Revision: Multivariate probability

Joint, marginal, conditional distributions

Optional video: Watch the below video for a brief review of multivariate functions.

A random vector $\boldsymbol{X} = \begin{pmatrix} X_1 & X_2 & \cdots & X_p \end{pmatrix}^\top \in \mathbb{R}^p$, $p \geq 2$ has a *joint cdf*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_p \leq x_p) = F_{\boldsymbol{X}}(x_1, x_2, \ldots, x_p).$$

In case of a *discrete* vector of observations $\boldsymbol{X}$ the *probability mass function* is defined as

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = P(X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p).$$

If a *density* $f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_p)$ exists such that

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f_{\boldsymbol{X}}(\boldsymbol{t}) dt_1 \dots dt_p \tag{0.12}$$

then $\boldsymbol{X}$ is a *continuous* random vector with a joint density function of p arguments $f_{\boldsymbol{X}}(\boldsymbol{x})$. From (0.12) we see that in this case $f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\partial^p F_{\boldsymbol{X}}(\boldsymbol{x})}{\partial x_1 \partial x_2 .. \partial x_p}$ holds.

The *marginal cdf of the first* $k < p$ components of the vector $\boldsymbol{X}$ is defined in a natural way as follows:

$$\begin{aligned} P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_k \leq x_k) &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_k \leq x_k, X_{k+1} \leq \infty, ..., X_p \leq \infty) \\ &= F_{\boldsymbol{X}}(x_1, x_2, \dots, x_k, \infty, \infty, \dots, \infty) \end{aligned} \tag{0.13}$$

The *marginal density* of the first k components can be obtained by partial differentiation in (0.13) and we arrive at

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x_1, x_2, \dots, x_p) dx_{k+1} \dots dx_p$$

For **any** other subset of $k < p$ components of the vector $\boldsymbol{X}$, their marginal cdf and density can be obtained along the same lines. X_i has marginal cdf $F_{X_i}(x_i), i = 1, 2, \dots, p$.

The *conditional density* $\boldsymbol{X}$ **when** $X_{r+1} = x_{r+1}, \dots, X_p = x_p$ is defined by

$$f_{(X_1,\dots,X_r|X_{r+1},\dots,X_p)}(x_1, \dots, x_r | x_{r+1}, \dots, x_p) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{X_{r+1},\dots,X_p}(x_{r+1}, \dots, x_p)} \tag{0.14}$$

The above conditional density is interpreted as the joint density of $X_1, \dots, X_r$ when $X_{r+1} = x_{r+1}, \dots, X_p = x_p$ and is only defined when $f_{X_{r+1},\dots,X_p}(x_{r+1}, \dots, x_p) \neq 0$.

In case $\boldsymbol{X}$ has p independent components then

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_p}(x_p) \tag{0.15}$$

holds and, equivalently, also

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = P_{X_1}(x_1) P_{X_2}(x_2) \cdots P_{X_p}(x_p),\ f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_p}(x_p) \tag{0.16}$$

holds. We note that in case of mutual independence the p components, all conditional distributions do **not** depend on the conditions and the factorisations

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{p} F_{X_i}(x_i),\ f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{p} f_{X_i}(x_i)$$

hold.

Moments & density transformation formula

Given the density $f_{\mathbf{X}}(\mathbf{x})$ of the random vector $\boldsymbol{X}$ the *joint moments of order* $s_1, s_2 \ldots, s_p$ are defined, in analogy to the univariate case, as

$$\mathrm{E}(\boldsymbol{X}_{\mathbf{1}}^{s_1} \ldots\ \boldsymbol{X}_{\boldsymbol{p}}^{s_p}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{s_1} \ldots\ X_p^{s_p} f_{\boldsymbol{X}}(x_1, \ldots, x_p) dx_1 \ldots dx_p \tag{0.17}$$

Note that if some of the s_i in (0.17) are equal to zero then in effect we are calculating the joint moment of a subset of the p random variables.

