GAUSSIAN MEASURES. A BRIEF SURVEY (*)

by Vladimir Sudakov (in St. Petersburg)(**)

Contents

0. Preliminaries.

The main content of these lectures is a brief survey of basic ideas in the general theory of Gaussian probability distributions in finite- and, in particular, in infinite-dimensional vector spaces. In these lectures we shall suppose the term "multi-dimensional" to mean "finite- or infinite-dimensional". The one-dimensional case is not excluded, but we shall examine mainly the case of very high or infinite dimension. Vector spaces over the field $\mathbb{R}$ of real numbers will always be

---

(*) Presentato al "Workshop di Teoria della Misura e Analisi Reale", Grado (Italia), 19 settembre-2 ottobre 1993.

(**) Indirizzo dell'Autore: Steklov Institute, St. Petersburg Branch, 197198, St. Petersburg (Russia).

considered unless another field ($\mathbb{C}$, if not $\mathbb{R}$) is specified. A preference will be given to coordinateless methods.

The term "probability measure", or "probability distribution", always means a nonnegative countable additive function $\mathbf{P}$ (or γ in case the measure $\mathbf{P}$ is Gaussian) defined on a σ-algebra $\mathfrak{A}$ of subsets of some set Ω, $\mathbf{P}(\Omega) = 1$. The triad $(\Omega, \mathfrak{A}, \mathbf{P})$ is called the *fundamental probability space* and often is provided with additional structures. For instance, it is meaningful to discuss whether or not a measure $\mathbf{P}$ is Gaussian only for vector spaces Ω, usually denoted in this case by E. The question of an agreement between both structures will be considered later.

Although we shall avoid going deeply into the history of the subject, some historical remarks are needed. First of all, the very term "Gaussian distribution", though accepted now almost everywhere, is not quite just since K. F. Gauss was not the first to introduce and to study "Gaussian" ("normal") distributions, at least in the one-dimensional case (nevertheless we shall follow the customary usage). Arising a normal distribution is the point of what is called "the central limit theorem". A special case of the central limit theorem (known as "Moivre theorem" about the limit behavior of the falling out coin heads number), though in "local" form, was first established by A. Moivre (1667–1754). His results were later extended by P. S. Laplace to "Moivre–Laplace theorem" dealing with a more general Bernoulli scheme (published in 1810). In addition, Laplace formulated this historically important theorem in the "integral" form, so that the limit normal distribution was described by him in a modern standard way by its density. At the same time Gauss was developing his least squares method, essentially based on the notion of joint distribution of a finite set of independent identically distributed Gaussian random variables. Gauss was also not the first to consider general joint distributions of independent, not necessary identically distributed, Gaussian random variables. G. A. A. Plana, a mathematician from Turin, probably was the first to do it. His paper appeared in 1813 and contained his earlier results related to the two-dimensional case. Another statistician who should be mentioned is R. Adrain. Still, Plana did not consider general multidimensional (nor general two-dimensional) Gaussian distributions.

General two-dimensional Gaussian distributions were for the first time introduced and investigated in details including coefficient of correlation by F. Galton (1889). Finally, Wiener stochastic process (N. Wiener, 1923) generating Wiener Gaussian measure on the space of continuous function of one variable was the first case of the infinite-dimensional Gaussian distribution ever considered.

Though A. N. Kolmogorov had published no paper dealing directly with Gaussian multi-dimensional distributions, the impact of his ideas can hardly be overestimated. His ideas permitted mathematicians to select the very subject of these lectures. The notion of metric entropy, important for our future discussions, also goes back to Kolmogorov. Many important results concerning properties of Gaussian multi-dimensional distributions were obtained during recent years; these results will be the subject of the discussion below.

1. Where do Gaussian distributions come from?

Do Gaussian distributions deserve such an attention? What are the reasons for the serious work of mathematicians in different countries during a few last decades which resulted in considerable new achievements in this field of the mathematical activity and profound changes in the very structure of the theory of Gaussian random variables and random functions? Is the continuing tendency to careful study of Gaussian distributions based mainly upon a comparative simplicity of their description? Or does the beauty of the subject seem for many of us to be so attractive? Or do Gaussian measures really constantly and independently arise from inside in serious physical and mathematical problems, and their study may indeed be considered as one of the fundamental topics of the theory since it inspires many other parts of the theory both in the results themselves and in the techniques of investigations? Surely, every of these reasons has its influence. Now some physically meaningful mathematical situations will be described such that the property of Gaussness strikingly arises even though no hint of such a property apparently can be found in the hypotheses prescribed.

The mentioned above one-dimensional case of the central limit theorem seems to be the first profound and simultaneously specific result in probability theory. Now it has been extended much further to include many versions of the infinity-dimensional case. Well known classical results as well as theorems of this kind obtained recently show how the interaction of many small independent or almost independent factors, having nothing in common with any Gaussness, acting additively, startlingly produces Gaussian distributions. This alone gives a good motivation for careful studying Gaussian measures and related questions *per se*.

Let us take another reason for naturally arising Gaussian distribution. This reason consists in the following characteristic property of Gaussian finite-dimensional distributions. Let p stand for the finite-dimensional probability density relative to a fixed Lebesgue measure λ of an absolutely continuous measure $\mathbf{P}$ on a finite-dimensional vector space E. The value of the integral

$$-\int p(\omega)\log p(\omega)\,\lambda(d\omega)$$

is called *the entropy* of the distribution $\mathbf{P}$ (with respect to the Lebesgue measure λ). In the case of discrete fundamental space Ω taken instead of vector space E, and the counting measure ν taken instead of the Lebesgue measure λ, this formula gives the value of the standard entropy of the (discrete) distribution $\mathbf{P}$ in the sense of Shannon information theory. Later (see part 2) we will see that every Gaussian measure γ is completely specified given its correlation characteristics: mean value (barycenter) $m_\gamma \in E$ and covariance quadratic form q_γ defined on the conjugate space E'.

THEOREM 1.1. *Among all measures on a vector finite-dimensional space* (E, λ) *with given barycenter* m *and covariance quadratic form* q*, the* (uniquely defined by them) *Gaussian measure and it alone has the maximal value of entropy.*

It seems to be reasonable to expect that measures with high values of their entropy are "more typical" than measures with low values of it.

Finally, let us mention here another very interesting property of probability distributions in finite-dimensional vector spaces of high dimension, which also leads to Gaussian distributions. For the purpose of simplicity of formulation, we will suppose that the vector spaces are Euclidean. Consider such Euclidean space E and *an arbitrary* measure $\mathbf{P}$ under *the only* restriction that its covariance quadratic form $q_{\mathbf{P}}$ (i.e., the trace of the $L^2(\mathbf{P})$-norm on E') exists and is bounded by a fixed quadratic form (say, by the square of the Euclidean norm in E'). Suppose a linear functional $f \in E'$ on the space E is being chosen at random with respect to the uniform probability distribution m on the surface of the unit ball $V_{E'}$ in the space E'. Then if the dimension of E is sufficiently high, there exists a one-dimensional distribution $\overline{\mathbf{P}}$ with a property to be an "almost typical" distribution for such linear functionals f being chosen at random. In other words, the distribution of f with the m-probability arbitrarily close to the unit is arbitrarily close (in some appropriate sense) to $\overline{\mathbf{P}}$. This typical distribution $\overline{\mathbf{P}}$ can always be taken as a *mixture of centered (one-dimensional) Gaussian measures.*

As an illustration one can take the unit cube in multi-dimensional (say, in 3-dimensional) Euclidean coordinate space E with the uniformly distributed probability measure $\mathbf{P}$ on it, and a one-dimensional subspace $L \subset E$ which should be thought of as being chosen "at random". Consider the image $\mathbf{P}_L = \mathbf{P} \circ \pi_L^{-1}$ of the measure $\mathbf{P}$ along the orthogonal projection $\pi\ E \to L$, i.e., really the distribution of the linear functional

$$f_L\ E \ni x \mapsto (x, e_L)_E,$$

where e_L is a unit vector from L. When L coincides with one of the coordinate axes (that is, of course, not typical for L) or is close to such an axis, the measure $\mathbf{P}_L$ is close to the measure uniformly distributed on an interval and hardly can be considered as one close to Gaussian measure. But according to the central limit theorem if many of coordinates are essentially involved (what is typical), then this distribution must be close to a Gaussian distribution. The classical central limit theorem is entirely based on the hypothesis of the independence (of coordinate functionals). The essence of the phenomenon under discussion consists in refusal of independence whatever. The price for this is evident: the typical distribution is *a* *mixture* of Gaussian one-dimensional centered distributions and not necessary can be token the pure Gaussian distribution, though suitable conditions for it can be given.

A precise formulation as well as some necessary discussion will be given later.

Theorem 1.1 not only establishes a characteristic property of Gaussian distribution. Together with other mathematical phenomena just mentioned, it helps understanding why Gaussian distributions are so wide-spread in the mathematical models of the reality. Some other characteristic properties without such ambitious pretensions will be discussed later.

One more remarkable property of the Gaussian distributions was used by J. Maxwell for establishing the distribution law for velocities of molecules. This three-dimensional distribution must possess the following properties:

1. The coordinate functionals are independent.
2. The space is isotropic, i.e., the distribution is rotation-invariant.

It can be proved that this distribution is centered Gaussian.

