## Question 1 – Joint Probability Mass Function

Consider the function $p_{X,Y}(\cdot,\cdot)$

| $x$ | $y$ | $p_{X,Y}(x,y)$ |
|---|---|---|
| -1.0 | 1.0 | 1/7 |
| 0.0 | 2.0 | 2/7 |
| 1.0 | 3.0 | 1/7 |
| 1.5 | 3.0 | 2/7 |
| 3.0 | 4.0 | 1/7 |

Determine the following:

(a) Show that $p_{X,Y}$ is a valid probability mass function.

**Solution:**
If $\sum\limits_{x,y} p_{XY} = 1$ then it is a valid probability mass function, therefore the calculation

$$\begin{aligned}\sum_{x,y} p_{XY} &= p_{XY}(-1.0,1.0) + p_{XY}(0.0,2.0) + p_{XY}(1.0,3.0) + p_{XY}(1.5,3.0) + p_{XY}(3.0,4.0)\\ &= \frac{1}{7} + \frac{2}{7} + \frac{1}{7} + \frac{2}{7} + \frac{1}{7}\\ &= 1\end{aligned}$$

So $p_{XY}$ is a valid probability mass function.

(b) $P(X < 2.5; Y < 3)$.

**Solution:**
The cases where $X < 2.5$ are $p_{XY}(-1.0, 1.0)$, $p_{X,Y}(0, 2.0)$, $p_{X,Y}(1, 3.0)$. and $p_{X,Y}(1.5, 3.0)$. The first two of these cases also satisfy the condition $Y < 3$, so they should be added together to get the probability required.

$$P(X < 2.5, Y < 3) = p_{XY}(-1.0,1.0) + p_{X,Y}(0,2.0) = \frac{1}{7} + \frac{2}{7} = \frac{3}{7} = 0.429$$

(c) $P(X < 2.5)$.

**Solution:**
Here all four cases when $X < 2.5$ are valid as there is no condition on $Y$. Adding these together gives the required probability.

$$\begin{aligned}P(X < 2.5) &= p_{X,Y}(-1.0,1.0) + p_{X,Y}(0,2.0) + p_{X,Y}(1,3.0) + p_{X,Y}(1.5,3.0)\\ &= \frac{1}{7} + \frac{2}{7} + \frac{1}{7} + \frac{2}{7}\\ &= 0.857\end{aligned}$$

(d) $P(Y < 3)$.

**Solution:** The cases where $Y < 3$ are the same as part (b), $p_{XY}(-1.0, 1.0)$ and $p_{X,Y}(0, 2.0)$.

$$P(Y < 3) = p_{XY}(-1.0,1.0) + p_{X,Y}(0,2.0) = \frac{1}{7} + \frac{2}{7} = \frac{3}{8} = 0.429$$

(e) $P(X > 0.8, Y > 3.7)$.

(f) The marginal expected values E(X), E(Y), and marginal variances V(X), V(Y).

**Solution:**
To work out the expected value of $X$ calculate $\sum xp_{XY}(x)$ as follows

$$\begin{aligned}E(X) &= \sum_x xp_{XY}(x,\cdot)\\ &= -1.0\times\frac{1}{7}+0.0\times\frac{2}{7}+1.0\times\frac{1}{7}+1.5\times\frac{2}{7}+3.0\times\frac{1}{7}\\ &= 0.8571429\end{aligned}$$

Similarly the same can be done for the expected value of $Y$

$$\begin{aligned}E(Y) &= \sum_y yp_{XY}(\cdot,y)\\ &= 1.0\times\frac{1}{7}+2.0\times\frac{2}{7}+3.0\times\frac{1}{7}+3.0\times\frac{2}{7}+4.0\times\frac{1}{7}\\ &= 2.5714286\end{aligned}$$

**Note** that 3.0 has been included twice due to appearing in the probability mass function twice. To calculate the variance of $X$, first calculate $E(X^2)$

$$\begin{aligned}E(X^2) &= \sum_x x^2p_{XY}(x,\cdot)\\ &= -1.0^2\times\frac{1}{7}+0^2\times\frac{2}{7}+1.0^2\times\frac{1}{7}+1.5^2\times\frac{2}{7}+3.0^2\times\frac{1}{7}\\ &= 2.2142857\end{aligned}$$

Now the variance can be calculated by

$$V(X) = E(X^2) - E(X)^2 = 2.2143 - 0.8571^2 = 1.4796$$

Similarly by calculating $E(Y^2)$

$$\begin{aligned}E(Y^2) &= \sum_y y^2p_{XY}(\cdot,y)\\ &= 1.0^2\times\frac{1}{7}+2.0^2\times\frac{2}{7}+3.0^2\times\frac{1}{7}+3.0^2\times\frac{2}{7}+4.0^2\times\frac{1}{7}\\ &= 7.4285714\end{aligned}$$

the variance of $Y$ can be calculated as

$$V(Y) = E(Y^2) - E(Y)^2 = 7.4286 - 2.5714^2 = 0.8163$$

This can be checked in R using the following code

```
> x = c(-1.0,0,1,1.5,3.0)
> y = c(1.0,2.0,3.0,3.0,4.0)
> p = c(1/7,2/7,1/7,2/7,1/7)
>
> EX = sum(x*p)
> EY = sum(y*p)
> EX2 = sum(x^2*p)
```

```
> EY2 = sum(y^2*p)
>
> VarX = EX2-EX^2
> VarY = EY2-EY^2
>
> c(EX,EY,VarX,VarY)

[1] 0.8571429 2.5714286 1.4795918 0.8163265
```

(g) Are $X$ and $Y$ independent random variables?

