Probability Refresher

Chapter 3: Random vectors

Tabea Rebafka

September 2022

Master X–HEC 2022–23

Outline

Random vectors I

- A **random vector** is the collection of a finite number of random variables $X_1, \ldots, X_d$ in a vector $\mathrm{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$.
- The **joint distribution** of $\mathrm{X} = (X_1, \ldots, X_d)$, denoted by $\mathbb{P}_{\mathrm{X}}$ or $\mathbb{P}_{(X_1,\ldots,X_d)}$, is the measure on $\mathbb{R}^d$ defined by

$$\mathbb{P}_{\mathrm{X}}(A) = \mathbb{P}\left((X_1, \ldots, X_d) \in A\right).$$

- The distribution of an element X_k of a random vector $\mathrm{X} = (X_1, \ldots, X_d)$ is called the **marginal distribution** of X_k.
- The **cumulative distribution function** $F_{\mathrm{X}} : \mathbb{R}^d \to [0, 1]$ of X is defined by

$$F_{\mathrm{X}}(t_1, \ldots, t_d) = \mathbb{P}(X_1 \leq t_1, \ldots, X_d \leq t_d), \quad (t_1, \ldots, t_d)^T \in \mathbb{R}^d.$$

Random vectors II

- The **characteristic function** $\Phi_{\mathsf{X}} : \mathbb{R}^d \to \mathbb{C}$ of X is defined as

$$\Phi_{\mathsf{X}}(\mathsf{t}) = \mathbb{E}\left[\mathrm{e}^{i\mathsf{t}^T\mathsf{X}}\right] = \mathbb{E}\left[\exp\left\{i\sum_{k=1}^{d} t_k X_k\right\}\right], \quad \mathsf{t} = (t_1, \ldots, t_d)^T \in \mathbb{R}^d.$$

- The **mean vector** of X is the column vector of the expectations of the elements of X given by

$$\mathbb{E}[\mathsf{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_d] \end{pmatrix}.$$

Joint and marginal densities I

Definition

We say that (X, Y) has **joint density** $f : \mathbb{R}^2 \to \mathbb{R}_+$ if for any set $A \subset \mathbb{R}^2$