2. What are Gaussian measures? I.

Of course, readers ought to be aware of the common notion of Gaussian measure in a finite-dimensional vector space E or, at least, in the coordinate space $\mathbb{R}^n$ (otherwise they would not understand the preceding discussion). Here we are going to gather definitions of some notions closely connected with the theory of Gaussian measures. As always, using different approaches to the object helps us to understand more clearly the nature of things. In the finite-dimensional case there may be different approaches to well known and evidently well and uniquely defined mathematical object. It is not quite so in the infinitely-dimensional case. Some efforts and a serious discussion will be needed in order to avoid the possible emergence of some extremely pathological objects which can drastically distort desirable important properties of measures from the considered class. In this section *only the finite-dimensional case is considered*, though many formulation may be and will be used in the very general case.

DEFINITION 2.1. A Borel probability measure γ on a finite-dimensional vector space E is called nondegenerate Gaussian if it has a density p_γ with respect to a Lebesgue measure λ on E of the form

$$p_\gamma(x) = C \exp(-\frac{1}{2} Q(x - m)),$$

where $Q = Q_\gamma$ is a positive quadratic form on E, the mean value $m \in E$ is an arbitrary vector, and $C = C_{\gamma,\lambda}$ is the normalizing constant. A space with Gaussian measure is called *a Gaussian measure space.*

The property of a probability measure to be Gaussian does not depend on the choice of Lebesgue measure λ, and a Gaussian measure γ being fixed, the quadratic form Q_γ does not depend on the choice of λ as well.

We will call the quadratic form Q_γ *the concentration form* of the Gaussian measure γ, and call the subset

$$\mathcal{E}_\gamma = \{\, x \in E : Q_\gamma(x - m) \leq 1 \,\}$$

the concentration ellipsoid of the measure γ. The case $m = 0$ corresponds to *centered* Gaussian measures. The transition from γ to a centered Gaussian measure with the same concentration form is called *centering.* We shall restrict ourselves mainly dealing with centered measures. The general case would involve only additional technical difficulties, at least in the finite-dimensional case. A measure $\mathbf{P}$ on a vector space E is called *centered* if every linear functional has zero-mean distribution with respect to $\mathbf{P}$.

It is easy to verify that in the case the Lebesgue measure λ on E is in agreement with the Euclidean metric defined on E by the quadratic form Q (i.e., the λ-measure of the unit cube is equal to the unit), the value of C is given by the following formula:

$$C = (2\pi)^{-\frac{1}{2}\dim E}.$$

In the case the concentration form Q_γ coincides with the square of Euclidean norm on E and, in addition, $m_\gamma = 0$, the Gaussian measure $\gamma = \gamma_E^0$ is called *the standard Gaussian measure in/on/of the Euclidean space E.* In the general case if the Lebesgue measure

λ is in agreement with the Euclidean norm on E, and $Q_\gamma(x) = (R_\gamma x, x)_E$, where R_γ is some linear operator, then

$$C_\gamma = (2\pi)^{-\frac{1}{2}\dim E}(\det R_\gamma)^{\frac{1}{2}}.$$

Given a nondegenerate Gaussian measure γ on a vector space E with the concentration form Q_γ, the corresponding quadratic form q_γ can be considered on the conjugate space E'. Namely, the form Q_γ considered as a metric on E defines a Euclidean space structure on E, the concentration ellipse $\mathcal{E}_\gamma$ being the unit ball. This, in turn, permits to identify the spaces E and E', and hence, to consider the function Q_γ as a function q_γ on E'.

Let us name the quadratic form q_γ defined on E' *the covariance form* of the Gaussian measure γ. Note that the covariance form does not depend on the mean value m of the measure γ. In case $E = \mathbb{R}^n$, $E' = \mathbb{R}^n$ the matrix of the covariance form coincides with *the covariance matrix* V_γ of the Gaussian measure γ. Hence, the covariance form can be defined without any assumption of nondegeneration for any probability measure $\mathbf{P}$ on a vector space E if $E' \subset L^2(\mathbf{P})$: for $f \in E'$ and the measure $\mathbf{P}$ supposed to be centered, the value $q(f) = q_{\mathbf{P}}(f)$ coincides by the definition with the value of $\int (f(x))^2 \mathbf{P}(dx)$. In other words, q is the restriction of the square of the L^2-norm generated by $\mathbf{P}$ to the subspace $E' \subset L^2(E, \mathbf{P})$. All this gives us a possibility to define an arbitrary (not necessary nondegenerate) centered Gaussian measure on a Euclidean space E proceeding from an arbitrary positive (not necessary strong positive) quadratic form $q = q_{\mathbf{P}}$ defined on the conjugate space E'. Such a construction can be carried out as follows.

Definition 2.2. Given a measure $\mathbf{P}$ on a vector space E, its *characteristic functional* $\chi = \chi_{\mathbf{P}}\ E' \to \mathbb{C}$ is the function defined on every linear measurable functional f on E as follows:

$$\chi_{\mathbf{P}}(f) \stackrel{\text{def}}{=} \int \exp(i\langle x, f\rangle)\mathbf{P}(dx).$$

(In fact, the property of linearity of f is not essential).

Characteristic functional of a measure $\mathbf{P}$ can be considered as its Fourier-transform and restores this measure completely. For a nondegenerating Gaussian measure γ on a finite-dimensional vector space its characteristic functional has the form

$$\chi_\gamma(f) = \exp(i\langle m_\gamma, f\rangle - \frac{1}{2}q_\gamma(f)),$$

where the quadratic form q_γ is strictly positive, i.e., it has the maximal rank: its extremal values do not vanish.

It is known that the function $\chi(f) = \exp(i\langle m, f\rangle - \frac{1}{2}q(f))$ is positively definite for every positive (not necessary strictly positive) quadratic form q on vector space $F \ni f$. According to the well known Bochner theorem it means that χ is the characteristic functional of a measure γ (Gaussian by definition) on the conjugate space F' which can be canonically identified with such a space E that $E' = F$.

In what follows the notation F will be used for a vector space of random variables.

Definition 2.3. A Borel probability measure γ on a finite-dimensional vector space E is called *Gaussian* if there exist an element $m_\gamma \in E$ and a quadratic form q_γ on E' such that its characteristic functional χ_γ has the form

$$\chi_\gamma(f) = \exp(\langle m_\gamma, f\rangle - \frac{1}{2}q_\gamma(x)).$$

The quadratic form q_γ is called covariance form of γ, and the element m_γ is called its mean value.

Sometimes instead of the correlation quadratic form q_γ one considers *the correlation operator* $K_\gamma E' \to E$ such that $q_\gamma(f) = \langle K_\gamma f, f\rangle$. In the case $E = \mathbb{R}^d$ the matrix of the correlation operator coincides with the covariance matrix of the Gaussian measure γ (and also with the matrix of its covariance form).

The notions of Gaussian random variable, Gaussian random vector etc. are defined in the usual way.

The mean value is the barycenter of γ. It means that for every $f \in E'$ we have

$$\int \langle x, f\rangle \gamma(dx) = \langle m_\gamma, f\rangle.$$

Suppose some finite parameterized set $\underline{f} = f_1, \dots, f_n$ of elements of $E' = F$ is fixed. (This does not exclude coincidence of some elements for different indices.) Consider the map

$$\underline{f}\, E \to \mathbb{R}^n, \quad E \ni x \mapsto (\langle x, f_1\rangle, \dots, \langle x, f_n\rangle) \in \mathbb{R}^n.$$

The image $\gamma\underline{f}^{-1}$ of the measure γ under the map $\underline{f}$ is a Gaussian measure ("the joint distribution of the random variables $f_1, \dots, f_n$"). Its mean value is the image of m_γ, and its covariance form $q_{\gamma\underline{f}^{-1}}$ is defined with a help of the conjugate map $\underline{f}^*\, \mathbb{R}^n \to F$, $(c_1, \dots, c_n) \mapsto \sum c_k f_k$:

$$\mathbb{R}^n \ni (c_1, \dots, c_n) \mapsto q_{\gamma\underline{f}^{-1}}(c_1, \dots, c_n) = q_\gamma(\sum c_k f_k).$$

Since $\underline{f}$ is a linear map of the general form, a very important property of Gaussian measures is established:

The image of a Gaussian measure under a linear map is always Gaussian.

Note that a coordinate system on the vector space E with a Gaussian measure γ being fixed, the matrix of a covariance form q_γ is the covariance matrix of the set of coordinate functionals (considered as random variables). In the case of centered measure it is the Gram matrix (in L^2-norm) for the set of coordinate functionals.

The condition of Gaussness is not essential here.

Note also that defining what the Gaussian measure is we tried to use no extra assumptions relative to the space E. The only structure used for our purpose was the structure of a vector space which itself generates the Borel structure on E. We avoid using any Euclidean structure on the vector space E as well as on its conjugate space till it is necessary.

At first sight it seems that the vector space structure is the minimal structure needed for reasonable definition of a Gaussian measure. Still a more careful consideration shows that in fact only the structure of an affine space is needed. (Remind that it just is the structure defined in the natural way on a shifted vector subspace: linear combinations of elements of an affine space are defined for linear combinations with zero sums of its coefficients only). A Borel probability measure on a finite-dimensional affine space A is Gaussian or not depending on form of its characteristic functional on the *vector* space A_1' of all polynomials on A of degree one, dimension of which exceeds the dimension of the affine space A by one. Or one can turn the affine space A into a vector space by fixing a zero-point. The most convenient way is to declare the mean value of the considered Gaussian measure as the zero-point, so we at once come to a centered Gaussian measure. Readers can consider details themselves. As in vector case, a Gaussian measure γ on an affine space is completely defined by its mean-value (barycenter) m_γ and covariance quadratic form q_γ. In the case of affine space no meaning has the notion of centered Gaussian measure.

