

Math 4261: Mathematical Statistics I

Lecture 3: Probability review II

Joint distributions and the multivariate normal

Ref: Blitzstein–Hwang

Fall 2026

Outline

Recap

Joint distributions and independence

Covariance matrices

Multivariate normal

Outline

Recap

Joint distributions and independence

Covariance matrices

Multivariate normal

Recap: The Gaussian distribution

Gaussian (or Normal)

- ▶ $X \sim \mathcal{N}(\mu, \sigma^2)$, if it has **PDF** supported on $\mathbb{R}$:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

- ▶ $X \sim \mathcal{N}(0, 1)$ is the **standard** Gaussian (or Normal)
- ▶ **CDF** of Gaussian often denoted $\Phi(x) = \mathbb{P}[X \leq x]$.
- ▶ **Linear functions of Gaussians are Gaussians:**

$$X \sim \mathcal{N}(\mu, \sigma^2) \implies \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

$$X \sim \mathcal{N}(0, 1) \implies \mu + \sigma X \sim \mathcal{N}(\mu, \sigma^2).$$

Recap: Quantiles and the quantile transform

For any $\alpha \in (0, 1)$, the α **quantile** of a distribution F is

$$q_\alpha := F^{-1}(\alpha) = \inf \{x \in \mathbb{R} : F(x) \geq \alpha\}.$$

The **upper- α point** is

$$q^{(\alpha)} := q_{1-\alpha} = F^{-1}(1 - \alpha).$$

Suppose F is a continuous, strictly increasing CDF. Then

$$\begin{aligned} U \sim \text{Unif}(0, 1) &\implies X = F^{-1}(U) \sim F, \\ X \sim F &\implies U = F(X) \sim \text{Unif}(0, 1). \end{aligned}$$

Simulation: starting from the uniform U , we can generate random variables from a desired distribution F , provided F^{-1} is tractable.

Outline

Recap

Joint distributions and independence

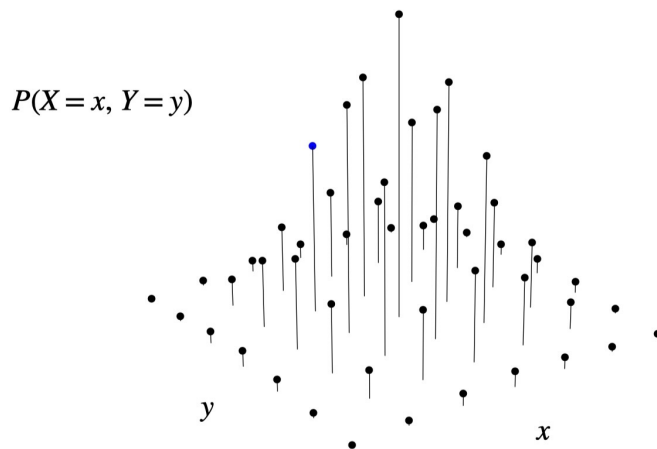
Covariance matrices

Multivariate normal

Discrete joint distributions

The **joint PMF** of a pair of discrete r.v.'s X and Y is a full specification of the probabilities that X and Y can take:

$$p_{X,Y}(x, y) = \mathbb{P}[X = x, Y = y].$$



Single r.v.: $p_X(x)$ = probability X takes the value x .

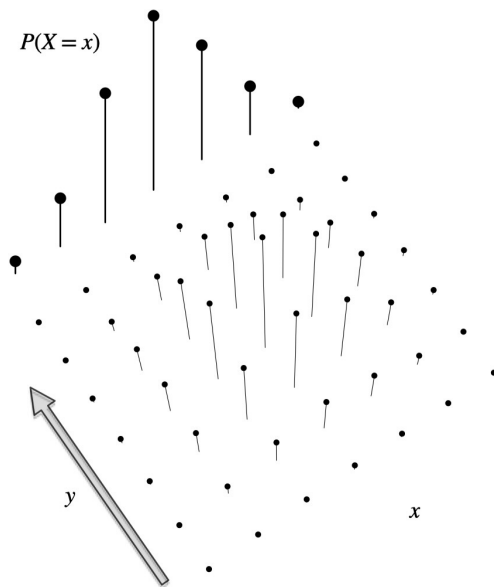
Joint distribution: $p_{X,Y}(x, y)$ = probability the **tuple** (X, Y) takes the value (x, y) .

