Gaussian vectors and spherical measure

Dimitri Meunier

20/04/2021

In those notes the theoretical sections on the characteristic function, the gaussian vectors and the uniform measure on the sphere make intensive use of the lecture notes Le Gall (2006) (in French!).

1 Random Vectors

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ a measurable space. Recall that a random variable is a measurable function $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$. The distribution of X denoted $\mathbb{P}_X$ is the probability measure on $(E, \mathcal{E})$ defined for all $A \in \mathcal{A}$ by $\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A))$ ($\mathbb{P}_X$ is the push-forward measure of $\mathbb{P}$ through X).

If $(E, \mathcal{E})$ is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, X is called a **real random variable** and if $(E, \mathcal{E})$ is $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, X is called a **real random vector**. In the latter case, we denote by X_i, $i = 1, \dots, d$ its coordinates, they are real random variables with distribution $\mathbb{P}_{X_i} = \pi^i_{\#}\mathbb{P}_X$ where π^i is the projection along axis i and $\#$ denotes the push-forward operator.

The Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is denoted λ_d, the unit sphere is $S^{d-1} = \{x \mid \|x\| = 1\}$ and the unit ball $B^d = \{x \mid \|x\| \leq 1\}$. The set of vectorial isometries on $\mathbb{R}^d$ is $\{\phi : \mathbb{R}^d \to \mathbb{R}^d \text{ linear } \mid \|\phi(x)\| = \|x\| \quad \forall x \in \mathbb{R}^d\}$. The set of associated matrices is the orthogonal group $\mathcal{O}(d) = \{A \in \mathbb{R}^{d \times d} | A^T A = I_d\}$. For a random vector X in $\mathbb{R}^d$, we recall the following,

- $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^T \in \mathbb{R}^d$
- $\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T] = \mathbb{E}[XX^T] - \mathbb{E}[X]\mathbb{E}[X]^T \in \mathbb{R}^{d \times d}$. $\mathbb{V}[X]_{ij} = Cov(X_i, X_j), \quad i, j = 1, \dots, d$

1.1 Characteristic function

The proof of the characterisation of Gaussian random vectors requires to introduce the characteristic function of a real random vector. We start by recalling the dominated convergence theorem and a corollary for continuity and differentiability of parametric integrals. We then define the characteristic function and state useful properties, before introducing Gaussian random vectors in the next section.

Theorem 1 (Dominated convergence theorem)**.** *Let* (f_n) *be a sequence of functions in* $\mathcal{L}^1(E, \mathcal{E}, \mu)$ *(respectively in* $\mathcal{L}^1_{\mathbb{C}}(E, \mathcal{E}, \mu)\big)$*. We assume that,*

1. *there is a measurable function* $f : E \to \mathbb{R}$ *(respectively in* $\mathbb{C}$*) such that,*

$$f_n(x) \longrightarrow f(x) \quad \mu\text{-}a.e,$$

2. *there is a measurable function* $g : E \longrightarrow \mathbb{R}_+$ *such that* $\int g d\mu < \infty$ *and for all* n*,*

$$|f_n| \leq g \quad \mu\text{-}a.e,$$

then $f \in \mathcal{L}^1(E, \mathcal{E}, \mu)$ *(resp.* $f \in \mathcal{L}^1_{\mathbb{C}}(E, \mathcal{E}, \mu)$*), and we have*

$$\lim_{n\to\infty} \int |f_n - f| d\mu = 0.$$

Theorem 2 (Continuity and differentiability of a parametric integral)**.** *Let* (U, d) *be a metric space and* $f : U \times E \longrightarrow \mathbb{R}$ *(or* $\mathbb{C}$*). Let* $u_0 \in E$*. We assume that,*