**Solution:**
To test for independence we need to calculate the marginal distributions. For $f_x$ the mariginal distribution can be read off the joint distribution probabily mass funciton table

| $x$ | $p_x$ |
|---|---|
| -1.0 | 1/7 |
| 0.0 | 2/7 |
| 1.0 | 1/7 |
| 1.5 | 2/7 |
| 3.0 | 1/7 |

For $f_Y$ we have to sum the probabilities with the same $y$ value. The marginal distribution $f_y$ is thus

| $y$ | $p_y$ |
|---|---|
| 1.0 | 1/7 |
| 2.0 | 2/7 |
| 3.0 | $1/7 + 2/7 = 3/7$ |
| 4.0 | 1/7 |

Recall to be independent

$$f_{x,y} = f_x f_y.$$

If we take the first entry in the distributions we find

$$f_{x,y}(-1.0, 1.0) = 1/7$$

$$\begin{aligned} f_x(-1.0)f_y(1.0) &= 1/7 \times 1/7 \\ &= 1/49 \\ &\neq f_{x,y}(-1.0, 1.0) \end{aligned}$$

As this does not hold, the variable are dependent.

(h) $P(X + Y \leqslant 5)$

**Solution:**
Here we need to work out the distribution of the sum of the variables

| $x$ | $y$ | $x+y$ | $p_{X,Y}(x,y)$ |
|---|---|---|---|
| -1.0 | 1.0 | 0.0 | 1/7 |
| 0.0 | 2.0 | 2.0 | 2/7 |
| 1.0 | 3.0 | 4.0 | 1/7 |
| 1.5 | 3.0 | 4.5 | 2/7 |
| 3.0 | 4.0 | 7.0 | 1/7 |

From this we can see that there are four case where the sum is less than 5. As $p_{x,y}(3.0, 4.0)$ is the only case where the sum is greater than 5 we simply take this from 1 and get the required probability using the law of total probability.

$$P(X+Y \leqslant 5) = 1 - P(X = 3.0, Y = 4.0) = 1 - \frac{1}{7} = \frac{6}{7}.$$

## Question 2 – Joint Probability Density Function

Determine the value of $c$ that makes the function

$$f(x,y) = ce^{-2x-3y}$$

a joint probability density function over the range $0 \le x \le \infty$ and $0 \le y \le x$.

**Solution:** To show that this is a valid joint probability density function we need to make sure than when calculated over the entire domain the joint cumulative distribution function equals 1. Therefore we can calculate $c$ as follows

$$\begin{aligned}
F(x,y) = 1 &= \int_0^\infty \int_0^x ce^{-2x-3y}\ dy\ dx \\
&= \int_0^\infty \left[-\frac{1}{3}ce^{-2x-3y}\right]_0^x\ dx \\
&= \int_0^\infty \left[-\frac{1}{3}ce^{-2x-3x} + \frac{1}{3}ce^{-2x}\right]\ dx \\
&= \left[\frac{1}{15}ce^{-5x} - \frac{1}{6}ce^{-2x}\right]_0^\infty \\
&= -\frac{1}{15}c + \frac{1}{6}c = \frac{1}{10}c \\
\therefore c = 10 &
\end{aligned}$$

Determine the following:

(a) $P(X < 1, Y < 2)$,

**Solution:** Here we use the joint cumulative distribution function again. As 2 is larger than 1 the function is not defined higher than $Y = 1$. Therefore

$$\begin{aligned}
P(X < 1, Y < 2) &= P(X < 1, Y < 1) \\
&= F(1,1) \\
&= \int_0^1 \int_0^x 10e^{-2x-3y}\ dx \\
&= \left[\frac{2}{3}e^{-5x} - \frac{5}{3}e^{-2x}\right]_0^1 \\
&= \frac{2}{3}e^{-5} - \frac{5}{3}e^{-2} - \frac{2}{3} + \frac{5}{3} \\
&= 0.7789332
\end{aligned}$$

