Written Assignment 06: Joint probability distributions

Farnoosh Koleini

2022-11-28

Assignment Goal

The overarching goal of this assignment is to deepen your understanding of bivariate and multivariate joint probability distributions.

Instructions

Please type your response to a question right below the question text. Compile this document to generate PDF output. Upload your PDF document to MS Teams. In the preamble above, change `V.N. Gudivada` to your name.

Questions

Marginal Probability Density Function (PDF)

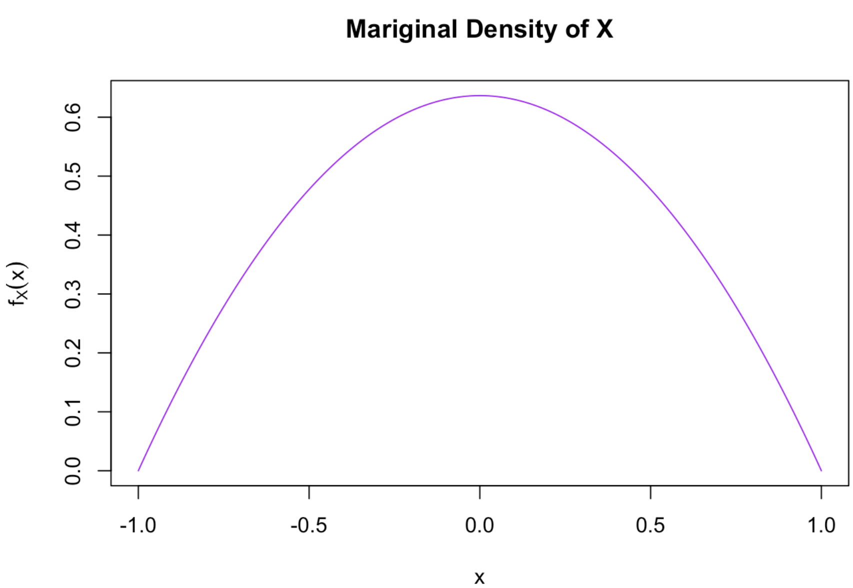
The random variables X and Y are uniformly distributed over the interior of a circle of radius 1 centered at the origin. Find the marginal probability density function $f_X(x)$.

Solution:

This question is about joint probability distribution. We have x and y axis. Create a circle and the origin on the (0,0). First we should find the joint probability density function. If you think about the area of the circle, it is going to be $r^2 $ and because r is 1 so area of the circle would be π. Therefore the density is $\frac{1}{area=\pi}$.

$f(x,y) = \frac{1}{\pi}$ for all area $x^2 + y^2 < 1$. This is a uniform density and every point in that area of the circle should follow this $area \times density = 1$. We need to find the marginal density, and for this we have to integrate out to get marginal density. $f_X(x) = \int \frac{dy}{\pi}$ and the upper and lower limit would be $-\sqrt{1-x^2}$ and $+\sqrt{1-x^2}$. The answer would be $ $ and for range $-1 < x < 1$. If you apply push in directions up and down. The marginal density of x would be -1 to 0 and 1 to 0. The point of this question is if you try to flat this cycle to special region.

```
library(latex2exp)
x <- seq(-1,1, by=0.01)
y <- (2*(1 - x^2)) / pi
plot(x,y, type="l", col="purple", main="Mariginal Density of X", xlab=TeX(r'($x$)'), ylab=TeX(r'($f_X(x)$)'))
```



■

Joint Cumulative Distribution Function (CDF)

Consider the random variables X and Y with joint PDF

$$f(x, y) = 2, \quad 0 < x < y < 1$$

Find $\mathbb{F}(x, y)$ for any point (x, y) in the support set $\mathcal{A}$. Use a calculus and a non-calculus (geometrical) approach to solve this problem.