1. *for all* $u \in U$*,* $x \longrightarrow f(u, x)$ *is measurable,*
2. $\mu(dx)$*-a.e* $u \longrightarrow f(u, x)$ *is continuous in* u_0*,*
3. *there is a function* $g \in \mathcal{L}^1(E, \mathcal{E}, \mu)$ *such that for all* $u \in U$*,*

$$|f(u, x)| \leq g(x) \quad \mu(dx)\text{-}a.e,$$

then, $F(u) = \int_E f(u, x)\mu(dx)$ *is well-defined for all* $u \in U$ *and continuous in* u_0*. Suppose now that* $U = I$ *is an open interval of* $\mathbb{R}$*. We assume that,*

1. *for all* $u \in I$*,* $x \longrightarrow f(u, x)$ *is in* $\mathcal{L}^1(E, \mathcal{E}, \mu)$*,*
2. $\mu(dx)$*-a.e.* $u \longrightarrow f(u, x)$ *is differentiable on* I*,*
3. *there is a function* $g \in \mathcal{L}^1_+(E, \mathcal{E}, \mu)$ *such that* $\mu(dx)$*-a.e,*

$$\forall u \in I, \quad \left|\frac{\partial f}{\partial u}(u, x)\right| \leq g(x),$$

then F *is differentiable on* I*, with*

$$F'(u_0) = \int \frac{\partial f}{\partial u}(u_0, x)\, \mu(dx)$$

Definition 1 (Characteristic function)**.** *If* X *is a real random vector, the characteristic function of* X *is the function* $\Phi_X : \mathbb{R}^d \longrightarrow \mathbb{C}$ *defined by*

$$\Phi_X(\xi) = \mathbb{E}[e^{i\langle \xi, X\rangle}] = \int_{\mathbb{R}^d} e^{i\langle \xi, X\rangle} \mathbb{P}_X(dx), \quad \xi \in \mathbb{R}^d$$

Φ_X *is the Fourier transform of the distribution* $\mathbb{P}_X$*.*

Remark. The characteristic function is defined as a parametric integral, and Theorem 2 shows that Φ_X is continuous (and bounded) on $\mathbb{R}^d$. Indeed, $|e^{i\xi\cdot x}| \leq 1 \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X)$.

Theorem 3. *The characteristic function of a real random vector X characterised its distribution. In other words, the Fourier transform defined on the space of probability measures on $\mathbb{R}^d$ is injective.*

Corollary 3.1 (Marginal distributions)**.** *Two real random vectors X and Y have the same distribution if and only if they have the same marginal distributions, i.e. $\mathbb{P}_X = \mathbb{P}_Y$ if and only if for all $\theta \in S^{d-1}$, $\mathbb{P}_{\langle X,\theta\rangle} = \mathbb{P}_{\langle Y,\theta\rangle}$.*

Proof. First assume that for all $\theta \in S^{d-1}$, $\mathbb{P}_{\langle X,\theta\rangle} = \mathbb{P}_{\langle Y,\theta\rangle}$, by injectivity of the Fourier transform on $\mathbb{R}$, $\Phi_{\langle X,\theta\rangle} = \Phi_{\langle Y,\theta\rangle}$. For all $\xi \in \mathbb{R}^d$ we can find $t \in \mathbb{R}$ and $\theta \in S^{d-1}$ such that $\xi = t\theta$ (take $t = \|\xi\|$ and $\theta = \xi/\|\xi\|$), then,

$$\Phi_X(\xi) = \Phi_X(t\theta) = \Phi_{\langle X,\theta\rangle}(t) = \Phi_{\langle Y,\theta\rangle}(t) = \Phi_Y(t\theta) = \Phi_Y(\xi).$$

By injectivity of the Fourier transform we conclude that $\mathbb{P}_X = \mathbb{P}_Y$. The other direction is proved similarly. $\square$

Corollary 3.2 (Independence)**.** *If X is a real random vector on $\mathbb{R}^d$, its coordinates are independant if and only if the characteristic function of X factorized as,*

$$\Phi_X(\xi_1, \ldots, \xi_d) = \prod_{i=1}^{d} \Phi_{X_i}(\xi_i)$$

Proof. It follows from the injectivity of the Fourier transform (Theorem 3) and the fact that the coordinates are independent if and only if $\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_d}$. $\square$

