1

Gaussian Measures in Euclidean Space

1.1. The change of variables formula

Let (X, S, μ) be a measure space, i.e. X is a non-empty set, S is a sigma-algebra on X and μ is a measure on S. Consider also a measurable space (Y, F), i.e. Y is another non-empty set and F is a sigma-algebra on Y. Let $T : X \to Y$ be a measurable transformation, which means that

$$\forall A \in F, \quad T^{-1}A \in S. \qquad [1.1]$$

Hereafter

$$T^{-1}A := \{x \in X : Tx \in A\} \qquad [1.2]$$

is preimage of A under T. (To simplify the notation, we write $T^{-1}A$ and Tx rather than $T^{-1}(A)$ and $T(x)$, respectively, if it does not cause confusion.)

Introduce a set function

$$\nu(A) = \mu(T^{-1}A), \quad A \in F. \qquad [1.3]$$

THEOREM 1.1.– (About induced measure) The set of function ν given in [1.3] is a measure on F.

PROOF.– The function ν is well defined due to [1.1]. We have to show that it is not identical to infinity, but it is non-negative and sigma-additive.

Indeed, $\nu(\emptyset) = \mu(T^{-1}\emptyset) = \mu(\emptyset) = 0$, and therefore, ν is not identical to infinity. For each $A \in F$, $\nu(A) \geq 0$, because μ is a non-negative set function.

Finally, $\{A_n, n \geq 1\}$ are disjoint sets from F. Then the preimages $\{T^{-1}A_n, n \geq 1\}$ are disjoint as well, and

$$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(T^{-1}\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} T^{-1}A_n\right) =$$
$$= \sum_{n=1}^{\infty} \mu\left(T^{-1}A_n\right) = \sum_{n=1}^{\infty} \nu(A_n).$$

Here, we used the sigma-additivity of μ. Thus, ν is a non-negative and sigma-additive set function on the sigma-algebra F, i.e. ν is a measure on F. □

DEFINITION 1.1.– *The set function ν given in* [1.3] *is called a measure induced by transformation T and is denoted as μT^{-1}.*

The notation prompts how to evaluate $\nu(A)$:

$$(\mu T^{-1})(A) = \mu(T^{-1}A), \quad A \in F.$$

For any measurable space (Y, F, ν), denote $L(F, \nu)$ the space of Lebesgue integrable functions on Y w.r.t. measure ν.

Let $f : Y \to \bar{\mathbb{R}}$ be an F-measurable function, i.e. for each Borel subset B of extended real line $\bar{\mathbb{R}}$, it holds $f^{-1}B \in F$.

THEOREM 1.2.– (The change of variables formula) Assume that either $f \geq 0$ or $f \in L(Y, \mu T^{-1})$. Then it holds

$$\int_X f(Tx)d\mu(x) = \int_Y f(y)d(\mu T^{-1})(y). \qquad [1.4]$$

PROOF.– Equality [1.4] is shown in a standard way: first for indicators, then for simple non-negative functions, then for $f \geq 0$, and finally, for $f \in L(Y, \mu T^{-1})$.

a) Let $A \in F$,

$$f(y) = I_A(y) = \begin{cases} 1, & \text{if } y \in A \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$I_A(Tx) = \begin{cases} 1, & \text{if } Tx \in A \\ 0, & \text{otherwise,} \end{cases}$$

$$I_A(Tx) = \begin{cases} 1, & \text{if } x \in T^{-1}A \\ 0, & \text{otherwise,} \end{cases}$$

$$I_A(Tx) = I_{T^{-1}A}(x).$$

Hence

$$\int_X I_A(Tx)d\mu(x) = \int_X I_{T^{-1}A}(x)d\mu(x) = \mu(T^{-1}A) = (\mu T^{-1})(A),$$

$$\int_Y I_A(y)d(\mu T^{-1})(y) = (\mu T^{-1})(A),$$

and [1.4] follows for the indicator function.

b) Let $f \geq 0$ be a simple F-measurable function. Then it admits a representation

$$f(y) = \sum_{k=1}^{m} a_k I_{A_k}(y), \quad y \in Y, \qquad [1.5]$$

with disjoint measurable sets $\{A_k, k = 1, \ldots, m\}$ and non-negative a_k. For the function [1.5], relation [1.4] follows due to part (a) of the proof and the linearity of the Lebesgue integral.

c) Let f be an arbitrary non-negative and F-measurable function. Then there exists a sequence $\{p_n(y), n \geq 1, y \in Y\}$ of non-negative, simple and F-measurable functions such that p_n converges to f pointwise and $p_n(y) \leq p_{n+1}(y)$, $n \geq 1$, $y \in Y$. By part (b) of the proof,

$$\int_X p_n(Tx)d\mu(x) = \int_Y p_n(y)d(\mu T^{-1})(y). \qquad [1.6]$$

Here, tend n to infinity. By the monotone convergence theorem, [1.6] implies [1.4].

d) Finally, let $f \in L(Y, \mu T^{-1})$,

$$f_+(y) := \max\{f(y), 0\}, \quad f_-(y) := -\min\{f(y), 0\}, \quad y \in Y.$$

By part (c) of the proof,

$$\int_X f_+(Tx)d\mu(x) = \int_Y f_+(y)d(\mu T^{-1})(y), \qquad [1.7]$$

$$\int_X f_-(Tx)d\mu(x) = \int_Y f_-(y)d(\mu T^{-1})(y). \qquad [1.8]$$

Subtract [1.8] from [1.7] and obtain [1.4] using the definition of Lebesgue integral.

$\square$

Problems 1.1

1) Let λ_2^T be a Lebesgue measure on $[0,T]^2$; $\pi_1(x_1,x_2) = x_1$, $(x_1,x_2) \in [0,T]^2$, $\pi_1 : [0,T]^2 \to \mathbb{R}$. Show that

$$(\lambda_2^T \pi_1^{-1})(A) = T \cdot \lambda_1(A \cap [0,T]), \quad A \in S_1,$$

where λ_1 is Lebesgue measure on $\mathbb{R}$ and S_1 is sigma-algebra of Lebesgue measurable sets on $\mathbb{R}$.

2) Let μ_1 and μ_2 be finite measures on Borel sigma-algebra $\mathcal{B}(\mathbb{R})$, and $\pi(x_1,x_2) = x_1$, $(x_1,x_2) \in \mathbb{R}^2$. Find the induced measure $(\mu_1 \times \mu_2)\pi_1^{-1}$.

3) For the objects of theorem 1.1, prove the following: if μT^{-1} is sigma-finite, then μ is sigma-finite as well. Does the opposite hold true?

4) Let $f : Y \to \bar{\mathbb{R}}$ be any F-measurable function. Show that the Lebesgue integral on the left-hand side of [1.4] is well defined if, and only if, the integral on the right-hand side of [1.4] is well defined. Moreover, in case where they are well defined, they coincide.

1.2. Invariance of Lebesgue measure

Consider a measure space (X, S, μ) and a measurable transformation $T : X \to X$.

DEFINITION 1.2.– *The measure μ is called invariant under T, or T-invariant, if $\mu T^{-1} = \mu$.*

REMARK 1.1.– Assume additionally that T is a bijection on X, and moreover T^{-1} is a measurable transformation as well. Then μ is T-invariant if and only if

$$\mu(B) = \mu(TB), \quad \forall B \in S. \qquad [1.9]$$

(Hereafter TB denotes image of B under T.)

PROOF.–

a) Let μ be T-invariant and $B \in S$. Because T^{-1} is measurable, $A := TB \in S$. It holds $B = T^{-1}A$, and $\mu(T^{-1}A) = \mu(A)$. Equality [1.9] follows.

b) Conversely, assume [1.9] and take any $A \in S$. Denote $B_0 = T^{-1}A$, $B_0 \in S$. Then $(\mu T^{-1})(A) = \mu(B_0) = \mu(TB_0) = \mu(A)$, and μ is T-invariant. □

EXAMPLE 1.1.– (Counting measure) Let $X = \{1, 2, ..., n\}$, $S = 2^X$ be the sigma-algebra of all subsets of X, and μ be the counting measure on X, i.e. $\mu(A) = |A|$, $A \in S$. (Hereafter $|A|$ is number of points in a set A; if A is infinite, $|A| = +\infty$.) Then

μ is invariant under any bijection π on X. Indeed, $\mu(\pi^{-1}A) = |\pi^{-1}A| = |A| = \mu(A)$, $A \in S$.

In this section, we show that Lebesgue measure λ_n on $\mathbb{R}^n$ is rotation and translation invariant. Hereafter, we suppose that Euclidean space $\mathbb{R}^n$ consists of column vectors $x = (x_1, \ldots, x_n)^\top$.

