CSCI-UA 9473
Additional note on MVN

Augustin Cosse

October 2020

Consider the multivariate Gaussian distribution

$$p(\boldsymbol{x}; \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})) \tag{1}$$

The distribution is encoded by the function `multivariate_normal` from the `random` module in `numpy` which is used to generated the points in Fig 1 below. Note that for the density to make sense, following from the definition of the covariance matrix, $\boldsymbol{\Sigma} = \mathbb{E}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T$, $\boldsymbol{\Sigma}$ must be positive semidefinite.

If we take $\boldsymbol{\Sigma}$ to be rank deficient, quite surprisingly, despite the fact that the density (1) is not defined, the function still returns samples. What `numpy` does is to consider an extension of the Multivariate distribution corresponding to a restriction of the distribution to the subspace spanned by the non zero eigenvectors of the coavariance matrix, and which can be defined from the pseudo inverse and pseudo determinant.

When $\text{rank}(\boldsymbol{\Sigma}) < D$, the inverse and the determinant are not defined. As a consequence, it does not make sense to look for a density such as (1) with respect to the Lebesgue measure on $\mathbb{R}^D$. The density can however be defined on a subspace. If we let $\mathcal{M}(\boldsymbol{\Sigma})$ to denote the linear manifold generated form the columns of $\boldsymbol{\Sigma}$ (i.e. $\mathcal{M}(\boldsymbol{\Sigma}) = \text{span}(\boldsymbol{\Sigma})$), let $\boldsymbol{V}$ denote a set of orthonormal vectors belonging to $\mathcal{M}(\boldsymbol{\Sigma})$ and $\boldsymbol{V}^\perp$ denote the set of orthonormal vectors such that $(\boldsymbol{V}^\perp)^T\boldsymbol{\Sigma} = \boldsymbol{0}$. For simplicity we can take $\boldsymbol{V}$ to contain the non zero eigenvectors of $\boldsymbol{\Sigma}$ and $\boldsymbol{V}_\perp$ to contain the zero eigenvectors of $\boldsymbol{\Sigma}$. Consider the transformation $f : \boldsymbol{x} \mapsto (\boldsymbol{u}, \boldsymbol{z})$ with $\boldsymbol{u} = \boldsymbol{V}^T\boldsymbol{x}$ and $\boldsymbol{z} = \boldsymbol{V}_\perp^T\boldsymbol{x}$. Then

$$\mathbb{E}\{\boldsymbol{z}\} = \boldsymbol{V}_\perp^T\boldsymbol{\mu}, \quad \text{Cov}(\boldsymbol{z}, \boldsymbol{z}) = \boldsymbol{V}_\perp^T\boldsymbol{\Sigma}\boldsymbol{V}_\perp = 0 \tag{2}$$

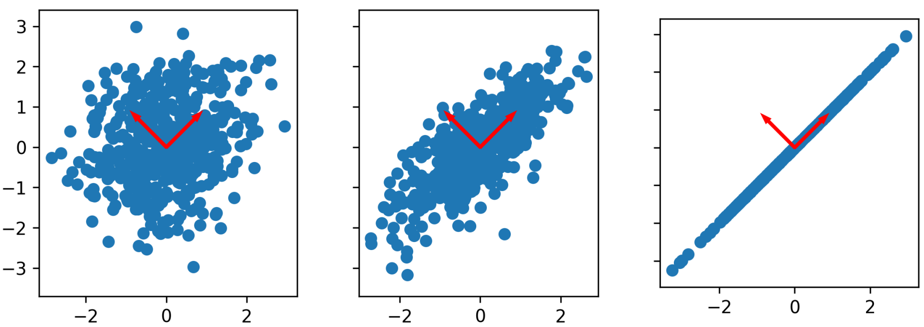

Figure 1: Points generated from the `multivariate_normal` function for mean $\boldsymbol{\mu} = [0, 0]$ and covariance $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

In particular the variable $\boldsymbol{z} = \boldsymbol{V}_\perp^T \boldsymbol{x}$ is thus deterministic and defined by $\boldsymbol{z} = \boldsymbol{V}_\perp^T \boldsymbol{\mu}$ with probability 1. Moreover, we have

$$\mathbb{E}\{\boldsymbol{u}\} = \boldsymbol{V}^T \boldsymbol{\mu}, \quad \mathrm{Cov}(\boldsymbol{u}, \boldsymbol{u}) = \boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V} \tag{3}$$

$\boldsymbol{u}$ is a linear transformation of a Multivariate Gaussian random variable. It therefore has a Gaussian distribution and we can write

$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{V}^T \boldsymbol{\mu}, \boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{B}) \tag{4}$$

