ENSIMAG – Grenoble INP – UGA
Introduction to Statistical Learning and Applications
Pedro L. C. Rodrigues
Alexandre Wendling

Year 2024-2025

pedro.rodrigues@inria.fr
alexandre.wendling@univ-grenoble-alpes.fr

TD 1: Exercises on multivariate statistics and regression

▶ Exercise 1

Let U and V be two independent random variables with uniform distribution over $[0, 1]$.

Let $X = U + V$ and $Y = U - V$.

(a) Compute the expectation and covariance matrix of $Z = \begin{pmatrix} X & Y \end{pmatrix}^T$.

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \Rightarrow \mathbb{E}[Z] = \begin{bmatrix} \mathbb{E}[X] \\ \mathbb{E}[Y] \end{bmatrix} = \begin{bmatrix} \mathbb{E}[U+V] \\ \mathbb{E}[U-V] \end{bmatrix} = \begin{bmatrix} \mathbb{E}[U] + \mathbb{E}[V] \\ \mathbb{E}[U] - \mathbb{E}[V] \end{bmatrix}$$

Remember that

$$\mathbb{E}[U] = \int_{\mathbb{R}} u \; p_U(u) \mathrm{d}u = \int_0^1 u \; \mathrm{d}u = \frac{1}{2} = \mathbb{E}[V] \quad \text{therefore} \quad \mathbb{E}[Z] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

We also have that

$$\mathrm{cov}(Z) = \boldsymbol{\Sigma}_Z = \mathbb{E}\Big[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])^\top\Big] = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{bmatrix}$$

Since U and V are independent, we can write

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}(U+V) = \mathrm{Var}(U) + \mathrm{Var}(V) \\ \mathrm{Var}(Y) &= \mathrm{Var}(U-V) = \mathrm{Var}(U) + \mathrm{Var}(V) \end{aligned}$$

and remember the definition of variance

$$\mathrm{Var}(U) = \int_{\mathbb{R}} (u - \mathbb{E}[u])^2 \; p_U(u) \mathrm{d}u = \frac{1}{12} = \mathrm{Var}(V)$$

Now we have to calculate the cross-covariance,

$$\mathrm{Cov}(X,Y) = \mathbb{E}_{XY}\Big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\Big] = \mathbb{E}\Big[(X-1)Y\Big] = \mathbb{E}\Big[XY\Big] - \mathbb{E}[Y] = \mathbb{E}[U^2 - V^2]$$

and finally

$$\mathrm{Cov}(X,Y) = \mathbb{E}[U^2] - \mathbb{E}[V^2] = 0 \quad \text{therefore} \quad \boldsymbol{\Sigma}_Z = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{6} \end{bmatrix}$$

(b) Prove that X and Y are uncorrelated but not independent.

From the result in (a) we see that $\mathrm{Cov}(X,Y) = 0$ so they are indeed uncorrelated.

To check whether two random variables are independent we have to first calculate their joint pdf $p_{XY}(x,y)$ and compare it to the marginal pdfs $p_X(x)$ and $p_Y(y)$. The random variables will be independent **if, and only if,** we can write $p_{XY}(x,y) = p_X(x)p_Y(y)$

First, recall that

$$\begin{aligned} X &= U + V \\ Y &= U - V \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} U &= \tfrac{1}{2}(X+Y) \\ Y &= \tfrac{1}{2}(X-Y) \end{aligned}$$

So let's first check what is the joint pdf. We can simply use the transformation method, which writes

$$p_{XY}(x,y) = p_{UV}(u,v) \times |\det{(()\, J)}|^{-1}$$

where J is the jacobian of the transformation from (U,V) to (X,Y):

$$J = \begin{bmatrix} \frac{\partial X}{\partial U} & \frac{\partial X}{\partial V} \\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow |\det(J)| = 2$$

Then

$$p_{XY}(x,y) = p_U\left(\tfrac{1}{2}(x+y)\right)p_V\left(\tfrac{1}{2}(x-y)\right) \times \frac{1}{2}$$