These remarks are especially meaningful when the conditional Gaussian distributions are being considered on elements of an affine measurable partition of a Gaussian measure space (all the definitions will be given later).

At the end of this section let us gather a few other important properties of finite-dimensional Gaussian measures.

1. Let $X_1, \ldots, X_n$ be independent random variables and let their two linear combinations

$$L_1 = \alpha_1 X_1 + \ldots + \alpha_n X_n \quad \text{and} \quad L_2 = \beta_1 X_1 + \ldots + \beta_n X_n$$

are independent. In this case for every k such that $\alpha_k \beta_k \neq 0$ the random variables X_k are Gaussian. (In particular, if for independent X and Y the sum $X + Y$ and the difference $X - Y$ are independent, then X and Y are Gaussian and have equal variances.)

This theorem can be extended to the vector-valued case, too.

2. If X and Y are two independent Gaussian random vectors $\Omega \to E$, where $(\Omega, \mathfrak{A}, \mathbf{P})$ is a fundamental probability space, then, by definition, their joint distribution is a Gaussian measure on $E \times E$. Hence, their sum is Gaussian, i.e., the convolution of Gaussian (multi-dimensional) distributions is also Gaussian. Given two Gaussian measures γ_1 and γ_2 on the vector space E with covariation forms q_1 and q_2, their convolution $\gamma_1 * \gamma_2$ has the covariance form $q_1 + q_2$.

3. Given a nondegenerating Gaussian measure in a d-dimensional vector space, the distribution of its concentration form with respect to this measure is well known χ_d^2-distribution (chi-square distribution with d degrees of freedom). The density of this distribution has the

form:

$$p_d(t) = C_d(\frac{t}{2})^{\frac{d}{2}-1}\exp(-\frac{1}{2}t).$$

Here $C_d = 2\Gamma(\frac{d}{2})$ is the normalizing constant. As it is easy to calculate, the mean value of this distribution is equal to d and its variance is equal to $2d$. It is easy to verify that if d is large, the distribution of the covariance form is concentrated relatively close to $d^{\frac{1}{2}}$. The distributions of $d^{-\frac{1}{2}}q_\gamma$ weakly converge to δ_1 (the probability measure concentrated at 1)

4. If $f_1, f_2 \in F$ are orthogonal in sense of Euclidean structure defined by q_γ, then these random variables are independent.

5. Suppose several Gaussian random variables (i.e., elements of F) are fixed, $L \subset F$ is a linear subspace spanned on these variables, and $f \in F$. Given values of the fixed random variables (and hence values of all the elements of L), consider the problem of the best prediction (in sense of L^2 distance) of the value of f. It means that we consider the vector space of *all* random variables such that their values are defined as soon as the values of elements of L are known. In other words, we consider the vector space of all measurable functions of the given random variables and look for the closest one in L^2-sense to f among them. Such a vector space can be described as the space of all functions on E measurable with respect to the σ-algebra $\sigma(L)$ generated by L. It is sufficient for our purpose to take only the subspace $L^2(E, \sigma(L), \gamma)$ of $L^2(E, \mathfrak{A}, \gamma)$. This σ-algebra $\sigma(L)$ consists of all measurable unions of shifts of the subspace $L^0 \subset E$ where L^0 is the polar of $L \subset F$ in E. In other words we consider the problem of finding the q_γ-orthogonal projection of $f \in F \subset L^2(E, \mathfrak{A}, \gamma)$ on $L^2(E, \sigma(L), \gamma)$. It is remarkable that in the Gaussian case this projection belongs to the very space L which is always only a special proper subspace of $L^2(E, \sigma(L), \gamma)$! It also means that given values of several Gaussian random variables, the best prediction for another Gaussian random variable (in the sense of minimal variance) can be always expressed as a *linear combination* (and not as a more complicated function) of these variables.

In fact, this apparently unexpected property can be easily explained. We will do it after introduction of notion of conditional distributions.

3. Category properties. Gaussian measure structure.

When dealing with measure theory it is very usual and useful not to distinguish between objects of the theory (sets, functions, maps, partitions etc) the difference of which can be described in terms of zero measure sets. Sets $A, B \subset \Omega$ are told to be equivalent if $\mathbf{P}(A \triangle B) = 0$. Functions (or random variables) X and Y defined on Ω are told to be equivalent if they differ only on a zero measure set (coincide "almost everywhere", or "almost surely"). In similar cases we will use the notation "mod0" and write $A = B \mathrm{mod} 0$, $X = Y \mathrm{mod} 0$. Going on, two probability measure spaces $(\Omega_1, \mathfrak{A}_1, \mathbf{P}_1)$ and $(\Omega_2, \mathfrak{A}_2, \mathbf{P}_2)$ are called isomorphic mod0 if there exist null sets $N_1 \in \mathfrak{A}_1$ and $N_2 \in \mathfrak{A}_2$ such that the measure subspaces defined by $\Omega_1 \setminus N_1 \subset \Omega_1$ and by $\Omega_2 \setminus N_2 \subset \Omega_2$ are isomorphic in the standard sense of measure theory. It means existing a bijection (one-to-one correspondence) between the sets $\Omega_1 \setminus N_1$ and $\Omega_2 \setminus N_2$ such that for every measurable subset of any of these sets its image under the bijection is also measurable and has the same value of measure. For instance, if $\mathbf{P}_1$ is a discrete measure concentrated on the set $\Omega_1 = \{ n^{-1},\ n = 1, \ldots \}$ such that $\mathbf{P}_1(\{n^{-1}\}) = 2^{-n}$, and $\Omega_2 = \mathbb{R}$, $\mathbf{P}_2$ being a purely atomic measure concentrated on the same points as $\mathbf{P}_1$ and with the same values at the corresponding one-element subsets, or $\mathbf{P}_1$ and $\mathbf{P}_2$ are two centered Gaussian measures possessing covariation forms with the same signature and defined on vector spaces of different dimensions, then we say that these measure spaces are isomorphic mod0, or isomorphic up to the equivalence.

In measure theory the notion of a measure-preserving transformation is of great importance. Often one does not distinguishes between mod0-equivalent transformations. Classes of mod0 equivalence of measure spaces as objects together with classes of mod0 equivalence of measure-preserving transformations as morphisms form a category of measure spaces considered up to the equivalence.

The notion of Gaussian measure is based on the structures of the measure theory as well as of the theory of vector spaces. Hence, in the theory of Gaussian measure spaces instead of measure-preserving transformations of the general form we should consider mod0 *linear measure-preserving* transformations. Factorization mod0 of linear maps will be described if we describe factorization mod0 of linear functionals on a Gaussian measure space $(E, \mathfrak{A}, \gamma)$. Evidently, for $f_1, f_2 \in E'$ $f_1 = f_2 \text{mod} 0$ if and only if we have $q_\gamma(f_1 - f_2) = 0$. In other words, the space E'_γ of all (mod0-equivalence classes of) γ-measurable linear functionals coincides with Euclidean (hence, separable) space generated by the metric q_γ:

$$E'_\gamma = E'/\{ f \in E' : q_\gamma(f) = 0\}.$$

(Here $\text{Ker} q_\gamma = \{ f \in E' : q_\gamma(f) = 0 \}$ is a vector subspace because of positiveness of q_γ). In the case $q_\gamma(f) = 0$ only for $f = 0$ or, in other words, for nondegenerate γ we have $E'_\gamma = E'$. Since the only invariant of q_γ and, hence, of E'_γ with respect to the vector space structure is the dimension (in other words, the signature) of q_γ, we come to the conclusion that *for every natural number d there exists a unique to within the* mod0 *measure preserving transformation Gaussian measure of dimension d.* The standard Gaussian measure in the d-dimensional Euclidean vector space is a representative of this d-dimensional measure. This statement seems to be almost evident for considered case of finite-dimensional E. We will see that it is also true in the general case if a definition of Gaussian measure is selected in a proper way.

We can say that the space E'_γ depends only on γ and does not depend on the space E. We shall use the notation F'_γ for this Euclidean space.

If two Gaussian measures γ_1 and γ_2 are considered on the same finite-dimensional vector space E, they are mutually absolutely continuous if and only if their covariance quadratic forms q_1 and q_2 are equivalent, i.e., Euclidean spaces generated by them coincide as vector spaces (these forms generate Euclidean structure on the same vector space $E'/\text{Ker} q_1 = E'/\text{Ker} q_2$). It means that null-subspaces of these two positive quadratic forms coincide.

Any two nondegenerate to δ finite-dimensional Gaussian measures are isomorphic mod0 in the sense of the isomorphism of measure spaces (not linear measure spaces, if their dimensions differ).

In the case of a linear (and hence measurable in the finite-dimensional case) map l of general form $(E, \mathfrak{A}, \gamma)$ to another vector space the image γl^{-1} of γ can be described as follows.

THEOREM 3.1. *Consider a vector space with a centered Gaussian measure* $(E, \mathfrak{A}, \gamma)$ *and a linear measurable map* l *from* E *to another vector space* E_1. *Then the image of* γ *is a centered Gaussian measure on* E_1 *defined by the correlation quadratic form* $q_{\gamma l^{-1}}$ *on* E_1' *which can be described as follows:*

$$q_{\gamma l^{-1}}(f_1) = q_\gamma(l^* f_1),$$

where l^* *is the conjugate map* $l^* E_1' \to E'$. *The concentration ellipsoid* $\mathcal{E}_1$ *of the image of Gaussian measure* γ *is the image* $l(\mathcal{E})$ *of the concentration ellipsoid of* γ.