Solution:

Here we need to find the cumulative distribution function. If we plot this special range, $0 < x < y < 1$ and the area where $x = y$, we will have $x < y$ and all other regions $x > y$. $f(x) = 2, 0 < x < y < 1$, to find CDF, we need to calculate $F_(x, y) = P(X \leq x, Y \leq y)$.We are gonna use double integral. If we want to do algebra first: $P(X \leq x, Y \leq y) = x(y - x) + \frac{1}{2}x \times x = 2[\frac{1}{2}x^2 + x(y - x)]$,

$F(0, 0) = 1/2(0) + 0(0 - 0) = 0$ So entire area should be equal to 1. $F(1, 1) = 1/2(1) + 1(1 - 1) = 1/2$.
$\int \int 2dtds = \int_0^x 2(y - 8)ds = 2xy - x^2$, $F(0, 0) = 0$, $F(1, 1) = 2 - 1 = 1$ If you go back and look at the geometric distribution, we have it.

```
x <- seq(0,1, by=0.01)
length(x)
```

```
## [1] 101
```

```
library(plotly)
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:latex2exp':
##
##     TeX
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

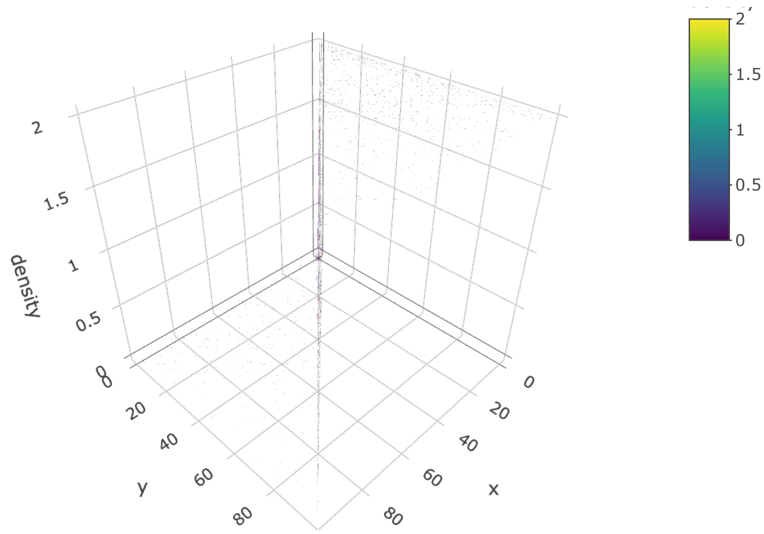
```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
x <- seq(0, 1, by=0.01)
# length(x)
y <- seq(0, 1, by=0.01)
# length(y)
density <- matrix(rep(0,10201), nrow=101, ncol=101)
# density
for(i in 1:(length(x)-1)){
  # print(i)
  for(j in (i+1):length(y)){
    density[j,i] <- 2
  }
}
print(density[2,3])
```

```
## [1] 0
```

```
biVarDensity <- plot_ly(z = ~density)
biVarDensity <- biVarDensity %>% add_surface()
biVarDensity
```

density



■

Joint Probability Density Function

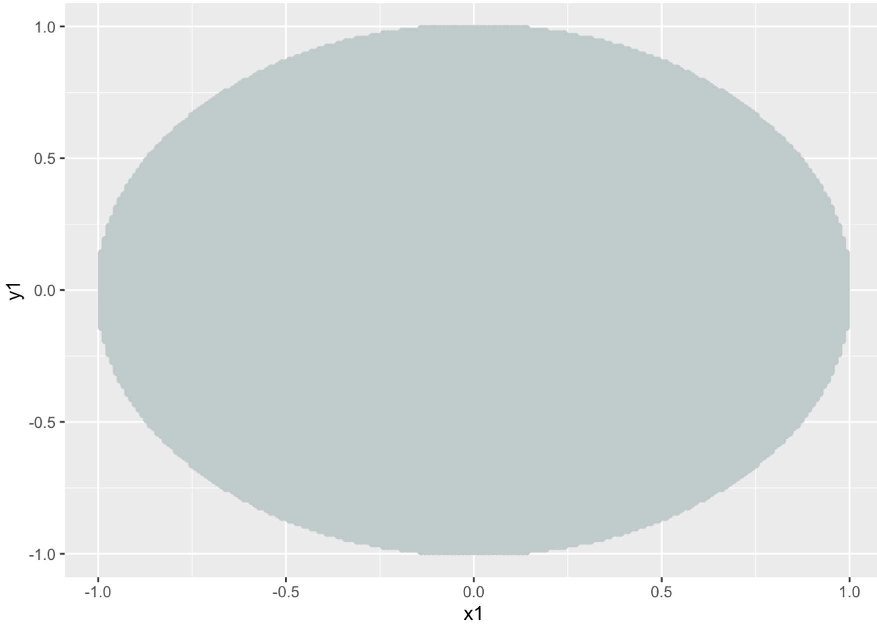
Let the random variables X and Y be uniformly distributed over the interior of the unit circle centered at $(0, 0)$.

a. Write the joint probability density function $f(x, y)$.

Solution:

We need to answer the exact value of x and y. The question wants us to find the joint probability density function. The support is less than 1. We have a unit circle here and $X^2 + Y^2 < 1$.