We thus identify a pdf which is constant on a region $\mathcal{S}$ defined by the two uniform marginals. Let's see what $\mathcal{S}$ looks like:

Since $0 \leq U \leq 1$ and $0 \leq V \leq 1$, then we have $0 \leq X+Y \leq 2$ and $0 \leq X-Y \leq 2$, therefore $0 \leq X \leq 2$ and $-X \leq Y \leq X$

So we can write the joint pdf as per

$$p_{XY}(x,y) = \begin{cases} \frac{1}{2}, & 0 \leq x \leq 2 \quad -x \leq y \leq x \\ 0, & \text{otherwise} \end{cases}$$

OK, we have the joint pdf. Let's see now what the marginal for X looks like.

Remember that $X = U + V$ with U and V independent. One way of calculating the pdf for the sum of two general RVs is to begin with the CDF and then taking its derivative. Let's see where this gets us:

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(U+V \leq x) = \iint\limits_{u+v \leq x} p_{UV}(u,v)\mathrm{d}u\mathrm{d}v = \int_{-\infty}^{+\infty}\int_{-\infty}^{x-u} p_U(u)p_V(v)\mathrm{d}u\mathrm{d}v$$

Rearranging things, we get

$$F_X(x) = \int_{-\infty}^{+\infty} p_U(u)\left(\int_{-\infty}^{x-u} p_V(v)\mathrm{d}v\right)\mathrm{d}u = \int_{-\infty}^{+\infty} p_U(u)F_V(x-u)\mathrm{d}u$$

So the pdf of X can be calculated as

$$p_X(x) = \frac{d}{dx}F_X(x) = \int_{-\infty}^{+\infty} p_U(x)\left(\frac{d}{dx}F_V(x-u)\right)\mathrm{d}u = \int_{-\infty}^{+\infty} p_U(x)p_V(x-u)\mathrm{d}u$$

For our specific case with two uniform distributions, this gives

$$p_X(x) = \int_0^1 p_V(x-u)\mathrm{d}u = \int_0^1 \mathbf{1}_{[0,1]}(x-u)\mathrm{d}u = \begin{cases} 0 & x \leq 0 \quad \text{or} \quad x \geq 2 \\ x & 0 \leq x \leq 1 \\ 2-x & 1 \leq x \leq 2 \end{cases}$$

We see right away that $p_X(x)$ is not a constant, so we don't even need to calculate the marginal for Y to conclude that the we can not factorize the joint pdf into the marginals. Conclusion, the two RVs are not independent.

▶ Exercise 2

Let $Z = \begin{pmatrix} X & Y \end{pmatrix}^T$ be a Gaussian vector with mean $\mu = \begin{pmatrix} 1 & 2 \end{pmatrix}^T$ and covariance $\Sigma = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$.

(a) Compute the probability density function of Z.

The pdf of a bivariate normal distribution is

$$p_Z(\mathbf{z}) = \frac{1}{2\pi \det\left(() \Sigma\right)} \exp\left(-\frac{1}{2}(\mathbf{z}-\mu)^\top \Sigma^{-1}(\mathbf{z}-\mu)\right)$$

With $\det\left(() \Sigma\right) = 1$ and $\Sigma^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$

We end up with

$$p_Z(x,y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(2x^2 + y^2 + 2xy - 8x - 6y + 10)\right)$$

(b) Using

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$$

compute the distribution of Y given $X = x$.

The conditional expectation can be written as

$$p_{Y|X=x}(y) = \frac{p_{XY}(x,y)}{p_X(x)} \quad \text{with} \quad p_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-1)^2\right)$$

Which after lots of simplifications gives us:

$$p_{Y|X=x}(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\Big(y-(3-x)\Big)^2\right) = \mathcal{N}(y \mid 3-x, 1)$$

(c) What is the best prediction of Y given $X = x$?

Remember from the CM1 that the best prediction of Y given $X = x$ is the conditional expectation as per

$$\hat{Y} = \mathbb{E}\Big[Y \mid X = x\Big] = 3 - x$$

▶ Exercise 3

Consider the regression problem discussed in class: we want to determine a function μ that takes a predictor X as input and gives the best estimate in terms of mean squared error for the observed variable Y.

