EPFL, Spring 2023
Guillaume Obozinski

EE-209 Eléments de Statistiques pour les Data Sciences

Feuille d'exercices 11

Exercise 11.1 We consider a pair of random variables X_1 and X_2 that follow jointly a multivariate Gaussian distribution, with joint probability density function:

$$p_{(X_1,X_2)}(x_1,x_2) = \frac{1}{2\pi\sqrt{\det(\mathbf{\Sigma})}} \exp\left(-\frac{1}{2}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}^{\top} \mathbf{\Sigma}^{-1} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} \right)$$

which is parameterized by a vector $(\mu_1,\mu_2)^\top$ and by a symmetric 2 by 2 matrix

$$\mathbf{\Sigma} := \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

It is possible to show that these parameters are such that:

$$\mathbb{E}[X_1] = \mu_1, \quad \mathbb{E}[X_2] = \mu_2, \quad \mathrm{Var}(X_1) = \sigma_1^2, \quad \mathrm{Var}(X_2) = \sigma_2^2, \quad \mathrm{cov}(X_1, X_2) = \sigma_{12}.$$

You can use all these identities in this exercise.

(a) Prove that for the joint probability distribution above if X_1 and X_2 are decorrelated, in the sense that $\sigma_{12} = 0$, then X_1 and X_2 are also independent.

(b) Show that the random variable $\tilde{X}_2 := X_2 - \frac{\sigma_{12}}{\sigma_1^2} X_1$ is decorrelated from X_1.

(c) Note that we can write $\begin{pmatrix} X_1 \\ \tilde{X}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{\sigma_{12}}{\sigma_1^2} & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$,

and for a multivariate Gaussian distribution a key property is that any linear transformation of a Gaussian vector (i.e., multiplication by a matrix) still produces a Gaussian vector, i.e. a vector of random variables following a joint Gaussian distribution. So, as a consequence, $(X_1, \tilde{X}_2)^\top$ follows a joint Gaussian distribution. Deduce from this and from the previous questions that X_1 and $\tilde{X}_2$ are independent.

(d) Deduce from the previous questions the values of $\mathbb{E}[\tilde{X}_2 \mid X_1]$ and $\mathbb{E}[X_2 \mid X_1]$.

(e) What can you say about the form of $\mathbb{E}[X_2|X_1]$? Can you make a connection with the simple linear regression?

(f) Using the same ideas is in the previous questions, compute $\mathrm{Var}(X_2|X_1)$.

(g) Is there a parallel that can be made between this conditional variance and the *variance of the residuals* in simple linear regression? To answer this question, you can consider the quantity $\varepsilon := X_2 - f(X_1)$ with $f(X_1) := \mathbb{E}[X_2|X_1] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X_1 - \mu_1)$, and compute $\mathrm{Var}(\varepsilon|X_1)$ and $\mathrm{Var}(\varepsilon)$.

(h) Show that ε and X_1 are independent. Given that $X_2 = f(X_1) + \varepsilon$, can you give an interpretation of ε?

Exercise 11.2 Assume that we have some noisy measurements $\{(x_1, z_1), \ldots, (x_n, z_n)\}$ where $x_i \in \mathbb{R}$ and $z_i \in \mathbb{R}$ that should ideally be on a circle, so that they should approximately satisfy the equation $(x - x_0)^2 + (z - z_0)^2 - c^2 = 0$ for some unknown values (x_0, z_0) and c corresponding to the center and the radius of the circle.

(a) Show that for any (x, z) on the circle, the quantity defined as $y := x^2 + z^2$ is an affine function of (x, z)

(b) Based on the answer to the previous question, explain how it is possible to cast the problem of estimating x_0, z_0 and c as a linear regression of the responses y_i on some vector of explanatory variables. In particular, state which explanatory variables you would consider, what is the design matrix, what are the parameters of the linear regression models and how they relate to the parameters we wish to estimate.

Exercise 11.3 ($*$) Identical twins... and Bayesian reasoning

Mary is pregnant and expecting twins. She would like to know if she will give birth to identical twins (vrai jumeaux) or fraternal twins (faux jumeaux). Identical twins have the same sex and so if they are identical twins, the probability that they are both boys is $\frac{1}{2}$ and the probability that they are both girls is $\frac{1}{2}$. On the other hand, if they are fraternal twins there is no correlation between the genders of the babies which are each independently a boy or a girl with probability $\frac{1}{2}$. In that case, the probability that they are both boys is $\frac{1}{4}$ and the probability that they are both girls is $\frac{1}{4}$. In the population, $\frac{1}{3}$ of twins are identical twins and $\frac{2}{3}$ are fraternal twins. Mary goes and visits her doctor and she does a sonogram which reveals that she is expecting two boys...

(a) Use Bayes' rule to determine what is the probability that Mary is expecting identical twins given the information provided by the sonogram, and the rest of the information provided.

(b) Mary does this calculation, and goes to get coffee with two friends, Joe and Brad, with whom she is sharing these news. Joe expresses his enthusiasm: "It's great that thanks to the use of Bayesian statistical inference, it is possible to combine prior and evidence and to simply apply Bayes' rule and obtain updated estimates of your chances of having identical twins !". Brad, who does not share the enthusiasm of Joe, comments immediately: "Well, the reason why Bayesian statistics work here, is because you have the chance to have a perfectly informative prior: you know what proportion of the twins are identical in the population. But in general in statistics you don't have such an informative prior. For example, here, if you did not know what was the proportion of twins which is identical, you would have to treat it as unknown, and to put a prior on the probability π for twins to be identical; for example, you could use a uniform prior for π on $[0, 1]$ and I bet you that you would not find the same answer!". Check if Brad is correct, and compute what is the probability that Mary is expecting identical twins, given the information provided by the sonogram, but this time assuming a uniform prior over the probability π on $[0, 1]$. Compare the two results. What was the prior on π that Mary used?

(c) Was Mary actually using Bayesian statistics in her initial computation as suggested by the comment of Joe?