Practice 4

Carlo Cavicchia

ccavicchia@luiss.it

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries his luck and answer each question randomly. Compute the probabilities that:

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries his luck and answer each question randomly. Compute the probabilities that:

- he gets the 5^{th} question right

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries his luck and answer each question randomly. Compute the probabilities that:

- he gets the 5^{th} question right

First of all, let us denote by R_i the event "Carlo gets the i^{th} question right". Since he answers randomly, the probability of answering correctly is $1/4$ for each answer, regardless of the answer.

Hence:

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries his luck and answer each question randomly. Compute the probabilities that:

- he gets the 5^{th} question right

First of all, let us denote by R_i the event "Carlo gets the i^{th} question right". Since he answers randomly, the probability of answering correctly is $1/4$ for each answer, regardless of the answer.

Hence:

The probability that he gets the 5^{th} question right is $P(R_5) = 1/4$

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries his luck and answer each question randomly. Compute the probabilities that:

- he gets the 1^{st} question right

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries his luck and answer each question randomly. Compute the probabilities that:

- he gets the 1^{st} question right

The same...

The probability that he gets the 1^{st} question right is $P(R_1) = 1/4$

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries his luck and answer each question randomly. Compute the probabilities that:

- he gets all of the questions right

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries his luck and answer each question randomly. Compute the probabilities that:

- he gets all of the questions right

As he answers completely at random, each answer is not influenced by the other but they can be considered all independent trials, hence the probability of event that he gets all question right can be factorised:

$$P(R_1 \cap R_2 \cap R_3 \cap R_4 \cap R_5) = P(R_1) \times P(R_2) \times P(R_3) \times P(R_4) \times P(R_5) = \left(\frac{1}{4}\right)^5 \approx 0.0009$$

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries her luck and answer each question randomly. Compute the probabilities that:

- Describe the random variable T that may represent the "number of correct answers", and use it to compute the probability that she gets at least two questions right.

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries her luck and answer each question randomly. Compute the probabilities that:

- Describe the random variable T that may represent the "number of correct answers", and use it to compute the probability that she gets at least two questions right.

We are in a situation of $n = 5$ independent trials, each of them with the same probability of success $p = 1/4$, hence we can use the Binomial distribution to model the "number of correct answers of Carlo". In a more compact notation: $T \sim$ Binomial$(n = 5, p = 1/4)$.

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries her luck and answer each question randomly. Compute the probabilities that:

› Describe the random variable T that may represent the "number of correct answers", and use it to compute the probability that she gets at least two questions right.

We are in a situation of $n = 5$ independent trials, each of them with the same probability of success $p = 1/4$, hence we can use the Binomial distribution to model the "number of correct answers of Carlo". In a more compact notation: $T \sim \text{Binomial}(n = 5, p = 1/4)$.

In order to compute the probability:

$$P(T \geq 2) = 1 - P(T < 2) = 1 - P(T = 0) - P(T = 1) =$$

$$1 - \binom{5}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^5 - \binom{5}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^4$$

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries her luck and answer each question randomly. Compute the probabilities that:

- Compute the expected value of the number of correct answers given by Carlo

Exercise 1

In a multiple choice exam, there are 5 questions, each of them with 4 different choices (A,B,C,D). Carlo has decided that he is too cool for school, and he didn't study at all. The day of the exam, he tries her luck and answer each question randomly. Compute the probabilities that:

› Compute the expected value of the number of correct answers given by Carlo

As we have already known, the expected value for a Binomial variable with parameters n and p is $n \times p$, hence

$$\mathrm{E}(T) = 5 \times 1/4 = 1.25$$

which means that on average Carlo will get 1.25 questions right (in case you were wondering whether this is a valid strategy or not, well...).

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Y = 2X - 1$.

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Y = 2X - 1$.

- Determine the c.d.f. of Y.

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Y = 2X - 1$.

By looking at the c.d.f., we observe that the support of the random variable X is the interval $\mathcal{X} = [0, 2)$ as we have that:

- $P(X < 0) = 0$ which means that it is not possible to observe values smaller than 0
- $P(X \leq 2) = 1$, or analogously $P(X > 2) = 1 - P(X \leq 2) = 0$, which means that there is 0 chance of observing a value larger than 2 for X. We now proceed to determine the support Y of the transformation $\mathcal{Y}$. One easy way to do so is to study at the graph of the function $Y = h(X) = 2X - 1$.