DEFINITION 1.3.– *Affine transformation of $\mathbb{R}^n$ is a mapping of a form $Tx = Lx + c$, with $L \in \mathbb{R}^{n\times n}$ and $c \in \mathbb{R}^n$. Such transformation is called non-singular if L is non-singular. Otherwise, if $\det L = 0$, then T of this form is called a singular affine transformation.*

REMARK 1.2.– Affine transformation $Tx = Lx + c$ is invertible if, and only if, it is non-singular. In this case, the inverse transformation is a non-singular affine transformation as well, and it acts as follows:

$$T^{-1}y = L^{-1}y - L^{-1}c, \quad y \in \mathbb{R}^n.$$

Remember that non-singular affine transformations on a plane include rotations, translations and axial symmetries.

THEOREM 1.3.– (Transformation of Lebesgue measure at Borel sets) Consider Lebesgue measure λ_n on Borel sigma-algebra $\mathcal{B}(\mathbb{R}^n)$. Let $Tx = Lx + c$ be a non-singular affine transformation on $\mathbb{R}^n$. Then

$$\lambda_n T^{-1} = \frac{1}{|\det L|}\lambda_n. \qquad [1.10]$$

PROOF.– Transformation T is continuous, and, therefore, Borel measurable. Hence the induced measure $\lambda_n T^{-1}$ on $\mathcal{B}(\mathbb{R}^n)$ is well defined.

a) For $a = (a_k)_1^n$ and $b = (b_k)_1^n$ with $a_k < b_k$, $k = 1, \ldots, n$ denote

$$[a, b] = \prod_{k=1}^n [a_k, b_k], \quad (a, b] = \prod_{k=1}^n (a_k, b_k]. \qquad [1.11]$$

Hereafter, $\prod_{k=1}^n A_k$ stands for Cartesian product of $A_1, \ldots, A_n$. Evaluate

$$\left(\lambda_n T^{-1}\right)([a, b]) = \lambda_n\left(T^{-1}[a, b]\right) = \int_{T^{-1}[a,b]} d\lambda_n = \int_{T^{-1}[a,b]} dx.$$

The latter integral is Riemann integral over the compact and Jordan measurable set $T^{-1}[a, b]$. The change of variables in the Riemann integral leads to the following:

$$\lambda_n\left(T^{-1}[a, b]\right) = \int_{[a,b]} \left|\frac{\partial y}{\partial x}\right|^{-1} dy = \frac{m_n([a, b])}{|\det L|} = \frac{\lambda_n([a, b])}{|\det L|}.$$

Here m_n is Jordan measure on $\mathbb{R}^n$.

b) Consider a set $(a, b]$ introduced in [1.11], and let $\{a_k(m), m \geq 1\}$ be a decreasing sequence that converges to a_k such that $a_k(m) < b_k, m \geq 1; k = 1, \ldots, n$. Denote $a(m) = (a_k(m))_{k=1}^n \in \mathbb{R}^n$. Then $A_m := [a(m), b]$ is a monotone sequence of sets that converges to $(a, b]$. The continuity of Lebesgue measure from below implies

$$\left(\lambda_n T^{-1}\right)((a, b]) = \lim_{m\to\infty} \lambda_n(T^{-1}A_m) = \lim_{m\to\infty} \frac{\lambda_n(A_m)}{|\det L|} = \frac{\lambda_n((a, b])}{|\det L|}.$$

Here, we used part (a) of the proof.

c) Thus, the two measures $\lambda_n T^{-1}$ and $\frac{\lambda_n}{|\det L|}$ in [1.10] coincide on the semiring $\mathcal{P}_n$ that consists of all bricks $(a, b]$ from [1.11]. Both measures are sigma-finite, and therefore, they coincide on $\sigma r(\mathcal{P}_n) = \mathcal{B}(\mathbb{R}^n)$, where $\sigma r(\mathcal{P}_n)$ is sigma-ring generated by $\mathcal{P}_n$. □

Now, we extend theorem 1.3 to Lebesgue measure λ_n on sigma-algebra S_n of Lebesgue measurable sets in $\mathbb{R}^n$.

LEMMA 1.1.– Non-singular affine transformation $Tx = Lx + c$ is (S_n, S_n)-measurable, i.e. for any $A \in S_n, T^{-1}A \in S_n$ as well.

PROOF.– It is known (see [HAL 13]) that

$$S_n = \{B \cup N | B \in \mathcal{B}(\mathbb{R}^n), N \subset N_0 \text{ with } N_0 \in \mathcal{B}(\mathbb{R}^n), \lambda_n(N_0) = 0\}. \quad [1.12]$$

Let $A \in S_n$, then $A = B \cup N$, with B and N described in [1.12]. It holds

$$T^{-1}A = \left(T^{-1}B\right) \cup \left(T^{-1}N\right), \quad T^{-1}B \in \mathcal{B}(\mathbb{R}^n), \quad [1.13]$$

$$T^{-1}N \subset T^{-1}N_0, T^{-1}N_0 \in \mathcal{B}(\mathbb{R}^n), \lambda_n\left(T^{-1}N_0\right) = \frac{\lambda_n(N_0)}{|\det L|} = 0. \quad [1.14]$$

Here, we used theorem 1.3 and the fact that T is a Borel function. Decompositions [1.13] and [1.14] show that $T^{-1}A \in S_n$. □

THEOREM 1.4.– (Transformation of Lebesgue measure) Let $Tx = Lx + c$ be a non-singular affine transformation on $\mathbb{R}^n$. For Lebesgue measure λ_n on S_n, it holds

$$\lambda_n T^{-1} = \frac{\lambda_n}{|\det L|}.$$

PROOF.– Consider $A \in S_n$ and decompose $T^{-1}A$ as in [1.13] and [1.14]. Because $\lambda_n(N) = \lambda_n(T^{-1}N) = 0$, we have by theorem 1.3:

$$\lambda_n\left(T^{-1}A\right) = \lambda_n\left(T^{-1}B\right) = \frac{\lambda_n(B)}{|\det L|} = \frac{\lambda_n(A)}{|\det L|}.$$

□

COROLLARY 1.1.– (Criterion for invariance of Lebesgue measure) Lebesgue measure λ_n on S_n is invariant under a non-singular affine transformation $Tx = Lx + c$ if, and only if, $\det L = \pm 1$.

In particular, λ_n is symmetric around the origin, and it is invariant under translations $Tx = x + c$ and orthogonal transformations $Tx = Ux$, where U is an orthogonal matrix (i.e. $U^{-1} = U^\top$), e.g. under symmetries w.r.t. hyperplanes that pass through the origin. In the planar case ($n = 2$), λ_2 is invariant under the transformation

$$Tx = \begin{pmatrix} 2x_1 \\ \frac{1}{2}x_2 \end{pmatrix}, \quad x \in \mathbb{R}^2.$$

Here, T is dilation along x_1-axis with coefficient 2 and contraction along x_2-axis with the same coefficient.

COROLLARY 1.2.– (Affine change of variables) Let $Tx = Lx + c$ be a non-singular affine transformation on $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be a Lebesgue measurable function, which is either non-negative or belongs to $L(\mathbb{R}^n, \lambda_n)$. Then it holds

$$\int_{\mathbb{R}^n} f(Tx) d\lambda_n(x) = \frac{1}{|\det L|} \int_{\mathbb{R}^n} f(y) d\lambda_n(y).$$

PROOF.– Apply theorems 1.2 and 1.4:

$$\int_{\mathbb{R}^n} f(Tx) d\lambda_n(x) = \int_{\mathbb{R}^n} f(y) d\left(\lambda_n T^{-1}\right)(y) = \frac{1}{|\det L|} \int_{\mathbb{R}^n} f(y) d\lambda_n(y). \quad \square$$

Problems 1.2

5) Let $Tx = |x|$, $x \in \mathbb{R}$. Find $\lambda_1 T^{-1}$.

6) Let $f : \mathbb{R} \to \mathbb{R}$ be a Lebesgue measurable function such that $\forall x, y \in \mathbb{R}$, $|f(x) - f(y)| \geq |x - y|$. Prove that f is (S_1, S_1)-measurable, where S_1 is sigma-algebra of Lebesgue measurable sets on real line.

7) Show that $\arctan x$ is an (S_1, S_1)-measurable function. Let $f : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to [0, +\infty]$ be a Lebesgue measurable function. Prove that

$$\int_{\mathbb{R}} f(\arctan x) d\lambda_1(x) = \int_{\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)} \frac{f(t)}{\cos^2 t} d\lambda_1(t).$$

8) Show that e^x is (S_1, S_1)-measurable function. Let $f : (0, \infty) \to [0, +\infty]$ be a Lebesgue measurable function. Prove that

$$\int_{\mathbb{R}} f(e^x) d\lambda_1(x) = \int_{(0, +\infty)} \frac{f(t)}{t} d\lambda_1(x).$$

9) Prove that $f(x) = ||x||$, $x \in \mathbb{R}^n$ is (S_n, S_1)-measurable function.