Proposition 1. *If X is a real random vector with finite second moments, then its characteristic function is C^2 and,*

$$\Phi_X(\xi) = 1 + i\mathbb{E}(X)\cdot\xi - \frac{1}{2}\xi^T\mathbb{E}\left(XX^T\right)\xi + o\left(\|\xi\|^2\right)$$

Proof. It follows from Theorem 2 and an order 2 Taylor expansion. $\square$

1.2 Gaussian vectors

We are now ready to introduce the formal definition of a Gaussian random vector. We first recall the density and characteristic function of a univariate Gaussian distribution.

Definition 2. *The standard normal (or Gaussian) distribution on $\mathbb{R}$ is the absolutely continuous measure (w.r.t to λ_1) with density,*

$$f(x) = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}.$$

A random variable that follows this distribution is denoted $X \sim \mathcal{N}_1(0,1)$. We say that $X \sim \mathcal{N}_1(\mu, \sigma^2)$ if $X = \mu + \sigma Z$ ($\mu \in \mathbb{R}$, $\sigma \geq 0$) where $Z \sim \mathcal{N}_1(0,1)$. If $\sigma > 0$, the change of variable formula shows that the density function of X is

$$f(x) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{x^2}{2\sigma^2}}.$$

If $\sigma = 0$, $\mathbb{P}_X = \delta_\mu$.

Proposition 2. *If $X \sim \mathcal{N}_1(\mu, \sigma^2)$,*

$$\Phi_X(\xi) = \exp\left(i\xi\mu - \frac{\sigma^2\xi^2}{2}\right), \quad \xi \in \mathbb{R}$$

Proof. It is sufficient to show that if $X \sim \mathcal{N}_1(0,1)$,

$$\Phi_X(\xi) = e^{-\frac{\xi^2}{2}}, \quad \xi \in \mathbb{R}$$

Since the sinus function is odd, we have,

$$\Phi_X(\xi) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} e^{i\xi x} dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(\xi x) dx$$

Then, since $|xe^{-x^2/2}\sin(\xi x)| \leq |x|e^{-x^2/2} \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda_1)$, by Theorem 2 and integration by parts,

$$\Phi'_X(\xi) = -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} \sin(\xi x) dx = -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \xi \cos(\xi x) dx = -\xi \Phi_X(\xi).$$

Φ_X is therefore solution of the differential equation $f'(\xi) = -\xi f(\xi)$, with initial condition $f(0) = 1$. We conclude that $\Phi_X(\xi) = \exp\left(-\xi^2/2\right)$. □

Definition 3. *Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ be a real random vector. X is a **Gaussian vector** if for all $\theta \in \mathbb{R}^d$, $\langle X, \theta \rangle$ has a univariate normal distribution.*

Remarks.

- It is equivalent to assume that the marginal normality holds for all $\theta \in S^{d-1}$.

- From the definition we see that if X is a vector of independent univariate Gaussian variables, X is a Gaussian vector. Indeed, for all $\theta \in \mathbb{R}^d$ and $\xi \in \mathbb{R}$,

$$\begin{aligned} \Phi_{\langle \theta, X \rangle}(\xi) &= \mathbb{E}\left\{e^{i\xi \sum_{l=1}^d \theta_l X_l}\right\} \\ &= \prod_{l=1}^d \mathbb{E}\left\{e^{i\xi\theta_l X_l}\right\} \\ &= \prod_{l=1}^d e^{\left(i\xi\theta_l\mu_l - \frac{1}{2}\xi^2\theta_l^2\sigma_l^2\right)} \qquad \text{if } X_l \sim \mathcal{N}_1\left(\mu_l, \sigma_l^2\right) \\ &= e^{i\xi \sum_{l=1}^d \theta_l\mu_l - \frac{1}{2}\xi^2 \sum_{l=1}^d \theta_l^2\sigma_l^2} \end{aligned}$$

Hence, by injectivity of the characteristic function,

$$\langle \theta, X \rangle \sim \mathcal{N}_1 \left(\sum_{l=1}^{d} \theta_l \mu_l, \sum_{l=1}^{d} \theta_l^2 \sigma_l^2 \right),$$

which shows that X is a Gaussian vector.

