Some basic properties and characterizations of Gaussian measures and variables

The note (to be completed!) collects, in a random order, some basic properties and classical characterizations of Gaussian measures and variables, which may be found in standard references in probability theory and mathematical statistics. In the text, γ_n, or $\mathcal{N}(0, \mathrm{Id})$, denote the standard Gaussian distribution on the Borel sets of $\mathbb{R}^n$ with density $\frac{1}{(2\pi)^{\frac{n}{2}}}\, e^{-\frac{1}{2}|x|^2}$, $x \in \mathbb{R}^n$, with respect to the Lebesgue measure λ_n.

Table of contents

Drafted by M. L. **v1** November 2023

1 Rotational invariance

The rotational invariance of Gaussian measures may be described by two main features.

1) If X is a random variable on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with law $\mathcal{N}(0, \mathrm{Id})$, and if O is an orthogonal matrix (transformation) on $\mathbb{R}^n$, then OX is also distributed according to $\mathcal{N}(0, \mathrm{Id})$. This is checked on the covariance matrix.

Maxwell's observation is that the standard Gaussian measure in $\mathbb{R}^n$ is the only probability measure which is both invariant under orthogonal transformations and is a product measure. More precisely, if μ is a probability measure on the Borel sets of $\mathbb{R}^n$ with those two properties, necessarily $\mu = \mathcal{N}(0, \sigma^2 \mathrm{Id})$ for some $\sigma > 0$. For a quick argument, after convolution with $\mathcal{N}(0, \varepsilon^2 \mathrm{Id})$ for some $\varepsilon > 0$, it may be assumed that μ has a smooth, strictly positive, density f with respect to the Lebesgue measure. But this density is both a function of $|x|^2$ by rotational invariance and a product (necessarily of the same one-dimensional density), which forces the Gaussian density (take the partial derivatives of $\log f$).

2) If X and Y are centered Gaussian vectors on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, independent and identically distributed, for every real θ, $X(\theta) = X \sin(\theta) + Y \cos(\theta)$ and $X'(\theta) = X \cos(\theta) - Y \sin(\theta)$ are Gaussian vectors, independent with the same law as X. In others words, the couples $(X(\theta), X'(\theta))$ have the same law as (X, Y). This may easily be checked on the covariances.

The Kac-Bernstein theorem [6, 4] ensures conversely that if X and Y are independent real random variables such that $X + Y$ and $X - Y$ are also independent, then both X and Y must have normal distributions. More generally, the statement holds true as soon as $X(\theta)$ and $X'(\theta)$ are independent for some θ which is not an integer multiple of $\frac{\pi}{2}$.

Cramér's theorem [5] (initially announced by P. Lévy) expresses that if X and Y are two independent (non-constant) real random variables such that $X + Y$ is normally distributed, then both X and Y follow a normal distribution.

2 Wick's formula

It is clear that, for a real random variable X on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, the collection of moments

$$\mathbb{E}(X^{2k+1}) \,=\, 0, \quad \mathbb{E}(X^{2k}) \,=\, \frac{(2k)!}{2^k k!}, \ k \geq 0,$$

characterize the law of X as the standard normal distribution $\mathcal{N}(0,1)$.

Wick's formula is a kind of multilinear extension which expresses any product of the coordinates of a Gaussian random vector by the covariance of this vector. Namely, if X is a centered Gaussian vector in $\mathbb{R}^n$, for any even collection $(\xi_1, \ldots, \xi_{2k})$ of linear functions on $\mathbb{R}^n$,

$$\mathbb{E}\big(\xi_1(X)\cdots\xi_{2n}(X)\big) \,=\, \sum \prod_{\ell=1}^{k} \mathbb{E}\big(\xi_{i_\ell}(X)\xi_{j_\ell}(X)\big)$$

where the sum runs over all unordered sequences of unordered pairs $\{i_1, j_1\}, \ldots, \{i_k, j_k\}$ where each of the integers $1, \ldots, 2k$ appears only once.

For example, if $X = (X_1, X_2, X_3, X_4)$ is a centered Gaussian vector in $\mathbb{R}^4$,

$$\begin{aligned}\mathbb{E}(X_1X_2X_3X_4) \,&=\, \mathbb{E}(X_1X_2)\,\mathbb{E}(X_3X_4) + \mathbb{E}(X_1X_3)\,\mathbb{E}(X_2X_4)\\ &\quad + \mathbb{E}(X_1X_4)\,\mathbb{E}(X_2X_3).\end{aligned}$$

Among other proofs, this type of identity may be obtained from the form of the Fourier transform

$$\mathbb{E}\big(e^{i\langle u,X\rangle}\big) \,=\, e^{-\frac{1}{2}\mathbb{E}(\langle u,X\rangle^2)}, \quad u = (u_1, u_2, u_3, u_4) \in \mathbb{R}^4.$$

Identification of the terms with all the u_ℓ's distincts in Taylor expansions of both sides at the order 4 yields the conclusion.

