Practice 3

Carlo Cavicchia

ccavicchia@luiss.it

Probability: random vectors

- Two-dimensional random variables (mass, density, cdf)
- Independent random variables
- Generalization to n dimensions

Exercise 1

It is known that approximately 2% of women of age 45 have breast cancer. The result of a mammography test is said to be positive if the test detects the presence of breast cancer. In an hospital, the mammography screening test has the following performances:

- 90% sensibility: the test is positive (correctly) in 90% of women with breast cancer;
- 5% false positives: the test is positive (wrongly) in 5% of women who do not have breast cancer.

Exercise 1

a. A women has a positive mammography. What is the probability that she has breast cancer?

What do we know?...

Exercise 1

a A women has a positive mammography. What is the probability that she has breast cancer?

$$P(+|C) = 90\%$$
$$P(+|\bar{C}) = 5\%$$

Exercise 1

a A women has a positive mammography. What is the probability that she has breast cancer?

What do we want to know?...

Exercise 1

a A women has a positive mammography. What is the probability that she has breast cancer?

$$P(C|+) = \frac{P(+|C)P(C)}{P(+|C)P(C) + P(+|\bar{C})P(\bar{C})} = \frac{0.9 \times 0.02}{(0.9 \times 0.02) + (0.05 \times 0.98)} = 0.26$$

Exercise 1

b Last week a group of 25 45-years-old women resulted positive to the test in the hospital. Determine expected value and variance of the random variable X = “total number of women in the group with a breast cancer”.

What kind of random variable is X?

Exercise 1

b Last week a group of 25 45-years-old women resulted positive to the test in the hospital. Determine expected value and variance of the random variable X = “total number of women in the group with a breast cancer”.

What kind of random variable is X?

$$\text{Binomial}(25, P(C|+))$$

Exercise 1

b Last week a group of 25 45-years-old women resulted positive to the test in the hospital. Determine expected value and variance of the random variable X = “total number of women in the group with a breast cancer”.

So,

Exercise 1

b Last week a group of 25 45-years-old women resulted positive to the test in the hospital. Determine expected value and variance of the random variable X = “total number of women in the group with a breast cancer”.

So,

$$\mathrm{E}(X) = 25 \times 0.26 = 6.5$$

Exercise 1

b Last week a group of 25 45-years-old women resulted positive to the test in the hospital. Determine expected value and variance of the random variable X = “total number of women in the group with a breast cancer”.

So,

$$\mathrm{E}(X) = 25 \times 0.26 = 6.5$$

$$\mathrm{V}(X) = 25 \times 0.26 \times 0.74 = 4.81$$

Exercise 1

c During last decade a group of 15213 women, all aged 45, resulted positive to the test in the hospital. Determine the approximate probability that less than 1000 had the breast cancer.

Let us start from...

Binomial(n,p) $\rightarrow$ Normal(np,$np(1-p)$) if $n \rightarrow \infty$

Exercise 1

c During last decade a group of 15213 women, all aged 45, resulted positive to the test in the hospital. Determine the approximate probability that less than 1000 had the breast cancer.

Let us start from...

Binomial(n,p) $\rightarrow$ Normal(np,$np(1-p)$) if $n \rightarrow \infty$

$$P(X < 1000) = P(Z < \frac{X - np}{\sqrt{np(1-p)}}) = P(Z < \frac{1000 - 3955.38}{54.10}) = P(Z < -54,73)$$

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Is it a valid pdf?

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Is it a valid pdf?

$$\int_2^{\infty} \frac{c}{x^3} dx = \left[-\frac{c}{2x^2} \right]_2^{\infty} = \frac{c}{8}$$

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Is it a valid pdf? Derive c.

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Is it a valid pdf? Derive c.

$$\int_2^\infty \frac{c}{x^3} dx = \frac{c}{8} = 1 \rightarrow c = 8$$

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Compute the cdf of X.

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Compute the cdf of X.

$$P(X < x) = \int_2^x \frac{c}{x^3} dx = \int_2^x \frac{8}{x^3} dx = \left[-\frac{4}{x^2} \right]_2^x = 1 - \frac{4}{x^2}$$

So,

$$F_X(x) = \begin{cases} 1 - \frac{4}{x^2} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Compute the expected value of X.

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Compute the expected value of X.

$$\mathrm{E}(X) = \int_2^\infty x \frac{c}{x^3} dx = \int_2^\infty \frac{8}{x^2} dx = \left[-\frac{8}{x} \right]_2^\infty = 4$$

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Let $Y = 2X + 3$. What is the density of Y?

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Let $Y = 2X + 3$. What is the density of Y?

Let us start from the cdf:

$$F_Y(y) = P(Y < y) = P(2X + 3 < y) = P(X < \frac{y-3}{2}) = 1 - \frac{16}{(y-3)^2}$$

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Let $Y = 2X + 3$. What is the density of Y?

Let us start from the cdf:

$$F_Y(y) = P(Y < y) = P(2X + 3 < y) = P(X < \frac{y-3}{2}) = 1 - \frac{16}{(y-3)^2}$$

So,

$$f_Y(y) = F_Y'(y) = \frac{32}{(y-3)^3}$$

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Can you compute expected value of Y?

