02417: Time Series Analysis

Math reference

Peder Bacher
DTU Compute

February 7, 2025

Outline

Multivariate random variables – distribution functions

- Definition (n-dimensional random variable; random vector)

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

- Joint distribution function (cumulative disitribution function (cdf)):

$$F(x_1, \ldots, x_n) = \mathrm{P}\{X_1 \leq x_1, \ldots, X_n \leq x_n\}$$

- Notation: See the modelling reference slides.

Exception! In this math reference:

- *Bold and capital means random vector (matrix in the other parts)*
- *Bold and small means observation vector (or fixed or "known" value) (vector in the other parts)*

Multivariate random variables - joint densities

- Joint distribution function (repeated from last slide):

$$F(x_1, \ldots, x_n) = \mathrm{P}\{X_1 \leq x_1, \ldots, X_n \leq x_n\}$$

- Joint density function - continuous case:

$$f(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$$

  - and back to the joint distribution function:

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n)\, dt_1 \ldots dt_n$$

- Joint density function - discrete case:

$$f(x_1, \ldots, x_n) = \mathrm{P}\{X_1 = x_1, \ldots, X_n = x_n\}$$

The Multivariate Normal Distribution

- The joint density (probability density function (pdf))

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

- $\boldsymbol{\Sigma}$ is the *covariance matrix*, it's symmetric and positive semi-definite
- Notation: $\boldsymbol{X} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Standard multivariate normal: $\boldsymbol{Z} \sim \mathsf{N}(\mathbf{0}, \boldsymbol{I})$
- If $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{T}\boldsymbol{Z}$, where $\boldsymbol{\Sigma} = \boldsymbol{T}\boldsymbol{T}^T$, then $\boldsymbol{X} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- If $\boldsymbol{X} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{X}$ then $\boldsymbol{Y} \sim \mathsf{N}(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T)$
- More relations between distributions in Sec. 2.7

Why this focus on normal distribution and its first and second moment?

Is any linear transformation of a multivariate-variate normal distribution also normal distributed?

Yes, any linear transformation of a multivariate normal distribution is also normally distributed.

Specifically, if you have a multivariate normal random vector $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and you apply a linear transformation $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$, where $\boldsymbol{A}$ is a matrix and $\boldsymbol{b}$ is a vector, then $\boldsymbol{Y}$ will also follow a multivariate normal distribution.

The new (first moment) mean vector and (second moment) covariance matrix of $\boldsymbol{Y}$ will be:

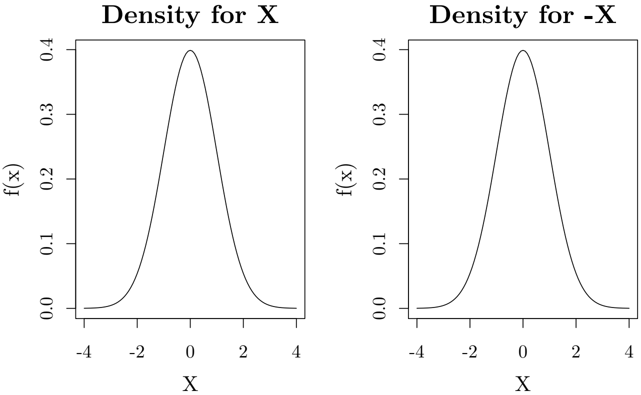
- Mean: $\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}$
- Covariance: $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T$

This property is quite useful in various applications, such as in statistics and machine learning.
Do you have a specific context or example in mind where you're applying this?

Yes, we are doing linear times series models, hence all the calculations are linear transformations (adding and multiplying with matrices): Great, now we know that all keeps being normal! and we keep track of everything with the mean and covariance :-)

Stochastic variables and distributions

- If $X \sim N(0,1)$, then $-X \sim N(0,1)$

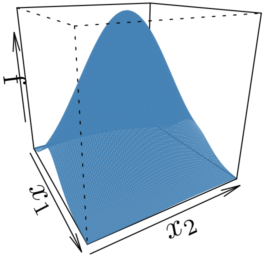


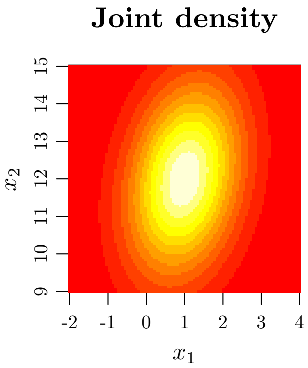
- X and $-X$ are different variables that have the same distribution