10) Let $f : [0, +\infty) \to [0, +\infty]$ be a Lebesgue measurable function. Prove that the measure

$$\mu(A) = \int_A f(||x||)\, d\lambda_n(x), \quad A \in S_n$$

is invariant under unitary operators in $\mathbb{R}^n$.

11) Let μ be a measure on S_n, which is finite at each bounded set from S_n, absolutely continuous w.r.t. λ_n and invariant under unitary operators in $\mathbb{R}^n$. Prove that there exists a Borel function $f : [0, +\infty) \to [0, +\infty)$ such that representation from Problem 10 holds true.

Hint. Given a locally Lebesgue integrable function g on $\mathbb{R}^n$, a point $x \in \mathbb{R}^n$ is Lebesgue point if

$$\lim_{r \to 0+} \frac{1}{\lambda_n(B(x, r))} \int_{B(x,r)} |g(y) - g(x)| d\lambda_n(y) = 0.$$

Hereafter, $B(x, r)$ is an open ball centered at x with radius r. Use the Lebesgue differentiation theorem [BOG 07] which states that, given any locally Lebesgue integrable function g on $\mathbb{R}^n$, almost every x is a Lebesgue point of g.

12) Let $g : \mathbb{R}^n \to \mathbb{R}$ be a Lebesgue measurable function such that $g(Tx) = g(x)$ $(\mathrm{mod}\ \lambda_n)$, for all unitary operators T in $\mathbb{R}^n$. Prove that there exists a Borel function $f : [0, +\infty) \to \mathbb{R}$, with $g(x) = f(||x||)$ $(\mathrm{mod} \lambda_n)$.

13) Let $\alpha > 0$ and $f \in L(\mathbb{R}, \lambda_1)$. Prove that $f(n^{1+\alpha} x) \to 0$ as $n \to \infty$ for almost all $x \in \mathbb{R}$. Extend this statement to functions from $L(\mathbb{R}^m, \lambda_m)$.

14) Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, $f \in L([0, +\infty), \lambda_1)$. Prove the following:

a) If f is an even function, then

$$\int_{\mathbb{R}} f d\lambda_1 = 2 \int_{[0,+\infty)} f d\lambda_1.$$

b) If f is an odd function, then $\int_{\mathbb{R}} f d\lambda_1 = 0$.

15) Let $f : [-1, 1] \to (0, +\infty)$ be a Lebesgue measurable function. Find the integral $\int_{[-1,1]} \frac{f(x)}{f(x)+f(-x)} d\lambda_1(x)$.

1.3. Absence of invariant measure in infinite-dimensional Hilbert space

Let H be a real Hilbert space, with Borel sigma-algebra $\mathcal{B}(H)$. In this section, we search for a measure λ on $\mathcal{B}(H)$ with the following properties:

i) λ is positive at each non-empty open set;

ii) λ is finite at each bounded Borel set;

iii) λ is invariant under each translation $Tx = x + c$, $x \in H$, with $c \in H$.

Remember that a linear operator U in H is called unitary if $||Ux|| = ||x||$, $x \in H$, and $R(U) = H$. An operator $U \in L(H)$ is unitary if, and only if, $U^* = U^{-1}$.

iv) λ is invariant under each unitary operator in H.

Note that Lebesgue measure λ_n on $\mathcal{B}(\mathbb{R}^n)$ possesses the properties (i)–(iv).

THEOREM 1.5.– (Absence of invariant measure in H) Let H be an infinite-dimensional real Hilbert space. Then:

a) There is no measure λ with properties (i)–(iii).

b) There is no measure λ with properties (i), (ii) and (iv).

PROOF.– Because $\dim(H) = \infty$, there exists an infinite orthonormal system $\{e_n, n \geq 1\}$ in H. For $k \neq m$, $||e_k - e_m|| = \sqrt{2}$, hence open balls $B\left(e_n, \frac{\sqrt{2}}{2}\right)$ are disjoint. For $x \in B\left(e_n, \frac{\sqrt{2}}{2}\right)$, it holds $||x|| \leq ||e_n|| + ||x - e_n|| < 1 + \frac{\sqrt{2}}{2} < 2$, and

$$B\left(e_n, \frac{\sqrt{2}}{2}\right) \subset B(0,2), \quad \bigcup_{n=1}^{\infty} B\left(e_n, \frac{\sqrt{2}}{2}\right) \subset B(0,2). \qquad [1.15]$$

a) Let λ have the properties (i)–(iii). For $k \neq m$, the translation $Tx = x + e_m - e_k$, $x \in H$ maps the ball $B\left(e_k, \frac{\sqrt{2}}{2}\right)$ onto $B\left(e_m, \frac{\sqrt{2}}{2}\right)$. Hence by (i) and (iii),

$$\lambda\left(B\left(e_k, \frac{\sqrt{2}}{2}\right)\right) = \lambda\left(B\left(e_m, \frac{\sqrt{2}}{2}\right)\right) = a > 0. \qquad [1.16]$$

Due to [1.15] and [1.16], we have

$$\lambda(B(0,2)) \geq \sum_{n=1}^{\infty} \lambda\left(B\left(e_n, \frac{\sqrt{2}}{2}\right)\right) = \sum_{n=1}^{\infty} a = +\infty, \quad \lambda(B(0,2)) = +\infty.$$

But this contradicts property (ii). Therefore, such a measure λ does not exist.

b) Now, assume that λ has the properties (i), (ii) and (iv). Construct a unitary operator U that maps $B\left(e_k, \frac{\sqrt{2}}{2}\right)$ onto $B\left(e_m, \frac{\sqrt{2}}{2}\right)$, with fixed $k \neq m$.

Let L be a subspace generated by $\{e_n, n = 1, 2, \dots\}$. Each $x \in H$ can be decomposed as

$$x = \sum_{n=1}^{\infty} (x, e_n)e_n + z, \quad z \in L^{\perp}.$$

The isometric operator

$$Ux = (x, e_k)e_m + (x, e_m)e_k + \sum_{n \neq k, n \neq m}^{\infty} (x, e_n)e_n + z$$

is a surjection, and hence is unitary. It maps $B\left(e_k, \frac{\sqrt{2}}{2}\right)$ onto $B\left(e_m, \frac{\sqrt{2}}{2}\right)$, and by properties (iv) and (i), it holds [1.16]. The rest of the proof follows the line of part (a) of the proof. □

As we see, in space l_2 of sequences and in the space $L_2[a, b]$ of functions there is no measure analogous to Lebesgue measure. Nevertheless, we will construct a measure in an infinite-dimensional Hilbert space, which is invariant under quite a large group of transformations. It will be a Gaussian measure.

Problems 1.3

16) Prove that there is no measure λ on $\mathcal{B}(l_\infty)$ with properties (i) and (ii) from section 1.3, where l_∞ is the space of real bounded sequences with the supremum norm.

17) Let X be a real normed space, with $\dim(X) = \infty$. Prove that there is no measure λ on $\mathcal{B}(X)$ with properties (i)–(iii) from section 1.3.

18) A linear bijection V on a normed space X is called isometry if $||Vx|| = ||x||$, $x \in X$. Prove that there is no measure λ on $\mathcal{B}(l_p)$ with properties (i) and (ii) from section 1.3 and such that λ is invariant under all isometries on l_p, $1 \leq p < \infty$.

19) Let $\varphi(t)$, $t \in [0, 1]$ be a continuous increasing function, with $\varphi(0) = 0$, $\varphi(1) = 1$, and $\varphi(t) < t$, $t \in (0, 1)$. In Banach space, $X = C[0, 1]$ introduce a transformation $(Tx)(t) = x(\varphi(t))$, $t \in [0, 1]$, $x \in X$. Prove that there is no measure λ on $\mathcal{B}(X)$ with properties (i) and (ii) from section 1.3 and such that it is T-invariant.

1.4. Random vectors and their distributions

Remember that a probability measure is a measure on sigma-algebra, which is equal to 1 at the total space. A measure space $(\Omega, \mathcal{F}, \mathsf{P})$ is called probability space if P is a probability measure, i.e. $\mathsf{P}(\Omega) = 1$.

1.4.1. *Random variables*

A random variable (r.v.) on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is just an $\mathcal{F}$-measurable function on Ω.

DEFINITION 1.4.– *Let $X = X(\omega)$ be a r.v. on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. The distribution of X is a probability measure μ_X defined as follows:*

$$\mu_X(B) = \mathsf{P}\{\omega : X(\omega) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}).$$

Note that μ_X is a measure induced by the mapping $X : \Omega \to \mathbb{R}$, i.e. $\mu_X = \mathsf{P}\, X^{-1}$ (see definition 1.1).