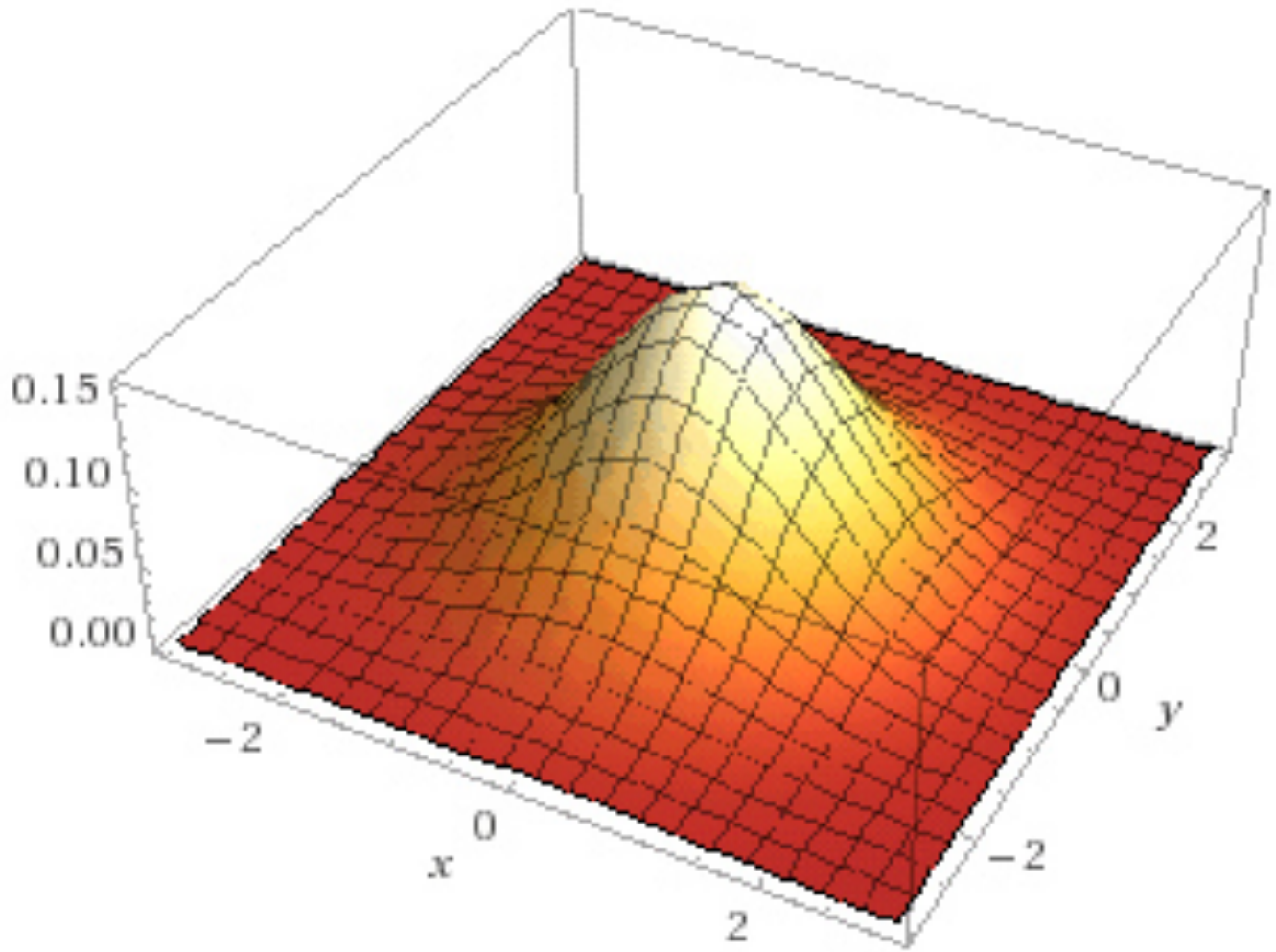
$$\mathbb{P}\left((X, Y) \in A\right) = \iint_A f(x, y)\mathrm{d}x\mathrm{d}y.$$

In this case, for every function ϕ,

$$\mathbb{E}\left[\phi(X, Y)\right] = \iint_{\mathbb{R}^2} \phi(x, y) f(x, y)\mathrm{d}x\mathrm{d}y.$$

This quantity is well-defined either if $\phi \geq 0$ or if $\iint |\phi| f < +\infty$.

Joint and marginal densities II



Joint density of (X, Y), where X and Y are independent $\mathcal{N}(0, 1)$.

Joint and marginal densities III

Definition

Let (X, Y) be a random vector. If the marginal distribution of X (or Y) is continuous, its density is called the **marginal density of X** (or Y).

Proposition

Let (X, Y) be a random vector with density $f(x, y)$. Then X has a density f_X given by

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\, dy$$

Likewise, Y has density f_Y given by $f_Y(y) \mapsto \int_{x \in \mathbb{R}} f(x, y) \mathrm{d}x$.

Joint and marginal densities IV

Exercise

Let (X, Y) be a random vector with density $f(x, y) = (x + y)\mathbb{1}_{[0,1]\times[0,1]}(x, y)$.

- Show that f is indeed a density.
- Compute the marginal density of X.

Joint and marginal densities V

It occurs that X and Y both have densities, while the random vector (X, Y) has not.

Exercise

- Let $X \sim \mathcal{E}(1)$ and $Y = 2X$.
- Note that the random vector (X, Y) lies on the line $D = \{y = 2x\}$ almost surely, that is $\mathbb{P}((X, Y) \in D) = 1$.
- Show that (X, Y) is not continuous.

Independence of random variables I

Definition

Let $X_1, X_2, \ldots$ be random variables defined on the same probability space $(\Omega, \mathbb{P})$.

- We say that $X_1, \ldots, X_n$ are (mutually) **independent** if for all events $B_1, \ldots, B_n$,

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i).$$

- We say that the random variables $X_1, X_2, \ldots$ are **independent and identically distributed (i.i.d.)** if, for every n, $X_1, \ldots, X_n$ are independent and the random variables X_i have all the same law.