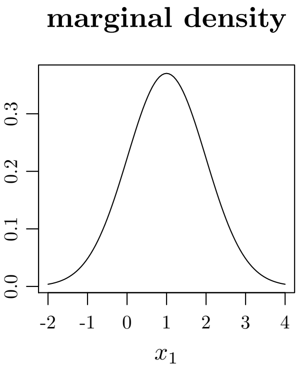
Marginal density function

- Sub-vector: $(X_1, \ldots, X_k)^T, \quad (k < n)$
- Marginal density function:

$$f_S(x_1, \ldots, x_k) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_{k+1} \ldots dx_n$$





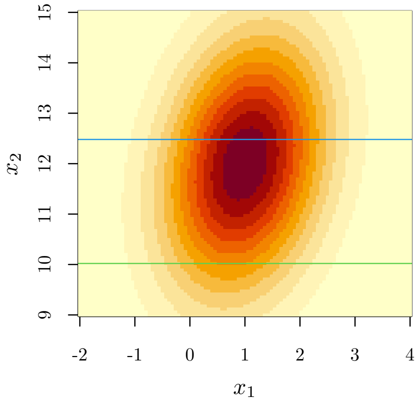


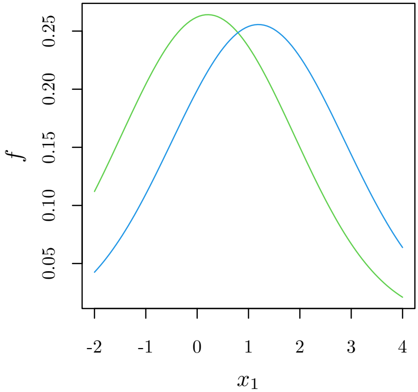
Conditional distributions

- The conditional density of X_1 given $X_2 = x_2$ is defined as ($f_{X_1}(x_1) > 0$):

$$f_{X_1|X_2=x_2}(x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)}$$

(joint density of (X_1, X_2) divided by the marginal density of X_2 evaluated at x_2)





Independence

- If knowledge of X does not give information about Y, we get that $f_{Y|X=x}(y) = f_Y(y)$
- This leads to the following definition of independence:

$$X,\ Y \text{ stochastically independent } \overset{def}{\Leftrightarrow}$$

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

Expectation

- Let X be a univariate random variable with density $f_X(x)$. The expectation of X is then defined as:
$$E[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \qquad \text{(continuous case)}$$
$$E[X] = \sum_{all\ \ x} x\, P(X = x) \qquad \text{(discrete case)}$$
- Expectation is a linear operator
- Calculation rule:
$$E[a + bX_1 + cX_2] = a + b\, E[X_1] + c\, E[X_2]$$

Moments and Variance

- n'th moment:

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\,dx$$

- n'th central moment:

$$E[(X - E[X])^n] = \int_{-\infty}^{\infty} (x - E[X])^n f_X(x)\,dx$$

- The 2'nd central moment is called the variance:

$$V[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

A multi-variate normal distribution is fully characterized by the first and second moment.

Covariance

- Covariance:

$$\mathrm{Cov}[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])] = E[X_1 X_2] - E[X_1]E[X_2]$$

- Variance and covariance:

$$V[X] = \mathrm{Cov}[X, X]$$

- Calculation rules:

$$\mathrm{Cov}[aX_1 + bX_2, cX_3 + dX_4] = \\ ac\,\mathrm{Cov}[X_1, X_3] + ad\,\mathrm{Cov}[X_1, X_4] + bc\,\mathrm{Cov}[X_2, X_3] + bd\,\mathrm{Cov}[X_2, X_4]$$

- The calculation rule can be used for the variance as well. For instance:

$$V[a + bX_2] = b^2\,V[X_2]$$
$$V[aX_1 + bX_2] = a^2\,V[X_1] + b^2\,V[X_2] + 2ab\mathrm{Cov}[X_1, X_2]$$

Moment representation

- All moments up to a given order.
- Second order moment representation:
  - Mean
  - Variance
  - Covariance (If relevant)

Expectation and Variance for Random Vectors

- Expectation: $E[\boldsymbol{X}] = [E[X_1], E[X_2], \ldots, E[X_n]]^T$
- Variance-covariance (matrix): $\boldsymbol{\Sigma}_{\boldsymbol{X}} = V[\boldsymbol{X}] = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T] =$

$$\begin{bmatrix} V[X_1] & \text{Cov}[X_1, X_2] & \cdots & \text{Cov}[X_1, X_n] \\ \text{Cov}[X_2, X_1] & V[X_2] & \cdots & \text{Cov}[X_2, X_n] \\ \vdots & & & \vdots \\ \text{Cov}[X_n, X_1] & \text{Cov}[X_n, X_2] & \cdots & V[X_n] \end{bmatrix}$$

- Correlation:

$$\rho_{ij} = \frac{\text{Cov}[X_i, X_j]}{\sqrt{V[X_i]\, V[X_j]}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

Correlation and Independence

- If X and Y are independent stochastic variables then $\mathrm{Cov}(X, Y) = 0$ and thus $\mathrm{Corr}(X, Y) = 0$.
- However, if $X \in N(0, 1)$, then

$$\begin{aligned}\mathrm{Cov}(X, X^2) =& E[X \cdot X^2] - E[X] \cdot E[X^2] = E[X^3] \\ =& \int x^3 f_X(x)\, dx = 0\end{aligned}$$

- Thus X and X^2 are uncorrelated, but $E[X^2|X = x] = x^2$.
- Independence implies no correlation, not the other way around.

