ANNALES SCIENTIFIQUES
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2
Série Mathématiques

JEAN-LOUIS PHILOCHE

A new intrinsic construction of the gaussian measure in $\mathbb{R}^d$; with application

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 98, série *Mathématiques*, nº 28 (1992), p. 167-175
<http://www.numdam.org/item?id=ASCFM_1992__98_28_167_0>

Ann. Sci. Univ. Blaise Pascal (Clermont II), Sér. Math., Fasc 28, pp. 167-175

A NEW INTRINSIC CONSTRUCTION OF THE GAUSSIAN MEASURE IN $\mathbb{R}^d$; WITH APPLICATION

1.INTRODUCTION. We give in this paper a new intrinsic construction of G_Q , the (centered) (1) gaussian probability measure of covariance Q in a d- dimensional space. Although our approach is linked to the type of construction starting with the definition of the density when Q is invertible, we still manage to construct G_Q with a convenient density, whether Q is invertible or not.

Then, using the tools and results introduced before, an application to statistics is given : a new intrinsic and (in a sense) natural, proof of the well-known theorem which states that the asymptotic distribution of the empirical chi-square is $\chi^2(d-1)$.

For each of these two parts we give a short survey of the literature on the subjet as well as references illustrating the different types of construction and proof which are essential to our paper.

(1) The presence of a mean in the (probability) law of a gaussian random vector X introduces only a translation term easy to handle. In the beginning of this paper we shall thus, without loss of generality , limit ourselves to the case where the mean is zero.

2. INTRINSIC CONSTRUCTION OF THE GAUSSIAN DISTRIBUTION. We wish to expose first what we shall do in this section :

After the introduction of a few notations we shall, via a proposition, introduce a very useful inner product, denoted $(.,.)_Q$ on ImQ, a subspace of the euclidian space $\mathfrak{X}$ in which everything happens.

Then we shall find out a suitable Lebesgue measure(1) which will enable us to construct the gaussian distribution G_Q, for any positive operator.

Let $\mathfrak{X}$ denote an inner product space of dimension d ; the inner product is $(.,.)$. $\mathcal{L}(\mathfrak{X})$ is the set of linear aperators on $\mathfrak{X}$. For every $T \in \mathcal{L}(\mathfrak{X})$, T^* denotes the adjoint of T. We shall give special interest to **positive operators,** that is $Q \in \mathcal{L}(\mathfrak{X})$ satisfying :

$$Q = Q^* \text{ and } (Qx,x) \geqslant 0 , \qquad x \in \mathfrak{X} . \tag{2.1}$$

Proposition 1. Let Q be a positive operator in $\mathfrak{X}$: one defines, on ImQ, an inner product, denoted $(.,.)_Q$, by the relation :

$$(Qx,Qy)_Q = (Qx,y), \qquad x,y \in \mathfrak{X} . \tag{2.2}$$

In case rkQ = d, $(.,.)_Q$ is defined on the whole $\mathfrak{X}$ by :

$$(x,y)_Q = (Q^{-1}x,y), \qquad x,y \in \mathfrak{X} . \tag{2.3}$$

For reasons given in the proof, we shall find convenient to say that $(.,.)_Q$ is the **reproducing inner product** associated to Q.

Proof. This proposition is a particular case, but in an intrinsinc version, of the theory of reproducing kernels introduced by Aronzsajn [1]. This tool has been popularized in statistics and Probability theory by E. Parzen [2], so we could skip the demonstration. Nevertheless, it is worth giving a quick proof in the finite dimensional case :

Let Q be a positive operator in $\mathfrak{X}$; with the symetry of Q, it is first easily checked that (Qx,y) only depends on Qx and Qy. Bilinearity and positivity of $(.,.)_Q$ are obvious. The fact that $(x,x)_Q$ is 0, on ImQ , iff x = 0 , results directly from KerQ = $\{x; x \in \mathfrak{X}, (Qx,x) = 0\}$ (for the elementary results in linear algebra admitted as such, an excellent reference is still P. Halmos [3]). The invertible case follows directly from the first part of the proposition. □

In the following section it will be useful to note the equality

$$(x,y) = (x,Qy)_Q , \qquad x \in \text{ImQ}, \ y \in \mathfrak{X} . \tag{2.4}$$

(1) Implicitely, the Borel σ-field will always be chosen on inner product spaces.

