UPPSALA UNIVERSITET

Föreläsningskommentarer

Multivariate Methods

Rami Abou Zahra

Inlämningsdatum
September 15, 2023

Contents

1. Introduction

Analysis dealing with simultaneous measurements on many variables.

We may want to do some statistical analysis on not only salary, but factor in things such as gender, wether or not one has been to uni etc.

One should always stride to use as much information as possible, you want to remove any chance to miss a pattern.

In general, if you arrive to a conclusion, think of why/what caused this and factor everything in your data and analysis.

1.1. **MANOVA.**

MANOVA is a method to measure if a data-set shares a similar mean. For example, with different flower types we may want to check if "does sweden has a similar income as norwegian citizens", comparing the sample from sweden to norwegian. We will get different numbers but that is something that we take into analysis.

1.2. **Regressionanalysis.**

Allows us to predict a variable y from an observation x. $x =$ bmi, while y is your blood pressure.

2. Sample & Random Matrices

2.1. Slide 3 - Expectation.

For a discrete random variable we use summation, for a continuous random variable we use integrals. What do we use for vectors/matrices?

$\Rightarrow$ We perform the operations elementwise in the matrix. Take $\mathbb{E}(X_{ij})$

2.2. Slide 4 - Covariance Matrix.

Recall

$$\mathrm{Cov}\,(X, Y) = \mathbb{E}(X - \mathbb{E}(X)(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \tag{1}$$

for scalars.
What about $\mathrm{Cov}\left(\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}\right)$?

We can pick any pair (X_i, Y_j) and compute $\mathrm{Cov}\,(X_i, Y_j)$ leading to the same as (1) but with X_i, Y_j instead.

In the case $\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, we get a 3×2 matrix where the i, jth elements corresponds to $\mathrm{Cov}\,(X_i, Y_j)$.

Think of it like

$$XY^T = \begin{pmatrix} X_1Y_1 & X_1Y_2 \\ X_2T_1 & X_2Y_2 \\ X_3Y_1 & X_3Y_2 \end{pmatrix} \tag{2}$$

Now look at $\mathbb{E}(XY^T)$, same as (2) but $\mathbb{E}(X_iY_j)$.
Then we can easily see that $\mathrm{Cov}\,(X, Y) = \mathbb{E}(XY^T) - \mu_X \mu_Y^T$

What if X is continuous and Y discrete?
What if $Y = X$?

$$\mathrm{Cov}\,(X_i, X_i) = \mathbb{E}(X_i^2) - (\mathbb{E}(X))^2 = \mathrm{Var}\,(X_i)$$

2.3. Slide 5 - Covariance Matrix.

Since in the scalar case $\mathrm{Cov}\,(X_i, X_j) = \mathrm{Cov}\,(X_j, X_i)$, then $\mathrm{Cov}\,(X, Y) = \sum$ = symmetric & positive definite.

Definition/Sats 2.1: Positive & Semi-definite

Definite matrix A:

$$A > 0 \Leftrightarrow x^T A x > 0$$

Semi-definite matrix A:

$$A \geq 0 \Leftrightarrow x^T A x \geq 0$$

2.4. Slide 6 - Linear Combination.

You can view the vector c as regression values for example

2.5. **Slide 7 - Linear Combination.**

Example:

$$\mathrm{Var}\left(X_1 + 2X_2 + 4X_3\right) \sim \mathrm{Var}\left(\begin{pmatrix}1 & 2 & 4\end{pmatrix}\begin{pmatrix}X_1 \\ X_2 \\ X_3\end{pmatrix}\right)$$

A tip for remembering where to put c^T, think of it like matching dimensions of left hand side and right hand side.

We only want to compute expectation for the random stuff, so we can chuck coefficients and constants out.

2.6. **Slide 9 - Independence.**

For simplicity, we define independence in the continuous case as $f(X,Y) = f(X)f(Y)$ and in the discrete case as $P(X,Y) = P(X)P(Y)$

Anmärkning: Jist because $\mathrm{Cov}\,(X,Y) = 0$ does not imply independence. Take the unit circle and the contour as pairs over (X,Y). It is clear that (X,Y) are dependant but their covariance is 0 since for every point on the circle you can reflect the X, Y and therefore, by $\mathrm{Cov}\,(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$, you would be adding a bunch of 0. Same goes for any function that can be reflected.

