CS 107, Probability, Spring 2020
Lecture 31

Michael Poghosyan
mpoghosyan@aua.am

AUA

20 April 2020

Content

- Important Multivariate Distributions
  - Discrete Multivariate Distributions
    - Multinoulli Distribution
    - Multinomial Distribution
  - Continuous Multivariate Distributions
    - Multivariate Uniform Distribution
    - Multivariate Normal Distribution

Important Multivariate Distributions
Discrete Multivariate Distributions

Multinoulli Distribution

Multinoulli Distribution

Multinoulli Distribution (the other name is the Categorical Distribution) is the multivariate analogue of the Bernoulli distribution (hence the name).
To define it, we fix numbers $p_1, p_2, ..., p_m$ such that

$$p_k \geq 0 \qquad \text{and} \qquad p_1 + p_2 + ... + p_m = 1.$$

Multinoulli Distribution

Multinoulli Distribution

We say that the m-dim random vector $\mathbf{X} = [X_1, X_2, ..., X_m]^T$ has a Multinoulli Distribution with parameters (probabilities) $p_1, p_2, ..., p_m$, and we'll write $\mathbf{X} \sim Multinoulli(p_1, p_2, ..., p_m)$, if

- each r.v. X_k is taking 2 values, 0 or 1
- the Joint PMF has the form

$$\begin{aligned} \mathbb{P}(X_1 = 1, X_2 = 0, ..., X_m = 0) &= p_1, \\ \mathbb{P}(X_1 = 0, X_2 = 1, ..., X_m = 0) &= p_2, \\ &\vdots \\ \mathbb{P}(X_1 = 0, X_2 = 0, ..., X_m = 1) &= p_m, \end{aligned}$$

and (of course) all other probabilities are 0.

Multinoulli Distribution

Say,

$$\mathbb{P}(X_1 = 1, X_2 = 1, ..., X_m = 0) = 0.$$

Idea: Assume we have an experiment, where

- one and exactly one of m possibilities (categories) A_1, A_2, ..., A_m can happen;
- $\mathbb{P}(A_k) = p_k$, $k = 1, ..., m$;
- the r.v. $X_k = 1$ if A_k happens, and $X_k = 0$ otherwise.

In other words, X_k is the Indicator of the event A_k, or, in other way, $\mathbf{X}$ is the one-hot encoding of the experiment's result.

Note: Please note that it is enough to give/describe only $m-1$ r.v.s out of $X_1, ..., X_m$, since the other one is uniquely determined by that $m-1$ ones.

Multinoulli Distribution: Example

Example: We have a box full of 10 black, 20 white and 50 red balls. We pick a ball at random. We can describe the outcome as a Multinoulli distributed r. vector, in the following way:

- Our events are:
$$\begin{aligned} A_1 &= \text{The ball is black;} \\ A_2 &= \text{The ball is white;} \\ A_3 &= \text{The ball is red;} \end{aligned}$$
- The probabilities are: $p_1 = \mathbb{P}(A_1) = \frac{1}{8}$, $p_2 = \mathbb{P}(A_2) = \frac{2}{8}$, $p_3 = \mathbb{P}(A_3) = \frac{5}{8}$;
- our r. vector $\mathbf{X} = [X_1, X_2, X_3]^T$ is showing which color (event) happened: $X_1 = 1$ means the ball taken was black, $X_2 = 1$ means it was white, and $X_3 = 1$, if it was red. So the result $(0, 0, 1)$ means the ball taken was red, and
$$\mathbb{P}\Big(\mathbf{X} = (0,0,1)\Big) = \mathbb{P}(X_1 = 0, X_2 = 0, X_3 = 1) = \frac{5}{8}.$$

Multinoulli Distribution: Example

Example: $Benoulli(p)$ distribution is just a particular case of the Multinoulli distribution, in fact, we can write

$$Bernoulli(p) = Multinoulli(p, 1-p).$$

To explain, assume we have an experiment, and an event called Success. Assume the $\mathbb{P}(Success) = p$. Then $\mathbb{P}(Failure) = 1-p$. Now, let the r. vector $\mathbf{X} = (X_1, X_2)$ be defined in the following way: $X_1 = 1$ and $X_2 = 0$, if Success happened, and $X_1 = 0$, $X_2 = 1$, if Failure was the result. So our r.v. $\mathbf{X}$ will have the following PMF:

$$\mathbb{P}(\mathbf{X} = (1,0)) = \mathbb{P}(Success) = p;$$
$$\mathbb{P}(\mathbf{X} = (0,1)) = \mathbb{P}(Failure) = 1-p;$$
$$\mathbb{P}(\mathbf{X} = (1,1)) = \mathbb{P}(\mathbf{X} = (0,0)) = 0.$$

Now, X_1 is our ordinary $Bernoulli(p)$ r.v., and knowing its value, we can uniquely determine X_2.