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Y = 2X - 1$.

Thus, for each value of the X, this graph show us which value will Y take. We see clearly for X that varies in $[0, 2)$, the corresponding Y varies in $[-1, 3)$, and since Y can take any value between $[-1, 3)$, this is its support $\mathcal{Y}$.

Knowning that the support of Y is $[-1, 3)$ means that $P(Y < -1) = 0$ and that $P(Y \leq 3) = 1$, which allows us to write the first bit of its c.d.f.:

$$F_Y(y) = \begin{cases} 0 & y < -1 \\ ?? & -1 \leq y < 3 \\ 1 & y \geq 3 \end{cases}$$

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Y = 2X - 1$.

$$F_Y(y) = \begin{cases} 0 & y < -1 \\ ?? & -1 \leq y < 3 \\ 1 & y \geq 3 \end{cases}$$

The remaining part can be computed by noticing that

$$P(Y \leq y) = P(2X - 1 \leq y) = P(X \leq \frac{y+1}{2}) = F_X\left(\frac{y+1}{2}\right)$$

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Y = 2X - 1$.

Thus,

$$F_Y(y) = \begin{cases} 0 & y < -1 \\ \left(\frac{(y+1)^2/4}{4}\right) & -1 \le y < 3 \\ 1 & y \ge 3 \end{cases}$$

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Y = 2X - 1$.

- Determine the c.d.f. of the new random variable $Z = X^2 - Y$.

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Y = 2X - 1$.

- Determine the c.d.f. of the new random variable $Z = X^2 - Y$.

In order to avoid working with a joint c.d.f. for the vector (X, Y) we rewrite Z in terms of the variable X alone:

$$Z = X^2 - Y = X^2 - 2X + 1 = (X - 1)^2$$

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Z = (X - 1)^2$.

Again, we start by defining $\mathcal{Z}$, the support of Z. Plotting the graph of $Z = (X - 1)^2$, we can clearly see that Z can take any value between 0 and 1, hence its support is going to be $(0, 1)$, which already tells us something about the c.d.f., and more specifically:

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ ?? & 0 \leq z < 1 \\ 1 & z \geq 1 \end{cases}$$

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Z = (X-1)^2$.

In order to fill in the last bit:

$$P(Z \leq z) = P((X-1)^2 \leq z) = P(-\sqrt{z}+1 \leq X \leq \sqrt{z}+1) =$$
$$F_X(\sqrt{z}+1) - F_X(-\sqrt{z}+1) = \left(\frac{1+\sqrt{z}}{2}\right)^2 - \left(\frac{1-\sqrt{z}}{2}\right)^2 =$$
$$\frac{1}{4}\left(1+2\sqrt{z}+z-(1-2\sqrt{z}+z)\right) = \sqrt{z}$$

Exercise 2

Consider the random variable X whose cumulative distribution function $F_X(x)$ is defined as follows:

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

and consider the new variable $Z = (X-1)^2$.

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ \sqrt{z} & 0 \leq z < 1 \\ 1 & z \geq 1 \end{cases}$$

Exercise 3

Consider the random variable X uniformly distributed between -1 and 1.

Exercise 3

Consider the random variable X uniformly distributed between -1 and 1.

- Determine the p.d.f and c.d.f. of X and of $Y = X^2$.

Exercise 3

Consider the random variable X uniformly distributed between -1 and 1.

- Determine the p.d.f and c.d.f. of X and of $Y = X^2$.

$$f_X(x) = \begin{cases} 1/2 & \text{if } -1 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Exercise 3

Consider the random variable X uniformly distributed between -1 and 1.

- Determine the p.d.f and c.d.f. of X and of $Y = X^2$.

$$f_X(x) = \begin{cases} 1/2 & \text{if } -1 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

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

Exercise 3

Consider the random variable X uniformly distributed between -1 and 1.

- Determine the p.d.f and c.d.f. of X and of $Y = X^2$.