- Secondly, if X is a Gaussian vector, for all $B \in \mathbb{R}^{r \times d}$ and $b \in \mathbb{R}^r$, $Y = BX + b$ is also a Gaussian vector. Indeed for all $\theta \in \mathbb{R}^r$, $\langle Y, \theta \rangle = \langle X, B^T \theta \rangle + \langle \theta, b \rangle$ follows a univariate normal distribution.

Theorem 4. *A random vector $X : \Omega \to \mathbb{R}^d$ is Gaussian if and only if, there exists a vector $\mu \in \mathbb{R}^d$ and a positive semi-definite matrix $K \in \mathbb{R}^{d \times d}$ such that,*

$$\Phi_X(\theta) = \exp\left(i\mu \cdot \theta - \frac{1}{2} \theta^t K \theta \right), \qquad \theta \in \mathbb{R}^d. \tag{1}$$

Furthermore, $\mu = \mathbb{E}[X]$ and $K = \mathbb{V}(X)$. If X is a random variable that admits the characteristic function above, we use the notation $X \sim \mathcal{N}_d(\mu, K)$.

Proof. Let X be a Gaussian vector, we first notice that for all $i = 1, \ldots, d$, $\mathbb{E}[|X_i|^p] < \infty$ $(1 \leq p < +\infty)$. Indeed, $X_i = \langle X, e_i \rangle$ follows a univariate normal distribution. Therefore, the expectation $\mu := \mathbb{E}[X]$ and covariance $K := \mathbb{E}[(X - \mu)(X - \mu)^T]$ exist. Let us fix $\theta \in \mathbb{R}^d$, since $Y := \langle X, \theta \rangle \sim \mathcal{N}_1(\mu^T \theta, \theta^T K \theta)$, we have,

$$\Phi_X(\theta) = \Phi_Y(1) = e^{i\langle \theta, \mu \rangle - \theta^T K \theta / 2} \qquad \text{by prop. 2}.$$

For the converse, assume that X is a random variable with a characteristic function as (1) with $\mu \in \mathbb{R}^d$ and $K \in \mathbb{R}_+^{d \times d}$. Then, for all $\theta \in \mathbb{R}^d$ and $\xi \in \mathbb{R}$,

$$\Phi_{\langle X, \theta \rangle}(\xi) = \Phi_X(\xi \theta) = e^{i\xi \langle \theta, \mu \rangle - \xi^2 \theta^T K \theta / 2}.$$

We recognise the characteristic function of a univariate Gaussian distribution, which implies that $\langle X, \theta \rangle$ follows a univariate normal distribution and we conclude that X is a Gaussian vector. It remains to prove that $\mu = \mathbb{E}[X]$ and $K = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$. Since X is a Gaussian vector, it is squared integrable and we can thus apply Proposition 1,

$$\Phi_X(\theta) = 1 + i\mathbb{E}(X) \cdot \theta - \frac{1}{2} \theta^T \mathbb{E}\left(X X^T \right) \theta + o\left(\|\theta\|^2 \right)$$

Hence, using that $\ln(1 + y) = y - y^2/2 + o(y^2)$,

$$\begin{aligned}
\ln \Phi_X(\theta) &= i\mathbb{E}(X)^T \theta - \frac{1}{2} \theta^T \mathbb{E}\left(X X^T \right) \theta + \frac{1}{2} \left(\mathbb{E}(X)^T \theta \right)^2 + o\left(\|\theta\|^2 \right) \\
&= i\mathbb{E}(X)^T \theta - \frac{1}{2} \theta^T \left(\mathbb{E}\left(X X^T \right) - \mathbb{E}(X)\mathbb{E}(X)^T \right) \theta + o\left(\|\theta\|^2 \right) \\
&= i\mathbb{E}(X)^T \theta - \frac{1}{2} \theta^T \mathbb{E}\left((X - \mathbb{E}[X])(X - \mathbb{E}[X])^T \right) \theta + o\left(\|\theta\|^2 \right)
\end{aligned}$$

On the other hand, by assumption,

$$\ln \Phi_X(\theta) = i\langle \theta, \mu \rangle - \frac{1}{2}\theta^t K \theta.$$

We conclude by identifying both expressions. $\square$

The theorem shows that a Gaussian vector is fully characterised by its two first moments!

