{E}\Big(C\mathbb{E}\big[(Y - \mathbb{E}[Y|X])\big|X\big]\Big) \\
&= \mathbb{E}\Big(C(\mathbb{E}[Y|X] - \mathbb{E}[Y|X])\Big) \\
&= \mathbb{E}\Big(C \cdot 0\Big) \\
&= 0
\end{aligned}$$

Collecting all the terms:

$$\begin{aligned}
\mathrm{Var}(Y) &= E_1 + 2E_2 + E_3 \\
&= \mathbb{E}[\mathrm{Var}(Y|X)] + 2(0) + \mathrm{Var}(\mathbb{E}[Y|X]) \\
&= \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X])
\end{aligned}$$

and the result has been proved. $\square$

3.18
If $\mathbb{E}[X|Y = y] = c$ for some constant c, then X and Y are uncorrelated.

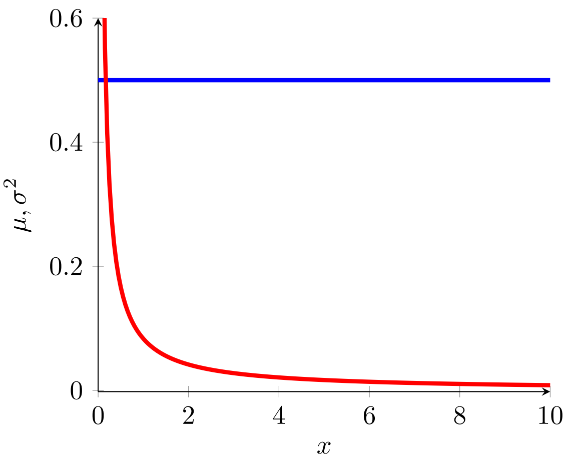
Proof. Assuming $\mathbb{E}[X|Y = y] = c$ for a constant c. Calculating the covariance:

$$\begin{aligned}
\operatorname{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[\mathbb{E}[XY|Y]] - \mathbb{E}[\mathbb{E}[X|Y]]\mathbb{E}[Y] \\
&= \mathbb{E}[\mathbb{E}[X|Y]Y] - \mathbb{E}[\mathbb{E}[X|Y]]\mathbb{E}[Y] \\
&= \mathbb{E}[cY] - \mathbb{E}[c]\mathbb{E}[Y] \\
&= c\mathbb{E}[Y] - c\mathbb{E}[Y] \\
&= 0
\end{aligned}$$

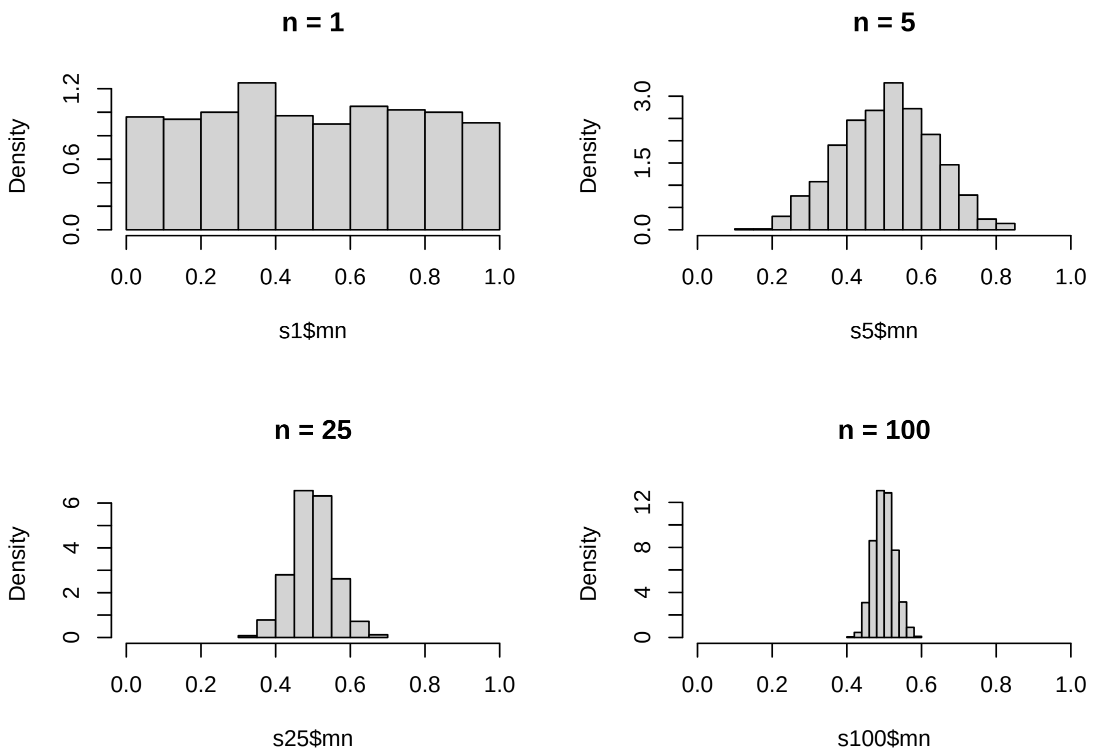
Since the covariance is 0, then it follows that the correlation is 0. $\square$