(b) $P(1 < X < 2)$,

**Solution:** Here we need to calculate the cumulative distribution function in the range $1 < x < 2$ and $0 < y < x$. We do so as follows

$$
\begin{aligned}
P(1 < X < 2) &= F(2, Y) - F(1, Y) \\
&= \int_1^2 \int_0^x 10e^{-2x-3y}\ dx \\
&= \left[\frac{2}{3}e^{-5x} - \frac{5}{3}e^{-2x}\right]_1^2 \\
&= \frac{2}{3}e^{-10} - \frac{5}{3}e^{-4} - \frac{2}{3}e^{-5} + \frac{5}{3}e^{-2} \\
&= 0.190571
\end{aligned}
$$

(c) $P(Y > 3)$,

**Solution:** As the joint cumulative distribution will give us the probability $P(Y < 3)$ and that this is only defined on the domain $0 < x < 3$ we need to take the compliment of this calculation.

$$
\begin{aligned}
P(Y > 3) &= 1 - F(X, 3) \\
&= \int_3^\infty \int_3^x 10e^{-2x-3y}\ dx \\
&= \int_3^\infty \left[-\frac{10}{3}e^{-2x-3y}\right]_3^x\ dx \\
&= \int_3^\infty -\frac{10}{3}e^{-5x} + \frac{10}{3}e^{-2x-9}\ dx \\
&= \left[\frac{2}{3}e^{-5x} - \frac{5}{3}e^{-2x}\right]_0^3 \\
&= \frac{2}{3}e^{-15} - \frac{5}{3}e^{-6} - \frac{2}{3} + \frac{5}{3} \\
&= 0.004131
\end{aligned}
$$

(d) $E[X]$,

**Solution:** Here we are calculating the marginal expectation of $X$. We do this as follows

$$
\begin{aligned}
E[X] &= \int_0^\infty \int_0^x x f(x,y)\ dy\ dx \\
&= \int_0^\infty \int_0^x 10xe^{-2x-3y}\ dy\ dx \\
&= \int_0^\infty 10x \left[\frac{e^{-2x-3y}}{-3}\right]_0^x\ dx \\
&= -\frac{10}{3}\int_0^\infty xe^{-5x} - xe^{-2x}\ dx \\
&= -\frac{10}{3}\left(\left[\frac{xe^{-5x}}{-5}\right]_0^\infty + \int_0^\infty \frac{e^{-5x}}{5}\ dx - \left[\frac{xe^{-2x}}{-2}\right]_0^\infty - \int_0^\infty \frac{e^{-2x}}{2}\ dx\right) \\
&= -\frac{10}{3}\left(\left[\frac{e^{-5x}}{-25}\right]_0^\infty - \left[\frac{e^{-2x}}{-4}\right]_0^\infty\right) \\
&= -\frac{10}{3}\left(\frac{1}{25} - \frac{1}{4}\right) \\
&= \frac{7}{10}
\end{aligned}
$$

(e) $E[Y]$,

**Solution:** Here we are calculating the marginal expectation of $Y$. We do this in a similar way to the previous part.

$$
\begin{aligned}
E[Y] &= \int_0^\infty \int_0^x y f(x,y)\ dy\ dx \\
&= \int_0^\infty \int_0^x 10ye^{-2x-3y}\ dy\ dx \\
&= 10\int_0^{infty} \left(\left[\frac{ye^{-2x-3y}}{-3}\right]_0^x + \int_0^x \frac{e^{-2x-3y}}{3}\ dy\right)\ dx \\
&= -\frac{10}{3}\int_0^\infty xe^{-5x} + \left[\frac{e^{-2x-3y}}{3}\right]_0^x\ dx \\
&= -\frac{10}{3}\int_0^\infty xe^{-5x} + \frac{e^{-5x}}{3} - \frac{e^{-2x}}{3}\ dx \\
&= -\frac{10}{3}\left(\left[\frac{xe^{-5x}}{-5}\right]_0^\infty + \int_0^\infty \frac{e^{-5x}}{5}\ dx - \left[\frac{e^{-5x}}{15}\right]_0^\infty + \left[\frac{e^{-2x}}{6}\right]_0^\infty\right) \\
&= -\frac{10}{3}\left(-\left[\frac{e^{-5x}}{25}\right]_0^\infty + \frac{1}{15} - \frac{1}{6}\right) \\
&= -\frac{10}{3}\left(\frac{1}{25} - \frac{1}{10}\right) \\
&= -\frac{10}{3}\left(\frac{2}{50} - \frac{5}{50}\right) \\
&= \frac{1}{5}
\end{aligned}
$$

(f) the marginal probability distribution of $X$,

**Solution:** Here we need to integrate the joint probability density function over all values of $Y$.

$$\begin{aligned} f_x(x) &= \int_0^x f(x,y)\ dy \\ &= \int_0^y 10e^{-2x-3y}\ dy \\ &= -\frac{10}{3}\left[e^{-2x-3y}\right]_0^x \\ &= -\frac{10}{3}\left(e^{-5x} - e^{-2x}\right) \end{aligned}$$