Corollary 4.1. *If X is a Gaussian vector, its coordinates are independant if and only if its covariance matrix is diagonal.*

Proof. Indeed from the last theorem, if K is diagonal, the characteristic function can be factorized in a product which characterises the independence (Corollary 3.2). $\square$

Definition 4 (Standard Gaussian random vector)**.** *X is called a **standard Gaussian** vector on $\mathbb{R}^d$ if its coordinates are i.i.d with distribution $\mathcal{N}_1(0, 1)$. By the last theorem, $X \sim \mathcal{N}_d(0, I_d)$.*

By independence of the coordinates we see that the density function of $X \sim \mathcal{N}_d(0, I_d)$ is

$$f(x) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}\|x\|_2^2}, \quad x \in \mathbb{R}^d.$$

Theorem 5. *Let X be a Gaussian vector, $X \sim \mathcal{N}_d(\mu, K)$ if and only if $X = K^{1/2}Z + \mu$, where $Z \sim \mathcal{N}_d(0, I_d)$ and the equality holds in distribution.*

Proof. Assume that $X \sim \mathcal{N}_d(\mu, K)$, K is a semi-definite positive matrix, hence there exists an orthogonal matrix U and a diagonal matrix D (with nonnegative diagonal elements) such that $K = UDU^T$[1]. If $Z \sim \mathcal{N}_d(0, I_d)$, since any affine transformation of a Gaussian vector is a Gaussian vector, $Y := K^{1/2}Z + \mu$ is a Gaussian vector. As mentioned previously, a Gaussian vector is characterised by its two first moments and $\mathbb{E}[Y] = K^{1/2}\mathbb{E}[Z] + \mu = \mu$ and $\mathbb{V}[Y] = K^{1/2}\mathbb{V}[Z]K^{1/2} = K$, we conclude that $X = Y$ in distribution.

By the same argument, the other direction is immediate. $\square$

Proposition 3. *If $X \sim \mathcal{N}_d(\mu, K)$, X admits a density if and only if K is invertible and in that case, its density function is*

$$f(x) = |2\pi K|^{-\frac{1}{2}} e^{-\frac{1}{2}\|x-\mu\|_{K^{-1}}^2},$$

where $\|.\|_A$ is the Mahalanobis distance (which is a norm for definite positive matrices).

Proof. We have seen that $X = K^{1/2}Z + \mu$, where $Z \sim \mathcal{N}_d(0, I_d)$ and the equality holds in distribution. The result follows from a change of variable on the density of the standard Gaussian vector through the C^1-diffeomorphism $\phi : x \in \mathbb{R}^d \to K^{-1/2}(x - \mu)$. $\square$

[1] Recall that the square root of a semi-definite positive matrix is defined as $K^{1/2} := UD^{1/2}U^T$ (the definition makes sense since $U^T = U^{-1}$, $K^{1/2}K^{1/2} = K$).

2 Spherical measure and normal distribution

In this section we show how one can define a canonical measure σ_d on $(S^{d-1}, \mathcal{B}(S^{d-1}))$ that is invariant to isometries. Similarly to the Lebesgue measure λ_d that is the unique — up to constants — translation-invariant measure on $\mathbb{R}^d$,, σ_d is the unique probability measure on S^{d-1} invariant to isometries.

