Exercise 4 - Random vector

The content of this script is only as a supplementary illustration to the exercise, it is not necessary to know at the exam. It is important to be able to calculate manually.

Martina Litschmannová, Adéla Vrtková, Michal Béreš

Example

Random vector $Z = (Y; X)^T$ has a probability function specified by the table </br>

$X \backslash Y$	1	2	3	4
3	0,01	0,02	0,03	0,25
5	0,04	0,16	?	0,05
7	0,12	0,07	0,06	0,01

a) Determine the missing value of the combined probability function,

In [1]:

```
data = c(0.01, 0.04, 0.12,
         0.02, 0.16, 0.07,
         0.03, 0,    0.06,
         0.25, 0.05, 0.01)
P = matrix(data , nrow=3, ncol=4) # possibly byrow=...
X = c(3, 5, 7)
Y = c(1, 2, 3, 4)
dimnames(P) = list(X,Y)
P
```

A matrix: 3 × 4 of type dbl

	1	2	3	4
3	0.01	0.02	0.03	0.25
5	0.04	0.16	0.00	0.05
7	0.12	0.07	0.06	0.01

In [2]:

```
sum(P)
```

0.82

In [3]:

```
# do not run this cell twice, otherwise you will set the value back to 0,
# Do you know why?
p_5_3 = 1 - sum(P)
P["5","3"] = p_5_3
P
```

A matrix: 3 × 4 of type dbl

	1	2	3	4
3	0.01	0.02	0.03	0.25
5	0.04	0.16	0.18	0.05
7	0.12	0.07	0.06	0.01

b) Specify the distribution function

Attention! The vector Z is $(Y, X)^T$ so the first parameter is the value Y and the second value X.

In [4]:

```
# F(2.8; 7.1)
# P(Y<2.8, X<7.1)
```

```
P[X<7.1, Y<2.8]
sum(P[X<7.1, Y<2.8])
```

A matrix: 3 × 2 of type dbl

	1	2
3	0.01	0.02
5	0.04	0.16
7	0.12	0.07

0.42

In [5]:

```
F = matrix(rep(0,4*5), nrow=4, ncol=5)
dimnames(F) = list(c('(-inf,3>', '(3,5>', '(5,7>', '(7,inf)'),
                   c('(-inf,1>', '(1,2>', '(2,3>', '(3,4>', '(4,inf)'))
F
```

A matrix: 4 × 5 of type dbl

	(-inf,1>	(1,2>	(2,3>	(3,4>	(4,inf)
(-inf,3>	0	0	0	0	0
(3,5>	0	0	0	0	0
(5,7>	0	0	0	0	0
(7,inf)	0	0	0	0	0

In [6]:

```
# we go through the rows and columns, we always take one value
# from the relevant row or column
x_vals = c(3,5,7,8)
y_vals = c(1,2,3,4,5)
for(i in 1:4){
    for(j in 1:5){
        x = x_vals[i]
        y = y_vals[j]
        F[i,j] = sum(P[X<x, Y<y])
    }
}
F
```

A matrix: 4 × 5 of type dbl

	(-inf,1>	(1,2>	(2,3>	(3,4>	(4,inf)
(-inf,3>	0	0.00	0.00	0.00	0.00
(3,5>	0	0.01	0.03	0.06	0.31
(5,7>	0	0.05	0.23	0.44	0.74
(7,inf)	0	0.17	0.42	0.69	1.00

c) Determine the marginal distribution

In [7]:

```
P
```

A matrix: 3 × 4 of type dbl

	1	2	3	4
3	0.01	0.02	0.03	0.25
5	0.04	0.16	0.18	0.05
7	0.12	0.07	0.06	0.01

In [8]:

```
P_x = rowSums(P)
P_x
```

3: 0.31 **5:** 0.43 **7:** 0.26

In [9]:

```
F_x = c(0, cumsum(P_x))
```

```
F_x
```

1: 0 **3:** 0.31 **5:** 0.74 **7:** 1

In [10]:
```
P_y = colSums(P)
P_y
```

1: 0.17 **2:** 0.25 **3:** 0.27 **4:** 0.31

In [11]:
```
F_y = c(0, cumsum(P_y))
F_y
```

1: 0 **1:** 0.17 **2:** 0.42 **3:** 0.69 **4:** 1

d) Conditional probabilities and conditional probability functions $P(x|y), P(y|x)$