Common multivariate moments and their linear transformations

Common multivariate moments

Now, let $\boldsymbol{X} \in \mathbb{R}^p$ and $\boldsymbol{Y} \in \mathbb{R}^q$ with densities as above. The following moments are commonly used:

Expectation:

$$\boldsymbol{\mu}_{\boldsymbol{X}} = \mathrm{E}(\boldsymbol{X}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \boldsymbol{x} f_{\boldsymbol{X}}(x_1, \ldots, x_p) dx_1 \ldots dx_p \in \mathbb{R}^p.$$

Variance–covariance matrix: (a.k.a. variance or covariance matrix)

$$\begin{aligned}\Sigma_{\boldsymbol{X}} &= \mathrm{Var}(\boldsymbol{X}) = \mathrm{Cov}(\boldsymbol{X}) = \mathrm{E}(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})^\top \\ &= \mathrm{E}\,\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}_{\boldsymbol{X}}^\top = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \in \mathscr{M}_{p,p}.\end{aligned}$$

Covariance matrix:

$$\begin{aligned}\Sigma_{\boldsymbol{X},\boldsymbol{Y}} &= \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})(\boldsymbol{Y} - \boldsymbol{\mu}_{\boldsymbol{Y}})^\top \\ &= \mathrm{E}\,\boldsymbol{X}\boldsymbol{Y}^\top - \boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}_{\boldsymbol{Y}}^\top = \begin{pmatrix} \sigma_{X_1Y_1} & \sigma_{X_1Y_2} & \cdots & \sigma_{X_1Y_q} \\ \sigma_{X_2Y_1} & \sigma_{X_2Y_2} & \cdots & \sigma_{X_2Y_q} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_pY_1} & \sigma_{X_pY_2} & \cdots & \sigma_{X_pY_q} \end{pmatrix} \in \mathscr{M}_{p,q}.\end{aligned}$$

Linear transformations of multivariate moments

Let $A \in \mathscr{M}_{p',p}$ and $B \in \mathscr{M}_{q',q}$ fixed and known. Then,

- $\boldsymbol{\mu}_{A\boldsymbol{X}} = A\boldsymbol{\mu}_{\boldsymbol{X}} \in \mathbb{R}^{p'}$
- $\Sigma_{A\boldsymbol{X}} = A\Sigma_{\boldsymbol{X}}A^\top \in \mathscr{M}_{p',p'}$
- $\Sigma_{A\boldsymbol{X},B\boldsymbol{Y}} = A\Sigma_{\boldsymbol{X},\boldsymbol{Y}}B^\top \in \mathscr{M}_{p',q'}$

As a corollary, if $\boldsymbol{X}'$,$\boldsymbol{Y}'$, A' and B' are variables and matrices with the same dimensions as originals (but possibly different values and distributions),

- $\mathrm{E}(A\boldsymbol{X} + A'\boldsymbol{X}') = A\boldsymbol{\mu}_{\boldsymbol{X}} + A'\boldsymbol{\mu}_{\boldsymbol{X}'}$
- $\mathrm{Var}(A\boldsymbol{X} + A'\boldsymbol{X}') = A\Sigma_{\boldsymbol{X}}A^\top + A\Sigma_{\boldsymbol{X},\boldsymbol{X}'}(A')^\top + A'\Sigma_{\boldsymbol{X}',\boldsymbol{X}}A^\top + A'\Sigma_{\boldsymbol{X}'}(A')^\top$
- $\mathrm{Cov}(A\boldsymbol{X} + A'\boldsymbol{X}', B\boldsymbol{Y} + B'\boldsymbol{Y}') = A\Sigma_{\boldsymbol{X},\boldsymbol{Y}}B^\top + A\Sigma_{\boldsymbol{X},\boldsymbol{Y}'}(B')^\top + A'\Sigma_{\boldsymbol{X}',\boldsymbol{Y}}B^\top + A'\Sigma_{\boldsymbol{X}',\boldsymbol{Y}'}(B')^\top$