Independence of random variables II

Proposition

- Let $X_1, X_2, \ldots, X_n$ be independent and integrable random variables. Then

$$\mathbb{E}[X_1 X_2 \ldots X_n] = \mathbb{E}[X_1]\,\mathbb{E}[X_2] \ldots \mathbb{E}[X_n]$$

Proposition

i. Let $X_1, X_2, \ldots, X_n$ be independent and $\phi_1, \ldots, \phi_n$ be some functions. Then

$X \perp\!\!\!\perp Y$

$\Rightarrow X^2 \perp\!\!\!\perp e^{-Y}$

$$\phi_1(X_1), \phi_2(X_2), \ldots, \phi_n(X_n)$$ independent $(\perp\!\!\!\perp)$

ii. If, moreover, $\phi_k(X_k) \in L^1$, then

$\mathbb{E}[X^2 e^{-Y}] = \mathbb{E}[X^2]\,\mathbb{E}[e^{-Y}]$

$$\mathbb{E}[\phi_1(X_1)\phi_2(X_2) \ldots \phi_n(X_n)] = \mathbb{E}[\phi_1(X_1)]\,\mathbb{E}[\phi_2(X_2)] \ldots \mathbb{E}[\phi_n(X_n)]$$

Independence of random variables III

Proposition

For any random variables $X_1, \ldots, X_n$ the following assertions are equivalent:

i. $X_1, \ldots, X_n$ are independent

ii. The cumulative distribution function F_{X} of the random vector $\mathrm{X} = (X_1, \ldots, X_n)$ can be factorized as follows

$\mathbb{P}(X_1 \le t_1, \ldots, X_n \le t_n)$

$$F_{\mathrm{X}}(t_1, \ldots, t_n) = F_{X_1}(t_1) \ldots F_{X_n}(t_n) \quad \forall\, t_1, \ldots, t_n \in \mathbb{R}$$

iii. The characteristic function Φ_{X} of the random vector $\mathrm{X} = (X_1, \ldots, X_n)$ can be factorized as follows

$e^{a+b} = e^a e^b$

$\mathbb{E}\left[\exp\left\{i \sum_{k=1}^{n} t_k X_k\right\}\right] = \mathbb{E}\left[\prod_{k=1}^{n} e^{i t_k X_k}\right]$

$$\Phi_{\mathrm{X}}(t_1, \ldots, t_n) = \prod_{k=1}^{n} \phi_{X_k}(t_k) \quad \forall\, t_k \in \mathbb{R}$$

Independence of random variables IV

$F_{(X,Y)}(x,y) \overset{\perp\!\!\!\perp}{=} F_X(x) F_Y(y)$

Proposition

$f_{(X,Y)}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} F_{(X,Y)}(x,y)$

$\overset{\perp\!\!\!\perp}{=} \frac{\partial}{\partial x} F_X(x) \cdot \frac{\partial}{\partial y} F_Y(y) = f_X(x) f_Y(y)$

Let X and Y be two random variables.

i. If X and Y are independent and continuous with densities f_X and f_Y, then the random vector (X, Y) is continuous with density

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

ii. Conversely, if (X, Y) has a density $f_{(X,Y)}$ and if there are functions g_1 and g_2 such that the joint density can be written as a product

$$f_{(X,Y)}(x,y) = g_1(x)g_2(y), \quad \text{for (almost) all } x, y \in \mathbb{R},$$

then X, Y are independent.

Independence of random variables V

$g_1(x) = 6x^2 \mathbb{1}_{[0,1]}(x) = c_1 f_X(x)$

$g_2(y) = y \mathbb{1}_{[0,1]}(y) = c_2 f_Y(y)$

$= g_1(x)\, g_2(y)$

Exercise

Let (X, Y) be a random vector with density $f(x, y) = \underbrace{6}_{= \frac{1}{c_1 c_2}} x^2 y \underbrace{\mathbb{1}_{(x,y)\in[0,1]^2}}_{= \mathbb{1}_{[0,1]}(x)\, \mathbb{1}_{[0,1]}(y)}$.

- Show that X and Y are independent. ✓
- Determine the marginal densities of X and Y.