(g) $P(X|Y = 1)$,

**Solution:** Here we use the marginal distribution of $X$ to calculated that $X|Y = 1$. However as this is a continuous distribution the probability of any particular value is 0. So

$$P(X|Y = 1) = 0.$$

(h) $E[X|Y = 2]$,

**Solution:** To calculate the conditional expectation when $Y = 2$ we use the following equation

$$E[X|Y = 2] = \int_0^\infty x f_{X|y=2}(x)\ dx = \int_0^\infty x\frac{f(x,2)}{f_Y(2)}\ dx$$

First we need to calculate the marginal distribution of $Y$.

$$\begin{aligned} f_Y(y) &= \int_y^\infty f(x,y)\ dy = \int_y^\infty 10e^{-2x-3y}\ dx \\ &= 10e^{-3y}\int_y^\infty e^{-2x}\ dx \\ &= 10e^{-3y}\left[-\frac{e^{-2x}}{2}\right]_y^\infty \\ &= 10e^{-3y}\left[-\frac{e^{-2\times\infty}}{2} + \frac{e^{-2y}}{2}\right] \\ &= 5e^{-5y} \end{aligned}$$

Now substitute $Y = 2$ into this equation

$$f_Y(2) = 5e^{-10}$$

We now substitute this value back into our original equation and integrate.

$$\begin{aligned} E[X|Y = 2] &= \int_0^\infty \frac{10xe^{-2x-6}}{5e^{-10}}\ dx \\ &= 2e^4\int_0^\infty xe^{-2x}\ dx \\ &= 2e^4\left[\frac{xe^{-2x}}{-2}\right]_0^\infty + \int_0^\infty \frac{e^{-2x}}{2}\ dx \\ &= 2e^4\left[-\frac{e^{-2x}}{4}\right]_0^\infty \\ &= 2e^4\frac{1}{4} \\ &= \frac{e^4}{2} \approx 27.299075. \end{aligned}$$

## Question 3 – Covariance and Correlation

Let

$$f_{x,y} = \frac{1}{2\pi} e^{-\frac{1}{2}\left[(x+1.5)^2+(y-5)^2\right]},$$

for $x, y \in \mathbb{R}$

(a) Show that $f$ is a joint density function, that is $f_{x,y} \geqslant 0$ and

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f_{x,y} \; dx \; dy = 1.$$

**Solution:** Looking at the probability density function we notice that it has an exponential function in it. We know $e^x > 0$ for all real numbers, so we can do the following calculation

$$e^w > 0$$

Set $w = -\frac{1}{2}\left((x+1.5)^2 + (y-3)^2\right)$. As $x, y \in \mathbb{R}$, $w \in \mathbb{R}$ and $w \leqslant 0$

$$e^{-\frac{1}{2}\left((x+1.5)^2+(y-3)^2\right)} > 0$$

Now multiply both sides by $\frac{1}{2\pi}$

$$\frac{1}{2\pi} e^{-\frac{1}{2}\left((x+1.5)^2+(y-3)^2\right)} > 0$$

Therefore $f_{x,y} \geqslant 0$. To show that the integral over the entire range equals one we need to integrate the function. We do this as follows

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}\left[(x+1.5)^2+(y-5)^2\right]} \; dx \; dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-3)^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x+1.5)^2} \; dx \; dy$$

Using the fact that the Gaussian Integral is $\int_{-\infty}^{\infty} e^{-a(x+b)^2} \; dx = \sqrt{\frac{\pi}{a}}$

$$\begin{aligned} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-3)^2} \sqrt{2\pi} \; dy \\ &= \frac{1}{2\pi} \sqrt{2\pi}\sqrt{2\pi} \\ &= 1. \end{aligned}$$

So this is a valid joint probability density.

(b) Determine the marginal density function $f_X(x)$.

**Solution:** To get the marginal density function for $X$ we have to integrate the joint probability density function over all values of $Y$

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[(x+1.5)^2+(y-3)^2\right]} \; dy$$

Using the Gaussian Integral

$$= \frac{\sqrt{2\pi}}{2\pi} e^{-\frac{1}{2}(x+1.5)^2}.$$

(c) Determine the conditional density function $f_x(x|y=2)$.

**Solution:** Again as this is a continuous joint probability function this density function does not exist.

(d) Determine $E(X)$.

**Solution:** Looking at the form of the probability density function we notice that is is a bivariate normal distribution with independent variables. Therefore $E[X] = -1.5$

(e) Obtain $P(X > 2, Y < 5)$.

**Solution:** Given that this is a bivariate normal distribution and we note that $\rho(X,Y) = 0$ we know that $X$ and $Y$ are independent. This means that

$$P(X > 2, Y < 5) = P(X > 2)P(Y < 5)$$

Examining the joint density function shows that $E[X] = -1.5$, $E[Y] = 5$ and $V[X] = V[Y] = 1$. So

$$\begin{aligned} P(X > 2) &= P\left(Z > \frac{2-(-1.5)}{1}\right) \\ &= P(Z > 3.5) \\ &= 2e-04 \end{aligned}$$

$$\begin{aligned} P(Y < 5) &= P\left(Z < \frac{5-5}{1}\right) \\ &= P(Z < 0) \\ &= 0.5 \end{aligned}$$

Multiplying these probabilities together we get

$$P(X > 2, Y < 5) = 1e-04$$

(f) Verify your answer in (d) and (e) using R code, translating the above joint probability density function into a bivariate normal distribution, using for example the R function: `dmvnorm`/`rmvnorm` from the `mvtnorm` package.