```
library(ggplot2)
x <- seq(-1, 1, by=0.01)
y <- seq(-1, 1, by=0.01)
x1 <- numeric()
y1 <- numeric()
for(i in 1:length(x)){
  for(j in 1:length(y)){
    if ( (x[i]^2 + y[j]^2) < 1){
      x1 <- c(x1, x[i])
      y1 <- c(y1, y[j])
    }
  }
}
# plot(x1,y1, col="azure3")
df <- data.frame(x1,y1)
ggplot(df, aes(x=x1, y=y1)) + geom_point(color="azure3")
```



b. Are X and Y independent?

The question here is we look at the product space and it should be a rectangle. X and Y are not independent.

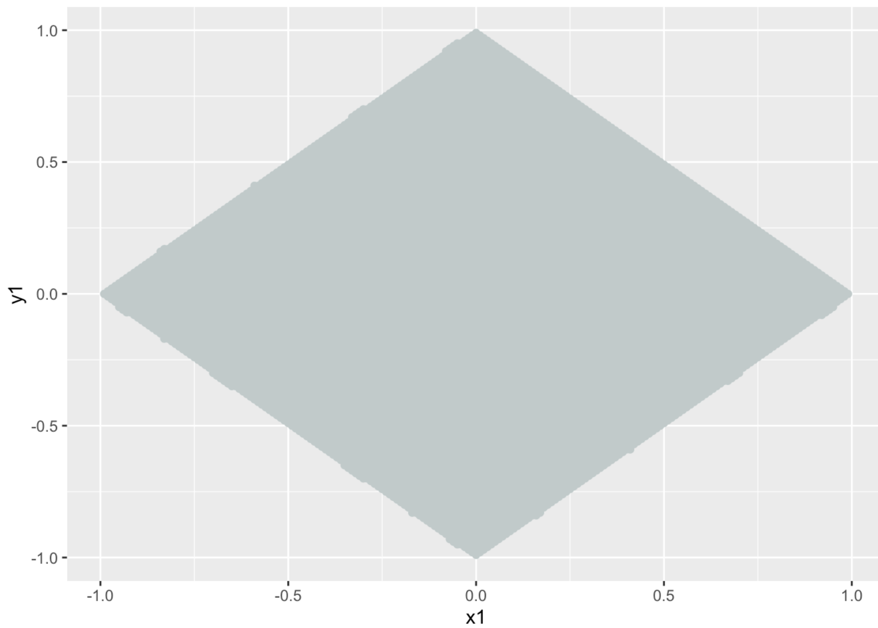
Solution:

c. Use no calculus to find the exact value of $\mathbb{P}(|X| + |Y| < 1)$.

Solution:

We need to know within this circle. Which area corresponds to absolute value of X and absolute value of Y is less than 1. You can actually see this is going to be a special area that we are looking at. This area is equal to 1. All the sites are kind of squared. Therefore, this probability area is $2 \times density = 1/\pi$

