Problems on random vectors (1/2)

Ignacio Cascos, Department of Statistics, Universidad Carlos III de Madrid

2018/19

1. Given the discrete random vector (X,Y) with the following probability mass function

$$p_{X,Y}(x,y)=\begin{cases} c(x^2+y^2) & \text{if } x\in\{1,2,4\},\ y\in\{1,3\} \\ 0 & \text{otherwise}\end{cases}.$$

   1. Compute the value of c. The value of the sum $\sum_{x\in\{1,2,4\}}\sum_{y\in\{1,3\}}(x^2+y^2)=72$, so $c=1/72$.
   2. Compute the probabilities: $P(Y<X)$, $P(Y>X)$, and $P(Y=X)$.

$$P(Y<X)=p_{X,Y}(2,1)+p_{X,Y}(4,1)+p_{X,Y}(4,3)=47/72$$

$$P(Y=X)=p_{X,Y}(1,1)=2/72$$

$$P(Y>X)=1-P(Y<X)-P(Y=X)=23/72.$$

   3. Compute $P(Y=3)$. $P(Y=3)=p_{X,Y}(1,3)+p_{X,Y}(2,3)+p_{X,Y}(4,3)=48/72=2/3$.
   4. Compute the marginal probability mass function $p_X(x)$ and $\mathbb{E}[X]$. $p_X(1)=12/72=1/6$, $p_X(2)=18/72=1/4$, $p_X(4)=42/72=7/12$ and $\mathbb{E}[X]=1\times 1/6+2\times 1/4+4\times 7/12=3$.

2. A dartboard is a wooden square of 40cm × 40cm. Inside the board there are several concentric circles, the greater among them having a diameter of 40cm. The darts of an unskilled player hit the board at a point that can be described by a random vector (X,Y) with joint density function

$$f_{X,Y}(x,y)=\begin{cases} 1/1600 & \text{if} & 0\le x,y\le 40 \\ 0 & \text{otherwise} & \end{cases},$$

where (X,Y) represent the coordinates (in cm) of the hitting point. The lower left corner of the board is represented by the origin, that is, the point (0,0).

   1. Whenever we hit one of the circles, we get some point. What is the probability of getting some point from a throw? The radius of the outer circle, represented by C is $R=40/2=20$, so

$$P((X,Y)\in C)=\iint_C f_{X,Y}(x,y)dxdy=\frac{1}{1600}\iint_C 1dxdx=\frac{\text{Area}(C)}{1600}=\frac{\pi R^2}{1600}=\frac{\pi}{4}.$$

   2. What is the mean vector of (X,Y)?, and the probability that (X,Y) assumes exactly that value?

$$\mathbb{E}[X]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{X,Y}(x,y)dydx=\int_0^{40}\frac{x}{1600}[y]_0^{40}dx=\infty_0^{40}\frac{x}{40}dx=20.$$

   Using the same arguments we show that $\mathbb{E}[Y]=20$, so the mean vector is $(20,20)^t$. Now, since (X,Y) is a continuous random vector, $P(X=20,Y=20)=0$.

   3. The inner circle has a diameter of 4cm. What is the probability of hitting it with a dart? The radius of the outer circle, represented by S is $r=4/2=2$, so

$$P((X,Y)\in S)=\iint_S f_{X,Y}(x,y)dxdy=\frac{1}{1600}\iint_S 1dxdx=\frac{\text{Area}(S)}{1600}=\frac{\pi r^2}{1600}=\frac{\pi}{400}.$$

3. Given the random variables X,Y,Z satisfying

$$X=3Y+2Z,\qquad \text{Var}[Y]=\text{Var}[Z]=1,\qquad \mathbb{E}[Y]=2,$$
$$\text{Cov}[Y,Z]=-0.5,\qquad \mathbb{E}[Z]=-3$$

Compute the following:

1. $\mathbb{E}[X]$. $\mathbb{E}[X] = \mathbb{E}[3Y + 2Z] = 3\mathbb{E}[Y] + 2\mathbb{E}[Z] = 0$.
2. $\mathrm{Var}[X]$. $\mathrm{Var}[X] = \mathrm{Var}[3Y + 2Z] = 3^2\mathrm{Var}[Y] + 2^2\mathrm{Var}[Z] + 2 \times 3 \times 2\mathrm{Cov}[Y, Z] = 7$.
3. $\mathrm{Cov}[X, Y]$. $\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[(3Y + 2Z)Y] = 3\mathbb{E}[Y^2] + 2\mathbb{E}[ZY] = 3\left(\mathrm{Var}[Y] + \mathbb{E}[Y]^2\right) + 2\left(\mathrm{Cov}[Y, Z] + \mathbb{E}[Y]\mathbb{E}[Z]\right) = 2$
4. $\rho_{X,Y}$, where ρ stands for correlation. $\rho_{X,Y} = \mathrm{Cov}[X, Y]/\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]} = 2/\sqrt{7}$