**Solution:** Examining the joint density function shows that $E[X] = -1.5$, $E[Y] = 5$ and $V[X] = V[Y] = 1$, with $Cov(X,Y) = 0$ and so they are independent.

```
> library(mvtnorm)
> library(tidyverse)
> mu = c(-1.5,5)
> Sigma = diag(2)
>
> x=seq(-7.5,4.5,by=0.01)
> y=seq(-1,11,by=0.01)
> x=rep(x,length(y))
> x=sort(x)
> y=rep(y,length(y))
> support=tibble(x,y)
>
> P = dmvnorm(support,mean=mu,sigma=Sigma)
```

```
> jpmf = bind_cols(support,p=P)
>
> with(jpmf,sum(x*p))/10000

[1] -1.5

> values = rmvnorm(10^6,mean=mu,sigma=Sigma)
> mean(values[,1]>2&values[,2]<5)

[1] 0.000122
```

In both cases the answers are confirmed.

## Question 4 – More Fun With Two Random Variables

Let $X$ and $Y$ be independent random variables with $E(X) = 3$, $V(X) = 4$, $E(Y) = 5$, $V(Y) = 9$. Let $g(X, Y) = -X + 2Y$. Determine the following:

(a) $E(g(X, Y))$,

**Solution:** Here we have a linear combination of random variables and so to calculate the expected value we do the following

$$\begin{aligned} E[g(X,Y)] &= E[-X + 2Y] \\ &= E[-X] + E[2Y] \\ &= -E[X] + 2E[Y] \\ &= -3 + 2 \times 5 \\ &= 7 \end{aligned}$$

(b) $V(g(X, Y))$,

**Solution:** Similarly for the variance the calculation is as follows, remembering that variance is a squared quantity.

$$\begin{aligned} V[g(X,Y)] &= V[-X + 2Y] \\ &= V[-X] + V[2Y] \\ &= V[X] + 4V[Y] \\ &= 4 + 4 \times 9 \\ &= 40 \end{aligned}$$

(c) $\rho(X, Y)$

**Solution:** As the random variables are independent $Cov(X, Y) = 0$. Therefore

$$\rho(X,Y) = \frac{Cov(X,Y)}{V[X]V[Y]} = 0$$

If you know additional that $X \sim U(a, b)$ and $Y \sim U(c, d)$ and are independent find:

(d) (a), (b), (c) based on the information given above

**Solution:** As the distribution of the random variables does not matter for the expected value and variance, the values will not change. So

   (a) $E[g(X,Y)] = 7$
   (b) $V[g(X,Y)] = 40$
   (c) $\rho(X,Y) = 0$

(e) $P(g(X,Y) < 18)$,

**Solution:** To work out this probability first we need to work out the starting and end points of our distribution. Given we know the expected value and variance we can work backward using the standard formula. So,

$$\begin{aligned}
E[U] &= \frac{a+b}{2} \\
7 &= \frac{a+b}{2} \\
14 - b &= a
\end{aligned}$$

$$\begin{aligned}
V[U] &= \frac{(b-a)^2}{12} \\
40 &= \frac{(b-a)^2}{12} \\
480 &= (b - 14 + b)^2 \\
&= 4(b-7)^2 \\
120 &= b^2 - 14b + 49 \\
0 &= b^2 - 14b - 71 \\
b &= \frac{14 \pm \sqrt{14^2 + 4 * 71}}{2} \\
&= \frac{14 \pm \sqrt{480}}{2} \\
&= 7 \pm 2\sqrt{30}.
\end{aligned}$$

taking the larger of the values

$$\begin{aligned}
b &= 7 + 2\sqrt{30} \\
a &= 7 - 2\sqrt{30}
\end{aligned}$$

Note here that $b \approx 17.954$, so $P(g(X,Y) < 18) = 1$.