```
library(ggplot2)
x <- seq(-1, 1, by=0.01)
y <- seq(-1, 1, by=0.01)
x1 <- numeric()
y1 <- numeric()
for(i in 1:length(x)){
  for(j in 1:length(y)){
    if ( (abs(x[i]) + abs(y[j])) < 1){
      x1 <- c(x1, x[i])
      y1 <- c(y1, y[j])
    }
  }
}
# plot(x1,y1, col="azure3")
df <- data.frame(x1,y1)
ggplot(df, aes(x=x1, y=y1)) + geom_point(color="azure3")
```

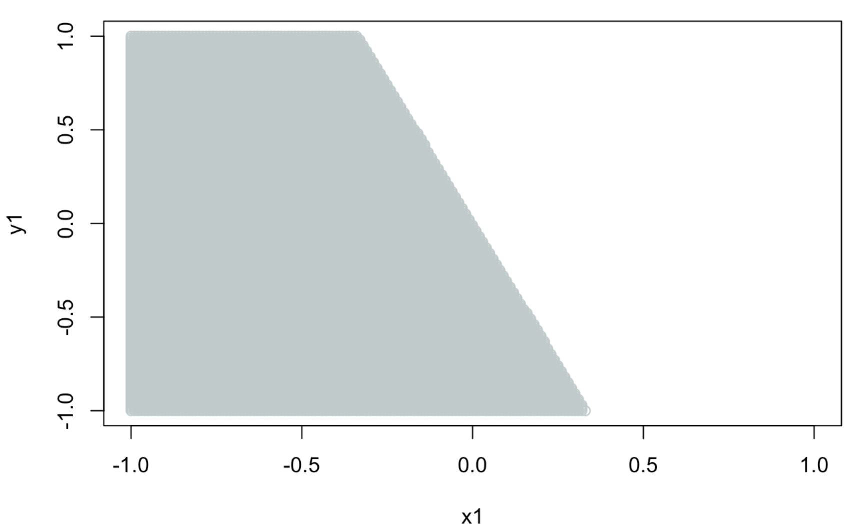


d. Use no calculus to find the exact value of $\mathbb{P}(3X + Y < 0)$.

Solution:

This probability would be an area, $\frac{1}{2}$. Now it is time to use some R codes to see the results. $\mathbb{P}(3X + Y < 0) = \pi/2 \times 1/\pi = 1/2$

```
library(ggplot2)
x <- seq(-1, 1, by=0.01)
y <- seq(-1, 1, by=0.01)
x1 <- numeric()
y1 <- numeric()
for(i in 1:length(x)){
  for(j in 1:length(y)){
    if ( (3*x[i] + y[j]) < 0){
      x1 <- c(x1, x[i])
      y1 <- c(y1, y[j])
    }
  }
}
plot(x1,y1, col="azure3", xlim=c(-1,1), ylim=c(-1,1))
```



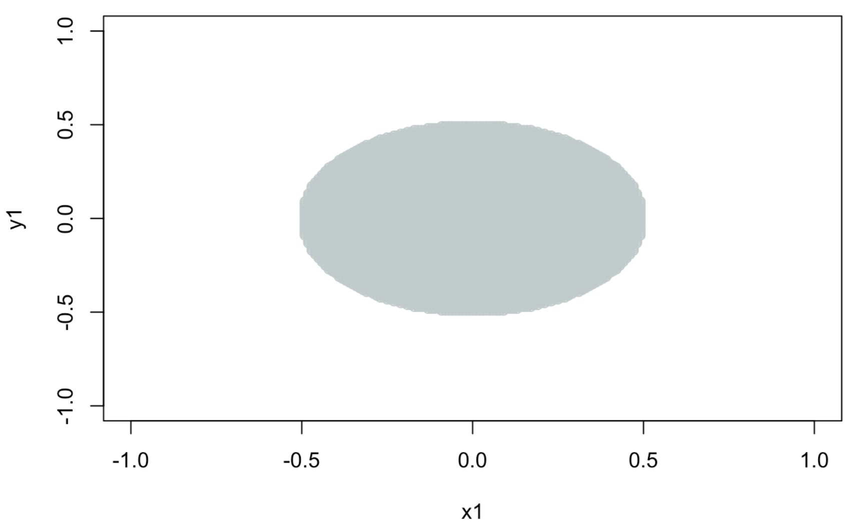
```
# df <- data.frame(x1,y1)
# ggplot(df, aes(x=x1, y=y1)) + geom_point(color="azure3")
```

e. Use no calculus to find the exact value of $\mathbb{P}\left(X^2 + Y^2 < 1/4\right)$.

Solution:

We have a unit circle which $x^2 + y^2 = 1$. $\mathbb{P}\left(X^2 + Y^2 < 1/4\right) = \pi(1/2)^2 = \pi/4$

```
library(ggplot2)
x <- seq(-1, 1, by=0.01)
y <- seq(-1, 1, by=0.01)
x1 <- numeric()
y1 <- numeric()
for(i in 1:length(x)){
  for(j in 1:length(y)){
    if ( (x[i]^2 + y[j]^2) < 1/4){
      x1 <- c(x1, x[i])
      y1 <- c(y1, y[j])
    }
  }
}
plot(x1,y1, col="azure3", xlim=c(-1,1), ylim=c(-1,1))
```