Following the way indicated in the introduction of this section, we have now in hand the suitable inner product on ImQ. We shall proceed by choosing the convenient[1] Lebesgue measure on ImQ.

Now, let $e = (e_1, e_2, \ldots, e_q)$ be a basis of ImQ ; it is convenient to use the same notation e to also denote the linear application of $\mathbb{R}^q$ into ImQ, which associates to

$$\xi = (\xi_i)_{1 \leq i \leq q} \in \mathbb{R}^q \longrightarrow e\xi = \sum_{i=1}^{d} \xi_i e_i \in \text{ImQ}. \qquad \textbf{(2.5)}$$

Let λ_q denote the usual Lebesgue measure of $\mathbb{R}^q$.

Definition 1. The Lebesgue measure on the inner product space ImQ, $(.,.)_Q$ is defined as the image of λ_q by the application e (defined above), where e is a basis of ImQ, **orthonormal for $(.,.)_Q$**.
Of course this definition is easily seen to be correct, because it is independent of the choice of the orthonormal basis. The specific Lebesgue measure chosen in this definition will **be denoted λ_Q**.

We are now in position to give the intrinsic definition of the gaussian random vectors (r.v) we are looking for :

Definition 2. Let Q be a positive operator in $\mathfrak{X}$, rkQ=q. A r.v x taki values in $\mathfrak{X}$ is said to be gaussian with Q as a covariance operator, if its (probability) law P_X is :

$$P_X(dx) = (2\pi)^{-q/2} \exp\left[-\frac{1}{2}(x,x)_Q\right] \lambda_Q(dx) . \qquad \textbf{(2.6)}$$

The notation $G_Q = P_X$ will be used.

Now that we have found a law, which is a "candidate" to achieve the construction we aim at, we must check up that (2.6) coïncides, when $\mathfrak{X} = \mathbb{R}^d$, with the similar distribution obtained by another classical route.
A few remarks must be done before :

Remarks.

1) Let e be a basis of ImQ , orthonormal for $(.,.)_Q$; the application e defined in (2.5) is clearly an isometry of $\mathbb{R}^q$ on ImQ.
Then we observe that G_Q is the image by e of G_q , the standard gaussian distribution on $\mathbb{R}^q$:

(1) On a euclidian space the Lebesgue measure, being invariant by translations, is unique up to a constant ; the point is precisely to make an unambiguous choice.

$$G_q(d\xi) = (2\pi)^{-q/2}\exp\left[-\frac{1}{2}(\xi,\xi)_q\right] \lambda_q(d\xi) \; ; \qquad \textbf{(2.7)}$$

for sake of clarity $(.,.)_q$ denotes the natural inner product of $\mathbb{R}^q$.

2) The Fourier transform[1] $\hat{G}_q$ of G_q is

$$\hat{G}_q(\eta) = \exp\left[-\frac{1}{2}\|\eta\|_q^2\right] . \qquad \textbf{(2.8)}$$

using again the isometry e easily gives a first expression of the Fourier transform of G_Q : for any $v \in \mathrm{Im}Q$,

$$\hat{G}_Q(v) = E\left(\exp\left[i(v,X)_Q\right]\right) = \exp\left(-\frac{1}{2}\|v\|_Q^2\right) . \qquad \textbf{(2.9)}$$

This formula gives the characteristic function of X , **inside ImQ** ; therefore we have not yet reached our goal which is to identify $E\exp[i(u,X)]$, for every $u \in \mathbb{R}^d \equiv \mathfrak{X}$; [here $(.,.)_d \equiv (.,.)$].

3) It should be noted that the law P_X concerned by (2.6) and (2.9) is always "spherical" as soon as a good inner product $(.,.)_Q$ is used in the space ImQ in which X really takes its values.