## Question 5 – Even More Fun With Bivariate Normal Distributions

Let $X$ and $Y$ be independent normally distributed with mean $\mu_X = 2$ and $\mu_Y = 3$ and standard deviations $\sigma_X = 3$ and $\sigma_Y = 5$, respectively. Determine the following:

(a) $P(3X + 6Y > 15)$,

**Solution:** First we need to work the distribution of the linear combination of the random variables is. We will call this variable $W$, and first we will calculate the mean.

$$\begin{aligned}
E[W] &= E[3X + 6Y] \\
&= E[3X] + E[6Y] \\
&= 3E[X] + 6E[Y] \\
&= 3 \times 2 + 6 \times 3 \\
&= 24.
\end{aligned}$$

Now we calculate the variance of the variable $W$, remembering that variance is a squared quantity

$$\begin{aligned} V[W] &= V[3X + 6Y] \\ &= V[3X] + V[6Y] \\ &= 9V[X] + 36V[Y] \\ &= 9 \times 9 + 36 \times 25 \\ &= 981. \end{aligned}$$

The standard deviation of $W$ is then the square root of the variance

$$sd[W] = \sqrt{981} = 31.3209.$$

Having worked this out, we can now move on to the probability. We need to standardise $W$ to use our Normal Distribution tables first, and then look up the value from there.

$$\begin{aligned} P(3X + 6Y > 15) &= P(W > 15) \\ &= P\left(Z > \frac{15 - 24}{\sqrt{142}}\right) \end{aligned}$$

As our tables give P(Z<z) we need to use the compliment of the probability

$$\begin{aligned} &= 1 - P\left(Z < \frac{-9}{\sqrt{142}}\right) \\ &= 1 - P(Z < -0.2873479) \\ &= 1 - 0.3869 \\ &= 0.6131. \end{aligned}$$

(b) $P(3X + 6Y < 30)$,

**Solution:** Here we follow a similar proceedure as the previous question

$$\begin{aligned} P(3X + 6Y < 30) &= P(W < 30) \\ &= P\left(Z < \frac{30 - 24}{\sqrt{981}}\right) \\ &= P(Z < 0.1916 \\ &= 0.5760. \end{aligned}$$

(c) $Cov(X, Y)$

**Solution:**
As the random variables $X$ and $Y$ are independent $Cov(X, Y) = 0$.

(d) Verify (a) and (b) using R code, where for each case you generate a million $X$'s and a million $Y$'s and simulate the linear combination $3X + 6Y$.

**Solution:**
First to set up the random variables $X$ and $Y$ we use the `rnorm(n,mu,sigma)` function in R

```
> xs=rnorm(10^6,2,3)
> ys=rnorm(10^6,3,5)
```

Now we create the linear combination of these values by simply applying the equation to them. We should also check the mean and the standard deviation.

```
> ws=3*xs+6*ys
> mean(ws)

[1] 23.97044

> sd(ws)

[1] 31.28386
```

For part (a) the following code is used where we use mean to count the number of times the statement is true out of all the possible values.

```
> mean(ws>15)

[1] 0.612213
```

and for part (b)

```
> mean(ws<30)

[1] 0.576769
```

(e) Assume now that the random variables come from another distribution (not Normal), but keep the same means and variances. Are your answers for (a), (b), (c) likely to change?

**Solution:** As (a) and (b) are probabilities which are dependent on the distribution used they will change. However as the answer for (c) does not depend on the distribution and is true for all independent variables it will not change.

This can be demonstrated if the distribution is change to a Uniform Distribution (the only one covered in the course that will allow the means and variances to remain the same). To calculate the start and end points, first calculate the relationship between $a$ and $b$ (the start and end points) using the Expected Value and Variance formulae.

$$\begin{aligned} E(X') = 2 &= \frac{a+b}{2} \\ 4 &= a+b \\ 4-b &= a \end{aligned}$$

$$\begin{aligned} V(X') = 9 &= \frac{(b-a)^2}{12} \\ 108 &= (b-a)^2 \end{aligned}$$

Combining these equations the values for $a$ and $b$ can be calculated

$$\begin{aligned}
108 &= (b-(4-b))^2 \\
&= (2b-4)^2 \\
&= 4(b-2)^2 \\
27 &= (b-2)^2 \\
9 &= b-2 \\
b &= 7 \\
a &= 6-b \\
&= 6-7 \\
&= -1
\end{aligned}$$

The same process can be done to create $Y'$ with $E(Y') = 3$ and $V(Y') = 5$, finding the end points to be $\left[3-5\sqrt{3}, 3+5\sqrt{3}\right]$. Using R to simulate these distributions as before:

```
> X = runif(10^6,min=-1,max=7)
> Y = runif(10^6,min=3-5*sqrt(3),max=3+5*sqrt(3))
>
> linComb = 3*X+6*Y
>
> mean(linComb > 15)

[1] 0.615762

> mean(linComb < 30)

[1] 0.528326
```

As predicted, these values are slightly different to those obtained in (a) and (b).

(f) Assume now that $X$ and $Y$ are normally distributed but are dependent with $Cov(X,Y) = 5$. Write an explicit expression using a double integral for $P(X < 5, Y > 3)$.