Marginals of Multinoulli Distribution

Now, assume $\mathbf{X} \sim Multinoulli(p_1, p_2, ..., p_m)$. Then it is easy to prove that

$$X_k \sim Bernoulli(p_k), \qquad k = 1, ..., m.$$

Multinomial Distribution

Multinomial Distribution

Distribution Name: Multinomial

Parameters: $n, p_1, \dots, p_m$ $\left(n \in \mathbb{N},\ p_k \geq 0,\ \sum_{k=1}^{m} p_k = 1\right)$

Multinomial Distribution

We say that the r. vector $\mathbf{X} = [X_1, X_2, \dots, X_m]^T$ has a Multinomial Distribution with probabilities $\mathbf{p} = [p_1, p_2, \dots, p_m]^T$, and we write

$$\mathbf{X} = [X_1, X_2, \dots, X_m]^T \sim Multinomial(n, p_1, p_2, \dots, p_m),$$

if its PMF is given by:

$$\mathbb{P}(X_1 = k_1, X_2 = k_2, \dots, X_m = k_m) = \binom{n}{k_1, k_2, \dots, k_m} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_m^{k_m}$$

for any $k_1, \dots, k_m \in \mathbb{N} \cup \{0\}$, with $k_1 + k_2 + \dots + k_m = n$.

Multinomial Distribution

Idea: The Multinomial distribution is the generalization of the Binomial distribution, and is modeling the independent repetition of Multinoulli Experiments. More precisely,

- Our Experiment consists of n times independent repetition (trials) of the same Simple Experiment;
- In our Simple Experiment, exactly one of m possibilities (categories) A_1, A_2, ..., A_m can happen;
- $\mathbb{P}(A_k) = p_k$;
- The r.v. X_k shows how many times the outcome A_k will appear in n trials.

Marginals of Multinomial Distribution

It can be proven that if

$$\mathbf{X} = [X_1, X_2, ..., X_m]^T \sim Multinomial(n, p_1, p_2, ..., p_m)$$

then the Marginal Distributions of r.v. X_k is Binomial, particularly,

$$X_k \sim Binom(n, p_k).$$

Exercise: Find the Joint Distribution of, say, (X_1, X_2).

Example

Example 31.1: Assume

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim Multinomial(5, 0.1, 0.7, 0.2).$$

a. Give some interpretation for X_k-s;
b. Calculate $\mathbb{P}(X_1 = 0, X_2 = 4, X_3 = 1)$ and interpret;
c. Calculate $\mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1)$.

Example

Example 31.2: 10 AUA instructors are choosing (independently), at random, one AUA student for some committee. We know that the proportion between the number of Bus/CSE/EC students is $8 : 3 : 1$. What is the Probability that among 10 chosen students, we will have exactly 3 Bus, 5 CSE and 2 EC students?

Important Multivariate Distributions

Continuous Multivariate Distributions

Multivariate Uniform Distribution

Multivariate Uniform Distribution

Distribution Name: $Unif$;
Parameters: Domain $D \subset \mathbb{R}^2$ with finite area (or, in general, $D \subset \mathbb{R}^n$, with finite measure);

Bivariate Uniform Distribution

We will say that the r. vector (X, Y) has a Uniform Distribution over the region $D \subset \mathbb{R}^2$, and we will write $(X, Y) \sim Unif(D)$, if the Joint PDF of (X, Y) has the form

$$f(x, y) = \begin{cases} \dfrac{1}{Area(D)}, & (x, y) \in D \\ 0, & \text{otherwise.} \end{cases}$$

Multivariate Uniform Distribution

Note: Bivariate Uniform Distribution $Unif(D)$ is the rigorous way to define 2D Geometric Probabilities: it is modeling experiments of choosing a point in D at random, *uniformly*. So if we will take 2 subregions of D with the same area, we have equal chances to pick a point at random from that subregions.