The concentration form $q_{\gamma l^{-1}}$ is canonically isomorphic to the restriction of the form q_γ to the polar $N^\circ \subset E'$ of the kernel $N = \mathrm{Ker} l \subset E$.

Reader can notice that there exist two kinds of "Gaussian" objects: those connected immediately with the space E (Gaussian measures γ itself; their concentration quadratic forms Q; σ-algebras of events; measurable partitions of Gaussian measure spaces etc.); and those connected immediately with Gaussian Euclidean spaces F_γ like correlation quadratic forms and corresponding Euclidean structures; random variables as elements of Gaussian spaces F_γ; optimal predictors and so on.

Given a finite-dimensional Gaussian measure space $(E, \mathfrak{A}, \gamma)$, we can consider Hilbert space $L^2(E, \mathfrak{A}, \gamma)$ of all quadratic integrable random variables, taking into account no vector space structure. The space $E_\gamma' = F_\gamma$ of all classes of equivalent Gaussian random variables is a proper finite-dimensional subspace of this Hilbert space L^2. The joint distribution of elements of any finite subset of E_γ is Gaussian by Theorem 3.1.

DEFINITION 3.2. Suppose $(\Omega, \mathfrak{A}, \mathbf{P})$ is a nonatomic probability measure space. A subspace $G \subset L^2(\Omega, \mathfrak{A}, \mathbf{P})$ is called *Gaussian* if the joint distribution of any its finite subset is Gaussian. Gaussian subspace is called *maximal* if it is a proper subspace of no other Gaussian subspace.

Given two Gaussian subspaces, their linear span is not necessarily a Gaussian subspace.

THEOREM 3.3. *The subspace $E'_\gamma \subset L^2(E, \mathfrak{A}, \gamma)$ is a maximal Gaussian subspace.*

Consider two maximal finite dimensional Gaussian subspaces G_1 and G_2 of $L^2(\Omega, \mathfrak{A}, \mathbf{P})$ where $\mathbf{P}$ is a nonatomic probability measure. An orthogonal operator $U\ L^2 \to L^2$ is called *an automorphism of the L^2-structure* if the joint distribution of any its finite subset coincides with the joint distribution of the U-image of this subset. In particular, for any automorphism T of the probability measure space $(\Omega, \mathfrak{A}, \mathbf{P})$ the conjugate orthogonal operator U_T,

$$(U_T f)(\omega) \stackrel{\text{def}}{=} f(T(\omega)),$$

is an authomorphism of the L_2-structure. Can these subspaces G_1 and G_2 be coincided by an automorphism of the L^2-structure? Of course, the necessary condition for it is the equality of their dimensions. Further, in the case $(\Omega, \mathfrak{A}, \mathbf{P}) = (E, \mathfrak{A}, \gamma)$ we have $\sigma(F_\gamma) = \mathfrak{A}$ (here $\sigma(L)$ denotes the σ-algebra generated by L).

THEOREM 3.4. *For any two Gaussian subspaces $G_1, G_2 \subset L^2(\Omega, \mathfrak{A}, \mathbf{P})$ of the same dimension and such that $\sigma(G_1) = \sigma(G_2) = \mathfrak{A}$ there exists an automorphism U of the L^2-structure such that $G_2 = UG_1$.*

The very interesting question whether or not every automorphism of the L^2-structure is conjugate to an automorphism T of the measure space $(\Omega, \mathfrak{A}, \mathbf{P})$, must be answered, generally speaking, negatively. Nevertheless, the answer becomes positive if the definition of the probability measure space is slightly modified in order to avoid pathologies. The proper class of probability measure spaces is formed by so called Lebesgue–Rokhlin measure spaces. They will be discussed later.

4. Multidimensional Hermite–Itô polynomials and orthogonal expansions.

Hermite polynomials $\mathrm{He}_n(x)$, $n = 0, \ldots$ form the system of orthonormal in the Hilbert space $L^2(\mathbb{R}, \mathfrak{B}, \gamma_0)$ polynomials for which the sequence of spanned subspaces coincides with the sequence of subspaces spanned on first monoms x^k (notations are not quite standardized). So,

$$\begin{aligned}
\mathrm{He}_0(x) &= 1\!\!\perp, \\
\mathrm{He}_1(x) &= x, \\
\mathrm{He}_2(x) &= 2^{-\frac{1}{2}}(x^2 - 1), \\
\mathrm{He}_{n+1}(x) &= (n+1)^{-\frac{1}{2}}(x\mathrm{He}_n(x) - \mathrm{He}_{n-1}(x)), \\
\mathrm{He}'(x) &= n^{\frac{1}{2}}\mathrm{He}_{n-1}(x), \\
\mathrm{He}_n(x) &= (-1)^n n^{-\frac{1}{2}} \exp(\tfrac{1}{2}x^2)\tfrac{d^n}{dx^n} \exp(-\tfrac{1}{2}x^2).
\end{aligned}$$

The sequence of the Hermite polynomials is obtained from the sequence $\{x^n\}$ by the usual procedure of sequential orthogonalization.

Consider now one-dimensional Gaussian measure space $(E, \mathfrak{A}, \gamma)$ and fix a unit vector $f \in E'_\gamma \subset L^2(\gamma)$. For every $n \in \mathbb{N}$ the function f^n is an element of $L^2(\gamma)$ as well as the unit function $1\!\!\perp$. The subspace $F = \mathrm{span}\{1\!\!\perp, f, f^2, \ldots, f^n\}$ of L^2 is the subspace of all polynomials on E of degree not exceeding n. (The notion of a polynomial on a vector space is well defined.) Now we come to orthogonal expansion

$$L^2(\gamma) = F_0 \oplus (F_1 \ominus F_0) \oplus (F_2 \ominus F_1) \oplus \ldots.$$

Evidently, for every $n \in \mathbb{N}$ we have $\mathrm{He}_n(f) \in F_n \ominus F_{n-1}$. The construction of this expansion is valid for arbitrary multi-dimensional Gaussian measure γ. In multi-dimensional case elements of the subspace $F_n \ominus F_{n-1} \subset L^2(\gamma)$ are called by definition Hermite–Itô polynomials of degree n. Hence, every element of $L^2(\gamma)$ admits a unique expansion in orthogonal Hermite polynomials. Of course, one can choose an orthonormal basis in every such a subspace consisting of Hermite polynomials of a fixed degree.

Let g be an arbitrary polynomial of degree n. The projection $\pi_{F_n \ominus F_{n-1}} g$ is called *the Hermite–Itô polynomial* of the random variable g and is denoted by $:g:$. Sometimes such polynomials are called

Wick polynomials, and $: \cdot :$ is called *the Wick symbol.* In the case $g = f_1 f_2 \dots f_n$, where $f_1, f_2, \dots, f_n$ are elements of E'_γ and hence are Gaussian, the Hermite–Itô polynomial for g is called the Hermite–Itô polynomial of the random variables $f_1, \dots, f_n$. If these random variables are orthogonal (and hence independent), their Hermite–Itô polynomial is merely the product of these variables.

While in the one-dimensional case of standard Hermitean polynomials the dimension of the space of all Hermite–Itô polynomials of given degree is equal to the unit, in case $1 < \dim F_\gamma < \infty$ dimensions of such Hermite–Itô polynomials spaces are larger than the unit, though finite. In the case of infinite-dimensional Gaussian measure, to be considered later, all Hermite–Itô subspaces are infinite-dimensional. In any case these subspaces are invariant with respect to the orthogonal transformations of $L^2(\gamma)$ conjugate to measure-preserving transformations of the measure space, generated by rotations. Every such transformation is conjugate to an orthogonal transformation of Euclidean (Hilbert in the infinite-dimensional case) space E'_γ, and every orthogonal transformation of E'_γ generates a measure-preserving transformation ("rotation") of the space $(E, \mathfrak{A}, \gamma)$ and, hence, has a uniquely defined extension on every Hermite–Itô subspace as well as on the whole $L_2(\gamma)$.

5. General measure theory: some nontraditional aspects.

Axiomatics of spaces with measure (of probability spaces) is very simple, and it helps us to grasp fundamental concepts of probability theory and to penetrate its deeper fields. Still the further we move, the more often we meet difficulties that arise mainly because such a simple and clear axiomatics turns out to be not quite adequate to the nature of things. It does not exclude some unpleasant, even disgusting pathologies, which must be eliminated by a more careful selection of the objects mathematicians deal with. Of course, such situations are by no means distinctive features of the theory of Gaussian measures.

A. N. Kolmogorov considered as the main achievement of his famous monograph "Grundbegriffe der Wahrscheinlichkeitsrechnung", 1933, the elaboration of the notion of conditional mathematical expectation of a random variable with respect to a given σ-subalgebra of events. His approach was based on Radon–Nikodym theorem about absolutely continuous measures. Kolmogorov's conditional expectations, however, cannot be considered as a quite adequate extension of notion of conditional distribution well known in classical probability theory, i.e., given an event of positive probability. It is known that Kolmogorov understood impossibility to define in the general case *a system of conditional probability distributions* on elements of a measurable (in whatever acceptable sense) continual partition generating σ-subalgebra as it can be easily done using the classical definition of conditional distribution for elements of finite or countable measurable partitions.