3.19
Studying the sample distribution for the Uniform(0,1) distribution. If $X \sim U(0, 1)$ then $\mathbb{E}[X] = 1/2$ and $\operatorname{Var}(X) = 1/12$. Will not plot f_X since it will simply be the straight line at $y = 1$.

From Theorem 3.17, we found that $\mathbb{E}[barX_n] = \mu = 1/2$ and $\operatorname{Var}(\overline{X}_n) = \sigma^2/n = 1/(12n)$. Plotting these as functions of n, mean as blue and variance as red. The mean remains at 1/2, but the variance will become a lot smaller as n increases.



For each $n = 1, 5, 25, 100$ we do a 1000 simulations of each n and make a histogram of the results that can be seen on the next page. The results are centered around the mean, 1/2, but the spread becomes smaller as we increase n. We are basically seeing the central limit theorem in action.




```r
simSampDist <- function(numSim) {
  retMean = rep(0, 1000); retVar = rep(0, 1000)
  for(i in 1:1000) {
    sim = runif(numSim)
    retMean[i]  = mean(sim)
    retVar[i]  = var(sim)
  }
  retVal = list()
  retVal$mn = retMean
  retVal$vr = retVal
  return(retVal)
}
s1 = simSampDist(1)
s5 = simSampDist(5)
s25 = simSampDist(25)
s100 = simSampDist(100)
# Creating plots
par(mfrow = c(2,2))
hist(s1$mn, main = "n = 1", xlim=c(0,1), prob = TRUE)
hist(s5$mn, main = "n = 5", xlim=c(0,1), prob = TRUE)
hist(s25$mn, main = "n = 25", xlim=c(0,1), prob = TRUE)
hist(s100$mn, main = "n = 100", xlim=c(0,1), prob = TRUE)
```

R

3.20
If a is a vector and X is a random vector with mean μ and variance Σ, then $\mathbb{E}[a^T X] = a^T \mu$ and $\text{Var}(a^T X) = a^T \Sigma a$. If A is a matrix then $\mathbb{E}[AX] = A\mu$ and $\text{Var}(AX) = A\Sigma A^T$.

PROOF.

$$\mathbb{E}[a^T X] = \begin{bmatrix} \mathbb{E}[a_1 X_1] \\ \mathbb{E}[a_2 X_2] \\ \vdots \\ \mathbb{E}[a_n X_n] \end{bmatrix} = \begin{bmatrix} a_1 \mathbb{E}[X_1] \\ a_2 \mathbb{E}[X_2] \\ \vdots \\ a_n \mathbb{E}[X_n] \end{bmatrix} = \begin{bmatrix} a_1 \mu_1 \\ a_2 \mu_2 \\ \vdots \\ a_n \mu_n \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}^T \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = a^T \mu$$

The variance of a vector is the covariance matrix.