Note: The generalization for the n-dim case is the following:

$$\mathbf{X} = [X_1, ..., X_n]^T \sim Unif(D), \quad \text{for} \quad D \subset \mathbb{R}^n,$$

if the Joint PDF of $\mathbf{X}$ is

$$f(\mathbf{x}) = \begin{cases} \dfrac{1}{Volume(D)}, & \mathbf{x} \in D \\ 0, & \text{otherwise.} \end{cases}$$

Multivariate Uniform Distribution

Note: In all cases, for 1D, 2D, and n-dim cases, the definition of Uniform Distribution on some region D is that the **PDF is constant on D, and is zero outside of D**, that is, it has the form

$$f(\mathbf{x}) = \begin{cases} K, & \mathbf{x} \in D \\ 0, & \mathbf{x} \notin D. \end{cases}$$

The constant K is determined from the property that the integral of PDF over the whole space needs to be 1.

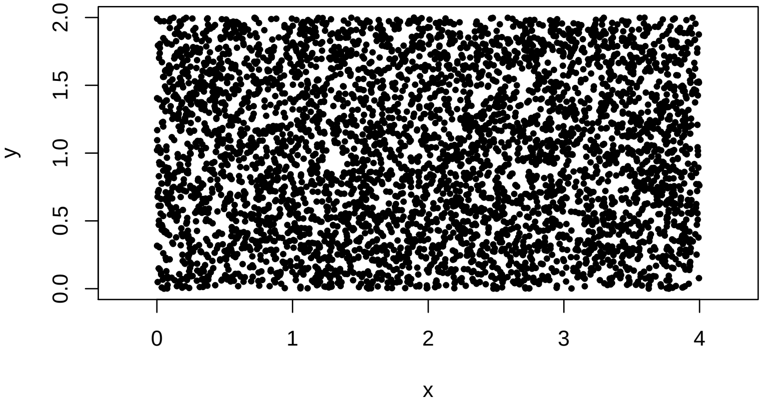


Figure: Random points (x,y) generated from $Unif([0,4] \times [0,2])$

2D Uniform Distrib and Geometric Probabilities

Above we have noted that the Uniform Distribution is the rigorous mathematical model behind Geometric Probabilities. Let us clarify this, in 2D.

2D Uniform Distrib and Geometric Probabilities

Assume $(X, Y) \in Unif(D)$, where $D \subset \mathbb{R}^2$ is a finite Area subset. This means that the Joint PDF of (X, Y) is

$$f(x, y) = \begin{cases} \dfrac{1}{Area(D)}, & (x, y) \in D \\ 0, & (x, y) \notin D. \end{cases}$$

Now, assume $A \subset D$. Then

$$\mathbb{P}\big((X, Y) \in A\big) = \iint_A f(x, y)\, dxdy \overset{A \subset D}{=\!=\!=} \frac{1}{Area(D)} \cdot \iint_A 1 \; dxdy$$

$$= \frac{Area(A)}{Area(D)}$$

Now, this means that we are calculating the probability that our point (X, Y) is in A by using Geometric Probabilities.

Bivariate Uniform Distribution, Example

Example 31.3: Assume $(X, Y) \sim Unif(D)$, where D is the triangle with vertices at $(0, 0)$, $(1, 0)$ and $(0, 1)$.

a. Find the Joint PDF of (X, Y);

b. Find the Marginal PDF of X and Y;

c. Calculate $\mathbb{P}(Y < 0.5 \cdot X)$;

d. Calculate $\mathbb{P}(0 \leq X \leq 0.5 \,|0 \leq Y \leq 0.5)$.

Bivariate Uniform Distribution, Example

Example 31.4: Assume $(X, Y) \sim Unif([a, b] \times [c, d])$. Find the Marginal Distributions of X and Y.

