Isotropy, Gaussian vector, spherical measure and concentration

Dimitri Meunier

13/04/2021

- Theorem 3.1.1: Concentration of the norm + deviation interpretation
- Definition of two first moments for vectors
- Isotropy and characterisation of isotropy
- Exercise 3.3.1: the spherically distributed random variable is isotropic

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 $\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A))$ (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 a **real** random variable and if $(E, \mathcal{E})$ is $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, X is 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.

+ Lp spaces, and Lebesgue measure

2 Gaussian vectors

Recall univariate density and univariate characteristic function. Recall characteristic function for vector and characterisaion. Independence.

Definition 1. *Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ be a 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 (use the characteristic function). Secondly, if X is a Gaussian vector, for all $B \in \mathbb{R}^{r \times d}$ and $b \in R^r$, $Y = BX + b$ 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 distribusion.

Theorem 1. *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 symmetric matrice $K \in \mathbb{R}^{d\times d}$ such that,*

$$\Phi_X(\theta) = \exp\left(-i\mu \cdot \theta - \frac{1}{2}\theta^t K\theta\right)$$

Furthermore, μ and K are the expectation and covariance of X. + name distribution

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

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

□

Definition 2 (Standard normal random vector)**.** *X is called a* ***standard Gaussian*** *vector on $\mathbb{R}^d$ if its coordinates are i.i.d with distribution $\mathcal{N}(0,1)$. We denote the distribution of X, $\mathcal{N}_d(0, I_d)$. + moments 1 and 2 + they characterise the law.*

Recall that the density function of the univariate standard normal distribution on $\mathbb{R}$ is $f(x) = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}$. Therefore, the density function of $X \sim \mathcal{N}_d(0, I_d)$ is, for all $x \in \mathbb{R}^d$,

$$f(x) = (2\pi)^{-\frac{d}{2}}e^{-\frac{1}{2}\|x\|_2^2}$$

Proposition 1. *If X is a Gaussian vector, its coordinates are independant if and only if the covariance if diagonal.*

Proof. Indeed from the theroem, if K is diagonal the characteristic function can be factorized which is a characterisation of independence. □

Theorem 2. *Let X be a Gaussian vector with mean μ and covariance 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 which is a positive symmetric matrice, hence there exists an orthogonal matrice U and a diagonal matrice D (with nonnegative diagonal elements) such that $K = UDU^T$. Recall that $K^{1/2} := UD^{1/2}U^T$, the definition makes sense since $U^T = U^{-1}$, $K^{1/2}K^{1/2} = K$.
Z is a Gaussian vector and we have seen that any affine transormation of a Gaussian vector is a Gaussian vector therefore $Y := K^{1/2}Z + \mu$ is a Gaussian vector. Since $\mathbb{E}[Y] = K^{1/2}\mathbb{E}[Z] + \mu = \mu$ and $\mathbb{V}[Y] = K^{1/2}\mathbb{V}[Z])K^{1/2} = K$. □

Remark. Isak: define moments of vector and Lp space for vectors (equivalence with each coordinates in standard Lp). Cartesian norm.

Proposition 2. *If X is a Gaussian vector with mean μ and variance K we use the notation $X \sim \mathcal{N}(\mu, K)$. X admits a density if and only if K is invertible. Its density is,*

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

+ Mahanobis distance.

Proof. Apply a change of variable to the density of the standard Gaussian density. □

3 Spherical Measure and Normal distribution

+ define group $O(n)$.

$S^{d-1} = \{x \mid \|x\| = 1\}$. **Goals:**

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

Similarly to the Lebesgue measure on $\mathbb{R}^d$ being the unique (up to constants) translation-invariant measure on $\mathbb{R}^d$, ω_d is the unique (up to constants) measure on S^{d-1} rotation-invariant.

Definition 3. *If $A \in \mathcal{B}\left(S^{d-1}\right)$, we define $\Gamma(A)$ the Borel set of $\mathbb{R}^d$ defined by*

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

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

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

is called the spherical measure.

Theorem 3. *ω_d is invariant to isometries and for any measurable function $f : \mathbb{R}^d \to \mathbb{R}_+$,*

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

Proposition 3. *The volume of the d-dimensional ball $B^d = \{x \mid \|x\| \leq 1\}$ is $\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}$*

Proof. It is an application of Fubini theorem. □

Therefore, $\omega_d(S^{d-1}) = d\lambda_d(B^d) = d\frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} = \frac{2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)}$, and the **uniform probability on the sphere** is,

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

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 change of variable formula leads to,

$$\int_{\mathbb{R}^d} f(x)dx = \omega_d(S^{d-1}) \int_0^\infty g(r)r^{d-1}dr \tag{2}$$

Proposition 4. *The measure σ_d is the unique probaility measure on the sphere S^{d-1} invariant to the action of vectorial isometries.*

+ Link to the Haar measure

Proposition 5 (Exercise 3.3.7 Vershynin: sampling on the unit sphere with a Normal distribution)**.** *Let us write $X \sim N_d\left(0, I_d\right)$ in polar form as*

$$X = r\theta$$

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

1. *the length R and direction S are independent random variables*
2. *the direction S 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 S where $(R, S) = (\|X\|_2, X/\|X\|_2)$ 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,

$$\begin{aligned}
\mathbb{E}[h(R,S)] &= \int_{\mathbb{R}^d} h(x/\|x\|, \|x\|)\rho(x)dx \\
&= \int_{S^{d-1}} \left(\int_0^\infty h(\gamma, r)\rho(r\gamma)r^{d-1}dr \right) d\omega_d(\gamma) \\
&= \int_{S^{d-1}} \left(\int_{\mathbb{R}} h(\gamma, r) \underbrace{\frac{e^{-r^2/2}}{(2\pi)^{d/2}} r^{d-1} 1_{r\geq 0}}_{=:g(\gamma.r)} dr \right) d\omega_d(\gamma)
\end{aligned} \tag{3}$$

g is the density of (R, S), we notice that $g(\gamma, r)$ is separable which implies the independence. Secondly g is constant in γ which implies that S is uniformely distributed on the sphere.

As a sanity check we can explicitely compute the constants (bonus). 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 of a **generalized gamma distribution** $\Gamma(d,\sqrt{2},2)$. Therefore, R follows a $\Gamma(d,\sqrt{2},2)$ distribution and the normalized density function[1] is,

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

Thus,

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

□

3.1 Gaussian concentration

Applying theorem 1 (Isak) to, $X \sim \mathcal{N}_d(0, I_d)$ we get, CONSTANTS (depends on d ??)

$$\mathbb{P}\left\{ \left| \|X\|_2 - \sqrt{d} \right| \geq t \right\} \leq 2\exp\left(-ct^2\right) \quad \text{for all } t \geq 0 \tag{4}$$

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

Say more?

4 Sub-Gaussian vectors

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