2.7. **Slide 10 - Random Sample.**

Example (Scalar case):
Let $\mathbf{x} \sim x_1 x_2 x_3 \cdots$ be a random sample from $N(\mu, \sigma^2)$

We look at what it means for scalar random variables to be independent:

$$F(X,Y) = F(X)F(Y)$$
$$f(x,y) = f(x)f(y)$$
$$p(x,y) = p(x)p(y)$$

The same principle goes for random vectors, eg:

$$X_{n\times p} = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}$$

Think of each row as a sample from a different place $\Rightarrow$ independence in row $\Rightarrow$ random sample.

Non-example: Looking at the pulse of 1 person is not an independent response since it is only about 1 person. Even if you sampled a bunch of values from the same person into a matrix, that would still be a non-independent sample since we only sample from 1 person.

Non-example: Let us assume there is a competition between Uppsala and Lund in Multivariate Analysis. Everyone in the class at Uppsala has had the same teacher, so the values collected from that class are not independent.

2.8. **Slide 12 - Some Notes on Sample Covariance Matrix.**

Unbiased becomes biased during non-linear & non-affine transformations.

Even for large n, sometimes you cannot ignore the difference between S_n and S (eg. determining exact distributions)

2.9. **Slide 17 - Sample Covariance Matrix.**

$$X - \frac{1}{n}\mathbf{1}\mathbf{1}^T X = (I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)X$$

So for $(X - \frac{1}{n}\mathbf{1}\mathbf{1}^T X)^T(X - \frac{1}{n}\mathbf{1}\mathbf{1}^T X)$:

$$X^T(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)^T(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)X = X^T\left[I - \frac{1}{n}\mathbf{1}\mathbf{1}^T - \frac{1}{n}\mathbf{1}\mathbf{1}^T - \frac{1}{n}\mathbf{1}\mathbf{1}^T + \frac{1}{n}\mathbf{1}\underbrace{\mathbf{1}^T\mathbf{1}}_{=n}\mathbf{1}^T\right]X$$

$$X^T(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T - \frac{1}{n}\mathbf{1}\mathbf{1}^T + \frac{1}{n}\mathbf{1}\mathbf{1}^T)X = X^T(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)X$$

$$X^T X - X^T\mathbf{1}\mathbf{1}^T X \Rightarrow S_n = \frac{1}{n}\underbrace{X^T X}_{\text{Data matrix}} - (\frac{1}{n}X^T\mathbf{1})(\frac{1}{n}\mathbf{1}^T X)$$

$$\mathrm{Cov}\,(X) = \mathbb{E}(XX^T)^n - \mathbb{E}(X)\mathbb{E}(X)^T$$

Anmärkning:
$\mathbf{1}$ is an $n \times 1$ vector of ones.

3. Multivariate Normal Distribution

3.1. **Slide 4-5 - From Univariate to Multivariate Normal.**
Recall that in the univariate case we had:

$$(x-\mu)\frac{1}{\sigma^2}$$

In the multivariate case, we swap x and μ for vectors instead.
Since variance matrix is expressed by $(x-\mu)^T\Sigma^{-1}(x-\mu)$, instead of σ^2 we have have

$$\frac{1}{\sigma\sqrt{2\pi}} \rightsquigarrow \frac{1}{(2\pi)^{p/2}\sqrt{\det(\Sigma)}}$$

Anmärkning:
Covariance matrix must be positive definite! Not semi.
There is no requirement for slide 4 with Σ
The $(2\pi)^{p/2}$ comes from multiplying $z_1 z_2 \cdots z_p$ p-times.

3.2. **Slide 6 - Special Case: Bivariate Normal.**

Anmärkning:
ρ denotes the correlation coefficient
$\sigma_{11}\&\sigma_{22}$ correspond to our variance
$\sigma_{12}\&\sigma_{21}$ correspond to our covariance

$$\text{Corr}(x_1,x_2) = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$$

3.3. **Slide 7 - Contour of Bivariate Normal Density.**
We change the correlation to see what happens.

3.4. **Slide 8 - Linear Combinations.**
For the univariate case, we had that if we scaled $X \sim N(\mu,\sigma^2)$ with an affine transformation, we got $aX + b \sim N(a\mu, a^2\sigma^2)$.

One thing that is good to keep in the back of the head is that the linear combination/affine transformation of normally distributed random variables will remain normal.