In mathematical terms, we have an optimization problem defined as

$$\mu = \underset{f \in \mathcal{F}}{\operatorname{argmin}}\ \mathbb{E}_{(X,Y)}\left[\Big(Y - f(X)\Big)^2\right]$$

where $\mathcal{F}$ is a space of functions with finite squared norm.

Show that the solution is $\mu(x) = \mathbb{E}_{Y|X}\Big[Y \mid X = x\Big]$

We can rewrite the loss function to minimize as per

$$\begin{aligned}
\mathbb{E}_{(X,Y)}\Big[(Y - f(X))^2\Big] &= \mathbb{E}_X\Big[\mathbb{E}_{Y|X}[(Y-f(X))^2 \mid X]\Big] \\
&= \mathbb{E}_X\Big[\mathrm{Var}(Y - f(X) \mid X) + \Big(\mathbb{E}_{Y|X}[(Y-f(X) \mid X]\Big)^2\Big] \\
&= \mathbb{E}_X\Big[\mathrm{Var}(Y \mid X) + \Big(\mathbb{E}_{Y|X}[(Y-f(X) \mid X]\Big)^2\Big] \\
&\geq \mathbb{E}_X\Big[\mathrm{Var}(Y \mid X)\Big]
\end{aligned}$$

so we see that to minimize the loss we should choose $f(x) = \mathbb{E}_{Y|X}[Y \mid X = x]$

▶ Exercise 4

Consider the Gaussian simple linear regression model presented in class

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \quad \text{with} \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

The estimates for the parameters of the model, $\hat{\beta}_0$ and $\hat{\beta}_1$, are obtained N paired samples (x_i, y_i).

(a) Show that the estimated parameters are unbiased.

From the CM we have the expressions

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ and $\hat{\beta}_1 = \frac{c_{XY}}{s_X^2}$ with $c_{XY} = \frac{1}{N}\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$ and $s_X^2 = \frac{1}{N}\sum_{i=1}^N (x_i - \bar{x})^2$

We first check the unbiasedness of $\hat{\beta}_1$

$\mathbb{E}[\hat{\beta}_1] = \mathbb{E}_{X_1,\ldots,X_N}\left[\mathbb{E}[\hat{\beta}_1 \mid X_1 = x_1, \ldots, X_N = x_N]\right]$ (explain why we first consider the conditional expectation)

$\mathbb{E}[\hat{\beta}_1 \mid X_1 = x_1, \ldots, X_N = x_N] = \frac{1}{s_X^2}\frac{1}{N}\sum_{i=1}^N \left(x_i - \bar{x}\right)\left(\mathbb{E}[Y_i \mid X_i = x_i] - \mathbb{E}[\bar{y} \mid X_1 = x_1, \ldots, X_N = x_N]\right)$
(the X are fixed but the Y are random variables)

Note that,

$$\mathbb{E}\left[Y_i \mid X_i = x_i\right] = \beta_0 + \beta_1 x_i + \mathbb{E}[\varepsilon_i] = \beta_0 + \beta_1 x_i$$

and that

$$\mathbb{E}\left[\bar{y} \mid X_1 = x_1, \ldots, X_N = x_N\right] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N (\beta_0 + \beta_1 x_i + \varepsilon_i)\right] = \beta_0 + \beta_1 \bar{x}$$

so we get

$$\mathbb{E}[\hat{\beta}_1 \mid X_1 = x_1, \ldots, X_N = x_N] = \frac{1}{s_X^2}\frac{1}{N}\sum_{i=1}^N \left(x_i - \bar{x}\right)\left(\beta_0 + \beta_1 x_i - \beta_0 - \beta_1 \bar{x}\right) = \beta_1$$

Then taking the expectacion along all possible datasets, we get

$$\mathbb{E}[\hat{\beta}_1] = \mathbb{E}_{X_1,\ldots,X_N}\left[\beta_1\right] = \beta_1$$

The bias for $\hat{\beta}_0$ is checked similarly.