It turned out that such impossibility can manifest itself only in connection with pathological situations. In the precise sense the nonexistence of the system of conditional distributions on elements of a measurable partition means that a nonmeasurable with respect to the Lebesgue measure subset is involved in the very nature of the considered measure space with countable generated initial σ-algebra.

Kolmogorov's pupil V. A. Rokhlin selected (1940, 1946) a very important class of measure spaces with countable generated σ-algebras such that its members were saved from any pathology and therefore possessed important properties not valid in the general case. Rokhlin called his measure spaces Lebesgue spaces. Now, after Rokhlin's death, the term “Lebesgue–Rokhlin spaces” has been more convenient.

We have no possibility to give here a full exposition of Rokhlin's theory. Rokhlin gave an axiomatic description of what he called Lebesgue spaces. Another description refers us to Lebesgue measure: the Lebesgue–Rokhlin nonatomical space is a probability measure space isomorphic mod0 to the unit interval with Lebesgue measure.

Definition 5.1. The probability measure space $(\Omega, \mathfrak{A}, \mathbf{P})$ with complete σ-algebra is called *Lebesgue–Rokhlin space*, or LR-space, if

1. σ-algebra $\mathfrak{A}$ is *countably generated with respect to* $\mathbf{P}$, i.e., it is the completion with respect to $\mathbf{P}$ of a σ-algebra spanned on a countable subset $\mathcal{B} = \{B_k\}$ of $\mathfrak{A}$ (*a basis*), and $\mathcal{B}$ separates points

of Ω. In other words, $(\Omega, \mathfrak{A}, \mathbf{P})$ is a separable probability measure space.

2. For some (and then for every) basis $\mathcal{B}$ the image of Ω under the canonical imbedding mod0:

$$j_{\mathcal{B}}\, \Omega \to \{0,1\}^{\mathbb{N}}, \quad j_{\mathcal{B}}(\omega) = \{\, 1\!\!\perp_{B_k}(\omega) : k \in \mathbb{N}\}$$

is a measurable subset of the measure space

$$(\{0,1\}^{\mathbb{N}}, \mathrm{compl}\mathfrak{B}, \mathbf{P} j_{\mathcal{B}}^{-1}, \mathbf{P} j_{\mathcal{B}}^{-1})$$

(compl means "completion"). In other words, the canonical imbedding $j_{\mathcal{B}}$ is an isomorphism mod0. (In this case the basis $\mathcal{B}$ is called complete mod0). The measure $\mathbf{P}$ of the LR-space is called Lebesgue–Rokhlin measure.

We now see the precise meaning of the assertion that every separable probability measure space which has no Lebesgue–Rokhlin property is isomorphic to a nonmeasurable with respect to the Lebesgue measure subset of external measure equal to the unit. In fact, every nonatomic Borel probability measure on the compact set $\{0,1\}^{\mathbb{N}}$ is isomorphic mod0 to the Lebesgue measure on an interval. Note that it follows that the property of a subset to be measurable is its intrinsic property and does not depend on how this subset is situated in a LR-space.

THEOREM 5.2. *Every complete separable metric space with probability measure* $\mathbf{P}$ *defined on the completed with respect to* $\mathbf{P}$ *Borel* σ*-algebra is a Lebesgue–Rokhlin space. Every Borel space* $(B, \mathfrak{B})$ *with a probability measure* $\mathbf{P}$ *on its* σ*-algebra* $\mathfrak{B}$ *completed with respect to* $\mathbf{P}$ *is a Lebesgue–Rokhlin space. Every discrete probability measure space* $(\Omega, \mathfrak{A}, \mathbf{P})$ *on* σ*-algebra* 2^{Ω} *of all subsets is a Lebesgue–Rokhlin space.*

In particular, polish space with a probability measure on the completed σ-algebra has the Lebesgue–Rokhlin property.

The class of Lebesgue–Rokhlin spaces is stable with respect to all usual operations. No wonder: it includes all nonpathological probability measure spaces. In particular, the quotient space relative to nonpathological equivalence relation (relative to measurable partition) and homomorphic image of a Lebesgue–Rokhlin space are Lebesgue–Rokhlin spaces, too. The projective limit of any family of LR-spaces exists, and has LR-property for countable families. For a Lebesgue–Rokhlin space it is impossible neither enrich nor narrow its σ-algebra conserving the LR-property.

The central point of the theory of LR-spaces is the theory of measurable partitions and, in particular, the theorem about existence of a family of conditional distributions on elements of a measurable partition.

DEFINITION 5.3. A partition ξ of a probability measure space is called *measurable* in case it is generated by a finite or countable family $\mathcal{B}$ of measurable subsets. (Two points belong to the same element of partition generated by $\mathcal{B}$ if no subset of this family separates them. The map $j_{\mathcal{B}}$ does not distinguish them.)

A partition ξ of a probability measure space is measurable if and only if its elements are the sets of constancy for some measurable function f. If so, we write $\xi = \zeta(f)$. Similarly, the notation $\zeta(f_1, \dots, f_n)$ will be used. Always ε denotes the partition into separate points, and ν denotes the trivial partition. Often we will not distinguish between partitions coinciding mod0 using the term *measurable* mod0 *partitions*.

Every measurable mod0 partition ξ defines a σ-subalgebra $\mathfrak{A}_\xi$. Moreover, there is a bijection between the set MP of all measurable mod0 partitions of a Lebesgue–Rokhlin space, and the set of all its σ-subalgebras of subsets.

The space MP of all measurable mod0 partitions is a lattice with respect to the natural relation of order:

$$\xi < \eta \iff \mathfrak{A}_\xi \subset \mathfrak{A}_\eta.$$

Any subset of the lattice MP has its supremum and infimum. We will use the notations $\vee$ for the supremum and $\wedge$ for infimum. Always

$$\mathfrak{A}_{\xi\wedge\eta} = \mathfrak{A}_\xi \cap \mathfrak{A}_\eta.$$

Given a LR-space $(\Omega, \mathfrak{A}, \mathbf{P})$ and its measurable partition ξ, the quotient space

$$(\Omega, \mathfrak{A}, \mathbf{P})/\xi = (\Omega/\xi, \mathfrak{A}/\xi, \mathbf{P}\pi_\xi^{-1})$$

is defined, where π_ξ stands for the canonical projection $\pi_\xi\, \Omega \to \Omega/\xi$ and $\mathfrak{A}/\xi$ stands for σ-algebra of all $\mathfrak{A}$-measurable ξ-sets. Usually we will write $\mathbf{P}/\xi$ instead of $\mathbf{P}\pi_\xi^{-1}$. The quotient space Ω/ξ is also defined mod0 for the class ξ of mod0 equivalent partitions.

Definition 5.4. Suppose $(\Omega, \mathfrak{A}, \mathbf{P})$ is a separable probability measure space and ξ is a measurable partition of it. Suppose a σ-algebra $\mathfrak{A}_C \subset \mathfrak{A} \cap C$ is defined on every element $C \in \Omega/\xi$ as well as a probability measure $\mathbf{P}_C^\xi$ such that

1. For every mod0 (with respect to the measure $\mathbf{P}/\xi$) element C the space $(C, \mathfrak{A}_C, \mathbf{P}_C^\xi)$ is a Lebesgue–Rokhlin space.

2. For every measurable $A \in \mathfrak{A}$

a. for every mod0 element $C \in \Omega/\xi$ the set $A \cap C$ is $\mathfrak{A}_C$-measurable;

b. the function $C \mapsto \mathbf{P}_C^\xi(A \cap C)$ is measurable on $(\Omega, \mathfrak{A}, \mathbf{P})/\xi$;

c. $\mathbf{P}(A) = \int_{\Omega/\xi} \mathbf{P}_C^\xi(A \cap C)(\mathbf{P}/\xi)(dC)$.

Then the family of measures $\{\, \mathbf{P}_C^\xi,\ C \in \Omega/\xi \,\}$ is called *the system of conditional probability measures* on elements C of the measurable partition ξ.

Here is the main theorem of the theory of conditional distributions.

Theorem 5.5. *For a partition ξ of a Lebesgue–Rokhlin space $(\Omega, \mathfrak{A}, \mathbf{P})$ to have a system of conditional distributions it is necessary and sufficient for this partition to be measurable.*

Definition 5.6. Measurable mod0 partitions ξ and η are independent if σ-subalgebras $\mathfrak{A}_\xi$ and $\mathfrak{A}_\eta$ are independent. If, in addition, $\xi \vee \eta = \varepsilon$, we say that ξ and η are independent complementations for each other.

Note that for independent measurable mod0 partitions ξ and η we always have $\xi \wedge \eta = \nu$. Besides, the space $(\Omega, \mathfrak{A}, \mathbf{P})/\xi$ is canonically isomorphic to every "conditional subspace" formed by a $\mathbf{P}/\eta$-typical element of η provided with its conditional measure.

Every measurable partition of a Lebesgue–Rokhlin space considered to within mod0 equivalency can be generated by a measurable function considered to within mod0 equivalence.