**Solution:** From the previous parts, $\mu_X = 2$, $\mu_Y = 3$, $\sigma_X = 3$ and $\sigma_Y = 5$. Now calculate the correlation coefficient $\rho$

$$\begin{aligned}
\rho &= \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \\
&= \frac{5}{3 \times 5} \\
&= \frac{5}{15} \qquad = \frac{1}{5}
\end{aligned}$$

For the bivariate normal distribution

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\}$$

Substituting in the values above and simplifying

$$\begin{aligned}
f_{X,Y}(x,y) &= \frac{1}{2\pi \times 3 \times 5\sqrt{1-\frac{1}{25}}} \times \exp\left\{\frac{-1}{2\left(1-\frac{1}{25}\right)}\left[\frac{(x-2)^2}{3^2} - \frac{2\times\frac{1}{25}(x-2)(y-3)}{2\times 5} + \frac{(y-3)^2}{5^2}\right]\right\} \\
&= \frac{1}{30\pi\sqrt{\frac{4}{5}}} \times \exp\left\{\frac{-5}{8}\left[\frac{(x-2)^2}{9} - \frac{(x-2)(y-3)}{125} + \frac{(y-3)^2}{25}\right]\right\} \\
&= \frac{\sqrt{5}}{60\pi} \times \exp\left\{-\frac{5(x-2)^2}{72} + \frac{(x-2)(y-3)}{200} - \frac{(y-3)^2}{40}\right\}
\end{aligned}$$

Then the definite integral is as follows

$$\begin{aligned}
P(X<2, Y>7) &= \int_{x=-\infty}^{5}\int_{y=3}^{\infty} \frac{\sqrt{5}}{60\pi} \times \exp\left\{-\frac{5(x-2)^2}{72} + \frac{(x-2)(y-3)}{200} - \frac{(y-3)^2}{40}\right\}\, dx\, dy \\
&= \frac{\sqrt{5}}{60\pi}\int_{x=-\infty}^{5}\int_{y=3}^{\infty} \exp\left\{-\frac{5(x-2)^2}{72} + \frac{(x-2)(y-3)}{200} - \frac{(y-3)^2}{40}\right\}\, dx\, dy
\end{aligned}$$

## Question 6 – Door Casing

There width of a casing for a door is normally distributed with a mean of 24 inches and a standard deviation of 1/8 inch. The width of a door is normally distributed with a mean of 23 and 7/8 inches and a standard deviation of 1/16 inch. Assume independence.

(a) Determine the mean and standard deviation of the difference between the width of the casing and the width of the door.

**Solution:**
Work out the difference we need to create a new random variable. If we take $X$ to be the width of the casing and $Y$ to be the width of the door then we can define $D = X - Y$. To find the mean of the difference we do the following calculation

$$\begin{aligned}
E[D] &= E[X-Y] \\
&= E[X] + E[-Y] \\
&= E[X] - E[Y] \\
&= 24 - 23\frac{7}{8} \\
&= \frac{1}{8} \text{ inch.}
\end{aligned}$$

To find the standard deviation of the difference we first need to work out the variance of the difference. Remember variance is a squared quantity. The calculation is as follows

$$\begin{aligned}
V[D] &= V[X-Y] \\
&= V[X] + V[-Y] \\
&= V[X] + (-1)^2 V[Y] \\
&= V[X] + V[Y] \\
&= \left(\frac{1}{8}\right)^2 + \left(\frac{1}{16}\right)^2 \\
&= \frac{1}{64} + \frac{1}{256} \\
&= \frac{5}{256}
\end{aligned}$$

The standard deviation is the square root of the variance and so

$$sd[D] = \frac{\sqrt{5}}{16} \text{ inch.}$$

(b) What is the probability that the width of the casing minus the width of the door exceeds 1/4 inch?

**Solution:**
Given our standard Normal distribution table is given as $P(Z < z)$ to calculate this probability we need to take the complement of $P\left(D < \frac{1}{4}\right)$. Remember to also standardise so we can use our table.

$$\begin{aligned}
P\left(D > \frac{1}{4}\right) &= 1 - P\left(D < \frac{1}{4}\right) \\
&= 1 - P\left(Z < \frac{\frac{1}{4} - \frac{1}{8}}{\frac{\sqrt{5}}{16}}\right) \\
&= 1 - P\left(Z < \frac{4-2}{\sqrt{5}}\right) \\
&= 1 - P(Z < 0.89) \\
&= 1 - 0.8145 \\
&= 0.1855
\end{aligned}$$

(c) What is the probability that the door does not fit in the casing?

**Solution:**
For the door to not fit in the casing the difference between them must be negative. To calculate the probability of this we need to first standardise and then look up the value in the standard Normal Table

$$\begin{aligned}
P(D < 0) &= P\left(Z < \frac{-\frac{1}{8}}{\frac{\sqrt{5}}{16}}\right) \\
&= P\left(Z < \frac{-2}{\sqrt{5}}\right) \\
&= P(Z < -0.89) \\
&= 0.1855.
\end{aligned}$$

## Question 7 – The Beauty of Proof

Suppose hat the joint probability distribution of the continuous random variables $X$ and $Y$ is constant on the rectangle $0 < x < a$ and $0 < y < b$ for $a$, $b \in \mathbb{R}^+$.
Show mathematically that $X$ and $Y$ are independent. *Hint:*

(a) *Recall* $\int_{\Omega_X} \int_{\Omega_Y} f(x, y)\ dy\ dx = 1$

(b) *Recall* $X$, $Y$ *are independent if* $f_x f_y = f_{xy}$.