Exercise 2

Let X be a random variable with density:

$$f_X(x) = \begin{cases} \frac{c}{x^3} & x \geq 2 \\ 0 & x < 2 \end{cases}$$

- Can you compute expected value of Y?

$$\mathrm{E}(Y) = \int_7^{\infty} y \frac{32}{(y-3)^3} dy = \left[\frac{16(3-2y)}{(y-3)^2} \right]_7^{\infty} = 11$$

Exercise 3

Let consider the following joint distribution:

$X \mid Y$	2	4	6	8
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{4}$
2	$\frac{1}{4}$	$\frac{1}{12}$	0	$\frac{1}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{24}$	0

- Find the marginals for X and for Y.

Exercise 3

Let consider the following joint distribution:

$X\|Y$	2	4	6	8
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{4}$
2	$\frac{1}{4}$	$\frac{1}{12}$	0	$\frac{1}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{24}$	0

- Find the marginals for X and for Y.

For instance:

$$P(X=1) = P(X=1,Y=2)+P(X=1,Y=4)+P(X=1,Y=6)+P(X=1,Y=8)$$

$$= \frac{1}{12} + \frac{1}{12} + \frac{1}{24} + \frac{1}{4} = \frac{11}{24}$$

Exercise 3

Let consider the following joint distribution:

$X\|Y$	2	4	6	8	
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{4}$	$\frac{11}{24}$
2	$\frac{1}{4}$	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{5}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{24}$	0	$\frac{1}{8}$
	$\frac{5}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$	1

- Can you compute expected value of X and Y?

Exercise 3

Let consider the following joint distribution:

$X\|Y$	2	4	6	8	
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{4}$	$\frac{11}{24}$
2	$\frac{1}{4}$	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{5}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{24}$	0	$\frac{1}{8}$
	$\frac{5}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$	1

- Can you compute expected value of X and Y?

$$\mathrm{E}(X) = \sum_{x \in \mathcal{X}} xP(X = x) = 1 \times P(X = 1) + 2 \times P(X = 2) + 3 \times P(X = 3) = \frac{5}{3}$$

Exercise 3

Let consider the following joint distribution:

$X \mid Y$	2	4	6	8	
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{4}$	$\frac{11}{24}$
2	$\frac{1}{4}$	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{5}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{24}$	0	$\frac{1}{8}$
	$\frac{5}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$	1

- Can you compute expected value of X and Y?

Exercise 3

Let consider the following joint distribution:

$X \vert Y$	2	4	6	8	
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{4}$	$\frac{11}{24}$
2	$\frac{1}{4}$	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{5}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{24}$	0	$\frac{1}{8}$
	$\frac{5}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$	1

- Can you compute expected value of X and Y?

$$\mathrm{E}(Y) = \sum_{y \in \mathcal{Y}} y P(Y = y) = 2 \times P(Y = 2) + \cdots + 8 \times P(Y = 8) = \frac{14}{3}$$

Exercise 3

Let consider the following joint distribution:

$X \mid Y$	2	4	6	8	
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{4}$	$\frac{11}{24}$
2	$\frac{1}{4}$	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{5}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{24}$	0	$\frac{1}{8}$
	$\frac{5}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$	1

- Are X and Y correlated?

Exercise 3

Let consider the following joint distribution:

$X \mid Y$	2	4	6	8	
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{4}$	$\frac{11}{24}$
2	$\frac{1}{4}$	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{5}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{24}$	0	$\frac{1}{8}$
	$\frac{5}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$	1

- Are X and Y correlated?

$$\operatorname{Cov}(X,Y) = \sum_{(x,y)\in(\mathcal{X},\mathcal{Y})} (x - \operatorname{E}(X))(y - \operatorname{E}(Y)) P(X = x, Y = y) = \cdots = -\frac{7}{9}$$

Exercise 3

Let consider the following joint distribution:

$X \mid Y$	2	4	6	8	
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{4}$	$\frac{11}{24}$
2	$\frac{1}{4}$	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{5}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{24}$	0	$\frac{1}{8}$
	$\frac{5}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$	1

- Are X and Y correlated?

They are negatively correlated and they are not independent.

Exercise 4

The percentage of smokers in a college near Rome is 13%, meaning that if I pick a student at random, the probability that he/she smokes is $p = 0.13$. What kind of random variable can we use to represent the random experiment of assessing whether or not a student is a smoker?

Exercise 4

The percentage of smokers in a college near Rome is 13%, meaning that if I pick a student at random, the probability that he/she smokes is $p = 0.13$. What kind of random variable can we use to represent the random experiment of assessing whether or not a student is a smoker?

If we consider a “Success” picking an actual smoker, we can formalize the output of our “student picking experiment” as a Bernoulli random variable, X, with parameter $p = 0.13$. The random variable X can assume only two values, 1 if the student we pick is a smoker and 0 if he is not. The set $\mathcal{X} = \{0, 1\}$, containing all the values that X may take, is called the support of the random variable X.