Bivariate Uniform Distribution, Example

Example 31.5: Write an **R** code to generate random points from $Unif(D)$, where

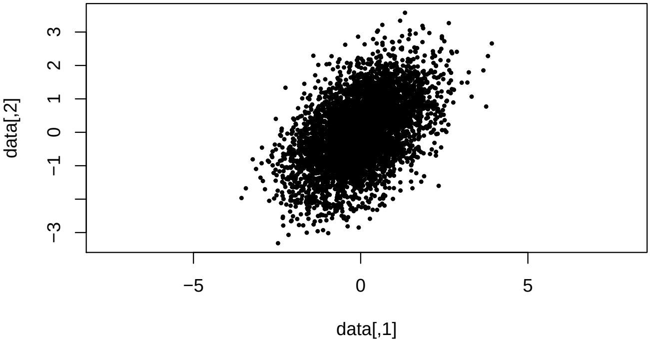
a. $D = [1, 3] \times [0, 4]$;

b. D is the triangle with vertices at $(0, 0)$, $(1, 0)$ and $(0, 1)$.

Multivariate Normal Distribution

Bivariate Normal (Gaussian) Distribution

Above we have seen some data generated from a Bivariate Uniform Distribution. Now assume we want to make a theoretical model of the distribution behind the following data:



Bivariate Normal (Gaussian) Distribution

The above data is not from a Uniform Distribution, since we have high density and low density regions. It is generated from the Bivariate Normal Distribution. So let us define the Normal Distribution in the 2D case, the Bivariate Normal Distribution.

Bivariate Normal (Gaussian) Distribution

To define the 2D Normal Distribution, we need two (multidimensional) parameters: a 2D vector $\boldsymbol{\mu}$, and a 2×2 matrix $\boldsymbol{\Sigma}$. So assume we are given a vector in $\mathbb{R}^2$:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix},$$

and a matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

which is Symmetric Positive Definite (SPD) matrix.

Reminder: SPD Matrices

The Positive Definiteness for a Symmetric Matrix can be described (defined) in several ways. Here are some equivalent descriptions: $\mathbf{\Sigma}$ is SPD iff

- $\mathbf{\Sigma}$ is symmetric and all its eigenvalues are positive;
- $\mathbf{\Sigma}$ is symmetric and all Leading Principal Minors are Positive, i.e., in 2D case,

$$\sigma_{12} = \sigma_{21}, \qquad \text{and} \qquad \sigma_{11} > 0, \qquad \sigma_{11}\cdot\sigma_{22} - \sigma_{12}^2 > 0;$$

- $\mathbf{\Sigma}$ is symmetric and the corresponding quadratic form is positive:

$$QF(\mathbf{x}) = QF_{\mathbf{\Sigma}}(\mathbf{x}) = \mathbf{x}^T \cdot \mathbf{\Sigma} \cdot \mathbf{x} > 0 \qquad \forall \mathbf{x} \neq \mathbf{0},$$

in the expanded form,

$$QF(x_1, x_2) = \sigma_{11}(x_1)^2 + 2\sigma_{12}x_1x_2 + \sigma_{22}(x_2)^2 > 0, \quad \forall (x_1, x_2) \neq (0, 0)$$

Bivariate Normal (Gaussian) Distribution

Bivariate Normal (Gaussian) Distribution

We say that the r. vector (X, Y) has a Bivariate Normal (or Gaussian) Distribution with the **mean** $\boldsymbol{\mu}$ and the **covariance matrix** $\boldsymbol{\Sigma}$, and we will write

$$(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

if the Joint PDF of (X, Y) is given by

$$f(\mathbf{x}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}} \cdot \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\mathbf{x}-\boldsymbol{\mu})\right\},$$