A Borel function $f : \mathbb{R} \to \mathbb{R}$ is called the probability density function (pdf) of a r.v. X if

$$\mathsf{P}\{X(\omega) \in B\} = \int_B f(t) d\lambda_1(t), \quad B \in \mathcal{B}(\mathbb{R}).$$

Actually, this means that the distribution $\mu_X \ll \lambda_1$, where the Lebesgue measure λ_1 is considered on $\mathcal{B}(\mathbb{R})$, and moreover the Radon–Nikodym derivative $\frac{d\mu_X}{d\lambda_1} = f(t)(\mathrm{mod}\lambda_1)$.

DEFINITION 1.5.– *A r.v. γ is called normal (or normally distributed) if it has a pdf of the form*

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

with parameters $m \in \mathbb{R}$ and $\sigma > 0$. This is denoted as follows: $\gamma \sim N(m, \sigma^2)$.

If $\gamma \sim N(m, \sigma^2)$, then it holds

$$\mathsf{E}\,\gamma = m, \quad \mathsf{D}\,\gamma = \sigma^2.$$

Hereafter, E stands for expectation and D stands for the variance of a r.v. Remember that

$$\mathsf{E}\,\gamma = \int_\Omega \gamma(\omega) d\,\mathsf{P}(\omega), \quad \mathsf{D}\,\gamma = \mathsf{E}(\gamma - m)^2 = \int_\Omega (\gamma(\omega) - m)^2 \, d\,\mathsf{P}(\omega).$$

By the change of variables formula (theorem 1.2), it holds

$$\mathsf{E}\,\gamma = \int_{\mathbb{R}} x d\mu_\gamma(x), \quad \mathsf{D}\,\gamma = \int_{\mathbb{R}} (x - m)^2 \, d\mu_\gamma(x),$$

and since γ has pdf equal ρ, we have

$$m = \mathsf{E}\,\gamma = \int_{\mathbb{R}} x\rho(x) dx, \quad \sigma^2 = \mathsf{D}\,\gamma = \int_{\mathbb{R}} (x - m)^2 \, \rho(x) dx.$$

The latter integrals are Lebesgue integrals.

DEFINITION 1.6.– *A r.v. γ has degenerate normal distribution $N(m, 0)$ if $\gamma(\omega) = m$, almost surely (a.s.). We denote it as $\gamma \sim N(m, 0)$.*

In this case, $\mathsf{E}\,\gamma = m, \mathsf{D}\,\gamma = 0$, and the distribution μ_γ is a point measure concentrated at m:

$$\mu_\gamma(B) = I_B(m), \quad B \in \mathcal{B}(\mathbb{R}).$$

Such a measure is called Dirac measure at point m and is denoted as δ_m.

DEFINITION 1.7.– *A r.v. γ is called Gaussian if it is either normally distributed (with positive variance) or has degenerate normal distribution (with zero variance).*

Thus, a Gaussian r.v. γ satisfies $\gamma \sim N(m, \sigma^2)$ with some $m \in \mathbb{R}$ and $\sigma \geq 0$. If $\sigma > 0$, then $\mu_\gamma \ll \lambda_1$, and if $\sigma = 0$, then $\mu_\gamma = \delta_m \perp \lambda_1$ (i.e. μ_γ is singular to λ_1).

DEFINITION 1.8.– *A r.v. $\gamma \sim N(0, 1)$ is called standard normal.*

Remember that characteristic function φ_ξ of a r.v. ξ is as follows:

$$\varphi_\xi(t) = \mathsf{E}\,e^{it\xi}, \quad t \in \mathbb{R}.$$

A normal r.v. $\gamma \sim N(m, \sigma^2)$ has characteristic function

$$\varphi_\gamma(t) = \exp\{imt - \frac{\sigma^2 t^2}{2}\}. \qquad [1.17]$$

If γ has degenerate normal distribution $N(m, 0)$, then

$$\varphi_\gamma(t) = \exp\{imt\}.$$

Thus, relation [1.17] holds true for any Gaussian r.v. $\gamma \sim N(m, \sigma^2)$, with $\sigma \geq 0$.

1.4.2. *Random vectors*

Remember that a random vector X distributed in $\mathbb{R}^n$ is $\mathcal{F}$-measurable mapping $X : \Omega \to \mathbb{R}^n$, where $(\Omega, \mathcal{F}, \mathsf{P})$ is the underlying probability space.

A mapping $X(\omega) = (X_1(\omega), \ldots, X_n(\omega))^\top$, $\omega \in \Omega$ is a random vector if, and only if, all $X_k(\omega)$, $k = 1, \ldots, n$ are random variables.

For a random vector X, its expectation is defined coordinate-wise:

$$\mathsf{E}\,X = (\mathsf{E}\,X_1, \ldots, \mathsf{E}\,X_n)^\top =: m, \quad m = (m_1, \ldots, m_n)^\top \in \mathbb{R}^n.$$

In case $\mathsf{E}\,X_k^2 < \infty$ for all $k = 1, \dots, n$, its variance–covariance matrix $\mathsf{Cov}(X) = (s_{ij})_{i,j=1}^n$ is defined as follows:

$$s_{ij} = \mathsf{Cov}(X_i, X_j) = \mathsf{E}\,(X_i - m_i)\,(X_j - m_j)\,, \quad i, j = 1, \dots, n. \qquad [1.18]$$

Hereafter, expectation E is considered as an operator that acts on the total product under its sign, and we omit brackets for brevity.

The variance–covariance matrix can be expressed as

$$\mathsf{Cov}(X) = \mathsf{E}\,(X - m)\,(X - m)^\top. \qquad [1.19]$$

Here, the expectation of a random matrix is a matrix composed of expectations of entries, according to [1.18]. The variance–covariance matrix is a positive semidefinite matrix, i.e. it is symmetric and the corresponding quadratic form is non-negative. The variance–covariance matrix of a random vector X exists if and only if $\mathsf{E}\,||X||^2 < \infty$.

LEMMA 1.2.– (Variance–covariance matrix under linear transform) Let X be a random vector in $\mathbb{R}^n$, with variance–covariance matrix S, and $A \in \mathbb{R}^{p\times n}$. Then AX is a random vector in $\mathbb{R}^p$, with

$$\mathsf{Cov}(AX) = ASA^\top. \qquad [1.20]$$

PROOF.– The mapping $AX : \Omega \to \mathbb{R}^p$ is $\mathcal{F}$-measurable, as a Borel function of random vector. Hence AX is a random vector in $\mathbb{R}^p$, and because $\mathsf{E}\,||AX||^2 \le ||A||^2 \cdot \mathsf{E}\,||X||^2 < \infty$, it has a variance–covariance matrix.

Hereafter, $||A||$ is Euclidean norm of matrix A,

$$||A|| = \sup_{x\neq 0} \frac{||Ax||}{||x||}.$$

Now, $\mathsf{E}(AX) = A(\mathsf{E}\,X) = Am$, where $m = \mathsf{E}\,X$, and

$$\mathsf{Cov}(AX) = \mathsf{E}\,(AX - Am)\,(AX - Am)^\top =$$

$$= \mathsf{E}[A\,(X - m)\,(X - m)^\top A^\top] = A \cdot \mathsf{E}\,(X - m)\,(X - m)^\top \cdot A^\top = ASA^\top.$$

Here, we used the linearity of the operator of taking expectation. □

COROLLARY 1.3.– (Moments of linear functional) Let $a \in \mathbb{R}^n$ and X be a random vector in $\mathbb{R}^n$, with mean value m and variance–covariance matrix S. Then

$$\mathsf{E}(a^\top X) = a^\top m, \quad \mathsf{D}(a^\top X) = a^\top S a.$$

PROOF.– The statement follows from lemma 1.2 and its proof if to put $A = a^\top \in \mathbb{R}^{1\times n}$. The random vector $a^\top X$ is just a r.v., and its variance–covariance matrix is just the variance. □

1.4.3. *Distributions of random vectors*

Let X be a random vector distributed in $\mathbb{R}^n$. Its distribution is introduced similarly to definition 1.4.

DEFINITION 1.9.– *The distribution of X is a probability measure μ_X defined as follows:*

$$\mu_X(B) = \mathsf{P}\{\omega : X(\omega) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}^n).$$

It is always possible to construct a random vector with a given distribution.

LEMMA 1.3.– Given a probability measure μ on $\mathcal{B}(\mathbb{R}^n)$, there exists a random vector X, with distribution $\mu_X = \mu$.