Marginal distributions

The PMF of X *regardless of* Y is the **marginal** dist. of X .

To compute the marginal from the joint, we **sum out** Y :

$$\mathbb{P}[X = x] = \sum_y \mathbb{P}[X = x, Y = y].$$



Example: BH Example 7.1.5

X = indicator r.v. of being a **smoker**

Y = indicator r.v. of having **lung cancer**

Suppose the following hypothetical joint distribution:

Complete the table and find the marginal distributions of X and Y .

	$Y = 1$	$Y = 0$	Total
$X = 1$	$\frac{5}{100}$	$\frac{20}{100}$	$\frac{25}{100}$
$X = 0$	$\frac{3}{100}$	$\frac{72}{100}$	$\frac{75}{100}$
Total	$\frac{8}{100}$	$\frac{92}{100}$	$\frac{100}{100}$

"Margins"

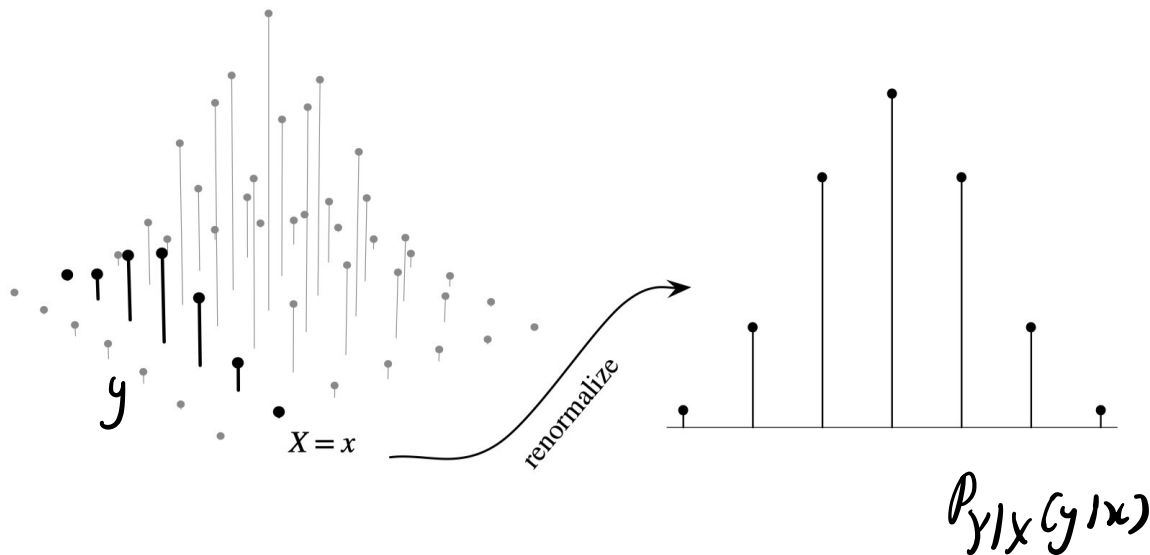


$$P[Y=1] = \sum_x P[Y=1, X=x] = P[Y=1, X=1] + P[Y=1, X=0]$$

Conditional distributions

The **conditional PMF** of Y given $X = x$ is:

$$p_{Y|X}(y|x) = \mathbb{P}[Y = y \mid X = x] = \frac{\mathbb{P}[Y = y, X = x]}{\mathbb{P}[X = x]} \quad \leftarrow \text{defined if } > 0$$



Example: BH Example 7.1.5

X = indicator r.v. of being a **smoker**

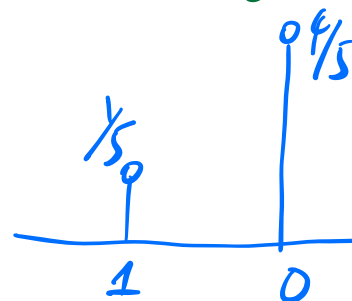
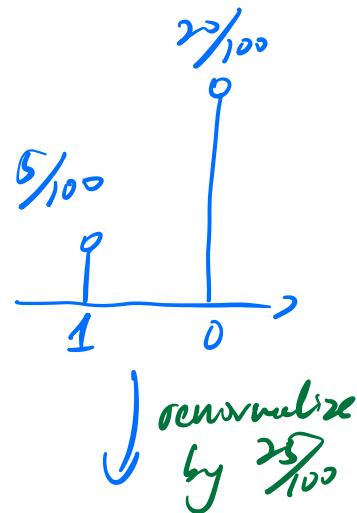
Y = indicator r.v. of having **lung cancer**