3 Fourier transform

The Fourier transform, of characteristic function in the probabilistic language, of the standard Gaussian measure γ_n is given by

$$\varphi_{\gamma_n}(u) \,=\, e^{-\frac{1}{2}|u|^2}, \quad u \in \mathbb{R}^n.$$

In other words, the density $f(x) = \frac{1}{(2\pi)^{\frac{n}{2}}}\, e^{-\frac{1}{2}|x|^2}$, $x \in \mathbb{R}^n$, of γ_n, is its own Fourier transform $\hat{f}$ (up to the standard Fourier normalizations).

4 Integration by parts and Stein's characterization

For a function $f : \mathbb{R} \to \mathbb{R}$, locally Lipschitz and such that xf and f' are integrable with respect to γ_1,

$$\int_{\mathbb{R}} xf \, d\gamma_1 \, = \, \int_{\mathbb{R}} f' d\gamma_1. \tag{1}$$

This immediately follows by integration by parts on the form of the Gaussian density

$$\int_{\mathbb{R}} xf \, d\gamma_1 \, = \, \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) x \, e^{-\frac{1}{2}x^2} d\lambda_1(x).$$

On $\mathbb{R}^n$, if $f : \mathbb{R}^n \to \mathbb{R}$ is smooth enough,

$$\int_{\mathbb{R}^n} xf \, d\gamma_n \, = \, \int_{\mathbb{R}^n} \nabla f \, d\gamma_n$$

as vector integrals.

Stein's observation, leading to various approximation results cf. [2], expresses that (1) ranging over a rich enough family of functions f characterizes γ_1 (over all probability measures with a first moment). Let indeed μ be a probability on $\mathbb{R}$ with a first moment such that $\int_{\mathbb{R}} xf \, d\mu = \int_{\mathbb{R}} f' d\mu$ for every, say, bounded continuous function $f : \mathbb{R} \to \mathbb{R}$. Applied to the real and imaginary parts of the family of functions $f(x) = e^{iux}$, $x \in \mathbb{R}$, $u \in \mathbb{R}$, this integration by parts formula yields that the Fourier transform $\varphi(u)$, $u \in \mathbb{R}$, of μ satisfies the differential equation $\varphi'(u) = -u \, \varphi(u)$, $u \in \mathbb{R}$, so that $\varphi(u) = e^{-\frac{1}{2}u^2}$, $u \in \mathbb{R}$, the Fourier transform of γ_1.

5 Heat and Mehler kernels

Let

$$h_t(x) \, = \, \frac{1}{(4\pi t)^{\frac{n}{2}}} \, e^{-\frac{1}{4t}|x|^2}, \quad t > 0, \; x \in \mathbb{R}^n, \tag{2}$$

be the standard heat kernel on $\mathbb{R}^n$, fundamental solution of the heat equation $\partial_t h_t = \Delta \, p_t$. In other words, h_t is the density of the normal law $\mathcal{N}(0, 2t)$.

The convolution semigroup $H_t f(x) = f * h_t(x)$, $t > 0$, solves

$$\partial_t H_t f \, = \, \Delta H_t f = H_t \Delta f$$

with initial data f. By the definition of h_t, the semigroup H_t, $t > 0$, admits the integral representation

$$H_t f(x) \, = \, \int_{\mathbb{R}^n} f\big(x + \sqrt{2t} \, y\big) d\gamma_n(x)$$

for all $t > 0$, $x \in \mathbb{R}^n$, and any suitable measurable function $f : \mathbb{R}^n \to \mathbb{R}$. At $t = \frac{1}{2}$, h_t is just the standard Gaussian density so that $H_{\frac{1}{2}} f(0) = \int_{\mathbb{R}^n} f d\gamma_n$ (while $H_0 f = f$).

There is however a related Gaussian kernel which has the advantage to be invariant with respect to γ_n (as the classical heat kernel is invariant under the Lebesgue measure λ_n).

