EPFL, Spring 2023
Guillaume Obozinski

EE-209 Eléments de Statistiques pour les Data Sciences

Feuille d'exercices 2

Exercise 2.1 Variance of the sum

(a) We consider a pair of random variables (X, Y) with X taking values in $\mathcal{X}$ and Y taking values in $\mathcal{Y}$ and with joint pmf $P_{(X,Y)}(x, y)$. Prove that

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + 2\,\mathrm{cov}(X, Y) + \mathrm{Var}(Y).$$

(b) Deduce from the previous question that if X and Y are independent, then

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Exercise 2.2 Law of total expectation

(a) We consider a pair of random variables (X, Y) with X taking values in $\mathcal{X}$ and Y taking values in $\mathcal{Y}$ and with joint pmf $P_{(X,Y)}(x, y)$. Let f be a function defined on $\mathcal{X}$. Prove that

$$\sum_{y\in\mathcal{Y}} \mathbb{E}[\, f(X) \mid Y = y\,]\, P_Y(y) = \mathbb{E}[f(X)].$$

This formula is called the *law of total expectation.*

(b) Note that $\mathbb{E}[f(X) \mid Y = y]$ takes a value that depends only on y. If we call this function h then we have $h(y) = \mathbb{E}[f(X) \mid Y = y]$. Sometimes it is useful to consider the random variable $h(Y)$. The convention is to write $\mathbb{E}[f(X)|Y]$ for $h(Y)$. With this notation note that by definition $\mathbb{E}\Big[\mathbb{E}[f(X)|Y]\Big] := \sum_{y\in\mathcal{Y}} \mathbb{E}[f(X)|Y = y]P_Y(y)$. So the *total expectation* formula from the previous question can be rewritten:

$$\mathbb{E}[\,\mathbb{E}[f(X)|Y]\,] = \mathbb{E}[f(X)].$$

It basically says that if you first compute a conditional expectation of $f(X)$ given something and recompute the expectation of what you obtained, you should obtain the same value as if you compute directly the expectation of $f(X)$. We will apply the *law of total expectation* in an exercise next week.

Exercise 2.3 Density, mean and mode of a continuous random variable We assume that the continuous random variable X has the probability density function (pdf)

$$f_X(x) = \begin{cases} c \times (1 - x^2) & x \in [-1, 1] \\ 0 & \text{otherwise.} \end{cases}$$

for some constant c.

(a) Find the constant c and verify that the function is a well-defined density.

(b) Draw the pdf and the cumulative distribution function (cdf). What is the maximum of the pdf? And for the cdf? What is the mode of the distribution?

(c) Compute the expectation of X. More generally, what is the expectation of a variable that is symmetric around 0? *(Hint: a symmetric function around 0 satisfy $f(x) = f(-x)$.)*

(d) Compute the variance of X.

Exercise 2.4 "Poisson" in the forest !?

In a forest, there are K species of animals. We assume that during one day the number of animals C_k of species k that a biologist who is expert of the fauna of this forest will observe follows a Poisson distribution with parameter $\lambda_k > 0$. Note that the Poisson rate λ_k can be interpreted as the mean frequency at which the animal can be observed.

We denote by:

- C_k the random variable giving the number of animals of species k the biologist got to see during a given day,
- N the random variable giving the total number of animals that the biologist got to see during that same day, so that $N = C_1 + \ldots + C_K$,

(a) What is the distribution of the random variable N?

(b) What is the distribution of $(C_1, \ldots, C_K)$ conditional to the event $\{N = n\}$?

(c) In particular, if we assume that the biologist only saw a single animal, what is the probability that it is from species k?