EXAMPLE 5.7. Consider at first the case of Gaussian measure γ in a finite-dimensional vector space E and a partition ξ into parallel shifts of a vector subspace L (let us call a measurable partition of this kind an affine partitions). Suppose the measure γ is not degenerate. In this case we can describe conditional distributions by their densities with respect to corresponding Lebesgue measure normalizing the restrictions of the density of the very measure γ to every shifted subspace L. But a better way consists in representation of the Euclidean space $(E, Q^{\frac{1}{2}})$ as a direct sum of two Euclidean subspaces with inherited Euclidean structure:

$$E = L \oplus L^{\perp}.$$

Now γ can be represented as a product-measure

$$\gamma = \gamma_L \times \gamma_{L^{\perp}}$$

of the standard Gaussian measures on these two orthogonal subspaces. Their covariance forms can be considered on the spaces

$$L'_{\gamma} = E'/L^{\circ}, \qquad (L^{\perp})'_{\gamma} = E'/(L^{\perp})^{\circ}$$

which are canonically isomorphic to the subspaces $(L^{\circ})^{\perp}$ and L° with inherited structures of Euclidean spaces. Now in order to define the conditional Gaussian distribution on a shifted copy of L one can take the pullback of γ_L along the canonical projection of shifted L to L. So, we come to conclusion that all conditional measures coincide (i.e., are canonically isomorphic) with γ_L and the quotient measure coincide with $\gamma_{L^{\perp}}$. Note that beginning with an affine partition we have got a quite symmetric picture for the quotient and the conditional measures.

EXAMPLE 5.8. Consider again a d-dimensional Gaussian measure γ and consider a measurable partition ζ into the rays going out from the zero-point. This partition is by no means affine, and its conditional measures are not Gaussian. By definition, the conditional distribution on every ray is the distribution of the random variable $(\chi_d^2)^{\frac{1}{2}}$ where the random variable inside the brackets has χ^2-distribution with d degrees of freedom. It is easy to verify that the densities of the considered conditional measure are

$$p_d(x) = C_d x^{d-1} \exp(-\frac{1}{2}x^2), \quad \text{where} \quad C_d = 2^{-\frac{d}{2}}\Gamma(\frac{d}{2}).$$

Evidently, the quotient space can be considered as the surface of the unit sphere, and the quotient distribution is the uniform distribution on this surface of the unit sphere.

EXAMPLE 5.9. *Multi-dimensional Cauchy distribution* $\varkappa$. This is the distribution of the ratio YX^{-1} where Y is a finite-dimensional Gaussian random variable and X is an independent one-dimensional Gaussian random variable. In other terms, d-dimensional Cauchy distribution on a vector space E is the image of a multidimensional Gaussian measure under central projection to a shifted hyperplane of dimension d (so that dimension of the Gaussian measure is equal to $d+1$). The characteristic functional of the Cauchy measure is defined on the space F of all linear functionals on E and has the form

$$\chi(f) = \exp(-q^{\frac{1}{2}}(f)),$$

q being a positive quadratic form on F. The Cauchy measure is rotation invariant (like Gaussian one) with respect to rotations preserving q.

Consider now the measurable partition ζ of E into rays described in the previous example. The Cauchy distribution turns out to be a mixture of Gaussian distributions. The mixing one-dimensional distribution has the density

$$\rho(\sigma) = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\exp(-\frac{1}{2\sigma^2}).$$

Therefore all the conditional measures of the d-dimensional Cauchy distribution are mixtures of ζ-conditional Gaussian distributions with densities

$$p_{\sigma,d}(r) = \frac{1}{\sigma^d (2\pi)^{d/2}} \exp(-\frac{1}{2} r^2).$$

The mixing distribution ρ does not depend on dimension. (Soon we will see why every rotation invariant infinite-dimensional probability measure is a mixture (a convolution) of rotation invariant Gaussian measures.)

EXAMPLE 5.10. For Gaussian random variables the optimal prediction by quadratic loss function is linear. We began considering this property of Gaussian vector spaces in the part 2. Now we can explain it quite clearly.

Given a random variable $Y \in L^2(\Omega, \mathfrak{A}, \mathbf{P})$, the best prediction of its value by quadratic loss function $z \to z^2$ is

$$\arg\min \int (Y - a)^2 \, d\mathbf{P} = \int Y(\omega) \mathbf{P}(d\omega) \stackrel{\text{def}}{=} E_{\mathbf{P}} Y.$$

Suppose that values of random variables $X_1, \ldots, X_n \in L^2(\Omega, \mathfrak{A}, \mathbf{P})$ are known, and we look for the best prediction of a random variable $Y \in L^2$. The correct formulation is: to find the orthogonal projection of Y to the subspace

$$L^2(\Omega, \sigma(X_1, \ldots, X_n), \mathbf{P}|_{\sigma(X_1, \ldots, X_n)}).$$

For the purpose of simplicity we suppose that the mean values of all these random variables are equal to zero. Without loss of generality, we can also suppose that the measurable partition $\zeta_{X_1, \ldots, X_n, Y}$ generated of all these functions is ε. Hence, we can linearize the measure space so that we get a Gaussian probability space $(E, \mathfrak{A}, \gamma)$, and our random variables turn into elements of E'_γ. Consider now the affine partition ξ defined by $X_1, \ldots, X_n$. Elements of this partition are one-dimensional affine subspaces, and conditional distributions are one-dimensional Gaussian. Restrictions of Y on every such one-dimensional subspace of E is an affine function, and its (conditional) mean value is equal to the value of Y in the barycenter of considered one-dimensional conditional distribution. All these barycenters form a hyperplane (a vector subspace), hence its position, i.e., the conditional expectation under consideration, is a linear function on the quotient space such that its elements are defined by fixing values of $X_1, \ldots, X_n$.

6. Gaussian measures: what they are? II.

Here we are going to detail the most sophisticated subject of the Gaussian distributions theory: the infinite-dimensional Gaussian distribution. First of all we should discuss a proper definition, then we shall consider some its properties. Some of this properties demand no particular attention since there is no great difference between the finite- and the infinite-dimensional cases. The property of linearity of the best prediction (see Example 5.9) gives a good illustration. Some other properties are almost evident in the finite-dimensional situation but become less trivial in the general case like the criterion for two Gaussian measures to be mutually absolutely continuous. Finally, there are properties like possible existence of a nonzero oscillation that the Gaussian measures manifest only in the infinite-dimensional case.

We have seen that for every $d \in \mathbb{N}$ there exists precisely one mod0 d-dimensional Gaussian measure. Such measure has the Lebesgue–Rokhlin property since the structure of a finite-dimensional space generates the structure of the Polish space. In the infinite-dimensional case we include the demand for the Gaussian measure space to be Lebesgue–Rokhlin probability measure space separately.

Let $(\Omega, \mathfrak{A}, \mathbf{P})$ be a Lebesgue–Rokhlin probability measure space and let $G \subset L^2(\Omega, \mathfrak{A}, \mathbf{P})$ be a closed Gaussian space generating σ-algebra, i.e., such that

$$\operatorname{compl}(\sigma(G); \mathbf{P}) = \mathfrak{A}$$

(or, the same, the measurable mod0 partition of the probability measure space generated by G is ε). The characteristic functional

$$f \mapsto \int \exp i\langle \omega, f\rangle \mathbf{P}(d\omega) = \exp(-\frac{1}{2}\|f\|^2)$$

is defined on the Hilbert space G.

THEOREM 6.1. *Suppose for every of two Lebesgue–Rokhlin probability measure spaces some vector space of measurable functions generating σ-algebra is fixed. Suppose that a linear bijection between these two vector spaces conserves the values of the characteristic functionals (and, hence, of all distributions). Then this bijection is conjugate to a uniquely defined isomorphism* mod0 *between the given Lebesgue–Rokhlin spaces.*

In particular, in the Gaussian case every automorphism of the Hilbert space G (every orthogonal transform of G) generates a (conjugate) automorphism of the Lebesgue–Rokhlin space $(\Omega, \mathfrak{A}, \mathbf{P})$. Measure space automorphisms of this kind are called *Gaussian.*

In the case $\zeta(G) \neq \varepsilon$ one should consider the quotient space $(\Omega, \mathfrak{A}, \mathbf{P})/\zeta(G)$ instead of $(\Omega, \mathfrak{A}, \mathbf{P})$.

The Gaussian random variables space G not only canonically generates a uniquely mod0 defined Lebesgue–Rokhlin probability space. This probability space can be provided with a *uniquely* mod0 *defined probability vector space structure*. It is quite clear in the finite-dimensional case. Indeed, d-dimensional quadratic covariance form q on a d-dimensional vector space F defines uniquely a Euclidean space E (conjugate with F) with the standard Gaussian measure γ, so that F is *the space of all linear* (hence, *measurable) functionals on* (E, γ). The two Lebesgue–Rokhlin spaces $(\Omega, \mathbf{P})$ and (E, γ) are canonically isomorphic mod0 since a basis $f_1, \ldots, f_d$ in F being fixed, the map

$$\Omega \ni \omega \mapsto x \quad \langle x, f_k \rangle = f_k(\omega) \quad \text{for} k = 1, \ldots, d$$

is mod0 isomorphism. Hence, the Gaussian space G induces a structure of a linear (d-dimensional) Gaussian vector probability space on the Lebesgue–Rokhlin space $(\omega, \mathfrak{A}, \mathbf{P})$ considered mod0. We say $(\Omega, \mathfrak{A}, \mathbf{P})$ admits *a linearization* corresponding to G.

In the infinite-dimensional case (G is a countable-dimensional Hilbert space) no standard Gaussian measure can be defined on the conjugate Hilbert space G' (nor nontrivial rotation-invariant probability measures whatsoever). Nevertheless, an orthonormal measures $\{ f_k : k \in \mathbb{N} \} \subset G$ being fixed, the joint distribution of this family of random variables can be described as a Borel probability distribution $\gamma_{\{f_k\}}$ on $\mathbb{R}^{\mathbb{N}}$, namely, the countable product of one-dimensional standard Gaussian measures.