Definition 5. *If $A \in \mathcal{B}\left(S^{d-1}\right)$, we define the* ***wedge*** *$\Gamma(A)$ as the Borel set of $\mathbb{R}^d$ defined by*

$$\Gamma(A) = \{rx; r \in [0,1] \text{ and } x \in A\}$$

For all $A \in \mathcal{B}\left(S^{d-1}\right)$, the measure,

$$\omega_d(A) = \lambda_d(\Gamma(A))$$

is called the ***spherical measure****.*

Proposition 4. *The volume of the d-dimensional ball $B^d = \{x \mid \|x\| \leq 1\}$ is*

$$\lambda_d(B^d) = \omega_d(S^{d-1}) = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}.$$

Definition 6 (Uniform probability distribution on the sphere)**.**

$$\sigma_d(A) := \frac{\Gamma\left(\frac{d}{2}+1\right)}{\pi^{d/2}} \lambda_d(\{rx : 0 \leq r \leq 1, x \in A\}) \tag{2}$$

defines the ***uniform probability distribution on the sphere****.*

Theorem 6 (Polar change of variable)**.** *For any measurable function $f : \mathbb{R}^d \to \mathbb{R}_+$,*

$$\begin{aligned}\int_{\mathbb{R}^d} f(x)dx &= \int_{S^{d-1}} \left(\int_0^\infty f(r\gamma)dr^{d-1}dr\right) d\omega_d(\gamma) \\ &= \omega_d(S^{d-1}) \int_{S^{d-1}} \left(\int_0^\infty f(r\gamma)dr^{d-1}dr\right) d\sigma_d(\gamma)\end{aligned}$$

If f is integrable on $\mathbb{R}^d$, the above equation also holds.

Remark. If f is radial, i.e. $f : \mathbb{R}^d \to \mathbb{R}_+$ and there exists $g : \mathbb{R} \to \mathbb{R}_+$ such that $f(x) = g(\|x\|)$ for all $x \in \mathbb{R}^d$, then the polar change of variable leads to,

$$\int_{\mathbb{R}^d} f(x)dx = \omega_d(S^{d-1}) \int_0^\infty g(r)dr^{d-1}dr \tag{3}$$

Theorem 7. *The probability measure σ_d is the unique probability measure on the sphere S^{d-1} invariant to the action of vectorial isometries, hence we call it the uniform probability measure on the sphere.*

Proposition 5 (Exercise 3.3.7 Vershynin (2018))**.** *Let us write $X \sim N_d(0, I_d)$ in polar form as*

$$X = R\theta$$

where $R = \|X\|_2$ is the length and $\theta = X/\|X\|_2$ is the direction of X. Prove the following:

1. *the length R and direction θ are independent random variables*
2. *the direction θ is uniformly distributed on the unit sphere S^{d-1}*
3. *(Bonus) the length R follows a generalized gamma distribution*

Proof. We note ρ the density of $X \sim \mathcal{N}_d(0, I_d)$. We want to compute the distribution of R and θ where $(R, \theta) = (\|X\|_2, X/\|X\|_2)$ is a random vector with values in $\mathbb{R} \times S^{d-1}$. For all measurable function $h : \mathbb{R} \times S^{d-1} \to \mathbb{R}$ positive or bounded,

$$\begin{aligned}
\mathbb{E}[h(R,\theta)] &= \int_{\mathbb{R}^d} h(\|x\|, x/\|x\|)\rho(x)dx \\
&= \int_{S^{d-1}} \left(\int_0^\infty h(r,\theta)\rho(r\theta)dr^{d-1}dr \right) d\omega_d(\theta) \\
&= \int_{S^{d-1}} \left(\int_{\mathbb{R}_+} h(r,\theta) \underbrace{\frac{e^{-r^2/2}}{(2\pi)^{d/2}} dr^{d-1} 1_{r\geq 0}}_{=:g(r,\theta)} dr \right) d\omega_d(\theta)
\end{aligned} \tag{4}$$

g is the density of (R, θ) and we notice that g is constant with respect to θ, it implies both that R and θ are independent and that θ is uniformely distributed on the sphere.