We shall end our construction, by proving that, in the case $\mathfrak{X} = \mathbb{R}^d$, G_Q is identical with $G_Q(d)$ the gaussian law with covariance matrix Q defined by any other method in the literature. For this we shall use characteristic functions :

Proposition 2. Let X be a r.v in $\mathbb{R}^d$, suppose that its probability law P_X is G_Q , then

$$E(\exp[i(u,X)_d]) = \exp\left[-\frac{1}{2}(Qu,u)_d\right] , \qquad u \in \mathbb{R}^d . \qquad \textbf{(2.10)}$$

Proof. As $X \in \mathrm{Im}Q$ (a.s), and with (2.4),

$$E\exp[i(u,X)_d] = E\exp[i(Qu,X)_Q] = \exp\left[-\frac{1}{2}(Qu,Qu)_Q\right] = \exp\left[-\frac{1}{2}(Qu,u)_d\right],$$

using (2.9) and (2.2). □

For further need we shall give two useful propositions :

(1) In Probability theory it is classical to define the Fourier transform of a probability μ in $\mathbb{R}^q$ by $\hat{\mu}(u) = \int \exp[i(u,x)_q]\mu(dx)$, $(u \in \mathbb{R}^q)$. If, as often, $\mu = P_X$ the law of some r.v X , $\hat{\mu}(u) = E\exp[i(u,X)_q]$. The right hand side of this last equality is called the characteristic function (c.f) of X.

Proposition 3. Suppose the probability distribution of a r.v X is G_Q, with rkQ = q , then :
$X \in ImQ$ a.s, and
the law of $\|X\|_Q^2$ is $x^2(q)$ (1).
The **proof** is obvious when one uses the isometry e in (2.5) and if one relies on the following convenient definition of $x^2(q)$:

Definition 3. Let X be a gaussian vector in $\mathbb{R}^q$, with I_q (the identity) for covariance matrix, then $x^2(d)$ is by definition the probability law of $\|X\|_d^2$.

Proposition 4. (Central limit theorem in $\mathbb{R}^d$).
Let $(X_k)_{k\geq 1}$ be i.i.d random vectors in $\mathbb{R}^d$. m ($m \in \mathbb{R}^d$) and Q are the mean an covariance matrix common to all X_k .

Let $S_n = \sum_{k=1}^{n} X_k$. Then, when $n \to \infty$, the law of $\frac{S_n - nm}{\sqrt{n}}$ converges to G_Q .

A proof of this proposition can be found, for example in [4].

3.SHORT SURVEY OF OTHER AVAILABLE CONSTRUCTIONS. It is well-known that, in the literature one finds many other constructions of the gaussian probability in $\mathbb{R}^d$. By "construction" we mean both a definition and the basic properties (in a certain order). There are three main types of constructions :

The density approach at least dates back to P. Levy [5]. In this construction one starts defining in $\mathbb{R}^d$ the density of $G(I_d)$, which is obvious, or the density of G(Q), when Q is non degenerate and positive (in this case the constant in front of the exponential has to be identified). Excellent exemples of this approach are given in W. Feller [6], A. Renyi [7], L. Breiman [4], and (in french) M. Métivier [8].

(1) $x^2(n)$ is a distribution, famous in statistics, where it is extensively used. As often this distribution can be defined in many ways, the integer n counts what is called the degrees of freedom. $x^2(n)$ has a density on $\mathbb{R}^+$, explicitely :

$$\left[2^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)\right]^{-1} e^{-\frac{x}{2}}\, x^{\frac{n}{2}-1} \;, \qquad (x > 0) \;.$$

The characteristic function approach seems to be less popular. This construction starts, straight away, by defining G(Q) in $\mathbb{R}^d$ as the distribution which characteristic function $\varphi(u)$ is :

$$\varphi(u) = \exp\left[-\frac{1}{2}(Qu,u)_d\right], \quad u \in \mathbb{R}^d .$$

Or course the fact that such a function is a c.f is part of the problem.
Two excellent illustrations of this approach are given in H. Cramer [9] and A.N. Shiryayev [10].