Exercise 4

Compute the expected value of the random variable $X \sim \text{Bernoulli}(p)$

Exercise 4

Compute the expected value of the random variable $X \sim$ Bernoulli(p)

By definition of expected variable, we have that

$$\mathrm{E}(X) = \sum_{x \in \mathcal{X}} x P(X = x) = 0 \times (1 - p) + 1 \times p = p$$

Exercise 4

Let us now take a sample of $n = 100$ units from the student population and let us count how many are smokers. Define and describe the random variable that we could use to describe the results of this random experiment.

Exercise 4

Let us now take a sample of $n = 100$ units from the student population and let us count how many are smokers. Define and describe the random variable that we could use to describe the results of this random experiment.

The result of this experiment can be formalized as a Binomial random variable $Y \sim \text{Binomial}(n, p)$, which represent the number of successes in n independent Bernoulli trials $X_1, \dots, X_n$, each with the same probability of success, p. Each of the X_i is equal to 1 only when the student is smoking, hence we could also think of Y as the sum of the Bernoulli random variables, i.e. $Y = \sum_{i=1}^{n} X_i$. In our sample we could find any number of smokers, ranging from 0 (a super healty group) to n. The support of the random variable Y is thus $\mathcal{Y} = \{0, 1, \dots, n\}$. In our case we have that each of the 100 student is a different Bernoulli, hence $n = 100$ and that the probability of success for each of them is $p = 0.13$. Our random variable will thus be $Y \sim \text{Binomial}(n = 100, p = 0.13)$.

Exercise 4

Compute the expected value of the random variable $Y \sim$ Binomial(n, p).

Exercise 4

Compute the expected value of the random variable $Y \sim$ Binomial(n, p).

By definition of expected variable, we have that

$$\mathrm{E}(Y) = \sum_{y \in \mathcal{Y}} y P(Y = y) = \sum_{y=1}^{n} y \binom{n}{y} p^y (1-p)^{n-y} =$$

Exercise 4

Compute the expected value of the random variable $Y \sim$ Binomial(n, p).

By definition of expected variable, we have that

$$\mathrm{E}(Y) = \sum_{y \in \mathcal{Y}} y P(Y = y) = \sum_{y=1}^{n} y \binom{n}{y} p^y (1-p)^{n-y} =$$

but...

Exercise 4

Compute the expected value of the random variable $Y \sim$ Binomial(n, p).

$$\mathrm{E}(Y) = \sum_{y \in \mathcal{Y}} y P(Y = y) = \sum_{y=1}^{n} y \binom{n}{y} p^y (1-p)^{n-y}$$

There is also an easy way out: Y can be thought of as the sum of n independent Bernoulli random variables.

Exercise 4

Compute the expected value of the random variable $Y \sim \text{Binomial}(n, p)$.

$$\mathrm{E}(Y) = \sum_{y \in \mathcal{Y}} y P(Y = y) = \sum_{y=1}^{n} y \binom{n}{y} p^y (1-p)^{n-y}$$

There is also an easy way out: Y can be thought of as the sum of n independent Bernoulli random variables.

$$\mathrm{E}(Y) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} p = np$$

Exercise 4

For home:: Compute the probability that in the sample of $n = 100$ students there is at least one smoker.

Exercise 5

Let us consider $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$. X and Y are independent.

Exercise 5

Let us consider $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$. X and Y are independent.

- Compute the joint density.

Exercise 5

Let us consider $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$. X and Y are independent.

- Compute the joint density.

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) = 2e^{-(x+2y)} \quad \text{if} \quad x \geq 0, y \geq 0$$

Exercise 5

Let us consider $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$. X and Y are independent.

- Let us consider the triangle, T_t with vertices in $(0,0)$,$(0,t)$ and (t,t), where $t > 0$. Compute $P((X,Y) \in T_t)$.

Exercise 5

Let us consider $X \sim$ Exp(1) and $Y \sim$ Exp(2). X and Y are independent.

- Let us consider the triangle, T_t with vertices in $(0,0)$,$(0,t)$ and (t,t), where $t > 0$. Compute $P((X,Y) \in T_t)$.

$$P((X,Y) \in T_t) = \int_{T_t} f_{X,Y}(x,y)dxdy = 2\int_0^t e^{-2y}\left[\int_0^y e^{-x}dx\right]dy$$

$$= 2\int_0^t e^{-2y}\left[1 - e^{-y}\right]dy = \frac{1}{3} - e^{-2t} + \frac{2}{3}e^{-3t}$$

Exercise 5

Let us consider $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(2)$. X and Y are independent.

- Compute $P(X \leq Y)$.

Exercise 5

Let us consider $X \sim$ Exp(1) and $Y \sim$ Exp(2). X and Y are independent.

- Compute $P(X \leq Y)$.

$$P(X \leq Y) = 2\int_0^{\infty} e^{-2y}\left[\int_0^{y} e^{-x}dx\right]dy = 2\int_0^{\infty} e^{-2y}\left[1 - e^{-y}\right]dy = \frac{1}{3}$$