Compute $\mathbb{P}[Y = 1 \mid X = 1]$ and $\mathbb{P}[Y = 1 \mid X = 0]$.

Cond. dist.
of Y given $X=1$

	$Y = 1$	$Y = 0$	Total
$X = 1$	$\frac{5}{100}$	$\frac{20}{100}$	$\frac{25}{100}$
$X = 0$	$\frac{3}{100}$	$\frac{72}{100}$	$\frac{75}{100}$
Total	$\frac{8}{100}$	$\frac{92}{100}$	$\frac{100}{100}$

$$\begin{aligned}\mathbb{P}[Y=1 \mid X=1] &= \frac{\mathbb{P}[Y=1, X=1]}{\mathbb{P}[X=1]} \\ &= \frac{5/100}{25/100} = \frac{1}{5}\end{aligned}$$



Dependencies among discrete r.v.'s

Rephrasing in terms of Joint-Marginal-Conditional:

X and Y **independent**

$$\iff p_{X,Y}(x,y) = p_X(x)p_Y(y) \quad \text{for all } x \text{ and } y$$

$$\iff p_{Y|X}(y|x) = p_Y(y) \quad \text{for all } x \text{ and } y \text{ such that } \mathbb{P}[X = x] > 0$$

$$\iff p_{X|Y}(x|y) = p_X(x) \quad \text{for all } x \text{ and } y \text{ such that } \mathbb{P}[Y = y] > 0$$

- ▶ Joint contains **as much** information as the marginals
- ▶ Conditional is equal to marginal.

Without independence

- ▶ Joint contains **more** information than the marginals.
- ▶ Need both marginals and conditionals to derive the joint:

$$p_{X,Y}(x,y) = p_{Y|X}(y|x) \cdot p_X(x)$$

Example: BH Example 7.1.8

X = indicator r.v. of being a **smoker**

Y = indicator r.v. of having **lung cancer**

Are X and Y independent? Explain.

①

	$Y = 1$	$Y = 0$	Total
$X = 1$	$\frac{5}{100}$	$\frac{20}{100}$	$\frac{25}{100}$
$X = 0$	$\frac{3}{100}$	$\frac{72}{100}$	$\frac{75}{100}$
Total	$\frac{8}{100}$	$\frac{92}{100}$	$\frac{100}{100}$

② Previously computed $P[Y=1|X=1] = \frac{1}{5} = 0.2$

but $P[Y=1] = \frac{8}{100} = 0.08$

Hence X not independent of Y .

Example: Conditional independence

Suppose there are two coins:

Let $A_i = \{H \text{ on } i^{\text{th}} \text{ flip}\}$

- ▶ Coin 1 is biased with probability 90% to land on heads.
- ▶ Coin 2 is fair.

I draw a coin and start flipping it. Are the outcomes of the coin flips **independent**?

$$\begin{aligned} P[A_i] &= \overset{\text{LTP}}{P[A_i | C_1] P[C_1] + P[A_i | C_2] P[C_2]} \\ &= \frac{9}{10} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = 0.7 \quad \text{for any } i \end{aligned}$$

$$P[A_2 | A_1] = \frac{P[A_1, A_2]}{P[A_1]} \overset{\text{LTP}}{=} \frac{\left(\frac{9}{10}\right)^2 \frac{1}{2} + \left(\frac{1}{2}\right)^2 \frac{1}{2}}{0.7} = \frac{0.53}{0.7} = 0.757$$

So $P[A_2 | A_1] > P[A_2]$ i.e. A_1, A_2 not independent.

$$\text{However } P[A_1, A_2 | C_{\#}] = P[A_1 | C_{\#}] P[A_2 | C_{\#}]$$

So A_1, A_2 cond. indpt. ^{any coin} given coin drawn.