To compute the c.d.f. of Y, let us start from the extremes of $\mathcal{Y}$: plotting the graph of $Y = X^2$, we can clearly see that Y can take any value between 0 and 1, hence its support is going to be $(0, 1)$, which already tells us something about the c.d.f., and more specifically:

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ ?? & 0 \leq y < 1 \\ 1 & y \geq 1 \end{cases}$$

Exercise 3

Consider the random variable X uniformly distributed between -1 and 1.

- Determine the p.d.f and c.d.f. of X and of $Y = X^2$.

$$P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \sqrt{y}$$

Exercise 3

Consider the random variable X uniformly distributed between -1 and 1.

- Determine the p.d.f and c.d.f. of X and of $Y = X^2$.

$$P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \sqrt{y}$$

thus,

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \sqrt{y} & 0 \leq y < 1 \\ 1 & y \geq 1 \end{cases}$$

Exercise 3

Consider the random variable X uniformly distributed between -1 and 1.

- Determine the p.d.f and c.d.f. of X and of $Y = X^2$.

Now,

$$f_Y(y) = F'_Y(y) = \frac{1}{2\sqrt{(y)}}$$

Exercise 3

Consider the random variable X uniformly distributed between -1 and 1.

- Determine the p.d.f and c.d.f. of X and of $Y = X^2$.

Now,

$$f_Y(y) = F'_Y(y) = \frac{1}{2\sqrt{(y)}}$$

so,

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2\sqrt{(y)}} & 0 \leq y < 1 \\ 1 & y \geq 1 \end{cases}$$

Exercise 4

Compute expected value and varianca of $U \sim \text{Unif}(a, b)$.

Exercise 4

Compute expected value and varianca of $U \sim$ Unif(a, b).

$$\mathrm{E}(U) = \int_a^b u \frac{1}{b-a} du = \left[\frac{u^2}{2(b-a)} \right]_a^b = \frac{a^2 + b^2}{2(b-a)} = \frac{a+b}{2}$$

Exercise 4

Compute expected value and varianca of $U \sim$ Unif(a, b).

$$\mathrm{E}(U) = \int_a^b u \frac{1}{b-a} du = \left[\frac{u^2}{2(b-a)} \right]_a^b = \frac{a^2+b^2}{2(b-a)} = \frac{a+b}{2}$$

$$\mathrm{V}(U) = \int_a^b (u - \mathrm{E}(U))^2 \frac{1}{b-a} du = \left[\frac{u^3}{3(b-a)} \right]_a^b - \left[\frac{u^2(a+b)}{2(b-a)} \right]_a^b + \left[\frac{u(a+b)^2}{4(b-a)} \right]_a^b =$$
$$\frac{b^3}{3(b-a)} - \frac{a^3}{3(b-a)} - \frac{b^2(a+b)}{2(b-a)} + \frac{a^2(a+b)}{2(b-a)} + \frac{b(a+b)^2}{4(b-a)} - \frac{a(a+b)^2}{4(b-a)} =$$
$$\frac{1}{12(b-a)} (b^3 - a^3 + 3a^2b - 3ab^2) = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}$$

Exercise 4

Compute expected value and varianca of $U \sim$ Unif(a, b).

Alternative way to compute the variance:

$$\mathrm{V}(U) = \int_a^b u^2 \frac{1}{b-a} du - [\mathrm{E}(U)]^2 = \left[\frac{u^3}{3(b-a)}\right]_a^b - \frac{(a+b)^2}{4} =$$

$$\frac{1}{12}\left[\frac{4(b^3-a^3)}{b-a} - 3(a+b)^2\right] = \frac{1}{12}\left[\frac{4(b-a)(b^2+ab+a^2)}{b-a} - 3(a^2+2ab+b^2)\right] =$$

$$\frac{1}{12}\left(4b^2+4ab+4a^2-3a^2-6ab-3b^2\right) = \frac{(b-a)^2}{12}$$

Exercise 5

Neuroblastoma, a rare form of malignant tumor, occurs in 11 out of a million children. In the 12439 children of Oak Park, Illinois, 4 cases of neuroblastoma are reported. Assumming there was nothing special about Oak Park where the chance of neuroblastoma is higher than normal, find the probability of seeing more than a single case of neuroblastoma in a population this size. What can you likely conclude?