Since the columns of $\boldsymbol{V}$ are given by the eigenvectors associated to the non zero eigenvalues of $\boldsymbol{\Sigma}$, the determinant $|\boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V}|$ is equal to the product of the non zero eigenvalues of the covariance matrix $\boldsymbol{\Sigma}$. As an illustration of this, consider a covariance whose eigenvalue decomposition is given by

$$\boldsymbol{\Sigma} = \boldsymbol{U} \begin{bmatrix} \lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{U}^T \tag{5}$$

with $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2]$ with $\boldsymbol{u}_1 \perp \boldsymbol{u}_2$. We then take $\boldsymbol{V} = \boldsymbol{u}_1$ and $\boldsymbol{V}_\perp = \boldsymbol{u}_2$. From this,

$$\boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V} = [1, 0] \begin{bmatrix} \lambda_1 & 0 \\ 0 & 0 \end{bmatrix} [1, 0]^T = \lambda_1 \tag{6}$$

The variable $\boldsymbol{u}$ is thus a rank$(\boldsymbol{\Sigma})$ dimensional Gaussian random variable with associated density

$$p(\boldsymbol{u}) = \frac{1}{(2\pi)^{\mathrm{rank}(\boldsymbol{\Sigma})/2} |\boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V}|^{1/2}} \exp(-\frac{1}{2} (\boldsymbol{u} - \boldsymbol{V}^T \boldsymbol{\mu})^T (\boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V})^{-1} (\boldsymbol{u} - \boldsymbol{V}^T \boldsymbol{\mu})) \tag{7}$$

In short, although we cannot define a density for $\boldsymbol{x}$, we can characterize $\boldsymbol{x}$ by means of its decomposition onto the two subspaces $\boldsymbol{V}$ and $\boldsymbol{V}_\perp$.

Moreover, if we let $\boldsymbol{\Sigma}^+$ to denote the pseudo inverse of $\boldsymbol{\Sigma}$, that is to say the matrix defined by inverting only the non zero eigenvalues of $\boldsymbol{\Sigma}$ as $\sum_{i|\lambda_i > 0} \lambda_i^{-1} \boldsymbol{v}_i \boldsymbol{v}_i^T = \boldsymbol{V} (\boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V})^{-1} \boldsymbol{V}^T$. From this we can write

$$\begin{aligned}
&(\boldsymbol{u} - \boldsymbol{V}^T \boldsymbol{\mu})^T (\boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V})^{-1} (\boldsymbol{u} - \boldsymbol{V}^T \boldsymbol{\mu}) && (8) \\
&= (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{V} (\boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V})^{-1} \boldsymbol{V}^T (\boldsymbol{x} - \boldsymbol{\mu}) && (9) \\
&= (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^+ (\boldsymbol{x} - \boldsymbol{\mu}) && (10)
\end{aligned}$$

And we can rewrite the density of $\boldsymbol{u} = \boldsymbol{V}^T \boldsymbol{x}$ as

$$p(\boldsymbol{u}) = \frac{1}{(2\pi)^{(D-k)/2} \sqrt{\lambda_1 \ldots \lambda_{D-k}}} \exp(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^+ (\boldsymbol{x} - \boldsymbol{\mu})) \tag{11}$$

To summarize, when rank$(\boldsymbol{\Sigma}) < D$, although it does not make sense to define a Normal density for $\boldsymbol{x}$ on the $\mathbb{R}^D$ Lebesgue measure, it remains possible to consider the variable $\tilde{\boldsymbol{x}}$ defined as

$$\tilde{\boldsymbol{x}} = \boldsymbol{V} \boldsymbol{u} + \boldsymbol{V}_\perp \boldsymbol{z} \tag{12}$$

with

- $\boldsymbol{u} = \boldsymbol{V}^T \boldsymbol{x}$, $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{V}^T \boldsymbol{\mu}, \boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V})$
- $\boldsymbol{z} = \boldsymbol{V}_\perp^T \boldsymbol{\mu}$
- $\mathrm{Cov}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}}) = \mathbb{E} \boldsymbol{V} \boldsymbol{u} \boldsymbol{u}^T \boldsymbol{V} + \mathbb{E} \boldsymbol{V} \boldsymbol{u} \boldsymbol{z}^T \boldsymbol{V}_\perp^T + \mathbb{E} \boldsymbol{V}_\perp \boldsymbol{z} \boldsymbol{u}^T \boldsymbol{V}^T = \boldsymbol{\Sigma}$

We can thus define a random variable which has the same covariance and mean as $\boldsymbol{x}$ but whose density is defined only on the subspace $\boldsymbol{V}$ (as it would not make sense to define it on the whole space).

The expression (11) is sometimes written by making use of the pseudo-determinant. The pseudo-determinant is formally defined as

$$|\boldsymbol{\Sigma}|_+ = \lim_{\alpha \to 0} \frac{|\boldsymbol{\Sigma} + \alpha \boldsymbol{I}|}{\alpha^{D - \mathrm{rank}(\boldsymbol{\Sigma})}} \tag{13}$$

from the pseudo-determinant, we can write (11) as

$$p(\boldsymbol{u}) = \frac{1}{(2\pi)^{\mathrm{rank}(\boldsymbol{\Sigma})/2} |\boldsymbol{\Sigma}|_+^{1/2}} \exp(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^+ (\boldsymbol{x} - \boldsymbol{\mu})) \tag{14}$$