Define the Mehler kernel, for $t > 0$, $x, y \in \mathbb{R}^n$, by

$$p_t(x, y) \,=\, p_t(y, x) \,=\, \frac{1}{(1 - e^{-2t})^{\frac{n}{2}}} \exp\bigg(- \frac{e^{-2t}}{2(1 - e^{-2t})} \big[|x|^2 + |y|^2 - 2e^t \langle x, y \rangle \big] \bigg). \tag{3}$$

It holds true that $\int_{\mathbb{R}^n} p_t(x, y) d\gamma_n(y) = 1$ for all $t > 0$ and $x \in \mathbb{R}^n$. The Mehler kernel satisfies besides the basic semigroup property with respect to γ_n,

$$\int_{\mathbb{R}^n} p_s(x, z)\, p_t(z, y) d\gamma_n(z) \,=\, p_{s+t}(x, y)$$

for all $s, t > 0$ and $x, y \in \mathbb{R}^n$.

The Mehler kernel generates the Ornstein-Uhlenbeck semigroup

$$P_t f(x) \,=\, \int_{\mathbb{R}^n} f(y)\, p_t(x, y) d\gamma_n(y)$$

for all $t > 0$, $x \in \mathbb{R}^n$, and any suitable measurable function $f : \mathbb{R}^n \to \mathbb{R}$, which after a suitable change of variable admits the integral representation

$$P_t f(x) \,=\, \int_{\mathbb{R}^n} f\big(e^{-t} x + \sqrt{1 - e^{-2t}}\, y \big) d\gamma_n(y). \tag{4}$$

With the natural extension $P_0 = \mathrm{Id}$, the family $(P_t)_{t \ge 0}$ defines a Markov semigroup, symmetric in $\mathrm{L}^2(\gamma_n)$ and invariant with respect to γ_n, that is $\int_{\mathbb{R}^n} f P_t g d\gamma_n = \int_{\mathbb{R}^n} g P_t f d\gamma_n$ and $\int_{\mathbb{R}^n} P_t f d\gamma_n = \int_{\mathbb{R}^n} f d\gamma_n$. These properties are actually a reformulation of the rotational invariance of Gaussian measures, expressing that under $\gamma_n \otimes \gamma_n$, the couples

$$\big(x \sin(\theta) + y \cos(\theta), x \cos(\theta) - y \sin(\theta) \big),$$

with $e^{-t} = \sin(\theta)$, are distributed as (x, y).

The infinitesimal generator $\mathrm{L} = \Delta - x \cdot \nabla$ of the Markov semigroup $(P_t)_{t \ge 0}$ fulfills the integration by parts formula

$$\int_{\mathbb{R}^n} f(-\mathrm{L} g) d\gamma_n \,=\, \int_{\mathbb{R}^n} \nabla f \cdot \nabla g \, d\gamma_n \tag{5}$$

for every smooth functions $f, g : \mathbb{R}^n \to \mathbb{R}$. The spectrum of the operator $-\mathrm{L}$ is $\mathbb{N}$, and the eigenvectors are the Hermite polynomials (cf. [3]). For example, in dimension 1, if $(h_k)_{k\in\mathbb{N}}$ denotes the sequence of Hermite polynomials (normalized in $\mathrm{L}^2(\gamma_1)$),

$$-\mathrm{L}h_k \,=\, k\,h_k, \quad k \in \mathbb{N}.$$

In particular, by the integration by parts formula (5),

$$k\int_{\mathbb{R}} h_k f\, d\gamma_1 \,=\, \int_{\mathbb{R}} (-\mathrm{L}h_k) f\, d\gamma_1 \,=\, \sqrt{k}\int_{\mathbb{R}} h_{k-1} f' d\gamma_1$$

for every smooth function $f : \mathbb{R} \to \mathbb{R}$, which extends (1) since $h_1(x) = x$, $h_0(x) = 1$, $x \in \mathbb{R}$.

6 Maximum of entropy

If P is a probability measure on the Borel sets of $\mathbb{R}^n$, with density f with respect to the Lebesgue measure, its *entropy* is

$$\mathrm{H}(P) \,=\, -\int_{\mathbb{R}^n} f\log(f) d\lambda_n$$

whenever the integral is well-defined, $\mathrm{H}(P) = +\infty$ if not. It is easily seen that $\mathrm{H}(\gamma_n) = \frac{n}{2}\log(2\pi e)$.

If f and g are probability densities, by Jensen's inequality with respect to the convex function $-\log$ and to the probability measure $f d\lambda_n$,

$$\int_{\mathbb{R}^n} f\log\Big(\frac{f}{g}\Big) d\lambda_n \,=\, \int_{\mathbb{R}^n} (-\log)\Big(\frac{g}{f}\Big) f\, d\lambda_n \,\geq\, -\log\bigg(\int_{\mathbb{R}^n} g\, d\lambda_n\bigg) \,=\, 0.$$

In other words, $\int_{\mathbb{R}^n} f\log(f) d\lambda_n \geq \int_{\mathbb{R}^n} f\log(g) d\lambda_n$.