These identities are also useful when $p = p' = q = q' = 1$ (i.e., scalars)

Density transformation formula

Assume, the p existing random variables $X_1, X_2, \ldots, X_p$ with given density $f_{\boldsymbol{X}}(\boldsymbol{x})$ have been transformed by a smooth (i.e. differentiable) one-to-one transformation into p new random variables $Y_1, Y_2 \ldots, Y_p$, i.e. a new random vector $\boldsymbol{Y} \in \mathbb{R}^p$ has been created by calculating

$$Y_i = y_i(X_1, X_2 \ldots, X_p), i = 1, 2, \ldots, p \tag{0.18}$$

The question is how to calculate the density $g_{\boldsymbol{Y}}(\boldsymbol{y})$ of $\boldsymbol{Y}$ by knowing the transformation functions $y_i(X_1, X_2 \ldots, X_p), i = 1, 2, \ldots, p$ and the density $f_{\boldsymbol{X}}(\boldsymbol{x})$ of the original random vector. Naturally, since the transformation (0.18) is assumed to be one-to-one, its inverse transformation $X_i = x_i(Y_1, Y_2 \ldots, Y_p), i = 1, 2, \ldots, p$ also exists and then the following density transformation formula applies:

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_p) = f_{\boldsymbol{X}}[x_1(y_1, \ldots, y_p), \ldots, x_p(y_1, \ldots, y_p)]|J(y_1, \ldots, y_p)| \tag{0.19}$$

where $J(y_1, \ldots, y_p)$ is the *Jacobian* of the transformation:

$$J(y_1, \ldots, y_p) = \left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right| \equiv \left|\begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_p} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_p}{\partial y_1} & \frac{\partial x_p}{\partial y_2} & \cdots & \frac{\partial x_p}{\partial y_p} \end{pmatrix}\right| \equiv \left|\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right|^{-1} \tag{0.20}$$

Note that in (0.19) the *absolute value* of the Jacobian is substituted.

Check your understanding

Complete the below exercises to check your understanding of concepts presented so far.

1. In an ecological experiment, colonies of 2 different species of insect are confined to the same habitat. The survival times of the two species (in days) are random variables X_1 and X_2 respectively. It is thought that X_1 and X_2 have a joint density of the form

$$f_{\boldsymbol{X}}(x_1, x_2) = \theta x_1 \mathrm{e}^{-x_1(\theta + x_2)} \quad (0 < x_1, x_2)$$

for some constant $\theta > 0$.

a) Show that $f_{\boldsymbol{X}}(x_1, x_2)$ is a valid density.

b) Find the probability that both species die out within t days of the start of the experiment.

c) Derive the marginal density of X_1. Identify this distribution and write down $\mathrm{E}(X_1)$ and $\mathrm{Var}(X_1)$.

d) Derive the marginal density of X_2, and the conditional density of X_2 **given** $X_1 = x_1$.

e) What evidence do you now have that X_1 and X_2 are not independent?

2. Let $\boldsymbol{X} = [X_1, X_2]^\top$ a random vector with $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}$ and

$$\mathrm{Var}(\boldsymbol{X}) = \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

a) Find $\mathrm{Cov}(X_1 - X_2, X_1 + X_2)$.

b) Find $\mathrm{Cov}(X_1, X_2 - \rho X_1)$.

c) Choose b to minimise $\mathrm{Var}(X_2 - bX_1)$.

Solutions to these exercises will be provided in the 'Exercises' discussion forum.

Demonstration: Matrices

Please start by watching the following demonstration by Dr Krivitsky.

Transcript

This demonstration can be completed using the provided RStudio or your own RStudio.

To complete this task select the 'Matrix_Examples.demo.Rmd' in the 'Files' section of RStudio. Follow the demonstration contained within the RMD file.