Conditional independence

Independence (of events and random variables) extends naturally to conditional scenarios.

- ▶ **Events** A and B are **conditionally independent** given an event C if

$$\mathbb{P}[A, B \mid C] = \mathbb{P}[A \mid C] \mathbb{P}[B \mid C].$$

- ▶ **Discrete random variables** X and Y are **conditionally independent** given an **event** C if

$$\mathbb{P}[X = x, Y = y \mid C] = \mathbb{P}[X = x \mid C] \mathbb{P}[Y = y \mid C].$$

- ▶ **Discrete random variables** X and Y are **conditionally independent** given an **r.v.** Z if

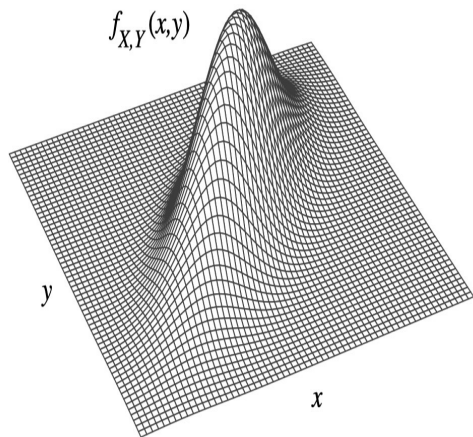
$$\mathbb{P}[X = x, Y = y \mid Z = z] = \mathbb{P}[X = x \mid Z = z] \mathbb{P}[Y = y \mid Z = z].$$

All conditional probabilities are understood where the conditioning event has positive probability.

Continuous r.v.'s: joint PDF

The **joint PDF** $f_{X,Y}(x,y)$ of a pair of continuous r.v.'s X and Y is a function on $\mathbb{R}^2$ with

$$\mathbb{P}[(X,Y) \in A] = \iint_A f_{X,Y}(x,y) \, dx \, dy.$$



Non-negative $f_{X,Y}(x,y) \geq 0$

Normalization

$$\iint_{\mathbb{R}^2} f_{X,Y}(x,y) \, dx \, dy = 1$$

Marginalization and independence

The **marginal PDFs** are obtained by integrating out the other variable:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx.$$

The variables are **independent** exactly when

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

for all x, y (up to changes on sets of area zero). That is,

joint PDF factorizes into product of marginal PDFs.

Conditioning

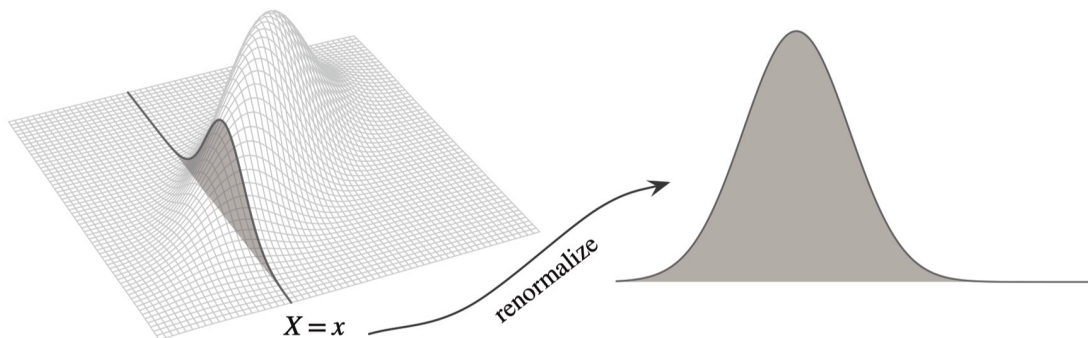
The **conditional PDF** of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad \text{provided } f_X(x) > 0.$$

That is

$\text{conditional PDF} = \frac{\text{Joint PDF}}{\text{Marginal PDF}}$

Tells us how to **update** the distribution of Y , having **observed that** $X = x$.