Let now P be a probability measure with density f and finite entropy $\mathrm{H}(P)$, satisfying $\int_{\mathbb{R}^n} |x|^2 f d\lambda_n \leq n$ $(= \int_{\mathbb{R}^n} |x|^2 d\gamma_n)$. If g is the density of the Gaussian distribution γ_n, by the preceding,

$$\begin{aligned}
\mathrm{H}(P) \,&=\, -\int_{\mathbb{R}^n} f\log(f) d\lambda_n \\
&\leq\, -\int_{\mathbb{R}^n} f\log(g) d\lambda_n \\
&=\, \frac{1}{2}\int_{\mathbb{R}^n} \big(|x|^2 + n\log(2\pi)\big) f\, d\lambda_n \\
&\leq\, \frac{n}{2}\big(1 + \log(2\pi)\big) \,=\, \mathrm{H}(\gamma_n)
\end{aligned}$$

since $\int_{\mathbb{R}^n} |x|^2 f d\lambda_n \leq n$.

As a conclusion, the standard Gaussian measure γ_n maximizes the entropy over all probability measures with density f of finite entropy satisfying $\int_{\mathbb{R}^n} |x|^2 f d\lambda_n \leq n$. By the case of equality in Jensen's inequality, the proof shows at the same time that the Gaussian measure γ_n is characterized in this way, a result commonly attributed to L. Boltzmann.

7 Chi-squared distribution

If $X_1, \ldots, X_n$ are independent variables with common law $\mathcal{N}(0,1)$, then $X_1^2 + \cdots + X_n^2$ follows the classical χ^2 law with n degree of freedom, expressed by the gamma distribution with density $\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$, $x \in (0, \infty)$, with respect to the Lebesgue measure.

8 Independence of empirical mean and variance

If $X = (X_1, \ldots, X_n)$ is random vector with law $\mathcal{N}(0, \mathrm{Id})$, the empirical mean and variance of the sample $(X_1, \ldots, X_n)$ defined by

$$\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k \quad \text{and} \quad S^2 = \frac{1}{n-1} \sum_{k=1}^{n} \big(X_k - \overline{X}\big)^2$$

are independent. Moreover, $(n-1)S^2$ has the same distribution as $\sum_{k=1}^{n-1} X_k^2$ (that is, follows a χ^2 law with $n-1$ degree of freedom).

To verify these properties, let O be a $n \times n$ orthogonal matrix, and $Y = OX$. Then $\sum_{k=1}^{n} X_k^2 = \sum_{k=1}^{n} Y_k^2$, and Y has the same distribution as X. Choosing O such that the coefficients of the last line are all equal to $\frac{1}{\sqrt{n}}$, so that $Y_n = \sqrt{n}\,\overline{X}$, it may be checked that

$$(n-1)S^2 = \sum_{k=1}^{n} X_k^2 - n\overline{X}^2 = \sum_{k=1}^{n} Y_k^2 - Y_n^2 = \sum_{k=1}^{n-1} Y_k^2.$$

Since $Y = (Y_1, \ldots, Y_n)$ has law $\mathcal{N}(0, \mathrm{Id})$, the coordinates $Y_1, \ldots, Y_n$ are independent standard normal real random variables. As a consequence, $\overline{X} = \frac{1}{\sqrt{n}} Y_n$ is independent from $(n-1)S^2 = \sum_{k=1}^{n-1} Y_k^2$, which proves the various claims.

9 Maximum of iid Gaussians

It is a classical exercise to check that if $X_1, \ldots, X_n$, $n \geq 1$, are independent standard normal random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then

$$c\sqrt{\log n} \,\leq\, \mathbb{E}\big(\max_{1\leq k\leq n} X_k\big) \,\leq\, C\sqrt{\log n} \tag{6}$$

for numerical constants $0 < c < C < \infty$.