```
# df <- data.frame(x1,y1)
# ggplot(df, aes(x=x1, y=y1)) + geom_point(color="azure3")
```

■

Independent random variables

Let the independent random variables X and Y have marginal PMFs

$$f_X(x) = \frac{1}{2}, \quad x = 1, 2$$

and

$$f_Y(y) = \frac{1}{3}, \quad y = 1, 2, 3$$

Find $\mathbb{P}(X = Y)$

Solution:

We need to find the probability of x equal to 1. We need to compute the joint distribution of X and Y then we can compute the probability of X=Y. So the first thing we need to do here is dependent random variables.

$f_X(x) = 1/2$ where $X = 1, 2$ and the marginal distributions of y is also $f_X(x) = 1/3$ where $Y = 1, 2, 3$. Therefore we need to find the probability joint distribution, $f(x, y) = 1/2 \times 1/3$ where $X = 1, 2$ and $Y = 1, 2, 3$. $P(X = Y) = P(X = 1, Y = 1) + P(X = 2, Y = 2)$, because they are independent, the joint distribution can be written as a product of marginal distributions. This can be written as $P(X = 1). P(Y = 1) + P(X = 2). P(X = 2)$. We have the marginal distributions and the probability X takes the value 1 is $1/2$. and the probability Y takes the value 1 is $1/3$. So the result would be: $1/2 \times 1/3 + 1/2 \times 1/3 = 1/6 + 1/6 = 1/3$

■

Independence determination

Determine, by inspection, whether X and Y with distribution described by $f(x, y)$ given below, are independent or dependent random variables.

a. $f(x, y) = \frac{1}{27}\left(x^3 + y\right), \qquad x = 0, 1, 2; y = 1, 2$

Solution:

We have the joint distribution $f(x, y) = \frac{1}{27}\left(x^3 + y\right), \qquad x = 0, 1, 2; y = 1, 2$. Now we need to say whether X and Y are dependent or independent. There are two steps to test this. The first is looking at product of the space. If the product space is not a rectangle, we stop and say they are not independent. If the product space is rectangle, then we need to second check which is writing the joint distribution as a product of the two functions where the first one is the function of X only and where the second one is the function of Y on;y. In this case, is the support product of space? X takes the value of 0, 1.

$f(x, y) = g(x). b(y)$, $\frac{1}{27}(x^3 + y) = g(x). b(y)$, X and Y are not independent.

b. $f(x, y) = \frac{1}{40}x^2 y, \qquad (x, y) = (1, 1), (3, 1), (1, 3), (3, 3)$

Solution:

X takes the value 1 and 3 also Y takes the value 1 and 3. We can actually say $f(x, y) = \frac{1}{40}x^2 y = g(x)h(y)$ We have a function g which is the function of x only and y is the function of y only. So X and Y are independent.

c. $f(x, y) = e^{-x}, \qquad x > 0; 0 < y < 1$

Solution:

All positive value is x and if you look at y, y is between 0 and 1. So therefore we can say the support is the product space. Second test would be to write the joint probability function of g(x) and h(y). $f(x, y) = e^{-x} = g(x)h(y)$. So the first one is $g(x) = e^{-x}$ and the second one is $h(y) = 1$. So X and Y here are independent.

d. $f(x, y) = 1/2, \qquad |x| + |y| < 1$

Solution:

If we plot the function it would be a diamond. Support is not the product space so you do not need to go to the next step. X and Y are not independent.

e. $f(x, y) = 1/4, \qquad -1 < x < 1; -1 < y < 1$

Solution:

The support is the product space so we just need to check the second condition or step. The first function is 1/4 and the second function would be 1 no matter what the value is for y. So X and Y are independent.

■

Joint moment generating functions

Let the random variables X and Y have joint moment generating function:

$$M(t_1, t_2) = \frac{1}{2}e^{t_1+t_2} + \frac{1}{4}e^{2t_1+t_2} + \frac{1}{12}e^{t_2} + \frac{1}{6}e^{4t_1+3t_2}$$