for any $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$

Bivariate Normal PDF plot

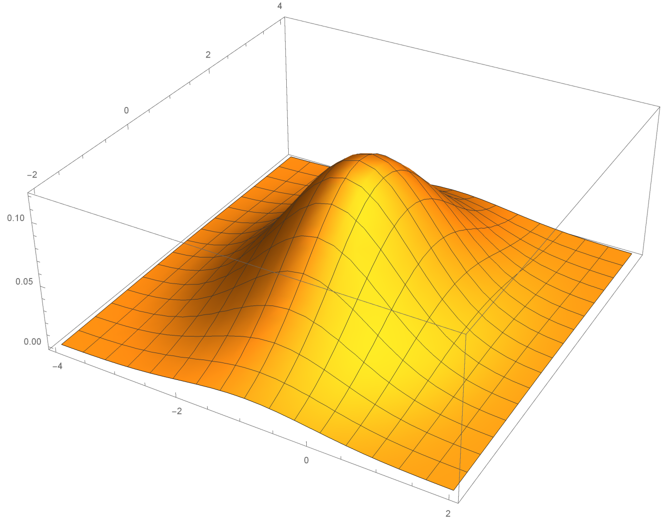


Figure: Bivariate Normal PDF. Maximum is at $\boldsymbol{\mu}$

Marginals of Multivariate Normal Distribution

It is remarkable that the Marginal Distributions of Multivariate Normal Distribution are again Normal. In particular, if

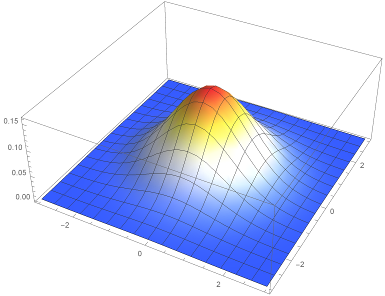
$$(X, Y) \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right)$$

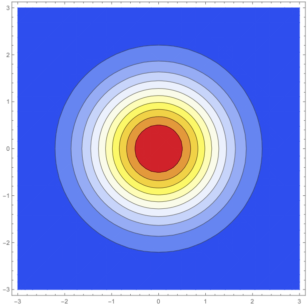
then

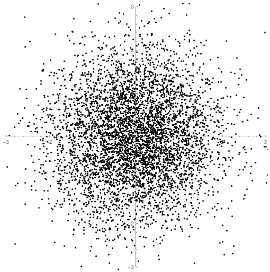
$$X \sim \mathcal{N}(\mu_1, \sigma_{11}) \qquad \text{and} \qquad Y \sim \mathcal{N}(\mu_2, \sigma_{22}).$$

Note: Geometrically,

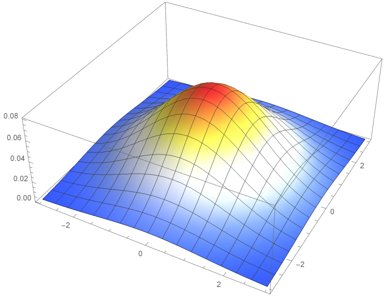
- X is centered at μ_1, with the variance σ_{11},
- Y is centered at μ_2, with the variance σ_{22},
- $\sigma_{12} = \sigma_{21}$ is showing the relationship between X and Y (it is called the covariance of X and Y).

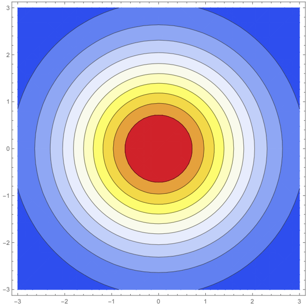


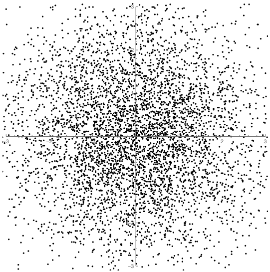




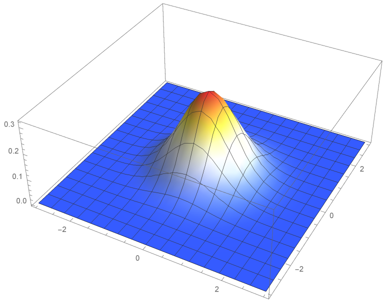
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

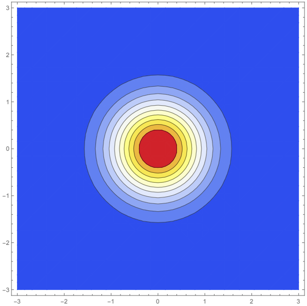


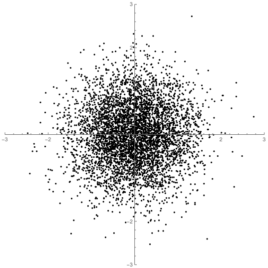




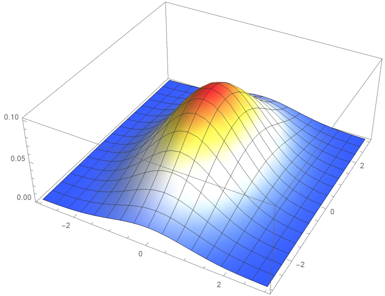
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