We shall call it *the standard Gaussian infinite-dimensional distribution.*

DEFINITION 6.2. Given a vector space E with a probability measure $\mathbf{P}$, a function f on E is called mod0 *linear measurable functional* if there exists *a vector* subspace $L \subset E$ of the full measure such that the restriction of f to L is a linear measurable functional. If two mod0 linear measurable functionals coincide on a full measure subspace $L \subset E$, they are said to be mod0 linear equivalent. A class of equivalence under such relation is called *defined up to the equivalence* mod0 *linear measurable functional.*

EXAMPLE 6.3. Consider the standard Gaussian measure γ on $\mathbb{R}^{\mathbb{N}}$ which corresponds to the Gaussian subspace of $L^2(\mathbb{R}^{\mathbb{N}}, \gamma)$ spanned on the family $\{e_k\}$ of coordinate functionals on $\mathbb{R}^{\mathbb{N}}$ and canonically isomorphic to l^2. Now $G = l^2$, $q(f) = \|f\|_{l^2}$ and every element $f \in l^2$ can be considered as a defined up to the equivalence mod0 linear measurable functional. Now in the infinite-dimensional case we cannot select a universal linear measurable subspace $L \subset \mathbb{R}^{\mathbb{N}}$, $g^0_{\mathbb{R}^{\mathbb{N}}}(L) = 1$ so that every functional $f \in l^2$ would be defined everywhere on L and all the linear operations in l^2, which would be isomorphic to $L^2(L, \gamma)$, would be carried pointwise, though for every $f \in l^2$ there is its own vector subspace of probability one where f is linear.

DEFINITION 6.4. Suppose $(E, \mathfrak{A}, \gamma)$ is a vector space with a probability measure γ on a completed σ-algebra generated by a set F (which always can be taken as linear and closed in measure) of defined up to the equivalence mod0 linear measurable functionals. Suppose also that $(E, \mathfrak{A}, \gamma)$ is a Lebesgue–Rokhlin space. Suppose that the characteristic functional on the space F has the form

$$\chi_\gamma(f) = \exp(-\frac{1}{2}q_\gamma(f)),$$

where q_γ is a quadratic form.

Then the triple $(E, \mathfrak{A}, \gamma)$ is called *a Gaussian vector probability space*, the quadratic form q_γ is called *the covariance quadratic form* *of* γ, and the closure in measure $\text{span}F$ of the linear hull of F is called *the Gaussian random variables space* compatible with the structure of the Gaussian vector probability space. It is denoted E'_γ.

THEOREM 6.5. *Given a Gaussian vector probability space* $(E, \mathfrak{A}, \gamma)$*, the Gaussian random variables space is uniquely defined by it and consists of all defined up to the equivalence* mod0 *linear measurable functionals on* $(E, \mathfrak{A}, \gamma)$.

The property of the space $E'_\gamma \subset L^2(\Omega, \mathfrak{A}, \mathbf{P})$ of all mod0 linear measurable functionals to coincide with $\text{span}F$ for every generating set F, i.e., such that $\zeta(F) = \varepsilon$ is not valid in the general case for non-Gaussian subspaces F. Note that in the Gaussian case the L^2-topology and the convergence-in-measure topology coincide on E'_γ.

The fundamental specificity of the finite-dimensional case (not only for Gaussian distributions) is that every element of the vector probability space defines a linear functional on the random variables space. In its turn, in the finite-dimensional case every mod0 element of a vector probability space $(E, \mathfrak{A}, \mathbf{P})$ can be described in terms of random variables space F. In the infinite-dimensional case almost every element of the Gaussian vector probability space can be eliminated by choosing an appropriate measurable vector subspace of measure one not including given element. Those exceptional elements of the Gaussian vector probability space that belong to the intersection of the set of all measurable full measure vector subspaces form an important subspace which can be considered as a "skeleton" of the Gaussian vector probability space.

THEOREM 6.6. *There exists a vector subspace* H_γ *of the Gaussian vector probability space* $(E, \mathfrak{A}, \gamma)$ *which is canonically isomorphic to the space* $(E'_\gamma)'$ *conjugate with the corresponding Gaussian random variables space provided with its standard Hilbert space structure. For every* $x \in H_\gamma$ *and every defined up to the equivalence linear* mod0 *measurable functional* $f \in E'_\gamma$ *the value of* $f(x) = \langle x, f\rangle$ *is well defined (i.e., does not depend on a selected version). This bilinear functional defines on* H_γ *a Hilbert space structure. The elements* x *of* H_γ *and only they have the property: the Gaussian shifted* *measure $\gamma(\cdot + x)$ is absolutely continuous with respect to $\gamma(\cdot)$. In the infinite-dimensional case $\gamma(H_\gamma) = 0$, otherwise $H_\gamma = E\,\mathrm{mod}0$. The restriction of every* mod0 *linear measurable functional to H_γ is completely defined linear functional on Hilbert space H_γ, and every linear functional on Hilbert space H_γ is the restriction of a* mod0 *linear measurable functional.*

DEFINITION 6.7. The Hilbert space $H_\gamma \subset E$ described by the previous theorem is called kernel, or skeleton, or concentration subspace of the Gaussian vector probability space $(E, \mathfrak{A}, \gamma)$. The quadratic form $Q_\gamma(x) = \|x\|^2_{H_\gamma}$ defined on H_γ is called the concentration form of γ.

EXAMPLE 6.8. *The Gaussian white noise.* Let $F = L^2([0,1])$ and let $q(f) = \|f\|^2_{L^2}$. Now we can put $H_\gamma = L^2([0,1])$, consider a pre-Hilbert norm $n(\cdot)$ on H_γ such that the imbedding operator is Hilbert–Schmidt, and take the completion of H_γ with respect to the norm n as E. The elements of this completion can be considered as distributions in sense of Schwartz. Obviously, $Q_\gamma(x) = \|x\|_{L^2}$.

EXAMPLE 6.9. *The Wiener measure.* Consider $E = C([0,1])$,

$$q(f) = \int\int \min(s,t) f(s) f(t)\, ds\, dt.$$

Here the kernel space consist of all functions $f \in E$ such that

$$Q_{\gamma_w}(f) = \int (f'(t))^2\, dt < \infty.$$

The standard d-dimensional Gaussian measure in Euclidean space is concentrated near the Euclidean sphere of radius $d^{\frac{1}{2}}$, so that for large values of d the standard Gaussian measure is very like the uniform distribution concentrated on the surface of such sphere. Consider now the infinite-dimensional case. The problem is: what is the limit case for the property just mentioned? Is the infinity-dimensional Gaussian measure really concentrated precisely on whatever sphere? This idea seems to be in a bad agreement with the idea of very rapid decreasing of Gaussian densities as the distance from the origin increases.

Consider the measurable partition ζ of the infinite-dimensional Gaussian vector probability space $(E, \mathfrak{A}, \gamma)$ into rays going out from the origin as it was done in Example 5.8.

THEOREM 6.10. *In the case of the infinite-dimensional Gaussian distribution* $\zeta = \varepsilon \bmod 0$.

So, we come to the conclusion that the typical conditional distribution is a δ-measure at a point of the ray. For instance, for the standard Gaussian measure on $\mathbb{R}^{\mathbb{N}}$ for almost all elements x of this space

$$\lim \frac{1}{2} \sum_{1}^{n} e_k^2(x) = 1.$$

This equality can be valid only at a single point on every ray. This point just shows the place where such conditional δ-measure is concentrated.

As a consequence we obtain

THEOREM 6.11. *Every infinite-dimensional rotation-invariant probability measure (i.e., a measure with rotation-invariant characteristic functional) is a mixture of centered rotation-invariant Gaussian measures.*

It is not so for the finite-dimensional case.

Now consider a condition for absolute continuity of two Gaussian measures. As in the finite-dimensional case, either two given Gaussian measures are mutually absolutely continuous, or they are singular. It follows from Theorem 6.10 that for the infinite-dimensional Gaussian measure its image under nontrivial homothety is singular with respect to the original measure. The following theorem deals with the general case.

THEOREM 6.12. *Suppose* γ_1 *and* γ_2 *are two Gaussian infinite-dimensional measures. Let their covariation forms* q_1 *and* q_2 *be defined on the same vector space, i.e., Hilbert norms generated by these* *quadratic forms are equivalent. Necessary and sufficient condition for the Gaussian measures* γ_1 *and* γ_2 *to be equivalent is that the spectrum of* q_2 *with respect to* q_1 *be discrete and* $\sum(\sigma_k^2 - 1)^2 < \infty$*, where* $\{\sigma_i^2\}$ *denotes the sequence of the extremal values of the quadratic form ratio.*

Theory of Gaussian measures uses many geometric notions. No wonder that some extensions of purely geometric theorems play important role in this theory. Consider here a very useful extension of the well-known isoperimetric property for Gaussian vector probability spaces.

