SpringerBriefs in Mathematics

Lectures on Gaussian Processes

Bearbeitet von
Mikhail Lifshits

1. Auflage 2012. Taschenbuch. x, 121 S. Paperback
ISBN 978 3 642 24938 9
Format (B x L): 15,5 x 23,5 cm
Gewicht: 213 g

Lectures on Gaussian Processes

1 Gaussian Vectors and Distributions

Theory of random processes needs a kind of normal distribution. This is why Gaussian vectors and Gaussian distributions in infinite-dimensional spaces come into play. By simplicity, importance and wealth of results, theory of Gaussian processes occupies one of the leading places in modern Probability.

1.1 Univariate Objects

A real random variable X is *normally distributed* or *Gaussian* with expectation $a \in \mathbb{R}$ and variance $\sigma^2 > 0$, if its distribution density with respect to Lebesgue measure is

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

We denote this distribution $N(a, \sigma^2)$ and write $X \sim N(a, \sigma^2)$. A normal distribution with zero variance $N(a, 0)$ is just the distribution concentrated at a point a.

If $X \sim N(a, \sigma^2)$, the characteristic function and Laplace transform of X are given by

$$\mathbb{E}e^{itX} = \exp\left\{iat - \frac{\sigma^2 t^2}{2}\right\},$$
$$\mathbb{E}e^{tX} = \exp\left\{at + \frac{\sigma^2 t^2}{2}\right\}.$$

By using the formula for characteristic function it is easy to check *stability* property: if the variables $X_1 \sim N(a_1, \sigma_1^2)$ and $X_2 \sim N(a_2, \sigma_2^2)$ are independent, then $X_1 + X_2 \sim N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$.

M. Lifshits, *Lectures on Gaussian Processes*, SpringerBriefs in Mathematics,
DOI: 10.1007/978-3-642-24939-6_1,

The family of normal variables and distributions is also invariant with respect to linear transformations: if $X \sim N(a, \sigma^2)$, then

$$cX + d \sim N(d + ca, c^2\sigma^2).$$

Expectation and variance of a normal random variable coincide with parameters of its distribution:

$$\mathbb{E}X = a, \quad \mathbb{V}ar\, X = \sigma^2.$$

Among normal distributions, the *standard* normal distribution $N(0, 1)$ plays a special role. Its distribution function is denoted by $\Phi(r)$. In other words,

$$\Phi(r) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r} \exp\left\{\frac{-x^2}{2}\right\} dx.$$

Let us notice a fast decay of the tails of normal distribution at infinity:

$$\Phi(-r) = 1 - \Phi(r) \sim \frac{1}{\sqrt{2\pi}r} \exp\left\{\frac{-r^2}{2}\right\}, \quad \text{as } r \to \infty.$$

A variable having any normal distribution $N(a, \sigma^2)$ can be obtained from the standard one by a linear transformation $X \mapsto Y = \sigma X + a$.

1.2 Multivariate Objects

A random vector $X = (X_j)_{j=1}^n \in \mathbb{R}^n$ is called *standard Gaussian*, if its components are independent and have a standard normal distribution. The distribution of X has a density

$$p(x) = \frac{1}{(2\pi)^{n/2}} \exp\left\{\frac{-(x, x)}{2}\right\}, \quad x \in \mathbb{R}^n.$$

There exist two equivalent definitions of a general Gaussian vector in $\mathbb{R}^n$.

Definition 1.1 A random vector $Y \in \mathbb{R}^n$ is called *Gaussian*, if it can be represented as $Y = a + LX$, where X is a standard Gaussian vector, $a \in \mathbb{R}^n$, and $L : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a linear mapping.

Definition 1.2 A random vector $Y \in \mathbb{R}^n$ is called Gaussian, if the scalar product (v, Y) is a normal random variable for each $v \in \mathbb{R}^n$.

Definition 1.1 easily yields Definition 1.2. Indeed,

$$(a + LX, y) = (a, y) + (X, L^*y) = (a, y) + \sum_{j=1}^{n} (L^*y)_j X_j$$

has a normal distribution due to stability property.

In the multivariate setting, a definition of Gaussian distribution through a particular form of the density makes no much sense, because in many cases (when the operator L is degenerated, i.e. its image does not coincide with $\mathbb{R}^n$) the density just does not exist.

We will stick to Definition 1.2 that is more convenient for further generalizations: in most interesting spaces there is no standard Gaussian vector X required in Definition 1.1.

Similarly to the univariate notation $N(a, \sigma^2)$, the family of n-dimensional Gaussian distributions also admits a reasonable parametrization. Recall that for any random vector $Z = (Z_j) \in \mathbb{R}^n$ one understands the expectation component-wise, i.e. $\mathbb{E}Z = (\mathbb{E}Z_j)$, while its *covariance operator* $K_Z : \mathbb{R}^n \mapsto \mathbb{R}^n$ is defined by

$$cov((v_1, Z), (v_2, Z)) = (v_1, K_Z v_2).$$

If all components of a vector Z have finite second moments, then the expectation $\mathbb{E}Z$ and covariance operator K_Z do exist. There is no restrictions on expectation value, while the covariance operator is necessarily non-negative definite and symmetric. In other words, there exists an orthonormal base (e_j) such that K has a diagonal form $K_Z e_j = \lambda_j e_j$ with $\lambda_j \geq 0$.