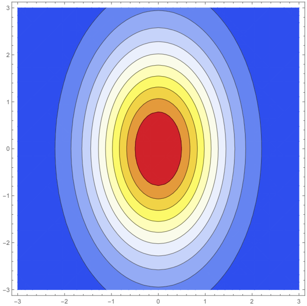


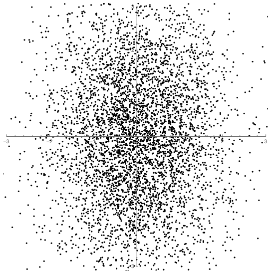




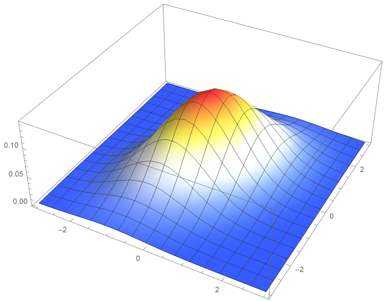
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

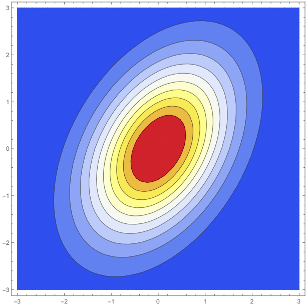


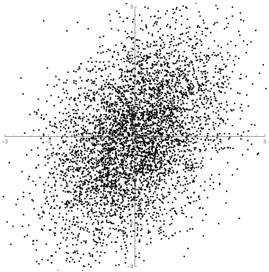




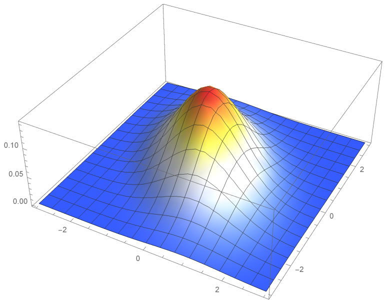
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 2.5 \end{bmatrix}$$

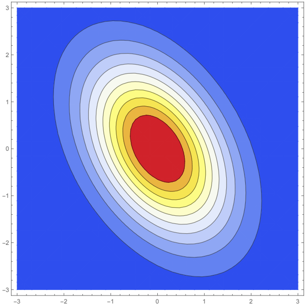


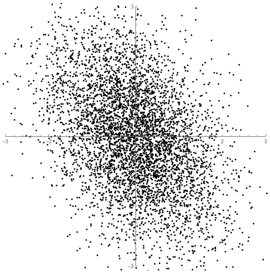




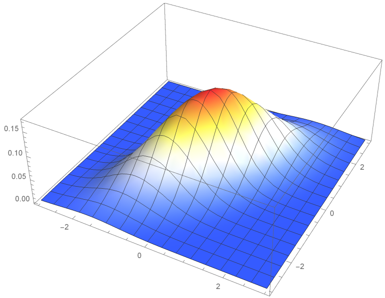
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{bmatrix}$$

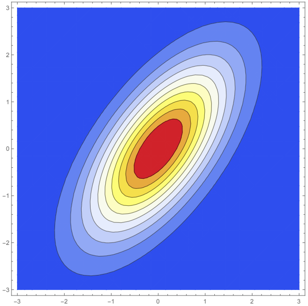


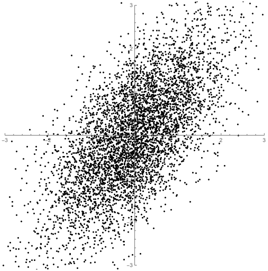




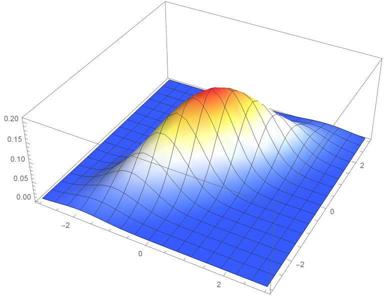
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1.5 \end{bmatrix}$$