THEOREM 6.13. (Isoperimetric inequality for Gaussian measures) *Consider a Gaussian vector probability space* $(E, \mathfrak{A}, \gamma)$*, and let* $\mathcal{E}_\gamma \subset H_\gamma \subset E$ *stand for the concentration ellipse. Then for any measurable* $A \subset E$ *and for any* $\varepsilon > 0$ *the following inequality holds:*

$$\Phi^{-1}(\gamma(A + \varepsilon\mathcal{E}_\gamma)) - \Phi^{-1}(\gamma(A)) \geq \varepsilon.$$

In particular, in the finite-dimensional case if T *is a half-space such that* $\gamma(T) = \gamma(A)$*, then for their* ε*-neighborhoods* T_ε *and* A_ε *we have* $\gamma(T_\varepsilon) \leq \gamma(A_\varepsilon)$.

("Given a value of volume, the minimal surface measure has the half-space".)

Sometimes it is important to have an information about properties of the distributions of functions on E from a given class. Classes of convex functions and of norm are often considered. Another important class of functions is the class of all Lipschitz functions. Evidently, in the one-dimensional case the class of distributions of Lipschitz functions (with Lipschitz constant equal to 1) coincides with the class GMC of all images of *one-dimensional* Gaussian measure under every possible contractions.

THEOREM 6.14. *For any Gaussian vector probability space the class of distributions of all Lipschitz functions with the constant equal to* 1 *coincides with* GMC.

Finally, we mention the "0–1 law" for Gaussian probability spaces.

THEOREM 6.15. *Every measurable vector subspace L of the Gaussian vector probability space has measure equal to zero or to one. If L does not include H_γ, then $\gamma(L) = 0$.*

The partition of the infinite-dimensional Gaussian vector probability space into shifts of H_k is "absolutely nonmeasurable": its measurable envelope is the trivial measurable partition ν.

7. Gaussian random processes and fields. Sample properties and large deviations.

This is a large subject. We shall touch only a few topics.

The standard definition of a random process or a random field as a family of random variables includes consideration of a parameter set. Often it is convenient to deal with sets of Gaussian random variables considering them as self-parameterized families (Gaussian families). We shell follow such idea.

Let $K \subset F_\gamma$ be a Gaussian family, $(E, \mathfrak{A}, \gamma)$ be the corresponding Gaussian vector probability space. One of the very important questions is the question of sample boundness of the Gaussian process K. The property of sample boundness means the existence of the random variable $\sup K \in L^0(E, \mathfrak{A}, \gamma)$ where, of course, sup means the lattice supremum in L^0. For Gaussian families either such supremum exists, or it is equal to 1 almost everywhere.

THEOREM 7.1. *For every $K \in F_\gamma$ either* $\sup K \in L^2(E, \mathfrak{A}, \gamma)$ *or* $\sup K = \infty$ *almost everywhere.*

DEFINITION 7.2. For $K \subset E$ the property of sample boundness of K is called GB-property.

The subset K will always be provided with structures induced from the Gaussian random variables space E. First of all, K is a separable metric space, so we can speak about continuity of sample functions and consider their modulo of continuity. So, the question of sample continuity of the Gaussian process K is the question whether or not the space $C(K)$ of continuous functions on K with a Borel Gaussian measure can be identified mod0 with the Gaussian vector probability space $(E, \mathfrak{A}, \gamma)$.

DEFINITION 7.3. For $K \subset F$ the property of sample continuity of K is called GC-property.

GB-property means that there exists a version of K with continuous sample functions.

To give a simple characterization of GB- and GC-sets is yet unsolved problem. We shall now describe a purely geometric characterization of GB- and GC-properties.

For convex bounded set $K \subset \mathbb{R}^n$ consider the polynomial of degree n

$$W(K, \varepsilon) = \mathrm{vol}_n(K + \varepsilon V) = w_0^{(n)}(K) + \varepsilon w_1^{(n)}(K) + \ldots + \varepsilon w_n^{(n)}(K),$$

where V stands for a ball in $\mathbb{R}^n$. Since V is not fixed, it is naturally to consider coefficients of the polynomial (Minkowski's "mixed volumes") to within a multiplication by a constant. It is possible to choose these factors a_n so that the value of $a_n w_{n-1}^{(n)}(K) = h_1(K)$ does not depend on an external space and is completely defined by proper geometry of K. We can also take that for the unit interval I we have $h_1(I) = 1$ and call $h_1(K)$ *one-dimensional half-perimeter* of K. Thus, for arbitrary finite-dimensional K the value $h_1(K)$ is defined and for infinite-dimensional subsets K of a Hilbert space we define the value of $h_1(K)$ as the supremum of values of one-dimensional half-perimeters of finite-dimensional subsets of K.

THEOREM 7.4. *For every $K \subset F_\gamma$ we have*

$$h_1(K) = (2\pi)^{\frac{1}{2}} \int \sup K \gamma(dx).$$

In particular, $K \in GB \iff h_1(K) < \infty$.

Geometric origin of the functional h_1 enables to demonstrate the following monotonicity property of h_1.

THEOREM 7.5. *Suppose* $x_1, \dots, x_n \in H$, $y_1, \dots, y_n \in H$ *and for every* i, j *we have* $\|x_i - x_j\| \le \|y_i - y_j\|$. *Then* $h_1(\{x_1, \dots, x_n\}) \le h_1(\{y_1, \dots, y_n\})$.

The Cauchy measure also possesses a monotonicity property.

THEOREM 7.6. *Under conditions of Theorem* 7.5

$$\varkappa(\{x_1, \dots, x_n\}^\circ) \le \varkappa(\{y_1, \dots, y_n\}^\circ).$$

Two-sided inequalities connecting values of $\gamma(K^\circ)$, $\varkappa(K^\circ)$ and $h_1(K)$ can be given.

As for GC-property, the existence of GC- but not GB-sets is closely connected with a phenomenon of oscillation.

DEFINITION 7.7. Suppose K is a Gaussian family and $f \in K$. The limit

$$\delta(f; K) = \lim_{\varepsilon} (\sup K \cap V_\varepsilon(f) - \inf K \cap V_\varepsilon(f))$$

is called *oscillation* of the Gaussian family K at the point $f \in K$.

THEOREM 7.8. *In the case* $K \in GB$ *the oscillation is well defined and for every* $f \in K$ *is constant* mod0.

Oscillation also can be described purely geometrically. Consider a special case.

THEOREM 7.9. *Let* $K \in GB$ *be a convex symmetric subset of* H_γ. *Necessary and sufficient condition for* K *to possess* GC*-property is that for some* (then for every) decreasing sequence of vector subspaces $L_k \subset H_\gamma$ with finite co-dimensions and zero-intersection

$$h_1(K \cap L_k) \to 0.$$

This condition can be replaced by $h_1(\pi_{L_k} K) \to 0$.

The value $\delta(0; K)$ is called the oscillation of K (for symmetric K).

The property of monotonicity of one-dimensional half-perimeter enables to prove rather satisfactory conditions for K to have GB- or GC-property in terms of metric entropy.

THEOREM 7.10. *For the convex symmetric precompact subset* K *of the Hilbert space* H *to belong to the class* GB *it is necessary that*

$$\limsup \varepsilon^2 \log_2 N(\varepsilon; K) < \infty.$$

Moreover, if $M(\varepsilon; \mathrm{conv} K) \geq 10$, *then*

$$h_1(K) \geq 0.65\varepsilon(\log M(\varepsilon; K))^{\frac{1}{2}}, \ \ \delta(K) \geq 0.31 \limsup \varepsilon(\log M(\varepsilon; K))^{\frac{1}{2}}$$

and, in particular, if $K \in GC$, *then*

$$\limsup \varepsilon^2 \log N(\varepsilon; K) = 0.$$

Moreover,

$$h_1(K) < 22 \sum_{-\infty}^{\infty} 2^{-k} (\log_2 N(2^{-k}; K))^{\frac{1}{2}}$$

and, in particular, if $\sum 2^{-k}(\log_2 N(2^{-k}; K))^{\frac{1}{2}} < \infty$, *then* $K \in GC$.

Here N and M are the cardinality of the minimal ε-net and the cardinality of the largest subset of K with pairwise distances not less than ε.

Let $K \subset F_\gamma$ be a convex symmetric GB-set. The function $\sup K$ is a measurable semi-norm. Let us take the case $\sup K$ is a norm. Consider distribution of this function. We know that $\sup K \in L^2$. Moreover, since it is a Lipschitz function, Theorem 6.14 gives an important information about the behavior of this distribution near the infinity. Generally speaking, it is similar to the behavior of a norm in the finite-dimensional situation. But near the origin we meet new phenomenon arising due to nonzero oscillation. There are such norms on the Gaussian vector probability space (with centered measure) that for some positive constant a for every $\varepsilon < a$ the γ-measure of the ε-neighborhood of any point is zero. The nearest example gives $K = \{c_k e_k\}$ where e_k are orthonormal and $c_k = (2\log(k+1))^{\frac{1}{2}}$, $a = 1$, $\delta(K) = 1$. It can be proven that for arbitrary convex functional on the Gaussian vector probability space the discrete part of its distribution contain no more then one atom. The continuous part of this distribution is concentrated to the right side of the atom (if it exists) and is absolutely continuous there.

References

[1] Sudakov V., *Geometric Problems in the Theory of Infinite-Dimensional Probability Distributions*, Proc. Steklov Inst. Math. **2** (1979), (Monograph. Original Russian edition: 1976).

[2] Adler R. J., *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, Institute of Mathematical Statistics. Lecture notes—monograph series **12** (1990) Hayward, CA.

[3] Ledoux M. and Talagrand M., *Probability in Banach Spaces: Isoperimetry and Processes*, Berlin, Heidelberg, New York Springer-Verlag (1991).