As a sanity check we can explicitely compute the constants. The part of the density that depends on r is $e^{-r^2/2}r^{d-1}1_{r\geq 0}$, it is the un-normalized density function of a **generalized gamma distribution**. Therefore, the density function of R is[2],

$$f_\gamma(r) = e^{-r^2/2}r^{d-1}\frac{2}{\Gamma(d/2)2^{d/2}}1_{r\geq 0}.$$

Thus for all $r \geq 0, \theta \in S^{d-1}$,

$$g(r,\theta) = f_\gamma(r) \times \frac{d\Gamma(d/2)2^{d/2}}{2(2\pi)^{d/2}} = f_\gamma(r) \times \frac{\Gamma(d/2+1)}{\pi^{d/2}} = f_\gamma(r) \times \omega_d(S^{d-1})^{-1}$$

□

[2] without knowing the generalized gamma density function, the normalisation constant can be obtained from the gamma density function by applying the change of variable $\phi(x) = \sqrt{x}$

3 Sub-Gaussian vectors

Definition 7 (Sub-gaussian random vectors)**.** *A random vector X in $\mathbb{R}^d$ is called sub-gaussian if the one-dimensional marginals $\langle X, \theta \rangle$ are sub-gaussian random variables for all $\theta \in \mathbb{R}^d$. The sub-gaussian norm of X is defined as*

$$\|X\|_{\psi_2} = \sup_{\theta \in S^{d-1}} \|\langle X, \theta \rangle\|_{\psi_2}$$

The sub-gaussian norm defines a proper norm,

- Since the univariate sub-gaussian norm is a norm, it is obvious that $\|X\|_{\psi_2} \geq 0$.
- If $\|X\|_{\psi_2} = 0$, then for all $\theta \in S^{d-1}$, $\|\langle X, \theta \rangle\|_{\psi_2} = 0$ and since the univariate sub-gaussian norm is a norm it implies $\langle X, \theta \rangle = 0$ a.s. For all $i = 1, \dots, d$, taking, $\theta = e_i$,, leads to $X_i = 0$ a.s. and thus $X = 0$ a.s.
- For the triangular inequality,

$$\|X+Y\|_{\psi_2} = \sup_{\theta \in S^{d-1}} \|\langle X+Y, \theta \rangle\|_{\psi_2} \leq \sup_{\theta \in S^{d-1}} \|\langle X, \theta \rangle\|_{\psi_2} + \|\langle Y, \theta \rangle\|_{\psi_2} \leq \|X\|_{\psi_2} + \|Y\|_{\psi_2}$$

This norm is also invariant to rotation. Indeed, if $U \in \mathcal{O}(d)$, for all $\theta \in S^{d-1}$, $\langle U^T X, \theta \rangle = \langle X, U\theta \rangle$, and $\|U\theta\|_2 = \theta^T U^T U \theta = \theta^T \theta = 1$. Therefore, $\langle U^T X, \theta \rangle = \langle X, \tilde{\theta} \rangle$, where $\tilde{\theta} \in S^{d-1}$.

Example 1. If $X \sim \mathcal{N}_d(0, \Sigma)$, $\|X\|_{\psi_2} = \lambda_{max}(\Sigma)$. Indeed for all $\theta \in S^{d-1}$, $\langle X, \theta \rangle \sim \mathcal{N}(0, \theta^T \Sigma \theta)$, and $\psi_{\langle X, \theta \rangle}(t) = \frac{t^2 \theta^T \Sigma \theta}{2}$ (the log moment generating function), hence for all $t \in \mathbb{R}$,

$$\arg\max_{\theta \in S^{d-1}} \psi_{\langle X, \theta \rangle}(t) = \frac{t^2 \lambda_{max}(\Sigma)}{2},$$

which implies $\|X\|_{\psi_2} = \lambda_{max}(\Sigma)$ (depending on the definition of the sub-gaussian univariate norm, there might be an absolute constant to add).