4. Data frame `cars` contains information on two numeric variables: `speed` (mph) and `dist` (feet).

   1. How many observations does `cars` contain? 50.
   2. What are the means of `speed` and `dist`? 15.4 and 42.98.
   3. What are the variances of `speed` and `dist`? 27.9591837 and 664.0608163.
   4. What are the covariance and correlation of `speed` and `dist`? 109.9469388 and 0.8068949.
   5. Assume the random vector (`speed`, `dist`) follows a bivariate normal distribution with mean as computed in 2. and covariance matrix as in 3. and 4. Compute (a) the probability that the distance to stop is greater than 80 feet, and (b) the probability that the distance minus twice the speed is less than 20. The random vector given by `speed` and `distance` follows a bivariate normal distribution with mean vector and covariance matrix

$$\mathrm{N}_2\left(\begin{pmatrix}15.4\\42.98\end{pmatrix}, \begin{pmatrix}27.9591837 & 109.9469388\\109.9469388 & 664.0608163\end{pmatrix}\right).$$

   The probability that the distance to stop is greater than 80 feet is the probability that the second component, X_2, which follows a $\mathrm{N}(42.98, \sqrt{664.0608163})$ distribution is greater than 80, $P(X_2 > 80) = 0.0754174$. Next we have to use the linear transformation $X_2 - 2X_1$ which is normally distributed with mean $\mathbb{E}[X_2 - 2X_1] = \mathbb{E}[X_2] - 2\mathbb{E}[X_1] = 12.18$ and variance $\mathrm{Var}[X_2 - 2X_1] = \mathrm{Var}[X_2] + 4\mathrm{Var}[X_1] - 4\mathrm{Cov}[X_1, X_2] = 336.1097959$. Finally we compute $P(X_2 - 2X_1 < 20) = 0.6651451$.
   6. Transform the speed to kilometers per hour and the distance to meters. What are the new mean vector and covariance matrix? The speed is multiplied times 1.60934 and the distance times 0.3048. Each mean is multiplied times the corresponding number, the variance times its square, and the covariances times their product. The resulting random vector follows distribution

$$\mathrm{N}_2\left(\begin{pmatrix}24.783836\\13.100304\end{pmatrix}, \begin{pmatrix}72.4135933 & 53.9319236\\53.9319236 & 61.6932686\end{pmatrix}\right).$$

```r
library(datasets)
attach(cars)
length(speed)
```

```
## [1] 50
```

```r
mean(speed)
```

```
## [1] 15.4
```

```r
mean(dist)
```

```
## [1] 42.98
```

```r
var(speed)
```

```
## [1] 27.95918
```

```r
var(dist)
```

```
## [1] 664.0608
```

```r
cov(speed,dist)
```

```
## [1] 109.9469
```

```r
cor(speed,dist)
```

```
## [1] 0.8068949
```

```r
1-pnorm(80,mean=mean(dist),sd=sd(dist))
```

```
## [1] 0.07541743
```

```r
pnorm(20,mean=mean(dist)-2*mean(speed),sd=sqrt(var(dist)+4*var(speed)-4*cov(dist,speed)))
```

```
## [1] 0.6651451
```

5. Four flatmates Alice, Bob, Charly, and Dave roll a 6-sided die every night in order to decide who washes the dishes after dinner. If the outcome is 1, Alice washes, it is 2, Bob does, while for 3 or 4 Charly must wash the dishes and for 5 or 6 it is Dave's turn.

    1. What is the probability that Alice and Bob wash the dishes once in a week (7 days) each, and Charly twice? We can use a multinomial random vector with four components, the first component corresponds to the number of days at which A washes the dishes in a wee, the second component to the number of days at which B washes the dishes, the third to the number of days at which C washes the dishes, and the fourth to the number of days at which D washes the dishes. The probability is 0.04801097.
    2. What is the probability that Alice and Bob wash the dishes once in a week each? We can use a multinomial random vector with three components, the first component corresponds to the number of days at which A washes the dishes in a wee, the second component to the number of days at which B washes the dishes, and the third to the number of days at which either of C or D wash the dishes. The probability is 0.1536351.

```r
dmultinom(c(1,1,2,3),size=7,prob=c(1/6,1/6,1/3,1/3))
```

```
## [1] 0.04801097
```

```r
dmultinom(c(1,1,5),size=7,prob=c(1/6,1/6,2/3))
```

```
## [1] 0.1536351
```