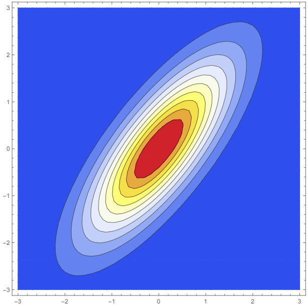


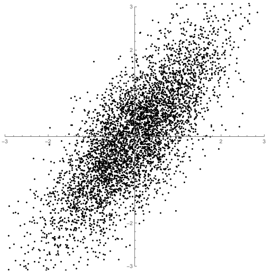




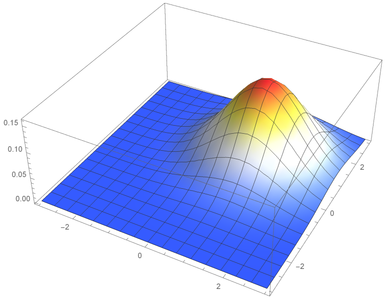
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1.5 \end{bmatrix}$$

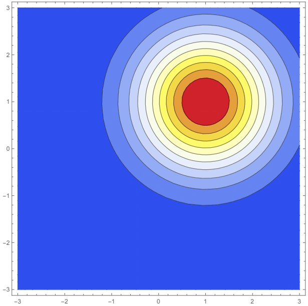


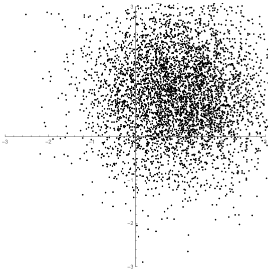




$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.95 \\ 0.95 & 1.5 \end{bmatrix}$$







$$\mu = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Bivariate Normal Distribution: Example

Example 31.6: We consider the Standard Bivariate Normal Distribution:

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = Id = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

a. Check that Σ is SPD;
b. Write the Joint PDF of a r. vector $(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$;
c. Calculate the Probability $\mathbb{P}(X^2 + Y^2 < 2)$;
d. Find the Marginal PDF of X and Y;
e. Plot the level curves of the Joint PDF.

Example 31.7: Assume

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 0 \\ 0 & 6 \end{bmatrix}.$$

a. Check that Σ is SPD;
b. Write the Joint PDF of a r. vector $(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$;
c. Calculate the Probability $\mathbb{P}(X \in [-1, 3], Y > 2)$;
d. Find the Marginal PDF of X and Y;
e. Plot the level curves of the Joint PDF;
f. Generate a random variate from this distribution in **R**.

Bivariate Normal Distribution: Example

Example 31.8: Assume

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}.$$

a. Check that Σ is SPD;
b. Write the Joint PDF of a r. vector $(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$;
c. Express the Probability $\mathbb{P}(|X| + |Y| < 3)$ in the form of a double integral;
d. Find the Marginal PDFs of X and Y;
e. Write the equation of the level curves of the Joint PDF;
f. Generate a random variate from this distribution in **R**.

Multivariate Normal (Gaussian) Distribution

Now, about the MultiVariate Normal Distribution. The definition goes in line with the 2D (BiVariate) case:
We assume we are given a vector in $\mathbb{R}^n$:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix},$$

and a Symmetric Positive Definite Matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

Multivariate Normal (Gaussian) Distribution

Multivariate Normal (Gaussian) Distribution

We say that the r. vector $\mathbf{X} = [X_1, X_2, ..., X_n]^T$ has a Multivariate Normal (or Gaussian) Distribution with the **mean $\boldsymbol{\mu}$** and the **covariance matrix $\boldsymbol{\Sigma}$**, and we will write

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

if the Joint PDF of $\mathbf{X}$ is given by

$$f(\mathsf{x}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}} \cdot \exp\left\{-\frac{1}{2}(\mathsf{x}-\boldsymbol{\mu})^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\mathsf{x}-\boldsymbol{\mu})\right\},$$

for any $\mathsf{x} \in \mathbb{R}^n$.

Marginals of Multivariate Normal Distribution

Again, one can prove that if

$$\mathbf{X} = [X_1, X_2, ..., X_n]^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

then

$$X_k \sim \mathcal{N}(\mu_k, \sigma_{kk}).$$