$$\mathbb{E}[\hat{\beta}_0 \mid X_1 = x_1, \ldots, X_N = x_N] = \mathbb{E}[\bar{Y} - \hat{\beta}_1 \bar{X} \mid X_1 = x_1, \ldots, X_N = x_N]$$

Which then gives us

$$\mathbb{E}[\bar{Y} \mid X_1 = x_1, \ldots, X_N = x_N] = \beta_0 + \beta_1 \bar{x}$$

and

$$\mathbb{E}[\hat{\beta}_1 \bar{X} \mid X_1 = x_1, \ldots, X_N = x_N] = \bar{x}\, \mathbb{E}[\hat{\beta}_1 \mid X_1 = x_1, \ldots, X_N = x_N] = \beta_1 \bar{x}$$

so in the end we get

$$\mathbb{E}[\hat{\beta}_0 \mid X_1 = x_1, \ldots, X_N = x_N] = \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0$$

(b) Show that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{N}\frac{1}{s_X^2} \quad \text{and} \quad \mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2}{N}\left(1 + \frac{\bar{X}^2}{s_X^2}\right)$$

where $\bar{X} = \frac{1}{N}\sum_{i=1}^N x_i$ and $s_X^2 = \frac{1}{N}\sum_{i=1}^N (x_i - \bar{X})^2$.

OK, now let's get the variances.

Remember that $\hat{\beta}_1 = \frac{c_{XY}}{s_X^2} = \left(\frac{1}{N}\sum_i x_i y_i - \bar{x}\bar{y}\right)\frac{1}{s_X^2}$ so we will first rewrite it so that things get easier later

Note that

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{and} \quad \bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{\varepsilon} \quad \text{where} \quad \bar{\varepsilon} = \frac{1}{N}\sum_i \varepsilon_i$$

and that

$$\frac{1}{N}\sum_i x_i y_i = \frac{1}{N}\sum_i x_i(\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 \bar{x} + \beta_1 \frac{1}{N}\sum_i x_i^2 + \frac{1}{N}\sum_i x_i \varepsilon_i$$

and

$$\bar{x}\bar{y} = \beta_0 \bar{x} + \beta_1 (\bar{x})^2 + \bar{x}\bar{\varepsilon}$$

Then

$$\hat{\beta}_1 = \left(\beta_1 s_X^2 + \frac{1}{N}\sum_i x_i \varepsilon_i - \bar{x}\bar{\varepsilon}\right)\frac{1}{s_X^2} = \beta_1 + \frac{1}{s_X^2}\frac{1}{N}\sum_i (x_i - \bar{x})\varepsilon_i$$

This way of rewriting the estimator $\hat{\beta}_1$ is very insightful, since now we can easily write that

$$\text{Var}(\hat{\beta}_1) = \mathbb{E}_{X_1,\dots,X_N}\Big[\text{Var}_X(\hat{\beta}_1 \mid X_1 = x_1, \dots, X_N = x_N)\Big] = \mathbb{E}_{X_1,\dots,X_N}\left[\frac{1}{s_X^4}\frac{1}{N^2}\sum_i (x_i - \bar{x})^2 \sigma^2\right]$$

so we get

$$\text{Var}(\hat{\beta}_1) = \frac{1}{N^2}\frac{1}{s_X^4} s_X^2 N s_X^2 \sigma^2 = \frac{1}{N}\frac{\sigma^2}{s_X^2}$$

What about $\hat{\beta}_0$?

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \Big(\beta_0 + \beta_1 \bar{x} + \bar{\varepsilon}\Big) - \Big(\beta_1 + \frac{1}{s_X^2}\frac{1}{N}\sum_i (x_i - \bar{x})\varepsilon_i\Big)\bar{x}$$

which gives us

$$\hat{\beta}_0 = \beta_0 + \bar{\varepsilon} - \frac{1}{s_X^2}\frac{1}{N}\sum_i (x_i - \bar{x})\bar{x}\varepsilon_i$$

Theferore,

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2}{N} + \frac{1}{N^2 s_X^4}\sum_i (x_i - \bar{x})^2 \bar{x}^2 \sigma^2 = \frac{1}{N}\sigma^2\Big(1 + \frac{\bar{x}^2}{s_X^2}\Big)$$