Exercise 5

Neuroblastoma, a rare form of malignant tumor, occurs in 11 out of a million children. In the 12439 children of Oak Park, Illinois, 4 cases of neuroblastoma are reported. Assumming there was nothing special about Oak Park where the chance of neuroblastoma is higher than normal, find the probability of seeing more than a single case of neuroblastoma in a population this size. What can you likely conclude?

Technically this is a binomial probability problem.

Exercise 5

Neuroblastoma, a rare form of malignant tumor, occurs in 11 out of a million children. In the 12439 children of Oak Park, Illinois, 4 cases of neuroblastoma are reported. Assumming there was nothing special about Oak Park where the chance of neuroblastoma is higher than normal, find the probability of seeing more than a single case of neuroblastoma in a population this size. What can you likely conclude?

However, we can also get a very good approximation of this probability using the Poisson distribution. Why?

Exercise 5

Neuroblastoma, a rare form of malignant tumor, occurs in 11 out of a million children. In the 12439 children of Oak Park, Illinois, 4 cases of neuroblastoma are reported. Assumming there was nothing special about Oak Park where the chance of neuroblastoma is higher than normal, find the probability of seeing more than a single case of neuroblastoma in a population this size. What can you likely conclude?

However, we can also get a very good approximation of this probability using the Poisson distribution. Why?

If $n \to \infty$ and $p \to 0$: Binom$(n, p) \to$ Poiss$(\lambda = np)$.

Exercise 5

Neuroblastoma, a rare form of malignant tumor, occurs in 11 out of a million children. In the 12439 children of Oak Park, Illinois, 4 cases of neuroblastoma are reported. Assumming there was nothing special about Oak Park where the chance of neuroblastoma is higher than normal, find the probability of seeing more than a single case of neuroblastoma in a population this size. What can you likely conclude?

Thus,

$$\lambda = np = 12439 \times \frac{11}{1000000} \approx 0.136829$$

Exercise 5

Neuroblastoma, a rare form of malignant tumor, occurs in 11 out of a million children. In the 12439 children of Oak Park, Illinois, 4 cases of neuroblastoma are reported. Assumming there was nothing special about Oak Park where the chance of neuroblastoma is higher than normal, find the probability of seeing more than a single case of neuroblastoma in a population this size. What can you likely conclude?

Thus,

$$\lambda = np = 12439 \times \frac{11}{1000000} \approx 0.136829$$

Then we use the complement to find $P(X > 1) = 1 - (P(0) + P(1))$:

$$1 - \frac{e^{-0.136829}(0.136829)^0}{0!} - \frac{e^{-0.136829}(0.136829)^1}{1!} \approx 0.00855$$

Exercise 5

Neuroblastoma, a rare form of malignant tumor, occurs in 11 out of a million children. In the 12439 children of Oak Park, Illinois, 4 cases of neuroblastoma are reported. Assumming there was nothing special about Oak Park where the chance of neuroblastoma is higher than normal, find the probability of seeing more than a single case of neuroblastoma in a population this size. What can you likely conclude?

As the probabilities found are so ridiculously small, we conclude (with a significant amount of certainty) there is something special about Oak Park that makes the chance of neuroblastoma higher than 11 out of a million.

Exercise 5

Neuroblastoma, a rare form of malignant tumor, occurs in 11 out of a million children. In the 12439 children of Oak Park, Illinois, 4 cases of neuroblastoma are reported. Assumming there was nothing special about Oak Park where the chance of neuroblastoma is higher than normal, find the probability of seeing more than a single case of neuroblastoma in a population this size. What can you likely conclude?

Remark:

That probability, computed with the Binom, results: 0.00854887.

Exercise 5

Neuroblastoma, a rare form of malignant tumor, occurs in 11 out of a million children. In the 12439 children of Oak Park, Illinois, 4 cases of neuroblastoma are reported. Assumming there was nothing special about Oak Park where the chance of neuroblastoma is higher than normal, find the probability of seeing more than a single case of neuroblastoma in a population this size. What can you likely conclude?

Remark:

That probability, computed with the Binom, results: 0.00854887.

Note how accurate the Poisson approximation was in this circumstance!