PROOF.– Take the measure space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu)$ as a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and define $X : \Omega \to \mathbb{R}^n$ as $X(\omega) = \omega$. Then

$$\mu_X(B) = \mathsf{P}\{\omega \in B\} = \mathsf{P}(B) = \mu(B), \quad B \in \mathcal{B}(\mathbb{R}^n). \qquad \square$$

Remember that random variables $X_1, \ldots, X_n$, which are defined on the same probability space, are independent if

$$\mathsf{P}\{X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n\} = \prod_1^n \mathsf{P}\{X_k \in B_k\},$$

for all $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$.

The latter relation can be written in terms of the distribution μ_X of random vector $X = (X_k)_1^n$ and marginal distributions μ_{X_k} of its components:

$$\mu_X(\prod_1^n B_k) = \prod_1^n \mu_{X_k}(B_k).$$

Here $\prod_1^n B_k$ denotes Cartesian product of the sets B_k.

It is clear that components of random vector $X = (X_k)_1^n$ are independent if, and only if, μ_X is a product of n probability measures, and in this case

$$\mu_X = \prod_1^n \mu_{X_k}.$$

Remember that characteristic function φ_X of a random vector X is defined as follows:

$$\varphi_X(t) = \mathsf{E}\, e^{i(X,t)}, \quad t \in \mathbb{R}^n.$$

One can rewrite $\varphi_X(t)$ using the change of variables formula (see theorem 1.2):

$$\varphi_X(t) = \int_\Omega e^{i(X(\omega),t)} dP(\omega) = \int_{\mathbb{R}^n} e^{i(z,t)} d(PX^{-1})(z),$$

$$\varphi_X(t) = \int_{\mathbb{R}^n} e^{i(z,t)} d\mu_X(z).$$

This prompts the following definition.

DEFINITION 1.10.– *Given a probability measure μ on $\mathcal{B}(\mathbb{R}^n)$, its characteristic function φ_μ is as follows:*

$$\varphi_\mu(t) = \int_{\mathbb{R}^n} e^{i(z,t)} d\mu(z), t \in \mathbb{R}^n.$$

Thus, φ_X and φ_{μ_X} coincide.

From standard course of probability theory, it is known that the cumulative distribution function (and therefore, the distribution) of a random vector is uniquely defined by its characteristic function.

LEMMA 1.4.– (Criterion for independence) Consider a random vector $X = (X_k)_1^n$. Its components are independent if, and only if, φ_X can be decomposed as follows:

$$\varphi_X(t) = \varphi_1(t_1)\varphi_2(t_2)\ldots\varphi_n(t_n), t \in \mathbb{R}^n,$$

where $\varphi_k : \mathbb{R} \to \mathbb{C}$ are some functions with $\varphi_k(0) = 1$, $k = 1, \ldots, n$, and in this case

$$\varphi_X(t) = \prod_1^n \varphi_{X_k}(t_k), \quad t \in \mathbb{R}^n.$$

PROOF.–

a) Let X_k be independent. Then random variables $e^{it_kX_k}$, $k = 1, \ldots, n$ are independent as well, and

$$\varphi_X(t) = \mathsf{E} \prod_1^n e^{it_kX_k} = \prod_1^n \mathsf{E}\, e^{it_kX_k} = \prod_1^n \varphi_{X_k}(t_k); \quad \varphi_{X_k}(0) = 1.$$

b) Assume that $\varphi_X(t) = \varphi_1(t_1)\varphi_2(t_2)\ldots\varphi_n(t_n)$, with $\varphi_k(0) = 1$, $k = 1, \ldots, n$. Let $\{e_k, k = 1, \ldots, n\}$ be the standard orthobasis in $\mathbb{R}^n$. Then

$$\varphi_{X_k}(t_k) = \mathsf{E}\, e^{it_k X_k} = \mathsf{E}\, e^{i(X, t_k e_k)} = \varphi_X(t_k e_k) = \varphi_k(t_k).$$

Let $Y = (Y_k)_1^n$ be a random vector with independent components and the same marginal distributions:

$$\mu_{Y_k} = \mu_{X_k}, \quad k = 1, \ldots, n.$$

(Such Y can be constructed if to apply lemma 1.3 to the measure $\mu = \prod_1^n \mu_{X_k}$.) Then, by part (a) of the proof,

$$\varphi_Y(t) = \prod_1^n \varphi_{Y_k}(t_k) = \prod_1^n \varphi_{X_k}(t_k) = \varphi_X(t), \quad t \in \mathbb{R}^n.$$

Therefore,

$$\mu_X = \mu_Y = \prod_1^n \mu_{Y_k} = \prod_1^n \mu_{X_k},$$

and the components of X are independent. □

In an obvious way, lemma 1.3 can be reformulated as a criterion for a probability measure on $\mathcal{B}(\mathbb{R}^n)$ to be a product of n probability measures on $\mathcal{B}(\mathbb{R})$.

DEFINITION 1.11.– *Given a probability measure μ on $\mathcal{B}(\mathbb{R}^n)$, its mean value m_μ and variance–covariance matrix* $\mathsf{Cov}(\mu) = (s_{ij})_{i,j=1}^n$ *are defined as follows:*

$$m_\mu = \int_{\mathbb{R}^n} x d\mu(x) := \left(\int_{\mathbb{R}^n} x_k d\mu(x) \right)_{k=1}^n = (m_k)_1^n,$$

$$s_{ij} = \int_{\mathbb{R}^n} (x_i - m_i)(x_j - m_j)\, d\mu(x), \quad i, j = 1, \ldots, n.$$

Definition 1.11 is consistent with the corresponding definition of the mean and variance–covariance matrix of a random vector. Indeed, for a random vector X, it holds

$$\mathsf{E}\, X = m_{\mu_X}, \quad \mathsf{Cov}(X) = \mathsf{Cov}(\mu_X),$$

i.e. expectation and variance–covariance matrix of a random vector are just the mean and variance–covariance matrix of its distribution.

Now, we interpret the bilinear form generated by $S = \mathsf{Cov}(X)$. Let $m = \mathsf{E}\,X$ and $u, v \in \mathbb{R}^n$; then

$$(Su, v) = v^\top \left(\mathsf{E}(X-m)(X-m)^\top\right) u = \mathsf{E}\left[v^\top (X-m)(X-m)^\top u\right],$$

$$(Su, v) = \mathsf{E}(X-m, u)(X-m, v).$$

For a probability measure μ on $\mathcal{B}(\mathbb{R}^n)$, with $m = m_\mu$ and $S = \mathsf{Cov}(\mu)$, we have, respectively:

$$(Su, v) = \int_{\mathbb{R}^n} (z-m, u)\,(z-m, v)\,d\mu(z), \quad u, v \in \mathbb{R}^n.$$

For the mean value, we have

$$(m, u) = \int_{\mathbb{R}^n} (z, u)\,d\mu(z), \quad u \in \mathbb{R}^n.$$

Those expressions are the first and the central second moments of measure μ.

Problems 1.4

20) A measure μ on $\mathcal{B}(\mathbb{R}^n)$ is called symmetric around the origin if $\mu(B) = \mu(-B)$, for all $\mathcal{B}(\mathbb{R}^n)$. Let μ be a probability measure on $\mathcal{B}(\mathbb{R}^n)$. Prove that μ is symmetric around the origin if, and only if, its characteristic function φ_μ takes real values only.

21) Let μ be a probability measure on $\mathcal{B}(\mathbb{R}^n)$. Prove that μ is invariant under all orthogonal transformations if, and only if, there exists a function $f : [0, +\infty) \to \mathbb{C}$ such that $\varphi_\mu(t) = f\left(||t||\right)$, $t \in \mathbb{R}^n$.

22) Let μ and ν be probability measures on $\mathcal{B}(\mathbb{R}^n)$ such that $\mu\left(\bar{B}(x,r)\right) = \nu(\bar{B}(x,r))$, for all closed balls $\bar{B}(x,r)$ in $\mathbb{R}^n$. Prove that $\mu = \nu$.