Expectation and Variance for Random Vectors

- The correlation matrix $\boldsymbol{R} = \boldsymbol{\rho}$ is an arrangement of ρ_{ij} in a matrix
- Covariance matrix between $\boldsymbol{X}$ (dim. p) and $\boldsymbol{Y}$ (dim. q):

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{XY}} &= C[\boldsymbol{X}, \boldsymbol{Y}] = E\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\nu})^T\right] \\ &= \begin{bmatrix} \mathrm{Cov}[X_1, Y_1] & \cdots & \mathrm{Cov}[X_1, Y_q] \\ \vdots & & \vdots \\ \mathrm{Cov}[X_p, Y_1] & \cdots & \mathrm{Cov}[X_p, Y_q] \end{bmatrix} \end{aligned}$$

- Calculation rules – see the book.
- The special case of the variance $C[\boldsymbol{X}, \boldsymbol{X}] = V[\boldsymbol{X}]$ results in $\boxed{V[\boldsymbol{A}\boldsymbol{X}] = \boldsymbol{A}\, V[\boldsymbol{X}]\boldsymbol{A}^T}$

Conditional expectation

$$E[Y|X=x] = \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\, dy$$

$E[Y|X] = E[Y]$ if X and Y are independent

$E[Y] = E\left[E[Y|X]\right]$

$E[g(X)Y|X] = g(X)E[Y|X]$

$E[g(X)Y] = E\left[g(X)E[Y|X]\right]$

$E[a|X] = a$

$E[g(X)|X] = g(X)$

$E[cX + dZ|Y] = cE[X|Y] + dE[Z|Y]$

Variance separation

- Definition of conditional variance and covariance:

$$
\begin{aligned}
V[\boldsymbol{Y}|\boldsymbol{X}] &= E\left[\left(\boldsymbol{Y} - E[\boldsymbol{Y}|\boldsymbol{X}]\right)\left(\boldsymbol{Y} - E[\boldsymbol{Y}|\boldsymbol{X}]\right)^T|\boldsymbol{X}\right] \\
C[\boldsymbol{Y},\boldsymbol{Z}|\boldsymbol{X}] &= E\left[\left(\boldsymbol{Y} - E[\boldsymbol{Y}|\boldsymbol{X}]\right)\left(\boldsymbol{Z} - E[\boldsymbol{Z}|\boldsymbol{X}]\right)^T|\boldsymbol{X}\right]
\end{aligned}
$$

- The variance separation theorem:

$$
\begin{aligned}
V[\boldsymbol{Y}] &= E\left[V[\boldsymbol{Y}|\boldsymbol{X}]\right] + V\left[E[\boldsymbol{Y}|\boldsymbol{X}]\right] \\
C[\boldsymbol{Y},\boldsymbol{Z}] &= E\left[C[\boldsymbol{Y},\boldsymbol{Z}|\boldsymbol{X}]\right] + C\left[E[\boldsymbol{Y}|\boldsymbol{X}], E[\boldsymbol{Z}|\boldsymbol{X}]\right]
\end{aligned}
$$

Linear transformations and projections

What is the difference between a linear transformation and a linear projection?

Great question! Both linear transformations and linear projections are concepts from linear algebra, but they have distinct characteristics:

Linear Transformation

A **linear transformation** is a function between two vector spaces that preserves the operations of vector addition and scalar multiplication. If T is a linear transformation from vector space V to vector space W, then for any vectors $\mathbf{u}, \mathbf{v} \in V$ and any scalar c, the following properties hold:

1. **Additivity**: $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
2. **Homogeneity**: $T(c\mathbf{u}) = cT(\mathbf{u})$

Linear Projection

A **linear projection** is a specific type of linear transformation **where the transformation maps a vector space onto a subspace**. If P is a projection, then applying P twice is the same as applying it once, i.e., $P^2 = P$. This means that once a vector is projected onto the subspace, applying the projection again does not change the vector.

Key Differences

1. **General vs. Specific**: All linear projections are linear transformations, but not all linear transformations are projections.
2. **Idempotence**: A linear projection is idempotent ($P^2 = P$), whereas a general linear transformation does not necessarily have this property.
3. **Purpose**: Linear transformations can map vectors between different spaces or dimensions, while linear projections specifically map vectors onto a subspace.