Using the estimated parameters, we can predict that for a given arbitrary value of X, say x (sometimes called the operation point), we have that on average Y will be

$$\hat{m}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

(c) Show that

$$\mathbb{E}\Big[\hat{m}(x)\Big] = \beta_0 + \beta_1 x$$

$$\hat{m}(x) = \hat{\beta}_0 + \hat{\beta}_1 x = (\bar{y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 x = \bar{y} + \hat{\beta}_1 (x - \bar{x})$$

Taking back the expression from item (b)

$$\begin{aligned}
\hat{m}(x) &= \bar{y} + \left(\beta_1 + \frac{1}{s_X^2}\frac{1}{N}\sum_i (x_i - \bar{x})\varepsilon_i\right)(x - \bar{x}) \\
&= \frac{1}{N}\sum_i \left(\beta_0 + \beta_1 x_i + \varepsilon_i\right) + \left(\beta_1 + \frac{1}{s_X^2}\frac{1}{N}\sum_i (x_i - \bar{x})\varepsilon_i\right)(x - \bar{x}) \\
&= \beta_0 + \beta_1\bar{x} + \bar{\varepsilon} + \beta_1 x - \beta_1\bar{x} + \frac{(x - \bar{x})}{N s_X^2}\sum_i (x_i - \bar{x})\varepsilon_i \\
&= \beta_0 + \beta_1 x + \bar{\varepsilon} + \frac{(x - \bar{x})}{N s_X^2}\sum_i (x_i - \bar{x})\varepsilon_i \\
\mathbb{E}[\hat{m}(x)] &= \beta_0 + \beta_1 x
\end{aligned}$$

(d) Show that the variance of $\hat{m}(x)$ conditioned on a given choice of datapoints $x_1, \ldots, x_N$ can be written as per

$$\mathrm{Var}_X\left(\hat{m}(x)\right) = \frac{\sigma^2}{N}\left(1 + \frac{(x - \bar{X})^2}{s_X^2}\right)$$

Describe how the variance changes for different choices of the operation point.

Finally, the conditional variance on a given choice of dataset $x_1, \ldots, x_N$ is

$$\mathrm{Var}_X\left(\hat{m}(x)\right) = \mathrm{Var}_X\left(\beta_0 + \beta_1 x + \bar{\varepsilon} + \frac{(x - \bar{x})}{N s_X^2}\sum_i (x_i - \bar{x})\varepsilon_i\right)$$

which can be simplified as per

$$\begin{aligned}
\mathrm{Var}_X\left(\hat{m}(x)\right) &= \mathrm{Var}\left(\frac{1}{N}\sum_i \varepsilon_i + \tfrac{(x-\bar{x})}{N s_X^2}\sum_i (x_i - \bar{x})\varepsilon_i\right) \\
&= \mathrm{Var}\left(\frac{1}{N}\sum_i \left[1 + \tfrac{(x-\bar{x})}{s_X^2}(x_i - \bar{x})\right]\varepsilon_i\right) \\
&= \mathrm{Var}\left(\frac{1}{N}\sum_i \left[1 + \tfrac{(x-\bar{x})}{s_X^2}(x_i - \bar{x})\right]\varepsilon_i\right) \\
&= \frac{1}{N^2}\sum_i \left[1 + \tfrac{(x-\bar{x})}{s_X^2}(x_i - \bar{x})\right]^2 \mathrm{Var}(\varepsilon_i) \\
&= \frac{\sigma^2}{N^2}\sum_i \left(1 + \frac{2(x - \bar{x})(x_i - \bar{x})}{s_X^2} + \frac{(x - \bar{x})^2(x_i - \bar{x})^2}{s_X^4}\right) \\
&= \frac{\sigma^2}{N}\sum_i \left(1 + \frac{(x - \bar{x})^2}{s_X^2}\right)
\end{aligned}$$

Notice that to get the unconditional variance we would have to take the expectation with respect to the x_i, which can be quite cumbersome and not that much insightful. We won't do that.