Example 2. If X is a random vector with (non-necessarily independent) sub-gaussian coordinates, X is still a sub-gausian vector but the norm can grow with the dimension! For all $\theta \in S^{d-1}$, by the triangular inequality,

$$\begin{aligned}\|\langle X, \theta \rangle\|_{\psi_2} &= \left\| \sum_{i=1}^d \theta_i X_i \right\|_{\psi_2} \leq \sum_{i=1}^d |\theta_i| \, \|X_i\|_{\psi_2} \\ &\leq \max_{i \leq d} \|X_i\|_{\psi_2} \|\theta\|_1 \leq \sqrt{d} \max_{i \leq d} \|X_i\|_{\psi_2}\end{aligned}$$

Where we have used that for all $x \in \mathbb{R}^d$, $\|x\|_1 \leq \sqrt{d}\|x\|_2$. Taking the supremum over the sphere, it shows that,

$$\|X\|_{\psi_2} \leq \sqrt{d} \max_{i \leq d} \|X_i\|_{\psi_2}$$

The bound grows with the dimension and quickly becomes vacuous, worse, it is tight! Indeed, consider a sub-gaussian real random variable Z and define $X := (Z, \ldots, Z)^T$. For all $\theta \in S^{d-1}$,

$$\|\langle X, \theta\rangle\|_{\psi_2} = \left\|\sum_{i=1}^{d} \theta_i Z\right\|_{\psi_2} = \left|\sum_{i=1}^{d} \theta_i\right| \|Z\|_{\psi_2} \leq \|\theta\|_1 \|Z\|_{\psi_2} \leq \sqrt{d}\, \|Z\|_{\psi_2}$$

Let us note that for $\theta = (d^{-1/2}, \ldots, d^{-1/2})$, $\theta \in S^{d-1}$ and $\|\theta\|_1 = \sqrt{d}$. Thus, taking the supremum over the sphere, it shows that,

$$\|X\|_{\psi_2} = \sqrt{d}\, \|Z\|_{\psi_2} >> \|Z\|_{\psi_2} = \max_{i \leq d} \|X_i\|_{\psi_2}.$$

Example 3. The last example shows that a real random vector with sub-gaussian coordinates does not necessarily behaves properly in high dimension. However, if we assume that X is a random vector with **independent mean-zero** sub-gaussian coordinates, we can derive a more satisfying (i.e. independent of the dimension) bound for the sub-gaussian vector norm. It is because we can use a stronger inequality than the triangular inequality in that setting.

Proposition 6 (Proposition 2.6.1 in Vershynin (2018) — Sums of independent sub-gaussians)**.** *Let $X_1, \ldots, X_d$ be independent, mean zero, sub-gaussian random variables. Then $\sum_{i=1}^{d} X_i$ is also a sub-gaussian random variable, and*

$$\left\|\sum_{i=1}^{d} X_i\right\|_{\psi_2}^2 \leq \sum_{i=1}^{d} \|X_i\|_{\psi_2}^2$$

where again, depending on the definition of the sub-gaussian univariate norm one takes, there might be an absolute constant to add.

Using this inequality, for all $\theta \in S^{d-1}$,

$$\begin{aligned}\|\langle X, \theta\rangle\|_{\psi_2}^2 = \left\|\sum_{i=1}^{d} \theta_i X_i\right\|_{\psi_2}^2 &\leq \sum_{i=1}^{d} \theta_i^2 \|X_i\|_{\psi_2}^2 \\ &\leq \max_{i \leq d} \|X_i\|_{\psi_2}^2\end{aligned}$$

Therefore, taking the supremum over the sphere, we get,

$$\|X\|_{\psi_2} \leq \max_{i \leq d} \|X_i\|_{\psi_2}$$

References

Jean-François Le Gall. Intégration, probabilités et processus aléatoires. ***Ecole Normale Supérieure de Paris***, 2006.

Roman Vershynin. ***High-dimensional probability: An introduction with applications in data science***, volume 47. Cambridge university press, 2018.