Applications

For us, conditional and marginal PDFs are useful for

1. Finding **joint PDF**

$$f_{X,Y} = f_{Y|X} \cdot f_X.$$

2. **Flipping the conditioning**, i.e. Bayes rule

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) \cdot f_X(x)}{f_Y(y)} \quad \text{for } f_Y(y) > 0$$

3. In tandem with a **continuous version of LOTP**

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) \, dy$$

Morally the same as the discrete versions with p 's instead of f 's, and with sums and integrals appropriately interchanged.

Outline

Recap

Joint distributions and independence

Covariance matrices

Multivariate normal

Two important quantities

Let X and Y be two random variables.

Covariance between X and Y is

$$\begin{aligned}\text{Cov}[X, Y] &= \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

The **Correlation** between X and Y is

$$\rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var } X} \sqrt{\text{Var } Y}}.$$

These measure the tendency for X and Y to **move in the same direction**.

Properties of covariance

Covariance is **translation invariant**:

$$\text{Cov}[X + a, Y + b] = \text{Cov}[X, Y].$$

Furthermore, covariance is **bilinear**:

$$\begin{aligned}\text{Cov}[aX, bY] &= ab \text{Cov}[X, Y], \\ \text{Cov}[X + Y, W + Z] &= \text{Cov}[X, W] + \text{Cov}[X, Z] \\ &\quad + \text{Cov}[Y, W] + \text{Cov}[Y, Z].\end{aligned}$$

Consequently,

$$\text{Cov}[a_1X_1 + \cdots + a_nX_n, b_1Y_1 + \cdots + b_mY_m] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}[X_i, Y_j].$$

We also have

- ▶ $\text{Cov}[c, X] = 0$ for any constant c ,
- ▶ $\text{Cov}[X, X] = \text{Var } X$,
- ▶ $\text{Cov}[X, Y] = \text{Cov}[Y, X]$.

Uncorrelated

X and Y are **uncorrelated** if

$$\text{Cov}[X, Y] = 0 \iff \rho(X, Y) = 0 \iff \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Independence is a **stronger** notion than uncorrelated:

- ▶ **Independent** $\implies$ **uncorrelated**: $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
- ▶ **Uncorrelated** $\not\Rightarrow$ **independent**.

Example

Choose a point (X, Y) uniformly at random to be one of the four vertices of a ‘diamond’ $(-1, 0)$, $(0, 1)$, $(1, 0)$, $(0, -1)$. Check that

$$\mathbb{E}[XY] = 0 = \mathbb{E}[X] = \mathbb{E}[Y].$$

But X and Y are **dependent**.

Variance of a sum: covariance matters

For **any** random variables $X_1, \dots, X_n$,

$$\text{Var}[X_1 + \dots + X_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n] + \sum_{i \neq j} \text{Cov}[X_i, X_j].$$

For two variables,

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \text{Cov}[X, Y].$$

Covariance matrix

Let $\mathbf{X} = (X_1, \dots, X_p)^\top$ be a random vector with mean vector

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}].$$

Its **covariance matrix** is

$$\begin{aligned}\boldsymbol{\Sigma} &= \text{Cov}[\mathbf{X}] = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top], \\ \Sigma_{ij} &= \text{Cov}[X_i, X_j].\end{aligned}$$

Equivalently,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \text{Cov}[X_1, X_1] & \cdots & \text{Cov}[X_1, X_p] \\ \vdots & \ddots & \vdots \\ \text{Cov}[X_p, X_1] & \cdots & \text{Cov}[X_p, X_p] \end{bmatrix}.$$

The diagonal entries are the **variances**; the off-diagonal entries are the **covariances**.

Covariance matrices encode linear combinations

For deterministic vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$,

$$\begin{aligned}\mathbb{E}[\mathbf{a}^\top \mathbf{X}] &= \mathbf{a}^\top \boldsymbol{\mu}, \\ \text{Cov}[\mathbf{a}^\top \mathbf{X}, \mathbf{b}^\top \mathbf{X}] &= \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{b}, \\ \text{Var}[\mathbf{a}^\top \mathbf{X}] &= \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}.\end{aligned}$$

Thus the covariance matrix encodes the variance of **every linear combination** of the coordinates of $\mathbf{X}$.