$$\text{Var}(a^T X) = \begin{bmatrix} \text{Var}(a_1 X_1) & \text{Cov}(a_1 X_1, a_2 X_2) & \cdots & \text{Cov}(a_1 X_1, a_n X_n) \\ \text{Cov}(a_2 X_2, a_1 X_1) & \text{Var}(a_2 X_2) & \cdots & \text{Cov}(a_2 X_2, a_n X_n) \\ \vdots & \vdots & \ddots & \\ \text{Cov}(a_n X_n, a_1 X_1) & \text{Cov}(a_m X_m, a_2 X_2) & \cdots & \text{Var}(a_n X_n) \end{bmatrix}$$

$$= \begin{bmatrix} a_1 \text{Var}(X_1) a_1 & a_1 \text{Cov}(X_1, X_2) a_2 & \cdots & a_1 \text{Cov}(X_1, X_n) a_n \\ a_2 \text{Cov}(X_2, X_1) a_1 & a_2 \text{Var}(X_2) a_2 & \cdots & a_2 \text{Cov}(X_2, X_n) a_n \\ \vdots & \vdots & \ddots & \\ a_n \text{Cov}(X_n, X_1) a_1 & a_n \text{Cov}(X_n, X_2) a_2 & \cdots & a_n \text{Var}(X_n) a_n \end{bmatrix}$$

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}^T \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Var}(X_n) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a^T \Sigma a$$

The same general idea applies to matrices as well, execpt the notation becomes more extensive. Will skip this for brevity. □

3.21
Let X and Y be random variables. If $\mathbb{E}[Y|X] = X$, then $\text{Cov}(X, Y) = \text{Var}(X)$.

PROOF. Applying the covariance identity and using iterated expectation:

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

and the result has been proved. □

3.22
Let $X \sim U(0,1)$ and $0 < a < b < 1$. Define:

$$Y = \begin{cases} 1 & 0 < x < b \\ 0 & \text{otherwise} \end{cases}, \qquad Z = \begin{cases} 1 & a < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Are Y and Z independent? One way of thinking about independence is whether knowing something about Y tells us anything about Z, or vice-versa. In this case it doesn't seem to be that way. If we only know that Z is one, there is no way to determine whether $x < b$ or $x > b$, so we don't know what Y will be. By this reasoning, Y and Z are independent. To calculate it, let's set $a = 1/4$ and $b = 3/4$. Then:

	$Y = 0$	$Y = 1$	
$Z = 0$	1/16	3/16	1/4
$Z = 1$	3/16	9/16	3/4
	1/4	3/4	1

To determine e.g. $\mathbb{P}(Y = 0 \cap Z = 0)$ we first determine $\mathbb{P}(Y = 0|Z = 0)$ which is $1/4$, and then we determine $\mathbb{P}(Z = 0)$ which is $1/4$ which makes $\mathbb{P}(Y = 0 \cap Z = 0) = 1/16$. Now we calculate that $\mathbb{P}(Y = 0)\mathbb{P}(Z = 0) = (1/4)(1/4) = 1/16 = \mathbb{P}(Y = 0 \cap Z = 0)$ etc. which verifies that we have independence.

(b) Finding $\mathbb{E}[Y|Z]$. By definition of the iterated expectation and using independence ($\mathbb{P}(Y = y|Z = z) = \mathbb{P}(Y = y)$).

$$\mathbb{E}[Y|Z = 0] = \sum_y y \cdot \mathbb{P}(Y = y|Z = 0) = 0 + (1)\mathbb{P}(Y = 1|Z = 0) = \mathbb{P}(Y = 1) = b$$
$$\mathbb{E}[Y|Z = 1] = \sum_y y \cdot \mathbb{P}(Y = y|Z = 1) = 0 + (1)\mathbb{P}(Y = 1|Z = 1) = \mathbb{P}(Y = 1) = b$$

The last step follows because we know $x \sim U(0,1)$, so the probability is equal to the length of the interval.

Because of independence, Z does not affect the outcome. And in fact, there is a result that says $\mathbb{E}[Y|Z] = \mathbb{E}[Y]$ when Y is independent of Z, which we have almost proved above.

