

Isotropy, Gaussian vector, spherical measure and concentration

Dimitri Meunier

13/04/2021

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 $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$. 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 .

1.1 Characteristic function

In order to introduce Gaussian random vectors we first recall useful properties of the characteristic function.

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

$$\Phi_X(\xi) = E[\exp(i\xi \cdot X)] = \int e^{i\xi \cdot x} \mathbb{P}_X(dx), \quad \xi \in \mathbb{R}^d$$

Φ_X is the Fourier transform of the distribution on X . The dominated convergence theorem shows that Φ_X is continuous (and bounded) on \mathbb{R}^d .

Theorem 1. *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.*

Proposition 1. *If X is a real random vector on \mathbb{R}^d , its coordinates are independent if and only if the characteristic function of X is*

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

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

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

1.2 Gaussian vectors

We first recall the density and the 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 $X = \mu + \sigma Z$ ($\sigma \geq 0$) where $Z \sim \mathcal{N}_1(0, 1)$. If $\sigma > 0$, the change of variable formula shows that, the density function of X is then,

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

Proposition 3. *If $X \sim \mathcal{N}_1(\mu, \sigma^2)$, then,*

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

We are now ready to introduce the definition of a Gaussian random vector.

Definition 3. *Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ 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.*

From this 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 $\xi \in \mathbb{R}$,

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

Hence, by injectivity of the characteristic function,

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

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 2. *A random vector $X : \Omega \rightarrow \mathbb{R}^d$ 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,*

$$\Phi_X(\xi) = \exp \left(i\mu \cdot \xi - \frac{1}{2} \xi^T K \xi \right), \quad \xi \in \mathbb{R}^d. \quad (1)$$

Furthermore, μ and K are the expectation and covariance of 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 $i = 1, \dots, d$, $\mathbb{E}[|X_i|^p] < \infty$ ($1 \leq p < +\infty$). 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,

$$\Phi_X(\theta) = \Phi_Y(1) = e^{i\theta^T \mu - \theta^T K \theta / 2}.$$

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 $\xi \in \mathbb{R}^d$,

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

We recognise the characteristic function of a univariate Gaussian distribution, which implies that $\langle X, \theta \rangle$ follows a univariate distribution. 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, we have seen that it is squared integrable, thus, we can apply Proposition 2,

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

Hence,

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

On the other hand, by assumption,

$$\ln \Phi_X(\xi) = i\mu^T \xi - \frac{1}{2} \xi^T K \xi,$$

and we conclude by identification. \square

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

Corollary 3.1. *If X is a Gaussian vector, its coordinates are independant if and only if the covariance is diagonal.*

Proof. Indeed from the theroem, if K is diagonal the characteristic function can be factorized in a product which characterised the independence. \square

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

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 3. *Let X be a Gaussian vector with mean μ and covariance K (i.e. $X \sim \mathcal{N}(\mu, K)$), then $X = K^{1/2}Z + \mu$, where $Z \sim \mathcal{N}_d(0, I_d)$ and the equality holds in distribution.*

Proof. K is a covariance matrix which is a semi-definite positive symmetric matrix, hence there exists an orthogonal matrix U and a diagonal matrix D (with nonnegative diagonal elements) such that $K = UDU^T$. Recall that the square root of a semi-definite positive symmetric matrix is defined as $K^{1/2} := UD^{1/2}U^T$ (the definition makes sense since $U^T = U^{-1}$, $K^{1/2}K^{1/2} = K$).

$Z \sim \mathcal{N}_d(0, I_d)$ is a Gaussian vector and we have seen that any affine transformation of a Gaussian vector is a Gaussian vector, therefore $Y := K^{1/2}Z + \mu$ 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 $\mathbb{V}[Y] = K^{1/2}\mathbb{V}[Z]K^{1/2} = K$, Q.E.D. \square

Proposition 4. *If $X \sim \mathcal{N}(\mu, K)$, X admits a density if and only if K is invertible and in that case, its density function is,*

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

$\|\cdot\|_A$ 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 change of variable on the density of the standard Gaussian density through the C^1 -diffeomorphism $\phi : x \in \mathbb{R}^d \rightarrow K^{-1/2}(x - \mu)$. \square

2 Spherical Measure and Normal distribution

In this section the sphere is denoted $S^{d-1} = \{x \mid \|x\| = 1\}$ and the unit ball $B^d = \{x \mid \|x\| \leq 1\}$. The set of vectorial isometries on \mathbb{R}^d is $\{\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d \text{ linear} \mid \|\phi(x)\| = \|x\| \forall x \in \mathbb{R}^d\}$. The set of associated matrices is the orthogonal group $O(d) = \{A \in \mathbb{R}^{d \times d} \mid A^T A = I_d\}$.

Goals:

- define a measure ω_d on $(S^{d-1}, \mathcal{B}(S^{d-1}))$ that is invariant to isometries, in order to have a canonical measurable space $(S^{d-1}, \mathcal{B}(S^{d-1}), \omega_d)$ on the sphere.
- introduce the change of variable in polar coordinates

The Lebesgue measure λ_d on \mathbb{R}^d is the unique (up to constants) translation-invariant measure on \mathbb{R}^d . Similarly ω_d is the unique (up to constants) measure on S^{d-1} invariant to isometries.

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

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

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

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

is called the **spherical measure**.

Theorem 4. ω_d is invariant to isometries and for any measurable function $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$,

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

+ integrable setting?

Proposition 5. The volume of the d -dimensional ball $B^d = \{x \mid \|x\| \leq 1\}$ is $\frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}$.

Proof. It is an application of the Fubini theorem. □

Therefore, $\omega_d(S^{d-1}) = d\lambda_d(B^d) = d \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)} = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}$, and we define the **uniform probability distribution on the sphere** as

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

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

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

Proposition 6. The measure σ_d is the unique probability measure on the sphere S^{d-1} invariant to the action of vectorial isometries.

+ Link to the Haar measure

Proposition 7 (Exercise 3.3.7). 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 $h : \mathbb{R} \times S^{d-1} \rightarrow \mathbb{R}$ positive or bounded,

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

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

As a sanity check we can explicitly 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 is¹,

¹without knowing the generalized gamma density function, the normalisation constant can be obtained from the gamma density function by applying the change of variable $\phi(x) = \sqrt{x}$

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

Thus,

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

□

2.1 Gaussian concentration

Applying the concentration result for an isotropic random vector, to $X \sim \mathcal{N}_d(0, I_d)$ we get,

$$\mathbb{P} \left\{ \left| \|X\|_2 - \sqrt{d} \right| \geq t \right\} \leq 2 \exp(-ct^2) \quad \text{for all } t \geq 0 \quad (5)$$

Using the notations of the last section, it says that $R \approx \sqrt{d}$ with high probability. Moreover, $X = RS \approx \sqrt{n}S \sim \text{Unif}(\sqrt{n}S^{d-1})$.

Say more?

3 Sub-Gaussian vectors