Every covariance matrix is **symmetric** and **positive semidefinite**:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^\top, \quad \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} = \text{Var}[\mathbf{a}^\top \mathbf{X}] \geq 0.$$

Derivation of variance of linear combination

$$\text{Var}[\mathbf{a}^\top \mathbf{X}] = \mathbf{a}^\top \Sigma \mathbf{a}$$

$$\begin{aligned}\text{Var}[\mathbf{a}^\top \mathbf{X}] &= \text{Var}\left[\sum_{i=1}^n a_i X_i\right] \\&= \sum_{i=1}^n \underbrace{\text{Var}[a_i X_i]}_{\substack{a_i^2 \text{Cov}[X_i, X_i] \\ = a_i^2 \Sigma_{ii}}} + 2 \sum_{i < j} \underbrace{\text{Cov}[a_i X_i, a_j X_j]}_{a_i a_j \Sigma_{ij}} \\&= \sum_{i,j} a_i a_j \Sigma_{ij} = \mathbf{a}^\top \Sigma \mathbf{a}\end{aligned}$$

Outline

Recap

Joint distributions and independence

Covariance matrices

Multivariate normal

What “jointly Gaussian” means

A random vector $\mathbf{X} = (X_1, \dots, X_p)^\top$ has a **multivariate normal distribution** if every linear combination

$$\mathbf{a}^\top \mathbf{X} = a_1 X_1 + \dots + a_p X_p$$

has a univariate Gaussian distribution.

Plugging in $\mathbf{a}$ to as the standard basis vectors $e_1, \dots, e_p$ shows that every X_j is marginally Gaussian.

The converse is false: having Gaussian marginals is **not** enough to make the variables jointly Gaussian.

Example: Gaussian marginals are not enough

Let $Z \sim \mathcal{N}(0, 1)$ and, independently, let

$$\mathbb{P}[S = 1] = \mathbb{P}[S = -1] = \frac{1}{2}.$$

Set $X = Z$ and $Y = SZ$. Show that X and Y are **marginally** standard Gaussian and uncorrelated, but (X, Y) is not jointly Gaussian.

$$\cdot) Y = SZ \stackrel{d}{=} Z \sim \mathcal{N}(0, 1).$$

$$\begin{aligned} \cdot) \text{Cov}[X, Y] &= \text{Cov}[Z, SZ] = \mathbb{E}[SZ^2] \\ &= \underbrace{\mathbb{E}S}_{=0} \cdot \underbrace{\mathbb{E}Z^2}_{=1} = 0 \end{aligned}$$

$$\cdot) X + Y = (1 + S)Z$$

$$P[X + Y = 0] = P[S = -1] = \frac{1}{2} \Rightarrow \begin{array}{l} X+Y \text{ not Gaussian} \\ \text{so } (X, Y) \text{ not} \\ \text{jointly Gaussian} \end{array}$$

The Gaussian exception

In general,

uncorrelated $\not\Rightarrow$ independent.

For **jointly Gaussian** random variables, however,

uncorrelated $\iff$ independent.

Thus, if $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then X_i and X_j are independent exactly when $\Sigma_{ij} = 0$. In particular, a diagonal covariance matrix means the coordinates are independent.

Multivariate normal

A collection of p random variables $\mathbf{Z}$ has a **multivariate normal distribution** with mean vector $\boldsymbol{\mu} \in \mathbb{R}^p$ and (symmetric) covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ written $\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e.

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1p} \\ \Sigma_{21} & \cdots & \Sigma_{2p} \\ \vdots & \ddots & \vdots \\ \Sigma_{p1} & \cdots & \Sigma_{pp} \end{bmatrix} \right),$$