1.5. Gaussian vectors and Gaussian measures

1.5.1. *Characteristic functions of Gaussian vectors*

DEFINITION 1.12.– *A random vector ξ in $\mathbb{R}^n$ is called Gaussian if for each $a \in \mathbb{R}^n$, inner product (ξ, a) is a Gaussian r.v.*

Consider a Gaussian random vector $\xi = (\xi_k)_1^n$ in $\mathbb{R}^n$. Its components $\xi_k = (\xi, e_k)$ are Gaussian random variables, $\xi_k \sim N(m_k, \sigma_k^2)$ with $\sigma_k \geq 0$; $k = 1, \dots, n$. Such random variables, the components of a Gaussian random vector, are called jointly Gaussian. It holds

$$\mathsf{E}\,||\xi||^2 = \sum_1^n \mathsf{E}\,\xi_k^2 = \sum_1^n (m_k^2 + \sigma_k^2) < \infty,$$

and therefore, $\mathsf{Cov}(\xi)$ is well defined. We have

$$\mathsf{E}\,\xi = (\mathsf{E}\,\xi_k)_1^n = (m_k)_1^n =: m, \quad m \in \mathbb{R}^n;$$

$$S := \mathsf{Cov}(\xi) = (s_{ij})_{i,j=1}^n, \quad s_{ij} = \mathsf{Cov}(\xi_i, \xi_j),$$

$$s_{ii} = \mathsf{D}\,\xi_i = \sigma_i^2, \quad 1 \le i, j \le n.$$

Then we write

$$\xi \sim N(m, S)$$

and say that ξ is a Gaussian random vector with mean m and variance–covariance matrix S. Here, S is a positive semidefinite $n \times n$ real matrix as a variance–covariance matrix of a random vector in $\mathbb{R}^n$.

LEMMA 1.5.– If $\xi \sim N(m, S)$ then for each $a \in \mathbb{R}^n$,

$$(\xi, a) \sim N(m_a, \sigma_a^2), \quad m_a = (m, a), \quad \sigma_a^2 = (Sa, a).$$

PROOF.– R.v. (ξ, a) is Gaussian according to definition 1.12. Its mean and variance are evaluated in corollary 1.3. □

LEMMA 1.6.– If $\xi \sim N(m, S)$ in $\mathbb{R}^n$, then

$$\varphi_\xi(t) = \exp\left\{ i(t, m) - \frac{(St, t)}{2} \right\}, \quad t \in \mathbb{R}^n.$$

PROOF.– It holds

$$\varphi_\xi(t) = \mathsf{E}\,e^{i(\xi, t)} = \varphi_{(\xi, t)}(1).$$

Now, use lemmas 1.5 and [1.17]

$$\varphi_\xi(t) = \exp\left\{ iz(m, t) - \frac{z^2 (St, t)}{2} \right\} \Bigg|_{z=1},$$

and the statement follows. □

REMARK 1.3.– If a random vector ξ has characteristic function given in lemma 1.6, with certain $n \times n$ real and symmetric matrix S, then S is positive semidefinite and $\xi \sim N(m, S)$.

PROOF.– For $a \in \mathbb{R}^n$, it holds

$$\varphi_{(\xi, a)}(u) = \mathsf{E}\,e^{i(\xi, ua)} = \varphi_\xi(ua),$$

$$\varphi_{(\xi, a)}(u) = \exp\left\{ iu(a, m) - \frac{(Sa, a)u^2}{2} \right\}, \quad u \in \mathbb{R}.$$

Since the absolute value of characteristic function does not exceed 1, matrix S is positive semidefinite, and moreover $(\xi, a) \sim N(m_a, \sigma_a^2)$, with $m_a = (a, m)$, $\sigma_a^2 = (Sa, a)$. Hence, ξ is a Gaussian vector, with some parameters $m_1 \in \mathbb{R}^n$ and $S_1 \in \mathbb{R}^{n\times n}$, S_1 is positive semidefinite. Then lemma 1.5 implies $(\xi, a) \sim N(\tilde{m}_a, \tilde{\sigma}_a^2)$, with $\tilde{m}_a = (a, m_1)$, $\tilde{\sigma}_a^2 = (S_1 a, a)$. We get $(a, m) = (a, m_1)$ and $(Sa, a) = (S_1 a, a)$ for all $a \in \mathbb{R}^n$. Thus, $m_1 = m$ and $S_1 = S$. $\square$

DEFINITION 1.13.– *Random vector $\gamma \sim N(0, I_n)$ is called standard, or canonical, Gaussian vector distributed in $\mathbb{R}^n$.*

The components $\gamma_1, \ldots, \gamma_n$ of standard Gaussian vector are uncorrelated jointly Gaussian standard random variables. In particular, $\mathsf{E}\,\gamma_k = 0$, $\mathsf{D}\,\gamma_k = 1$, and for all $i \neq j$, $\mathsf{Cov}(\gamma_i, \gamma_j) = \mathsf{E}\,\gamma_i\gamma_j = 0$.

THEOREM 1.6.– (About components of standard Gaussian vector)

1) Let $\gamma = (\gamma_k)_1^n$ be standard Gaussian vector. Then

$$\varphi_\gamma(t) = e^{-\frac{||t||^2}{2}}, \quad t \in \mathbb{R}^n$$

and $\gamma_1, \ldots, \gamma_n$ are independent and identically distributed (i.i.d.) $N(0,1)$ random variables.

2) If $\gamma_1, \ldots, \gamma_n$ are i.i.d. $N(0,1)$ random variables, then $\gamma = (\gamma_k)_1^n$ is standard Gaussian vector.

PROOF.–

1) The formula for φ_γ follows from lemma 1.5 with $m = 0$ and $S = I_n$. Therefore,

$$\varphi_\gamma(t) = \prod_{k=1}^{n} e^{-\frac{t_k^2}{2}}.$$

Now, by lemma 1.3 the components of γ are independent.

2) Let $\gamma_1, \ldots, \gamma_n$ be i.i.d. $N(0,1)$ random variables. Then $\gamma = (\gamma_k)_1^n$ is a random vector, and for each $a \in \mathbb{R}^n$, $(\gamma, a) = \sum_{k=1}^n a_k\gamma_k$ is a Gaussian r.v. as a sum of independent Gaussian random variables. Therefore, γ is Gaussian. We have

$$\mathsf{E}\,\gamma = (\mathsf{E}\,\gamma_k)_1^n = 0,$$

$$\mathsf{Cov}\,\gamma = (s_{ij})_{i,j=1}^n, \quad s_{ij} = \mathsf{E}\,\gamma_i\gamma_j = \delta_{ij}, \quad 1 \le i, j \le n.$$

Hereafter, δ_{ij} is Kronecker delta, $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$, otherwise. Hence $\gamma \sim N(0, I_n)$. $\square$

REMARK 1.4.– (About uncorrelated Gaussian variables) Theorem 1.6 can be extended as follows: jointly Gaussian random variables $\xi_1, \ldots, \xi_n$ are independent if, and only if, they are uncorrelated. Based on theorem 1.6, this can be proven by consideration of normalized random variables $\eta_i = \frac{\xi_i - \mathsf{E}\,\xi_i}{\sqrt{\mathsf{D}\,\xi_i}}$ (before we cancel that ξ_i are constant a.s.).

1.5.2. *Expansion of Gaussian vector*

Lemma 1.5 can be generalized for any linear transformation of a Gaussian vector.

LEMMA 1.7.– (Linear transform of Gaussian vector) Let $\xi \sim N(m, S)$ be a Gaussian vector in $\mathbb{R}^n$ and $A \in \mathbb{R}^{p\times n}$. Then $A\xi$ is a Gaussian vector in $\mathbb{R}^p$ and

$$A\xi \sim N(Am, ASA^\top).$$

PROOF.– For $a \in \mathbb{R}^p$, it holds $(A\xi, a) = (\xi, A^\top a)$. It is a Gaussian r.v. due to definition 1.12. Now, the statement follows from lemma 1.2 and its proof. □

Now, we want to show that for any $m \in \mathbb{R}^n$ and positive semidefinite matrix $S \in \mathbb{R}^{n\times n}$, a random vector $\xi \sim N(m, S)$ can be obtained as an affine transformation of standard Gaussian vector.

Let $A \in \mathbb{R}^{n\times n}$ be a positive semidefinite matrix. Then it has orthonormal eigenbasis $e_1, \ldots, e_n$ and corresponding non-negative eigenvalues $\lambda_1, \ldots, \lambda_n$. The matrix can be expanded as follows:

$$A = \sum_1^n \lambda_k e_k e_k^\top.$$

DEFINITION 1.14.– *Let A be a positive semidefinite matrix as described above. $A^{\frac{1}{2}} = \sqrt{A}$ is a matrix satisfying*

$$\sqrt{A} = \sum_1^n \sqrt{\lambda_k} e_k e_k^\top.$$

It is a unique positive semidefinite matrix with its square equal to A. The matrix $\sqrt{A}$ has the same eigenbasis $e_1, \ldots, e_n$ and corresponding eigenvalues $\sqrt{\lambda_1}, \ldots,$ $\sqrt{\lambda_n}$.