In [12]:
```
# P(Y>2.1|X<5.3)
# P(Y>2.1 ∧ X<5.3)/P(X<5.3)
sum(P[X<5.3, Y>2.1])
sum(P[X<5.3,])
sum(P[X<5.3, Y>2.1])/sum(P[X<5.3,])
```

0.51

0.74

0.689189189189189

In [13]:
```
# P(X=5|Y=1)
# P(X=5 ∧ Y=1)/P(Y=1)
P['5','1']/sum(P[,'1'])
P['5','1']/sum(P_y['1'])
```

0.235294117647059

0.235294117647059

$$P(x|y) = \frac{P(X=x,Y=y)}{P_Y(y)}$$

In [14]:
```
P_xy = P # it's the same size, so we'll steal the formatting
X_lab = c('3', '5', '7')
Y_lab = c('1', '2', '3', '4')
for(x in X_lab){
    for(y in Y_lab){
        P_xy[x, y] = P[x, y]/P_y[y]
    }
}
P_xy
colSums(P_xy)
```

A matrix: 3 × 4 of type dbl

	1	2	3	4
3	0.05882353	0.08	0.1111111	0.80645161
5	0.23529412	0.64	0.6666667	0.16129032
7	0.70588235	0.28	0.2222222	0.03225806

1: 1 **2:** 1 **3:** 1 **4:** 1

$P(y|x)$

In [15]:
```
P_yx = P # it's the same size, so we'll steal the formatting
for(x in X_lab){
    for(y in Y_lab){
        P_yx[x, y] = P[x, y]/P_x[x]
    }
}
```

```
P_yx
rowSums(P_yx)
```

A matrix: 3 × 4 of type dbl

	1	2	3	4
3	0.03225806	0.06451613	0.09677419	0.80645161
5	0.09302326	0.37209302	0.41860465	0.11627907
7	0.46153846	0.26923077	0.23076923	0.03846154

3: 1 **5:** 1 **7:** 1

e) basic characteristics of random variables X and Y

In [16]:
```
E_X = sum(X*P_x)
E_X
E_XX = sum(X*X*P_x)
D_X = E_XX - E_X^2
D_X
```

4.9

2.27

In [17]:
```
E_Y = sum(Y*P_y)
E_Y
E_YY = sum(Y*Y*P_y)
D_Y = E_YY - E_Y^2
D_Y
```

2.72

1.1616

f) conditional mean E(X|Y=2)

In [18]:
```
# P(x|Y=2)
P_xy[,'2']
E_X_Y2 = sum(X*P_xy[,'2'])
E_X_Y2
```

3: 0.08 **5:** 0.64 **7:** 0.28

5.4

g) covariance and correlation

In [19]:
```
X_Y = P # matrix where in each column is the value x * y
for(x in X){
    for(y in Y){
        X_Y[toString(x), toString(y)] = x*y
    }
}
X_Y
```

A matrix: 3 × 4 of type dbl

	1	2	3	4
3	3	6	9	12
5	5	10	15	20
7	7	14	21	28

In [20]:
```
X_Y*P
```

A matrix: 3 × 4 of type dbl

	1	2	3	4

	1	2	3	4
3	0.03	0.12	0.27	3.00
5	0.20	1.60	2.70	1.00
7	0.84	0.98	1.26	0.28

In [21]:

```
# or we can use matrix multiplication
X %*% t(Y)
```

A matrix: 3 × 4 of type dbl

3	6	9	12
5	10	15	20
7	14	21	28

In [22]:

```
# mean value of E(X * Y)
E_XY = sum(X_Y*P)
E_XY
```

12.28

In [23]:

```
# covariance
cov_XY = E_XY-E_X*E_Y
cov_XY
```

-1.048

In [24]:

```
# correlation
cov_XY/sqrt(D_X*D_Y)
```

-0.64538676102769