Independence of random variables VI

Independent continuous random variables never take the same value:

Proposition

Let $X_1, X_2, \ldots$ be a sequence of independent random variables with densities $f_k, k = 1, 2 \ldots$. Then the random variables $X_1, X_2, \ldots$ are all pairwise distinct with probability one, that is,

$$\mathbb{P}(X_i \neq X_j, \forall i \neq j) = 1$$

Sums of independent random variables I

Let X and Y be random variables. What can we say about $X + Y$?

- By linearity of the expectation

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

- If X and Y are **independent**,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

$$\mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2$$

$$= \mathbb{E}[X^2] + \cancel{2\underbrace{\mathbb{E}[XY]}_{= \mathbb{E}X\,\mathbb{E}Y \perp\!\!\!\perp}} + \mathbb{E}[Y^2] - (\mathbb{E}[X])^2 - \cancel{2\mathbb{E}X\mathbb{E}Y} - (\mathbb{E}[Y])^2$$

Sums of independent random variables II

Theorem

Let X and Y be independent random variables with densities f and g. Then the sum $X + Y$ has a density given by

$$f_{X+Y}(t) = f * g\ (t),$$

where $u \mapsto f * g\ (u)$ is the **convolution function** of f and g defined as

$$f * g\ (u) = \int_{x \in \mathbb{R}} f(x) g(u - x) \mathrm{d}x.$$

Sums of independent random variables III

Exercise

- Let X and Y be i.i.d. with the exponential distribution $\mathcal{E}(1)$, i.e. with density $f(x) = \mathrm{e}^{-x}\mathbb{1}_{x\geq 0}$.
- Compute the density of the sum $U = X + Y$.

Sums of independent random variables IV

The characteristic function is an efficient tool to handle sums of independent random variables.

Proposition

Let X and Y be independent random variables. Then the characteristic function of the sum $X + Y$ is given by

$$\Phi_{X+Y}(t) = \phi_X(t)\,\phi_Y(t)$$

$$= \mathbb{E}[\exp\{i(X+Y)t\}] = \mathbb{E}[e^{iXt} \cdot e^{iYt}]$$

$$\overset{\perp\!\!\!\perp}{=} \mathbb{E}[e^{iXt}]\,\mathbb{E}[e^{iYt}]$$

Sums of independent random variables V

- Let X and Y be independent and normally distributed with $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$.
- Then $X + Y$ is also a gaussian random variable with

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

- For the proof uses the characteristic function of the normal distribution $\mathcal{N}(\mu, \sigma^2)$ given by

$$\Phi_X(t) = \exp\left(it\mu - \frac{t^2\sigma^2}{2}\right).$$

Covariance I

$\mathbb{E}[X^2] < \infty$
$\mathbb{E}[Y^2] < \infty$

Definition

Let X and Y be random variables in L^2. The **covariance** of X and Y is defined by

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] \\ &= \underbrace{\mathbb{E}[XY]}_{= \mathbb{E}X\,\mathbb{E}Y \text{ if } \perp\!\!\!\perp} - \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

$Y = c$ constant $= c\,\mathbb{E}[X] - \mathbb{E}X \cdot c = 0$

Covariance II

Proposition

Let $X, Y, X_i, Y_j \in L^2$.

i. (**Symmetry**) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.
ii. (**Bilinearity**) For any constants $a_i, b_i \in \mathbb{R}$,

$$\mathrm{Cov}\left(\sum_i a_i X_i, \sum_j b_j Y_j\right) = \sum_i a_i \sum_j b_j \; \mathrm{Cov}(X_i, Y_j)$$

iii. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
iv. If X and Y are independent, $\mathrm{Cov}(X, Y) = 0$
v. For any constant c, $\mathrm{Cov}(X, c) = 0$ $\quad \mathrm{Cov}(X+c, Y) = \mathrm{Cov}(X, Y)$
vi. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$