LEMMA 1.8.– (Representation of Gaussian vector via standard Gaussian vector) Let $m \in \mathbb{R}^n$, $S \in \mathbb{R}^{n\times n}$ be a positive semidefinite matrix and $\gamma \sim N(0, I_n)$. Then

$$\xi := m + \sqrt{S}\gamma \sim N(m, S).$$

PROOF.– According to lemma 1.7

$$\sqrt{S}\gamma \sim N(0, \sqrt{S} I_n (\sqrt{S})^\top), \quad \sqrt{S}\gamma \sim N(0, S).$$

This implies the statement. □

Since $\gamma \sim N(0, I_n)$ can be constructed based on i.i.d. $N(0,1)$ random variables, lemma 1.8 shows that Gaussian $N(m, S)$ random vectors do exist, for any $m \in \mathbb{R}^n$ and any positive semidefinite matrix $S \in \mathbb{R}^{n\times n}$.

Now, we expand a Gaussian vector $\xi \sim N(m, S)$ using i.i.d. $N(0,1)$ random variables $\gamma_1, \dots, \gamma_r$, with $r = rk(S)$.

THEOREM 1.7.– (Decomposition of Gaussian vector) Let $m \in \mathbb{R}^n$ and $S \in \mathbb{R}^{n\times n}$ be a positive semidefinite matrix, with $rk(S) = r \geq 1$. Let $\lambda_1, \dots, \lambda_r$ be positive eigenvalues of S and $e_1, \dots, e_r$ be the corresponding orthonormal system of eigenvectors.

1) For $\xi \sim N(m, S)$, there exist i.i.d. $N(0,1)$ random variables $\gamma_1, \dots, \gamma_r$ on the underlying probability space such that

$$\xi = m + \sum_1^r \sqrt{\lambda_k}\gamma_k e_k, \quad \text{a.s.}$$

2) If $\gamma_1, \dots, \gamma_r$ are i.i.d. $N(0,1)$ random variables, then

$$\eta := m + \sum_1^r \sqrt{\lambda_k}\gamma_k e_k \sim N(m, S).$$

PROOF.– We complete the orthonormal system to an orthobasis $e_1, \dots, e_n$. Denote $\lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_n = 0$; they are the rest eigenvalues of S.

1) Introduce $X = \xi - m$, $X \sim N(0, S)$. Now, $X = \sum_1^n (X, e_k)e_k$, $\mathsf{E}(X, e_k) = 0$, and for all $k \geq r+1$, $\mathsf{D}(X, e_k) = \lambda_k = 0$. Hence $(X, e_k) = 0$ a.s., for all $k \geq r+1$. Thus,

$$X = \sum_1^r \sqrt{\lambda_k}\frac{(X, e_k)}{\sqrt{\lambda_k}} e_k, \quad \text{a.s.}$$

Random variables $\gamma_k := \frac{(X, e_k)}{\sqrt{\lambda_k}}$, $k = 1, \dots, r$ are jointly Gaussian (because vector $\gamma = (\gamma_k)_1^r$ is a linear transformation of Gaussian vector X) and $\mathsf{E}\,\gamma_k = 0$,

$$\mathsf{Cov}(\gamma_k, \gamma_j) = \frac{1}{\sqrt{\lambda_k \lambda_j}}(Se_k, e_j) = \frac{\lambda_k}{\sqrt{\lambda_k\lambda_j}}\delta_{kj} = \delta_{kj}.$$

According to theorem 1.6(1), $\gamma_1, \dots, \gamma_r$ are i.i.d. $N(0,1)$ random variables. Now,

$$X = \sum_1^r \sqrt{\lambda_k}\gamma_k e_k, \quad \text{a.s.}$$

and the statement follows.

2) For $a \in \mathbb{R}^n$, $(\eta, a) = (m, a) + \sum_1^r \sqrt{\lambda_k}\gamma_k(e_k, a)$ is a Gaussian r.v. as a sum of independent Gaussian random variables. Therefore, η is Gaussian. Next,

$$\mathsf{E}\,\eta = m, \quad \mathsf{Cov}(\eta) = \mathsf{E}\left(\sum_{k=1}^r \sqrt{\lambda_k}\gamma_k e_k\right)\left(\sum_{j=1}^r \sqrt{\lambda_j}\gamma_j e_j^\top\right) =$$

$$= \sum_{k,j=1}^r \sqrt{\lambda_k\lambda_j}(\mathsf{E}\,\gamma_k\gamma_j)e_k e_j^\top = \sum_{k=1}^r \lambda_k e_k e_k^\top = S.$$

Thus, $\eta \sim N(m, S)$. □

1.5.3. *Support of Gaussian vector*

DEFINITION 1.15.– *For a random vector X in $\mathbb{R}^n$, denote by G a union of all balls $B(x, r)$, with $\mathsf{P}\{X \in B(x, r)\} = 0$. The set $\mathbb{R}^n \setminus G$ is called support of X and denoted as supp X.*

It is clear that $\mathsf{P}\{X \in G\} = 0$. Moreover, G is the largest open set with this property. *supp* X is the smallest closed set such that X belongs to this set with probability 1. Since $G \neq \mathbb{R}^n$, *supp* $X \neq \emptyset$.

EXAMPLE 1.2.– Let that a r.v. ξ has Poisson distribution with parameter $\lambda > 0$. Then *supp* $\xi = \{0, 1, 2, \ldots, n, \ldots\}$.

THEOREM 1.8.– (About support of Gaussian vector) For random vector $\xi \sim N(0, S)$, *supp* $\xi = R(S)$, where $R(S)$ is the range of S.

PROOF.–

a) If $S = 0$, then $\xi = 0$ a.s., and *supp* $\xi = \{0\} = R(S)$.

Now, let $rk(S) = r \geq 1$, and let $\lambda_1, \ldots, \lambda_r$ be positive eigenvalues of S and $e_1, \ldots, e_r$ be the corresponding orthonormal system of eigenvectors. According to theorem 1.7(1), there exist i.i.d. $N(0, 1)$ random variables $\gamma_1, \ldots, \gamma_r$ on the underlying probability space such that

$$\xi = \sum_1^r \sqrt{\lambda_k}\gamma_k e_k, \quad \text{a.s.,}$$

and with probability one $\xi \in span(e_1, \ldots, e_r) = R(S)$.

b) Take arbitrary $x \in R(S)$ and $\varepsilon > 0$. Show that

$$\mathsf{P}\{\xi \in B(x, \varepsilon)\} > 0. \quad [1.21]$$

Indeed, $x = \sum_1^r a_k e_k$ and there exists $\delta > 0$ such that

$$M := \{y = \sum_1^r b_k e_k : |b_k - a_k| < \delta, k = 1, \dots, r\} \subset B(x, \varepsilon).$$

Then

$$\mathsf{P}\{\xi \in M\} = \mathsf{P}\left\{\gamma_k \in \left(\frac{a_k - \delta}{\sqrt{\lambda_k}}, \frac{a_k + \delta}{\sqrt{\lambda_k}}\right), k = 1, \dots, r\right\} =$$
$$= \prod_1^r \mathsf{P}\left\{\gamma_k \in \left(\frac{a_k - \delta}{\sqrt{\lambda_k}}, \frac{a_k + \delta}{\sqrt{\lambda_k}}\right)\right\} > 0,$$
$$\mathsf{P}\{\xi \in B(x, \varepsilon)\} \geq \mathsf{P}\{\xi \in M\} > 0.$$

Thus, [1.21] holds true. This fact and the relation shown in part (a) imply that $supp\, \xi = R(S)$. □

1.5.4. *Gaussian measures in Euclidean space*

Now, we reformulate the results of sections 1.5.1–1.5.3 for distributions of Gaussian vectors.

DEFINITION 1.16.– *A probability measure μ on $\mathcal{B}(\mathbb{R}^n)$ is called Gaussian measure in $\mathbb{R}^n$, if there exists a Gaussian random vector ξ in $\mathbb{R}^n$ such that its distribution $\mu_\xi = \mu$.*

Let $\xi \sim N(m, S)$ and $\mu = \mu_\xi$. In view of section 1.4.3, we have the following:

$$m = \mathsf{E}\,\xi = m_\mu, \quad S = \mathsf{Cov}(\xi) = \mathsf{Cov}(\mu),$$

i.e. m is mean value of μ and S is variance–covariance matrix of μ.

Next, according to lemma 1.6 the characteristic function of μ equals

$$\varphi_\mu(t) = \varphi_\xi(t) = \exp\left\{i(t, m) - \frac{(St, t)}{2}\right\}, \quad t \in \mathbb{R}^n.$$

Since parameters μ and S define uniquely the characteristic function of a Gaussian measure, they define uniquely the Gaussian measure μ itself. There is one-to-one correspondence between the set of all Gaussian measures in $\mathbb{R}^n$ and the set of couples $(m; S)$, where $m \in \mathbb{R}^n$ and S is a positive semidefinite $n \times n$ matrix.