The definition by linearity : a r.v $X = (X_k)_{1\leq k\leq d}$ with values in $\mathbb{R}^d$ is said to be gaussian, if any linear form $u_1X_1+\ldots+u_dX_d$ has a gaussian distribution in $\mathbb{R}$.
We shall not comment here this apparently non-constructive definition. Obviously such a definition is very appropriate to gaussian processes (it also works perfectly well for random vectors). This approach is illustrated in four excellent books (two of them in french) :
C.R. Rao [11], J. Neveu [12], A. Monfort [13] and P. Brémaud [14].

4.CONVERGENCE TO χ^2 . In this section we shall use both the reproducing inner product $(.,.)_Q$ on ImQ and the central limit theorem (proposition 4), in order to give a new proof of the well-known result stating that under some conditions the distribution of $D_n(p)$ (the so-called "empirical χ^2") converges to $\chi^2(d-1)$. The whole result will be denoted :

$$D_n(p) = \sum_{i=1}^{d} \frac{(N_n(i)-np_i)^2}{np_i} \xrightarrow{L} \chi^2(d-1), \text{ as } n \to +\infty . \qquad \textbf{(4.1)}$$

(Of course we shall define every ingredient of this formula in proper time).

Beforehand a few preliminaires :
Let I be a finite set (d = card I) and $p = (p_i,\ i \in I)$ a probability on I, such as $p_i \neq 0$, for every $i \in I$. Let $\{\varepsilon_i,\ i \in I\}$ be the natural basis of $\mathbb{R}^I$, and X a r.v with values in $\mathbb{R}^I$, the law of which is defined by

$$P[X = \varepsilon_i] = p_i, \qquad i \in I . \qquad \textbf{(4.2)}$$

Of course,

$$EX = (p_i\ ;\ i \in I) = p \in \mathbb{R}^I\ ; \qquad \textbf{(4.3)}$$

whereas the covariance Q of X satisfies

$$Q_{ij} = p_i\,\delta_{ij} - p_ip_j, \qquad i,j \in I, \qquad \textbf{(4.4)}$$

where δ is Knonecker.
We point out two facts :
As usual , $X \in EX + \text{ImQ}$ (a.s) ; **(4.5)**
Q is the covariance matrix of the n-multinomial law in the special cas n = 1.

Now we shall proceed and identify the reproducing inner product associated to the covariance Q of the above r.v X. Q can also be seen as the matrix with typical element given by (4.4).

Proposition 5.
(i) The subspace ImQ is exactly constituted by the y satisfying $\sum_i y_i = 0$.

(ii) The inner product $(y,y')_Q = \sum_i \frac{y_i y_i'}{p_i}$, $y,y' \in ImQ$. **(4.6)**

Proof. For $x \in \mathbb{R}^I$ and with (4.4), we see that $(Qx)_i = p_i x_i - p_i \sum_j p_j x_j$. So it is clear that $\sum_i (Qx)_i = 0$. On the other way let y be an element satisfying $\sum_i y_i = 0$. Such a y belongs to ImQ, since the equations $y_i = (Qx)_i$ (we want to solve in x) admit a solution

$$x = (x_i) \text{ with } x_i = \frac{y_i}{p_i} . \quad \textbf{(4.7)}$$

Hence ImQ, exactly identified, has d-1 for rank.
The proof of (ii) is straightforward :
Let $y = Qx$ and $y' = Qx'$ be in ImQ , (2.2) and (4.6) give successively :

$$(y,y')_Q = (x,Qx') \text{ , and}$$

$$(x,Qx') = \sum_i \frac{y_i y_i'}{p_i} . \quad \square$$

Now we are in a position to give a precise version of the proposition of the convergence to $\chi^2(d-1)$.

Proposition 6. (Theorem of convergence to χ^2).
Let $(X_k)_{k \geq 1}$ be an i.i.d sequence of random vectors in $\mathbb{R}^I$ (cardI = d) , the common distribution is the law of X defined in (4.2). Let $N_n(i) = \sum_{k=1}^{n} 1_{[X_k = i]}$.
Then, when $n \to +\infty$
the distribution of $\sum_{i=1}^{d} \frac{(N_n(i) - np_i)^2}{np_i}$ converges to $\chi^2(d-1)$. **(4.8)**

Proof. The $(X_k)_{k \geq 1}$ fulfills the hypothesis of proposition 4.