If you choose to complete the example in your own RStudio, upload the following file:

Matrix_Examples.demo.Rmd

After working through the demonstration, complete the associated 'Challenge' on the next slide.

The contents of the RMD file are also displayed below:

Constructing matrices in R

R provides a number of tools for constructing and manipulating matrices. The following is a quick overview, particularly where they are distinguished from data frames:

- `diag(n)` (`n` scalar integer) is a `n`×`n` Identity matrix. Note that the function behaves differently depending on whether `n` has length 1 or not.
- `diag(x)` (`x` a vector of length `n`) is a `n`×`n` diagonal matrix whose diagonal elements come from vector `x`; can lead to unpredictable behaviour when the length of `x` is not known in advance and could potentially be 1.

- `diag(x, n)` is a `n`×`n` diagonal matrix whose column elements come from vector `x`; works consistently unlike the previous case.
- `matrix(1,n,n)` is a `n`×`n` matrix of ones
- `matrix(1,n,p)` is a `n`×`p` matrix of ones
- `matrix(x,n,p)` (`x` a scalar or a vector) is a `n`×`p` filled with values of `x` *by columns* with `x` recycled as needed to length `n*p`
- `matrix(1,n)` is an `n`×`1` column vector of ones
- `matrix(1,ncol=n)` is an 1×n row vector of ones
- `cbind(x1,x2,x3)` bind dimensionless or column vectors or matrices `x1`, `x2`, and `x3` into a matrix as columns
- `rbind(x1,x2,x3)` bind dimensionless or row vectors or matrices `x1`, `x2`, and `x3` into a matrix as rows

Matrix arithmetic and operations

- Arithmetic operations `+`, `-`, `*`, `/`, etc. perform operations elementwise, as do many functions such as `exp()` and `sin()`.
- Matrix product is `%*%`: not `*`!

Matrix functions

- `dim(X)`, `nrow(X)`, `ncol(X)` obtain matrix dimensions
- `t(X)` transpose of `X`
- `c(X)` convert a matrix (or a vector with dimension) into a dimensionless vector; for a matrix, this is equivalent to stacking the columns.
- `solve(X)` inverse of `X`: not `X^-1`!
- `crossprod(X,Y)` is equivalent to `t(X)%*%Y` but much faster; if `Y` is omitted, `t(X)%*%X` is computed.
- `chol(X)` Cholesky decomposition: if `U <- chol(X)`, then `t(U)%*%U` reconstructs `X`
- `eigen(X)` Eigendecomposition: if `e <- eigen(X)`, then `e$vectors%*%diag(e$values,length(e$values))%*%t(e$vectors)` reconstructs `X`

Rowwise and columnwise computation

- `apply(X, MARGIN, FUN)` for each row (if `MARGIN=1`) or column (if `MARGIN=2`), evaluate function `FUN` and return the resulting vector. (If `FUN` itself returns a vector, `apply()` returns a matrix.)
- `sweep(X, MARGIN, STATS, FUN)` for each row (if `MARGIN=1`) or column (if `MARGIN=2`) `x`, evaluate `FUN(x, STATS[i])` and replace the original row.

- `colMeans`, `rowMeans`, `colSums`, `rowSums`: exactly as it sounds.
- `scale(X)`: centre the matrix so that its column means are 0 and its column mean-squared value is 1; further arguments can be used to do one or the other only.

Challenge: Matrices

If you choose to complete this task in your own RStudio, upload the following file:

Matrix_Examples.challenge.Rmd

Click on the 'Matrix_Examples.challenge.Rmd' in the 'Files' section to begin.

This activity and the solution will be discussed at the Collaborate session this week. In the meantime, share and discuss your results in the 'Tutorials' discussion forum.

The solution will also be available here on Friday of this week by clicking on the 'Solution' tab in the top right corner.

The output of the RMD file is also displayed below:

Further reading

Johnson and Wichern:

- 2.5–2.7

Härdle and Simar:

- 4.1–4.3