The upper-bound actually holds in a wider generality. Let $X_1, \ldots, X_n$, $n \geq 2$, be centered Gaussian variables on $(\Omega, \mathcal{A}, \mathbb{P})$ with respective variances $\sigma_1^2, \ldots, \sigma_n^2$. Then, for any $0 < p < \infty$, there is a constant $C_p > 0$ only depending on p such that

$$\Big[\mathbb{E}\big(\max_{1\leq k\leq n} |X_k|^p\big)\Big]^{1/p} \,\leq\, C_p \max_{1\leq k\leq n} \sigma_k \sqrt{\log n}. \tag{7}$$

As a quick proof, by homogeneity, it may be assumed that $\max_{1\leq k\leq n} \sigma_k \leq 1$. For $p \geq 2$, let $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$ be the convex function equal to $e^{\frac{1}{4}x^{2/p}}$ on the complement of the interval $[0, (2p-4)^{p/2}]$, and equal to $e^{\frac{1}{4}(2p-4)}$ on this interval. By Jensen's inequality,

$$\exp\bigg(\frac{1}{4}\Big[\mathbb{E}\big(\max_{1\leq k\leq n} |X_k|^p\big)\Big]^{2/p}\bigg) \,\leq\, \Psi\Big(\mathbb{E}\big(\max_{1\leq k\leq n} |X_k|^p\big)\Big) \,\leq\, \mathbb{E}\Big(\Psi\big(\max_{1\leq k\leq n} |X_k|^p\big)\Big).$$

Now

$$\begin{aligned}
\mathbb{E}\Big(\Psi\big(\max_{1\leq k\leq n} |X_k|^p\big)\Big) &\leq e^{\frac{1}{4}(2p-4)} + \mathbb{E}\big(e^{\frac{1}{4}\max_{1\leq k\leq n} X_k^2}\big) \\
&\leq e^{\frac{1}{4}(2p-4)} + \sum_{k=1}^n \mathbb{E}\big(e^{\frac{1}{4}X_k^2}\big) \\
&\leq e^{\frac{1}{4}(2p-4)} + n\sqrt{2}\,,
\end{aligned}$$

from which the claim follows.

Towards the lower-bound in (6), since the $X_1, \ldots, X_n$ are independent with common law γ_1, the distribution function of the random variable $\max_{1\leq k\leq n} X_k$ is $\Phi(t)^n$, $t \in \mathbb{R}$ (where Φ is the distribution function of γ_1), so its law has density $n\Phi(x)^{n-1}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2}$, $x \in \mathbb{R}$, with respect to the Lebesgue measure λ_1 on $\mathbb{R}$. Hence

$$\mathbb{E}\big(\max_{1\leq k\leq n} X_k\big) \,=\, \int_{\mathbb{R}} x\, n\Phi(x)^{n-1}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2} d\lambda_1(x) \,=\, \int_{\mathbb{R}} n(n-1)\Phi(x)^{n-2}\frac{1}{2\pi}\,e^{-x^2} d\lambda_1(x)$$

where the second equality follows from the integration by parts formula (1). In particular, if $n = 2$, $\mathbb{E}(\max(X_1, X_2)) = \frac{1}{2\sqrt{\pi}}$, so that in the following it may be assumed that n is large enough (larger than some fixed n_0). Now, for any $\alpha \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}\big(\max_{1\leq k\leq n} X_k\big) &\geq \int_\alpha^\infty n(n-1)\Phi(x)^{n-2}\,\frac{1}{2\pi}\,e^{-x^2}d\lambda_1(x) \\
&\geq n(n-1)\Phi(\alpha)^{n-2}\int_\alpha^\infty \frac{1}{2\pi}\,e^{-x^2}d\lambda_1(x) \\
&= n(n-1)\Phi(\alpha)^{n-2}\,\frac{1}{2\sqrt{\pi}}\,\big[1-\Phi\big(\sqrt{2}\alpha\big)\big].
\end{aligned}$$

Take then $\alpha = \alpha_n = \sqrt{2\log n - \log\log n}$, $n \geq n_0$. Recall the classical tail estimates (cf. [1])

$$\Big(\frac{1}{t} - \frac{1}{t^3}\Big)\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}t^2} \;\leq\; 1-\Phi(t) \;\leq\; \frac{1}{t}\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}t^2}$$

for every $t > 0$. By the upper-bound, $\Phi(\alpha_n) \geq 1 - \frac{c}{n}$ for some numerical $c > 0$, while by the lower-bound, $1 - \Phi(\sqrt{2}\alpha_n) \geq \frac{c}{n^2}\sqrt{\log n}$. The claim follows.

References

[1] Some basics on Gaussian measures and variables. *The Gaussian Blog.*

[2] Stein's method and applications. *The Gaussian Blog.*

[3] Hermite polynomials. *The Gaussian Blog.*

[4] S. Bernstein. On a property which characterizes a Gaussian distribution. *Proceedings of the Leningrad Polytechnic Institute* 217, 21–22 (1941).

[5] H. Cramér. Über eine Eigenschaft der normalen Verteilungsfunktion. *Math. Z.* 41, 405–414 (1936).

[6] M. Kac. On a characterization of the normal distribution. *Amer. J. of Math.* 61, 726–728 (1939).