Let us look at what happens when we look at the multivariate case:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \qquad Y_1 \sim N \qquad Y_2 \sim N$$

$$\Rightarrow \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, A\Sigma A^T\right)$$

From result 4.2, we can get the result of multi-linear combinations

3.5. **Slide 10 - Normal and Chi-Square.**

If X has a linear combination will it still be p-degreees of freedom? Answer is surprisingly yes!

$$\Sigma^{-1} = \Sigma^{-1/2}\Sigma^{-1/2} \qquad X \sim N_p(\mu,\Sigma)$$

$$\Rightarrow Z = \Sigma^{-1/2}(x-\mu) = \underbrace{\Sigma^{-1/2}}_{A} x \underbrace{-\Sigma^{-1/2}\mu}_{d} \sim N_p(0, \Sigma^{-1/2}\Sigma\Sigma^{-1/2})$$

$$(x-\mu)^T\Sigma^{-1}(x-\mu) = Z^t Z = \sum_{j=1}^{p} Z_i$$

3.6. **Slide 11 - Subset of Variables.**

Using result 4.4, we can choose subsets however we want, it will stay normal.

3.7. **Slide 12 - Example: Subset of Variables.**

From the slide we have the following:
Suppose that:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \right)$$

Find the distribution of $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ *as well as the distribution of*

$$\begin{bmatrix} X_1 & X_3 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$$

In the first one, what we really essentially are looking for is the following:

$$\begin{bmatrix} X_1 \\ \cancel{X_2} \\ X_3 \end{bmatrix} \sim N_3 \left(\begin{bmatrix} 0 \\ \cancel{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \cancel{\sigma_{12}} & \sigma_{13} \\ \cancel{\sigma_{21}} & \cancel{\sigma_{22}} & \cancel{\sigma_{23}} \\ \sigma_{31} & \cancel{\sigma_{32}} & \sigma_{33} \end{bmatrix} \right)$$

If we want $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$, then:

$$\begin{bmatrix} X_1 \\ X_3 \end{bmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix} \right)$$

So:

$$\begin{bmatrix} X_1 & X_3 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_3 \end{bmatrix} \sim \chi_2$$

It is really important to remember that linear combinations of normal variables, are still normal variables. Since linear combinations can be regarded as linear/affine transformations, the "crossing out the X_2" part of the computation is really just matrix-multiplication, since:

$$\begin{bmatrix} X_1 \\ X_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{A} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

3.8. **Slide 13 - Subset of Variables.**

Anmärkning:
Since what we really care about is what happens during the transpose, sometimes we write Σ_{12} for $\Sigma_{12} = \Sigma_{21} = 0$

3.9. **Slide 15 - Marginal Normal and Joint Distribution.**

Usually, if they are independent, they are normal.

3.10. **Slide 23 - Likelihood of Normal Random Sample.**

$$a^T B a = \operatorname{tr}(a^T B a) = \operatorname{tr}(B a a^T)$$

Of course, in order to maximize the likelihood we sometimes need to find the derivative of the matrix/vector.

Example:

$$\underbrace{\begin{bmatrix} x_1 & x_2 \end{bmatrix}}_{x^T} \underbrace{\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} b_{11}x_1 + b_{12}x_2 \\ b_{21}x_1 + b_{22}x_2 \end{bmatrix}$$

$$\Rightarrow b_{11}x_1^2 + b_{12}x_1x_2 + b_{21}x_1x_2 + b_{22}x_2^2 = f(x_1, x_2)$$

Now we can just collect the partials in a vector (or a matrix if we end up with a matrix):

$$\begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \\ \dfrac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2b_{11}x_1 + b_{12}x_2 + b_{21}x_2 \\ 2b_{22}x_2 + b_{12}x_1 + b_{21}x_1 \end{bmatrix} = \begin{bmatrix} 2b_{11} & b_{12} + b_{21} \\ b_{12} + b_{21} & 2b_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

3.11. **Slide 32 - Limit of MLE.**

$$\underbrace{\frac{n}{n-1}}_{\substack{n\to\infty \\ \to 1}} \underbrace{(\mu_1 - \hat{X}_i)}_{\to 0} \underbrace{(\hat{X}_k - \mu_k)}_{\to 0} \to \frac{1}{n-1} \sum \approx \sigma_{ik}$$