Let $\lambda_1, \dots, \lambda_n$ be eigenvalues of S (they are non-negative) and $e_1, \dots, e_n$ be the corresponding eigenbasis of S. Then

$$\varphi_\mu(t) = \exp\left\{i \sum_1^n (t, e_k)(m, e_k) - \frac{1}{2} \sum_1^n \lambda_k (t, e_k)^2\right\},$$

$$\varphi_\mu(t) = \prod_1^n \exp\left\{i(t,e_k)(m,e_k) - \frac{1}{2}\lambda_k(t,e_k)^2\right\}.$$

We treat (t, e_k) as coordinates t_k of vector $t \in \mathbb{R}^n$, and the same for $(m, e_k) = m_k$. Thus,

$$\varphi_\mu(t) = \prod_1^n \exp\left\{it_k m_k - \frac{1}{2}\lambda_k t_k^2\right\}.$$

In view of lemma 1.3, we get a decomposition of μ:

$$\mu = \prod_1^n \mu_k,$$

where μ_k is a Gaussian measure on real line, with mean m_k and variance λ_k, $k = 1, \ldots, n$. Thus, each Gaussian measure in Euclidean space is just a product of Gaussian measures on real line.

In case, $rk(S) = r$, $1 \le r \le n-1$ we may and do assume that $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$. Then $\mu_k = \delta_{m_k}$(Dirac measure at point m_k), $k \ge r+1$, and we get the expansion

$$\mu = \prod_1^r \mu_k \times \prod_{r+1}^n \delta_{m_k} = \prod_1^r \mu_k \times \delta_z, \quad z = (m_{r+1}, \ldots, m_n)^\top.$$

Here, δ_z is Dirac measure on $\mathcal{B}(\mathbb{R}^{n-r})$ at point z.

DEFINITION 1.17.– *Standard Gaussian measure g in $\mathbb{R}^n$ is the distribution of standard Gaussian vector in $\mathbb{R}^n$.*

The measure g has zero mean and the variance–covariance matrix is equal to I_n. Its characteristic function is

$$\varphi_g(t) = e^{-\frac{||t||^2}{2}}, t \in \mathbb{R}^n.$$

It holds

$$g = \prod_1^n g_k,$$

where $g_1 = \ldots = g_n$ is standard Gaussian measure on real line. This means that

$$g_i(B) = \int_B \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \quad B \in \mathcal{B}(\mathbb{R}), \quad i = 1, \ldots, n.$$

DEFINITION 1.18.– *(See definition 1.15) For a probability measure μ on $\mathcal{B}(\mathbb{R}^n)$, denote by G a union of all balls $B(x,r)$, with $\mu(B(x,r)) = 0$. The set $\mathbb{R}^n \setminus G$ is called support of μ and is denoted as supp μ.*

Since $\mathbb{R}^n$ is separable, G is a countable union of balls $B(x_i, r_i)$, with $\mu(B(x_i, r_i)) = 0$. Hence, $\mu(G) = 0$. Moreover, G is the largest open set with this property. Note that $supp\ \mu$ is the smallest closed set such that the value of μ at this set equals 1. Since $G \neq \mathbb{R}^n$, $supp\ \mu \neq \emptyset$.

Theorem 1.8 implies the following: if μ is a Gaussian measure with mean 0 and variance–covariance matrix S, then $supp\ \mu = R(S)$. In particular for standard Gaussian measure g in $\mathbb{R}^n$, $supp\ g = \mathbb{R}^n$.

Now, we study the invariance of Gaussian measures under linear transformations. Let X be a random vector in $\mathbb{R}^n$ with distribution μ_X and $T : \mathbb{R}^n \to \mathbb{R}^n$ be a Borel function. Random vector TX has distribution $\mu_{TX} = \mu_X T^{-1}$. Therefore, μ_X is invariant under T (see definition 1.2) if, and only if, $X \stackrel{d}{=} TX$. Hereafter, $X \stackrel{d}{=} Y$ means that random vectors X and Y are identically distributed, i.e. $\mu_X = \mu_Y$.

Remember definition 1.14 of $A^{1/2} = \sqrt{A}$ for a positive semidefinite matrix A.

THEOREM 1.9.– (Invariance of Gaussian measure) Let $U \in \mathbb{R}^{n\times n}$ and μ be a Gaussian measure in $\mathbb{R}^n$, with zero mean and non-singular variance–covariance matrix S. The measure μ is U-invariant if, and only if, the matrix $S^{-1/2}US^{1/2}$ is orthogonal. In particular, standard Gaussian measure g in $\mathbb{R}^n$ is U-invariant if, and only if, U is an orthogonal matrix.

PROOF.– Let X be a Gaussian random vector, with $\mu_X = \mu$. Then $X \sim N(0, S)$ and by lemma 1.7, $UX \sim N(m, USU^\top)$. Now, $UX \stackrel{d}{=} X$ if, and only if,

$$USU^\top = S \quad \Leftrightarrow \quad \left(S^{-1/2}US^{1/2}\right)\left(S^{-1/2}US^{1/2}\right)^\top = I_n.$$

But this is equivalent to the orthogonality of matrix $S^{-1/2}US^{1/2}$.

In case $\mu = g$, it holds $S = I_n$. Thus, g is U-invariant if, and only if, U is an orthogonal matrix. □

The main statement of theorem 1.9 can be interpreted as follows. For a positive definite $S \in \mathbb{R}^{n\times n}$, we introduce a new inner product in $\mathbb{R}^n$,

$$(x, y)_S = \left(S^{-1}x, y\right), \quad x, y \in \mathbb{R}^n.$$

The corresponding norm is

$$||x||_S = \sqrt{(x,x)_S} = \sqrt{||S^{-\frac{1}{2}}x||^2} = ||S^{-\frac{1}{2}}x||, \quad x \in \mathbb{R}^n.$$

Now, a Gaussian measure μ, with zero mean and non-singular variance–covariance matrix S, is U-invariant if, and only if, U is unitary operator w.r.t. the inner product $(x, y)_S$.

Indeed, $S^{-1/2}US^{1/2}$ is an orthogonal matrix if, and only if,

$$||S^{-1/2}US^{1/2}x|| = ||x||, \quad x \in \mathbb{R}^n.$$

Now, make a change of variable $y = S^{1/2}x$, $y \in \mathbb{R}^n$. Then we get an equivalent condition

$$||S^{-1/2}Uy|| = ||S^{-1/2}y|| \quad \Leftrightarrow \quad ||Uy||_S = ||y||_S, \quad y \in \mathbb{R}^n.$$

The latter equality means that the linear transformation U is orthogonal w.r.t. the inner product $(x,y)_S$. One can say that a Gaussian measure μ changes the geometry of Euclidean space. Standard Gaussian measure g stays in correspondence with standard geometry of Euclidean space. Compared to Lebesgue measure λ_n (see theorem 1.3 and corollary 1.1), the measure g has fewer invariant transformations. Theorem 1.6 shows that λ_n cannot be extended to an infinite-dimensional Hilbert space H. But we will see that Gaussian measure can be constructed in H, with quite a large group of invariant transformations.

Problems 1.5

23) Let $A = (a_{ij})_{i,j=1}^n$ and $B = (b_{ij})_{i,j=1}^n$ be positive definite matrices. Prove that the matrix $C = (a_{ij}b_{ij})_{i,j=1}^n$ is positive definite as well.

24) Let f and g be pdfs, with cumulative distribution functions F and G, respectively. Prove the following:

a) For each $\alpha \in (-1,1)$,

$$h(x,y) := f(x)g(y) + \alpha f(x)(1-2F(x))g(y)(1-2G(y)), \quad (x,y) \in \mathbb{R}^2$$

is pdf, with marginal densities $f(x)$ and $g(y)$.

b) Assume additionally that f and g are even functions and $\int_{\mathbb{R}} |x|f(x)dx < \infty$, $\int_{\mathbb{R}} |y|f(y)dy < \infty$. Let $(X;Y)^\top$ be a random vector with pdf equal to $h(x,y)$. If $\alpha \in (0,1)$, then X and Y are positively correlated, and if $\alpha \in (-1,0)$, then X and Y are negatively correlated.

25) Based on problem (24), construct Gaussian random variables X and Y which are not jointly Gaussian.

26) Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and $X \sim N(0,S)$ in $\mathbb{R}^n$. Denote eigenvalues of $S^{1/2}AS^{1/2}$ as $\lambda_1, \dots, \lambda_n$. Prove that $I_\alpha := \mathsf{E}\exp\{\alpha(AX,X)\} < \infty$ if, and only if, $\alpha\lambda_k < \frac{1}{2}$, $k = 1,\dots,n$. Show that in this case $I_\alpha = \frac{1}{\sqrt{\prod_{k=1}^n (1-2\alpha\lambda_k)}}$.

27) Let $X \sim N(0, I_n)$. Find for which real α

$$I_\alpha := \mathsf{E}\,||X||^{-\alpha} < \infty.$$