for all real values of t_1 and t_2.

a. Find $\mathbb{V}[X]$

Solution:

$\mathbb{V}[X] = E[X^2] - (E[X])^2$.

$$M(t) = \frac{1}{2}e^{t} + \frac{1}{4}e^{2t} + \frac{1}{12} + \frac{1}{6}e^{4t}$$

Now we need to find the variance of X we can differentiation the moment generating function one time and set t equal to 0. First derivative would be

$$M_X'(t) = \frac{1}{2}e^t + \frac{2}{4}e^{2t} + \frac{4}{6}e^{4t}$$

Now we need to evaluate this in 0, this would be: $1/2 + 1/2 + 4/6 = 5/3$.

$$M_X''(t) = \frac{1}{2}e^t + \frac{4}{4}e^{2t} + \frac{16}{6}e^{4t}$$

Now we need to evaluate this based on t = 0, it would be $E[X] = 1/2 + 1 + 8/3 = 25/6$

$\mathbb{V}[X] = E[X^2] - (E[X])^2 = 25/6 - (5/3)^2 = 25/18 = 1.38$

b. Find $\mathbb{P}(X < Y)$

Solution: Here we need to find P(X < Y), what we can actually do is look at the joint moment generating function:

$$M(t_1, t_2) = \frac{1}{2}e^{t_1+t_2} + \frac{1}{4}e^{2t_1+t_2} + \frac{1}{12}e^{t_2} + \frac{1}{6}e^{4t_1+3t_2}$$

If We look at the function f(x,y), it would be 1/2 where x=1, y =1 and 1/4 where x = 2, y = 1 and then 1/12 where x = 0, y = 1 and lastly 1/6 where x= 4 and y =3. $\mathbb{P}(X < Y)$ just when happens that X=0 and Y = 1/12.

■

Variance

Let the continuous random variables X and Y be uniformly distributed over the triangular-shaped support region with vertices $(0, 2), (1, 0)$, and $(0, -1)$

a. Find $\mathbb{E}[X]$

We need to find the area of the triangle, base is 3 and height is 1. Therefore it is going to be $f(x, y) = 2/3$ where x > 0, y > x-1 and y < -2x +2. $(y - y_1) = m(x - x_1)$ so $m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{0+1}{1-0} = 1$ This has the positive slope. $(y + 1) = 1(x - x_1)$, $x = y + 1$ we can rewrite this as $y = x - 1$ and $y > x - 1$. Next one is finding the equation of the second line which is $y - 2 = -2(x - 0)$ and $y > -2x + 2$. We can define our support which is going to be all pairs x and y. $A = (x, y)|x > 0, y > x - 1, y < -2x + 2$ By integrating out we can say the joint density would be $f_X(x) = \int \frac{2}{3}dy$ in the range of x-1 and -2x+2. We calculated the marginal density for X now we can find the marginal density for Y as well. So, the first segment would be $f_Y(y)$ for range of -1< y< 0 it would be $\int 2/3dx$ and for range 0<y<2 it would be $\int 2/3dx$. Also we have $f_Y(y) = \frac{2(y+1)}{3}$ where -1 < y < 0 and $\frac{2-y}{3}$ where 0 < y < 2.

Solution:

$\int_0^1 2(1 - x)dx = x^2 - \frac{2x^3}{3} = \frac{1}{3}$

b. Find $\mathbb{V}[X]$

Solution:

$\mathbb{V}[X] = E[X^2] - (E[X])^2$
$E[X^2] = \int x^2.2(1 - x)dx = [\frac{2x^3}{3} - \frac{1x^4}{2}] = \frac{1}{6}$

$\mathbb{V}[X] = 1/6 - (1/3)^2 = 1/6 - 1/9 = 1/18$

c. Find ρ

Solution:

We have a lot of work to do. If we remember ρ is going to be $\rho = \frac{E[XY]-E[X]E[Y]}{\sqrt{V[X]}\sqrt{V[Y]}}$.

$E[Y] = \int_{-1}^{0} \frac{y\times2(y+1)}{3}dy + \int_0^2 y \times \frac{2-y}{3}dy = 2/9 - 1/3 + 4/3 - 8/9 = 1/3,$

$E[Y^2] = \int_{-1}^{0} \frac{y^2\times2(y+1)}{3}dy + \int_0^2 y^2 \times \frac{2-y}{3}dy = 1/2$

$\mathbb{V}[X] = E[X^2] - (E[X])^2 = 1/2 - 1/9 = 7/18,$

$E[XY] = \int\int xy\frac{2}{3}dydx = \int \frac{xy^2}{3}dx = \int_0^1 (x^3 - 2x^2 + x)dx = [\frac{x^4}{4} - \frac{2x^3}{3} + \frac{x^2}{2}] = 1/12$

$\rho = \frac{1/12-1/9}{\sqrt{7/18}} = -0.1890$

d. Find $\mathbb{V}[X - Y]$

Solution:

$V[X - Y] = E[(X - Y)^2] - (E[(X - Y)]) = E[X^2 - 2XY + Y^2] - (E[X] - E[Y])^2 = E[X^2] - 2E[XY] + E[Y^2] - (E[X] - E[Y])^2 = E[X^2] - 2E[XY] = 1$

e. Write a Monte Carlo simulation that supports your solution to part (d)