3.23
Finding the moment generating function for the Poisson, Normal and Gamma distributions. Recall the definition of the moment generating function:

$$M(t) = \mathbb{E}[e^{tX}] = \begin{cases} \displaystyle\sum_x e^{tx} p(x) & \text{discrete with pmf } p(x) \\ \displaystyle\int_{-\infty}^{\infty} e^{tx} f(x)dx & \text{continuous with pdf } f(x) \end{cases}$$

As usual with the Poisson distribution, the following identity is useful:

$$\sum_{k=0}^{\infty} \frac{a^n}{n!} = e^a$$

• **Poisson**. Assuming some arbitrary λ.

$$\begin{aligned}
M(t) &= \mathbb{E}[e^{tX}] \\
&= \sum_{n=0}^{\infty} e^{-\lambda} \frac{e^{tn}\lambda^n}{n!} \\
&= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= e^{\lambda e^t - \lambda} \\
&= e^{\lambda(e^t - 1)} \\
&= \exp\big(\lambda(e^t - 1)\big)
\end{aligned}$$

• **Normal**. Assuming some arbitrary μ and σ^2. The usual way of deriving the MGF for a normal distribution is to first do it for the standard normal distribution, since this is a lot easier. Denote this by $M_Z(t)$. We will 'compelte the square'.

$$\begin{aligned}
M_Z(t) &= \mathbb{E}[e^{tZ}] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2}x^2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(tx) \exp\left(-\frac{1}{2}x^2\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2 + tx\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2 + tx - \frac{t^2}{2} + \frac{t^2}{2}\right) dx \\
&= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2 - 2tx + t^2)\right) dx \\
&= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x - t)^2\right) dx \\
&= e^{t^2/2}
\end{aligned}$$

The integration is done by substituting $u = x - t$ which leads to the usual standard pdf, which integrates to 0. Hence, we have the MGF for the standard normal distribution. Using the MGF for the standard normal distribution, we can find it for a general normal distribution.

$$\begin{aligned}
M_X(t) &= \mathbb{E}[e^{tX}] = \mathbb{E}[e^{t(\mu+\sigma Z)}] \\
&= \mathbb{E}[e^{t\mu + t\sigma Z)}] = e^{t\mu}\mathbb{E}[e^{t\sigma Z)}] \\
&= e^{t\mu} M_Z(t\sigma) = e^{t\mu} e^{(t\sigma)^2/2} \\
&= \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}
\end{aligned}$$

So tedious!!

• **Gamma**. Assuming some arbitrary α and β.

$$
\begin{aligned}
M(t) &= \mathbb{E}[e^{tX}] \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty e^{tx} x^{\alpha-1} \exp(-x/\beta) dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} \exp(tx - x/\beta) dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} \exp(-(\frac{1}{\beta} - t)x) dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \frac{\Gamma(\alpha)}{\left(\frac{1}{\beta} - t\right)^\alpha} \\
&= \frac{1}{\left[\beta\left(\frac{1}{\beta} - t\right)\right]^\alpha} \\
&= \frac{1}{[1 - \beta t]^\alpha} \\
&= (1 - \beta t)^{-\alpha}
\end{aligned}
$$

where we used the definition of the Gamma function like in an earlier exercise.

3.24
Let $X_1, \ldots, X_n \sim \text{Exp}(\beta)$. Find the MGF of X_i. Prove that $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \beta)$.

Finding the MGF for the exponential distribution with parameter β.

$$
\begin{aligned}
M(t) &= \mathbb{E}[e^{tX}] \\
&= \frac{1}{\beta} \int_0^\infty \exp(tx) \exp(-x/\beta) dx \\
&= \frac{1}{\beta} \int_0^\infty \exp(-(1/\beta - t)x) dx \\
&= \frac{1}{\beta} \left(\frac{1}{\frac{1}{\beta} - t} \right) \\
&= \frac{1}{\beta} \left(\frac{\beta}{1 - \beta t} \right) \\
&= \frac{1}{1 - \beta t}
\end{aligned}
$$

All X_i have the same MGF. If we now assume they are independent, and define $Y = \sum_{i=1}^n X_i$, then according to Lemma 3.31:

$$
M_Y(t) = \prod_{i=1}^n M_X(t) = \prod_{i=1}^n \left(\frac{1}{1 - \beta t} \right) = \frac{1}{(1 - \beta t)^n} = (1 - \beta t)^{-n}
$$

which is the MGF for a Gamma distribution with parameters n and β, as we can see from the previous exercise. Hence, $Y \sim \text{Gamma}(n, \beta)$.