Covariance III

Definition

Let $\mathrm{X} = (X_1, \ldots, X_d)$ be a random vector such that $X_k \in L^2$ for $k = 1, \ldots, d$. The **covariance matrix** $\mathrm{Cov}(\mathrm{X})$ of X is the $d \times d$ matrix with entries $\mathrm{Cov}(X_i, X_j)$. More precisely,

$$\mathrm{Cov}(\mathrm{X}) = (\mathrm{Cov}(X_i, X_j))_{i,j} = \begin{array}{cc} & j \\ i & \begin{pmatrix} & \vdots & \\ \ldots & \mathrm{Cov}(X_i, X_j) & \ldots \\ & \vdots & \end{pmatrix} \end{array}.$$

Covariance IV

Example

- Let $X_1, \ldots, X_d$ be i.i.d.
- The covariance matrix of $\mathrm{X} = (X_1, \ldots, X_d)$ is given by

$$\mathrm{Cov}(\mathrm{X}) = \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix}, = \sigma^2 I_d$$

identity matrix

$= \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \in \mathbb{R}^{d \times d}$

where $\sigma^2 = \mathrm{Var}(X_1)$.

Covariance V

Using vector and matrix notation, we obtain an expression of the covariance matrix that is easier to handle.

Proposition

$$\mathrm{Cov}(\mathsf{X}) = \mathbb{E}\left[(\mathsf{X} - \mathbb{E}[\mathsf{X}])(\mathsf{X} - \mathbb{E}[\mathsf{X}])^T\right].$$

Proposition

i. Any covariance matrix $\mathrm{Cov}(\mathsf{X})$ is symmetric.
ii. Any covariance matrix $\mathrm{Cov}(\mathsf{X})$ is positive semi-definite, i.e. for all $\mathsf{t} = (t_1, \ldots, t_d) \in \mathbb{R}^d$ we have $\mathsf{t}^T \mathrm{Cov}(\mathsf{X})\mathsf{t} \geq 0$.

Covariance VI

For linear transformations of a random vector, there are simple formulas for the mean and the covariance matrix.

Proposition

Let $X \in \mathbb{R}^d$ be a random vector with finite covariance matrix. Let M be a $p \times d$ matrix and $a \in \mathbb{R}^p$ a vector. Then the mean and the covariance vector of the random vector $MX + a \in \mathbb{R}^p$ are given by

$$\mathbb{E}[MX + a] = M\mathbb{E}[X] + a \quad \text{and} \quad \mathrm{Cov}(MX + a) = \mathrm{Cov}(MX) = M\,\mathrm{Cov}(X)M^T$$

$$= \mathbb{E}\left[(MX - \mathbb{E}MX)(MX - \mathbb{E}MX)^T\right]$$

$$= \mathbb{E}\left[M(X - \mathbb{E}X) \cdot \underbrace{(M(X - \mathbb{E}X))^T}_{= (X - \mathbb{E}X)^T M^T}\right]$$

$$= M\,\mathrm{Cov}(X)\,M^T$$

Correlation I

Definition

Let $X, Y \in L^2$ and $\mathrm{Var}(X) > 0$ and $\mathrm{Var}(Y) > 0$. The **correlation** or **Pearson's correlation coefficient** between X and Y is defined by

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

We say that X and Y have

- **positive correlation** if $\rho_{X,Y} > 0$ and
- **negative correlation** if $\rho_{X,Y} < 0$.

Correlation II

- If $X \perp\!\!\!\perp Y$, then $\rho_{X,Y} = 0$.
- The correlation is invariant with respect to affine transformations: for any $a \neq 0$, $b \neq 0$ $c, d \in \mathbb{R}$,

$$\rho_{aX+c,bY+d} = \text{sign}(ab)\rho_{X,Y},$$

where $\text{sign}(u) = \mathbb{1}\{u > 0\} + \mathbb{1}\{u < 0\}$.

Correlation III

Theorem

- $-1 \leq \rho_{X,Y} \leq 1$.
- $\rho_{X,Y} = 1 \iff \exists a > 0, b \in \mathbb{R}$ such that $Y = aX + b$ $a.s.$
- $\rho_{X,Y} = -1 \iff \exists a < 0, b \in \mathbb{R}$ such that $Y = aX + b$ $a.s.$

The correlation $\rho_{X,Y}$ measures the degree of **linear** dependence between X and Y.