We write $Y \sim N(a, K)$ if Y is a Gaussian vector with expectation a and covariance operator K. In particular, for a standard Gaussian vector we have $X \sim N(0, E_n)$, where $E_n : \mathbb{R}^n \mapsto \mathbb{R}^n$ is the identity operator.

The suggested notation generates legitime questions:

- Does $N(a, K)$ exist for all $a \in R^n$ and all non-negative definite and symmetric operators K ?
- Is the distribution $N(a, K)$ unique ?
- Is it true that any Gaussian distribution has a form $N(a, K)$?

Let us answer positively on all these questions. Indeed, let $a \in R^n$, and let K be a non-negative definite and symmetric linear operator. Consider a base (e_j) corresponding to the diagonal form of K (see above) and define $L = K^{1/2}$ by relations $Le_j = \lambda_j^{1/2} e_j$. Consider a random vector $Y = a + LX$. We have already seen that Y is Gaussian. It is almost obvious that it has expectation a and covariance operator K.

The uniqueness of $N(a, K)$ follows from the fact that a pair (a, K) determines the distribution of (v, Y) as $N((v, a), (v, Kv))$, hence by classical Cramér–Wold theorem the entire distribution is determined uniquely. Notice by the way that the distribution of (v, Y) yields a formula for the characteristic function

$$\mathbb{E}e^{i(v,Y)} = \exp\left\{ i(v, a) - \frac{(v, Kv)}{2} \right\}.$$

Finally, all components of a Gaussian vector are normal random variables, hence they have finite second moments. Therefore, any Gaussian vector has an expectation and a covariance operator, i.e. any Gaussian distribution can be written in the form $N(a, K)$.

Exercise 1.1 Assume that a vector Y satisfies Definition 1.2. Prove that it also satisfies Definition 1.1.

Exercise 1.2 Let all components of a random vector Y be normal random variables. Does it follow that Y is a Gaussian vector?

As in the one-dimensional case, the Gaussian property is preserved by summation of independent random vectors (stability property) and by a linear transformation. If the vectors $X_1 \sim N(a_1, K_1)$ and $X_2 \sim N(a_2, K_2)$ are independent, then $X_1 + X_2 \sim N(a_1 + a_2, K_1 + K_2)$.

If $L : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a linear operator, $h \in \mathbb{R}^n$, and $X \sim N(a, K)$, then

$$LX + h \sim N(h + La, LKL^*).$$

Norm distribution of a Gaussian vector

Let $X = (X_j)_{j=1}^n \in \mathbb{R}^n$ be a standard Gaussian vector. The density formula yields

$$\mathbb{P}\{||X|| \le R\} = c \int_0^R r^{n-1} \exp\{-r^2/2\} \mathrm{d}r,$$

with a normalizing factor $c = 2^{1-n/2} \Gamma(n/2)^{-1}$. We also know that

$$\mathbb{E}||X||^2 = \sum_{j=1}^n \mathbb{E}X_j^2 = n.$$

Moreover, if n is *large*, a substantial part of mass of the standard Gaussian distribution is concentrated in a band of constant width around $\sqrt{n}$. Indeed, we can apply the law of large numbers and the central limit theorem to the sum

$$||X||^2 = \sum_{j=1}^n X_j^2,$$

thus

$$\frac{1}{n} \sum_{j=1}^n X_j^2 \Rightarrow 1, \quad \text{hence } \frac{||X||}{\sqrt{n}} \Rightarrow 1.$$

For a band of width r we have

$$\begin{aligned}
&\mathbb{P}\left\{\big|||X|| - \sqrt{n}\big| \le r\right\} \\
&= \mathbb{P}\left\{ \frac{(\sqrt{n} - r)^2 - n}{\sqrt{n}} \le \frac{\sum_{j=1}^n X_j^2 - n}{\sqrt{n}} \le \frac{(\sqrt{n} + r)^2 - n}{\sqrt{n}} \right\} \\
&\to 2\Phi\left(\frac{2r}{\sqrt{\mathbb{V}ar(X_j^2)}}\right) - 1 = 2\Phi\left(\sqrt{2}\, r\right) - 1.
\end{aligned}$$

These calculations show that *in high-dimensional spaces the standard Gaussian distribution is similar to the uniform distribution on the sphere of a corresponding radius*.

1.3 Gaussian Objects in "Arbitrary" Linear Spaces

Let $\mathscr{X}$ be a linear topological space (its additional required properties will be mentioned below). Let $\mathscr{X}^*$ denote the dual space of continuous linear functionals on $\mathscr{X}$. We denote (f, x) the duality between the spaces $\mathscr{X}$ and $\mathscr{X}^*$, i.e. (f, x) stands for the value of a functional $f \in \mathscr{X}^*$ on an element $x \in \mathscr{X}$. A random vector X taking values in $\mathscr{X}$ is a measurable mapping $X : (\Omega, \mathscr{F}, \mathbb{P}) \mapsto \mathscr{X}$. A σ-field on $\mathscr{X}$ should be sufficiently large to provide measurability of all continuous linear functionals.