```
nrep <- 1000000
xx <- runif(nrep)
yy <- runif(nrep, -1,2)
index <- which(yy >xx - 1 & yy < 2 - 2*xx)
x <- xx[index]
y <- yy[index]
cat(sprintf("Mean of x: %f\n", mean(x)))
```

```
## Mean of x: 0.333710
```

```
cat(sprintf("Variance of x: %f\n", var(x)))
```

```
## Variance of x: 0.055545
```

```
cat(sprintf("cor(x, y): %f\n", cor(x, y)))
```

```
## cor(x, y): -0.187407
```

```
cat(sprintf("var(x - y): %f\n", var(x-y)))
```

```
## var(x - y): 0.498516
```

Solution:

■

Functions of random variables

A bag contains 5 balls numbered 1,2,3,4,5. Arno selects three balls at random and with replacement from the bag. The numbers on the balls selected are X_1, X_2, and X_3. Find the population mean and population variance of:

$$Y = X_1 + X_2 + X_3$$

Support your results by executing a Monte Carlo simulation experiment.

```
nrep <- 100000
y <- rep(0, nrep)
for(i in 1: nrep){
  x <- sample(5, 3, replace = TRUE)
  y[i] <- sum(x)
}
cat(sprintf("Mean of y: %f\n", mean(y)))
```

```
## Mean of y: 8.983590
```

```
cat(sprintf("Variance of y: %f\n", var(y)))
```

```
## Variance of y: 6.006741
```

Solution:

These are independent and identically distributed.(X1, X2, and X3 all are iid) Selecting the all 5 balls are equally likely.
$E[X_1] = \frac{a+b}{2} = \frac{1+5}{2} = 3 = E[X_2] = E[X_3]$, $V[X_1] = \frac{(b-a+1)^2-1}{12} = 2$, $V[X_1] = V[X_2] = V[X_3] = 2$,

$E[Y] = E[X_1 + X_2 + X_3] = E[X_1] + E[X_2] + E[X_3] = 3 + 3 + 3 = 9$,
$V[Y] = V[X_1 + X_2 + X_3] = V[X_1] + V[X_2] + V[X_3] = 2 + 2 + 2 = 6$

■

Variance-covariance matrix

Let the random variables X, Y and Z have a tri-variate probability distribution with population mean vector $\mu = (0, 1, 2)'$ and variance-covariance matrix:

$$\Sigma = \begin{bmatrix} 7 & 3 & 4 \\ 3 & 8 & 5 \\ 4 & 5 & 9 \end{bmatrix}$$

Let $W = 2X - Y + 3Z$.

a. Find $\mathbb{E}[W]$

Solution:

$E[W] = E[2X - Y + 3z] = E[2X] - E[Y] + E[3Z] = 2E[X] - E[Y] - 3E[Z] = 2(0) - 1 + 3 \times 2 = 5$

b. Find $\mathbb{V}[W]$

Solution:

$V[W] = V[2X - Y + 3Z] = V[2X] + V[Y] + Y[3Z] + 2Cov(2X - Y) + 2Cov(2X, 3Z) + 2Cov(8 - Y, 3Z) = 4V[X] + V[Y] + 9V[Z] - 4 \times Cov(X, Y) + 1$

■

Multivariate normal distribution

How many parameters are required to specify the n-dimensional multivariate normal distribution?

Solution:

We have n*n matrix also we have n elements on diagonal. This matrix is symmetric. First we need to get $\mu = \mu_1, \mu_2, \mu_3, \ldots$ Also we have variance co variance matrix. The number of parameters we need is, we need n population means and it is gonna be $\frac{n+n(n+1)}{2} = \frac{2n+n^2+n}{2} = \frac{n(n+3}{2}$,

There is a second approach here: n pop means and n pop variances. Also you have n number of random variables. The number of pairs is n choose from 2. Therefore you can sum up all of three. It is gonna be $ $. For bivariate the solution would be $ = 5 $ These 5 parameters are μ_1 , μ_2 , σ_1 , σ_2 , and ρ.

■