When $\boldsymbol{\Sigma}$ is positive definite, its joint PDF is

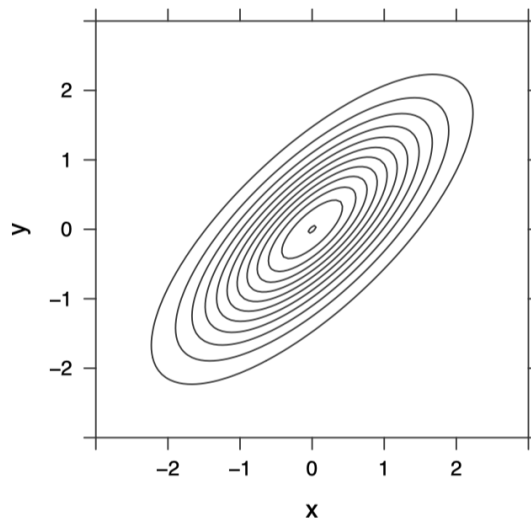
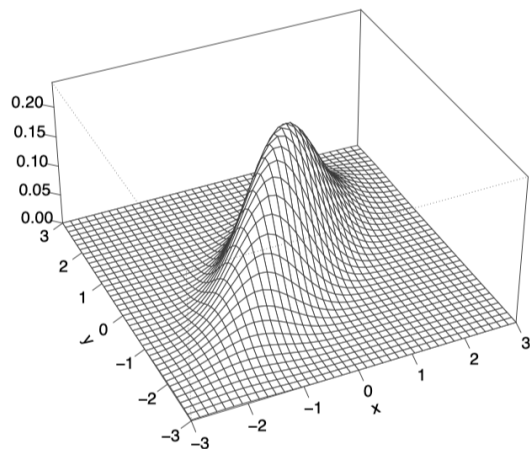
$$f(\mathbf{z}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})} \text{ where } \mathbf{z} = (z_1, \dots, z_p).$$

- ▶ Marginally $Z_j \sim \mathcal{N}(\mu_j, \Sigma_{jj})$. (Diagonal entries of $\boldsymbol{\Sigma}$ are variances)
- ▶ $\text{Cov}[Z_i, Z_j] = \Sigma_{ij} = \Sigma_{ji}$. (Off-diagonals of $\boldsymbol{\Sigma}$ are covariances)
- ▶ **Jointly Gaussian property:** any linear combination of the Z_j 's has a Gaussian distribution.

Multivariate normal

Visualization of $\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in $p = 2$. PDF

$$f(\mathbf{z}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}$$



Standard multivariate normal

An important special case is the **standard multivariate normal** which describes the joint distribution of

$$Z_1, \dots, Z_p \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1).$$

The notation is

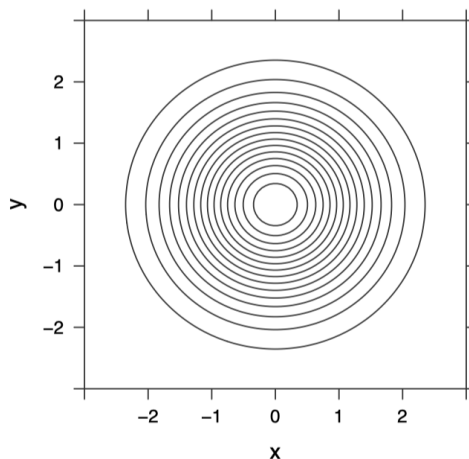
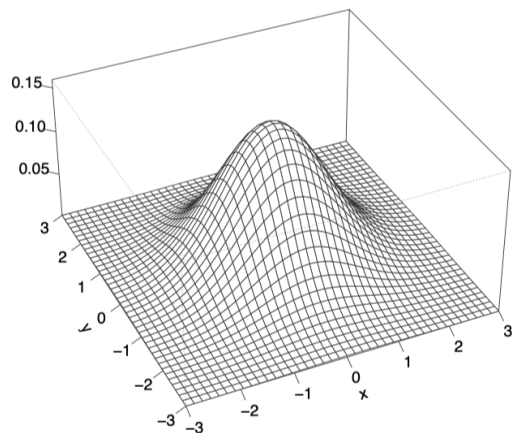
$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p \times p})$$

which corresponds to having $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_{p \times p}$ (identity matrix).

Joint PDF:

$$f(z_1, \dots, z_p) = \prod_{i=1}^p \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2} = \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2}(z_1^2 + \dots + z_p^2)}.$$

Symmetry of the standard multivariate normal



Observe that this joint PDF

$$f(z_1, \dots, z_p) = \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2}(z_1^2 + \dots + z_p^2)}$$

depends on $z = (z_1, \dots, z_p)$ only via its **length** $\sqrt{z_1^2 + \dots + z_p^2} = \|z\|_2$.

Thus the standard multivariate normal distribution is **symmetric** with respect to rotations/reflections about the origin.