Gaussian vectors, their expectations and covariance operators are defined in a same way as in the finite-dimensional case.

A random vector $X \in \mathscr{X}$ is called *Gaussian*, if (f, x) is a normal random variable for all $f \in \mathscr{X}^*$.

A vector $a \in \mathscr{X}$ is called *expectation* of a random vector $X \in \mathscr{X}$, if $\mathbb{E}(f, X) = (f, a)$ for all $f \in \mathscr{X}^*$. We write $a = \mathbb{E}X$ in this case. A linear operator $K : \mathscr{X}^* \mapsto \mathscr{X}$ is called *covariance operator* of a random vector $X \in \mathscr{X}$, if $cov((f_1, X), (f_2, X)) = (f_1, Kf_2)$ for all $f_1, f_2 \in \mathscr{X}^*$. We write $K = cov(X)$ in this case. Covariance operator is symmetric,

$$(f, Kg) = (g, Kf), \quad \forall f, g \in \mathscr{X}^*,$$

and non-negative definite, i.e.

$$(f, Kf) \geq 0, \quad \forall f \in \mathscr{X}^*.$$

From the definition of Gaussian vector, we see that it only makes sense when the space of continuous linear functionals on $\mathscr{X}$ is rich enough. For example, if $\mathscr{X}^* = \{0\}$, then any vector satisfies this definition rendering it senseless. Therefore, usually one of three situations of increasing generality is considered.

(1) $\mathscr{X}$ is a separable Banach space, for example, $\mathbb{C}[0, 1]$, $L_p[0, 1]$ etc.
(2) $\mathscr{X}$ is a complete separable locally convex metrizable topological linear space, for example, $\mathbb{C}[0, \infty)$, $\mathbb{R}^\infty$ etc.
(3) $\mathscr{X}$ is a locally convex linear topological space and a vector X is such that its distribution is a Radon measure.

In cases (1) and (2) every finite measure is a Radon measure, thus case (3) is the most general one.

In the subsequent exposition, we always assume by default that one of these assumptions is satisfied (i.e. at least assumption (3) holds), and call them *usual assumptions*.

As in finite-dimensional case, we assert that X has a distribution $N(a, K)$, if X is a Gaussian vector with expectation a and covariance operator K.

The same questions arise again:

- Does $N(a, K)$ exist for all $a \in \mathscr{X}$ and all symmetric non-negative definite operators $K : \mathscr{X}^* \mapsto X$?
- Is the distribution $N(a, K)$ unique?
- Is it true that any Gaussian distribution has a form $N(a, K)$?

The answers will be slightly different from those given in the previous subsection. As for the first question, the existence of $N(a, K)$ depends only on K. Indeed, if a random vector X has a distribution $N(a_1, K)$, then the vector $X + a_2 - a_1$ has a distribution $N(a_2, K)$. On the other hand, the following exercise shows that the distribution $N(0, K)$ does not necessarily exist for a symmetric non-negative definite operator K.

Exercise 1.3 Let $\mathscr{X}$ be an infinite-dimensional separable Hilbert space. Then $\mathscr{X}^* = \mathscr{X}$ and identity operator $E : \mathscr{X} \mapsto \mathscr{X}$ is a symmetric non-negative definite operator. Prove that the distribution $N(0, E)$ does not exist. Hint: would a random vector X have distribution $N(0, E)$, it would satisfy an absurd identity $\mathbb{P}(||X||^2 = \infty) = 1$.

Finding a criterion for existence of $N(0, K)$ is highly non-trivial problem, and the solution depends on the space $\mathscr{X}$. This question is deliberately omitted in these lectures (except for the Hilbert space case), because we are rather interested in investigation of objects that certainly exist.

Fortunately, *under usual assumptions* we can give positive answers on two remaining questions. Namely, every Gaussian vector possesses an expectation and a covariance operator, see [117] for details. Therefore, its distribution belongs to the family $\{N(a, K)\}$.

Furthermore, a pair (a, K) determines the distribution of a variable (f, x) as $N((f, a), (f, Kf))$, and we find the characteristic function

$$\mathbb{E}e^{i(f, X)} = \exp\left\{i(f, a) - \frac{(f, Kf)}{2}\right\}.$$

Any Radon distribution in $\mathscr{X}$ is determined by its characteristic function. Therefore, distribution $N(a, K)$ is unique.

2 Examples of Gaussian Vectors, Processes and Distributions

Example 2.1 (*Standard Gaussian measure in* $\mathbb{R}^\infty$) Consider the space of all sequences $\mathbb{R}^\infty$ equipped with topology of coordinate convergence. It becomes a complete separable metric space by introducing an appropriate distance, e.g.