Therefore, denoting $N_n = \sum_{k=1}^{n} X_k$, as $n \to +\infty$, the distribution of $\frac{N_n - np}{\sqrt{n}}$ converges to G_Q , where Q is the covariance matrix (4.4).

Clearly $N_n - np \in \mathrm{Im}Q$ a.s (see remark (4.5)). So, taking the square of the norm associated to the inner product $(.,.)_Q$ we see (with the use of proposition 3) that, as $n \to +\infty$:
the distribution of $\frac{1}{n}\|N_n - np\|_Q^2$ converges to $\chi^2(d-1)$ (1) .
By using part (ii) of proposition 5, we can explicit the square of the norm above and obtain :

$$\frac{1}{n}\|N_n - np\|_Q^2 = \sum_{i=1}^{d} \frac{(N_n(i) - np_i)^2}{np_i} ,$$

this gives the desired result. ▫

5. SHORT SURVEY OF OTHER AVAILABLE PROOFS. In the literature one can find different proofs of this χ^2 theorem. They are essentially of two types :
The characteristic function approach starts with taking the r.v $N_n = (N_n(i))_{1 \leq i \leq d}$ which has been defined in proposition 6. The distribution of N_n is multinomial, hence his c.f is well-known. One deduces from this the c.f of the r.v $\left(\frac{N_n(i) - np_i}{\sqrt{np_i}}\right)_{1 \leq i \leq d}$. As $n \to +\infty$, one obtains then, essentially with calculus, the limit of this c.f which appears to be the c.f of a singular gaussian distribution. The end of the proof is close. In their book H. Cramer [9] and M. Fisz [15] exactly follow this line of proof.
The central limit theorem approach is a short way to find the gaussian limit of the r.v $\frac{N_n - np}{\sqrt{n}}$. Then the question is to find the image of this distribution by the continuous application :

$$f(x) = \sum_{i=1}^{d} \frac{x_i^2}{p_i} , \qquad x = (x_k)_{1 \leq k \leq d} \in \mathbb{R}^d .$$

This can be done differently according to the style of the author. This line of presentation is, nowadays, often chosen ; see, for exemple [15] and [16] (this reference is in french).

(1) We have implicitely used the continuity of the function $x \to \|x\|_Q^2$.

References

1. N. Aronszajn, "Theory of reproducing kernels", Trans. Am. Math. Soc., 68 337-404 (1950).

2. E. Parzen, "Time series analysis papers", 477 and ff, Holden day (1967).

3. P. Halmos, "Finite-dimensional vector spaces", Van Nostrand (1958).

4. L. Breiman, "Probability", Addison Wesley (1968).

5. P. Levy, "Processus stochastiques et mouvement brownien", Gauthier-Villars, Paris (1948).

6. W. Feller, "An introduction to Probability Theory and its Applications", Vol 2 Wiley (1966).

7. A. Renyi, "Foundations of probability", Holden-day (1970).

8. M. Métivier, "Notions fondamentales de la théorie des probabilités", Dunod, Paris (1972).

9. H. Cramer, "Mathematical Methods of Statistics", Princeton University Press (1961), Ninth printing.

10. A.N. Shiryayev, "Probability", Springer-Verlag N.Y. (1984).

11. C.R. Rao, "Linear Statistical Inference and its Applications", Wiley (1965).

12. J. Neveu, "Processus Aléatoires Gaussiens", Presses de l'Université de Montréal (1968).

13. A. Monfort, "Cours de Probabilités", Economica, Paris (1980).

14. P. Brémaud, "Introduction to probabilistic modeling", Springer-Verlag New-York (1988).

15. M. Fisz, "Probability theory and Mathematical Statistics", Wiley (1963).

16. A. Monfort, "Cours de statistique mathématique", Economica, Paris (1982).

Jean-Louis PHILOCHE
INSEE/INA - PG
75000 - PARIS

Manuscrit reçu le 18 Juin 1992