**Solution:** First we should use the first hint and calculate the cumulative distribution function over the entire sample space.

$$\begin{aligned}
F(a, b) &= \int_0^a \int_0^b c\ dy\ dx \\
&= \int_0^a [cy]_0^b\ dx \\
&= \int_0^a cb\ dx \\
&= [cbx]_0^a \\
&= cba
\end{aligned}$$

As this should equal one as a valid joint probability distribution $cba = 1$. Now using the second hint we know that

$$f_{xy} = c$$

Now calculating the marginal distributions we get

$$\begin{aligned}
f_x &= \int_0^b c\ dy \\
&= cb
\end{aligned}$$

$$\begin{aligned}
f_y &= \int_0^a c\ dx \\
&= ca
\end{aligned}$$

Calculating the product of these we get

$$f_x f_y = cbca = cbac$$

As we have shown before with the cumulative joint distribution we have $cba = 1$ so

$$f_x f_y = c = f_{xy}.$$

## Question 8 – Risk Analysis

A marketing company performed a risk analysis for a manufacturer of synthetic fibres and concluded that new competitors present no risk 13% of the time, moderate risk 72% of the time, and high risk otherwise.
Eight international companies are planning to open new facilities for the manufacture of synthetic fibres with the next three years. Assume the companies are independent. Let $X$, $Y$, $Z$ denote the number of new competitors that will pose no, moderate and high risk for the

interested company, respectively.
The range of the joint probability distribution of $X$, $Y$ and $Z$ is

$$\Omega = \{(x, y, z) : x + y + z = 8, x, y, z \in [0.8]\}$$

To calculate the probability mass function for these variables, use the R code given in the file `STAT2201-A3-2019a-Q8.r`.

```
> library(tidyverse)
> norisk = 0.13
> modrisk = 0.72
> hirisk = 1-norisk-modrisk
> probs= c(norisk,modrisk,hirisk)
> jpmf<-tibble(X=c(rep(0,81),rep(1,81),rep(2,81),rep(3,81),rep(4,81),rep(5,81),
+ rep(6,81),rep(7,81),rep(8,81)),
+ Y=rep(c(rep(0,9),rep(1,9),rep(2,9),rep(3,9),rep(4,9),rep(5,9),rep(6,9),
+ rep(7,9),rep(8,9)),9),
+ Z=rep(0:8,81))
> jpmf2<-filter(jpmf,X+Y+Z==8)
> M<-as.matrix(jpmf2)
> colnames(M)<-NULL
> P<-apply(M,1,dmultinom,prob=c(norisk,modrisk,hirisk))
> jpmf2=bind_cols(jpmf2,p=P)
> #print(jpmf2,n=Inf) #to print pmf.
```

(a) Determine $P(X = 1, Y = 3, Z = 1)$.

**Solution:** Note, the sum of $X$, $Y$ and $Z$ here does not equal eight and therefore is not in the range of the joint probability distribution.Therefore the probability will be zero.

To confirm this and calculate this probability we need to count how many cases in the probability mass function have one company with no risk, three companies with moderate risk and one with high risk. We do this as follows:

```
> with(filter(jpmf2,X==1&Y==3&Z==1),sum(p))

[1] 0
```

(b) Determine $P(Y \leqslant 2)$.

**Solution:** Here we are looking for all the cases where $Y \leqslant 2$, and then sum their probabilities together. To do this in R we do the following.

```
> with(filter(jpmf2,Y<=2),sum(p))

[1] 0.007809707
```

(c) Determine $P(Z \geqslant 2|Y = 1, X \geqslant 4)$.

**Solution:** Here we are looking for probability that the number of high risk companies are greater than equal to two, given the there are four no risk companies and one moderate risk company. First we need to filter our joint probability mass function and then count how many of the cases have two or more high risk companies.

```
> p1 = with(filter(jpmf2,Y==1&X>=4&Z>=2),sum(p))
> p2 = with(filter(jpmf2,Y==1&X>=4),sum(p))
> pout = p1/p2
> print(pout)

[1] 0.9000361
```

(d) Determine $P(Y \leqslant 2, Z \leqslant 1 | X = 5)$.

**Solution:** Here we are interested in cases where

```
> p3 = with(filter(jpmf2,Y<=2&Z<=1&X==5),sum(p))
> p4 = with(filter(jpmf2,X==5),sum(p))
> pout2 = p3/p4
> print(pout2)

[1] 0.3542581
```

(e) Determine $E[Z|X = 5]$.

```
> Z2=filter(jpmf2,X==5)
> with(Z2,sum(Z*p))

[1] 0.0007081998
```