What symmetry means precisely

Let $Q \in \mathbb{R}^{p \times p}$ be **orthogonal**: $Q^\top Q = QQ^\top = I_p$. For $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_p)$, the precise claim is

$$Q\mathbf{Z} \stackrel{d}{=} \mathbf{Z}.$$

In a later class, we will see that the multivariate **change-of-variables formula** gives, for $\mathbf{Y} = Q\mathbf{Z}$,

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det Q|} f_{\mathbf{Z}}(Q^{-1}\mathbf{y}).$$

 *det of the Jacobian*

What symmetry means precisely

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$$

$$\mathbf{y} = \mathbf{Q}\mathbf{Z}$$

$$|\det \mathbf{Q}| = 1$$

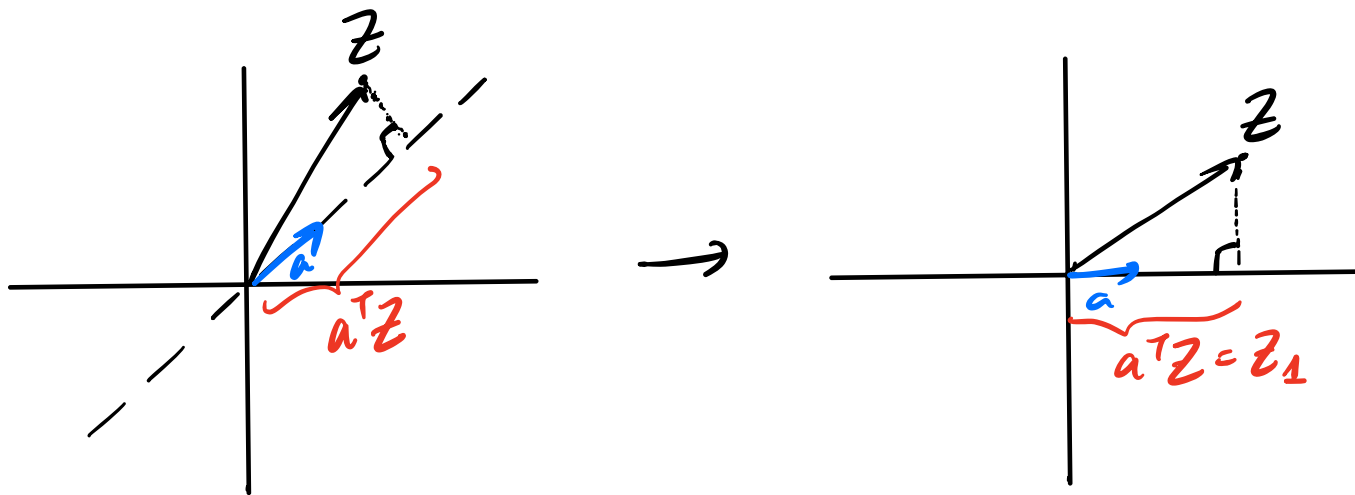
$$\begin{aligned} f_{\mathbf{y}}(\mathbf{y}) &= f_{\mathbf{Z}}(\mathbf{Q}^T \mathbf{y}) = c_p \cdot \exp\left\{-\frac{1}{2} \underbrace{\|\mathbf{Q}^T \mathbf{y}\|_2^2}_{\|\mathbf{y}\|_2^2}\right\} \\ &= f_{\mathbf{Z}}(\mathbf{y}) \end{aligned}$$

Consequences of symmetry—I

Let $\mathbf{Z} = (Z_1, \dots, Z_p) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p \times p})$ and let $\mathbf{a} \in \mathbb{R}^p$ be any vector having length 1. Then

$$\mathbf{a}^\top \mathbf{Z} \sim \mathcal{N}(0, 1)$$

because this distribution is the same as for $(1, 0, \dots, 0)^\top \mathbf{Z} = Z_1$.



Consequences of symmetry—I

In symbols, $\exists R$ orthonormal s.t.

$$Ra = (1, 0, \dots, 0)^T$$

Hence $a^T z = a^T \underbrace{R^T R}_{\mathbf{I}} z$

$$= (Ra)^T \underbrace{\mathbf{I}}_{\boxed{=}} (Ra)^T z = z_1$$