Equivalence of Jointly Gaussian Random Vector Definitions

Patrick Yin

June 3, 2020

1 Introduction

In this note, I prove the equivalence of the multiple different definitions of jointly Gaussian random vectors. This is a pretty optional note since I go more in depth than really needed for any practical purpose.

2 Sum of Independent Normal Random Variables

Before we prove for equivalence, we need to first hash out a lemma that proves the that the sum of independent normal random variables is also normal.

Lemma 1 (Sum of Two Standard Independent Normal Random Variables)**.** *The sum of two standard independent normal random variables is still normal. That is, let* $Z = aX_1 + bX_2$ *such that* $X_i \sim N(0, 1)$ *for* $i = 1, 2$*. Then* $Z \sim N(0, a^2 + b^2)$.

Proof. Since X_1 and X_2 are independent,

$$f(x_1, x_2) = f(x_1)f(x_2) = \frac{1}{2\pi}e^{-(x_1^2+x_2^2)/2} \tag{1}$$

Since $f(x_1, x_2)$ is symmetric around the origin, $f(T(x_1, x_2)) = f(x_1, x_2)$ if T is a rotation of the plane $\mathbb{R}^2$ around the origin. Then, for any $t \in \mathbb{R}$ there exists a set $A \subseteq \mathbb{R}^2$ such that

$$\begin{aligned}
\mathbb{P}[Z \leq t] &= \mathbb{P}[aX_1 + bX_2 \leq t] \\
&= \mathbb{P}[(X_1, X_2) \in A] \\
&= \mathbb{P}[(X_1, X_2) \in T(A)] \\
&= \mathbb{P}[X_1 \leq \frac{t}{\sqrt{a^2 + b^2}}] \\
&= \mathbb{P}[\sqrt{a^2 + b^2}X_1 \leq t]
\end{aligned} \tag{2}$$

The fourth line of equation 2 can be derived by applying a transformation that rotates the original line to be parallel to the x_2 axis on the positive side of x_1. In other words, we rotate the line $ax_1 + bx_2 = t$ to be $x = \frac{t}{\sqrt{a^2+b^2}}$.
With equation two, we can conclude that Z has the same distribution as $\sqrt{a^2+b^2}X_1$, so $Z \sim N(0, a^2 + b^2)$. $\square$

Lemma 2 (Sum of Two Independent Normal Random Variables)**.** *The sum of two independent normal random variables is still normal. That is, let* $Z = aX_1 + bX_2$ *such that* $X_i \sim N(\mu_i, \sigma_i^2)$ *for* $i = 1, 2$*. Then* $Z \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$*.*

Proof. $Z_1 = (X_1 - \mu_1)/\sigma_1$ and $Z_2 = (X_2 - \mu_2)/\sigma_2$ are standard independent normal random variables. So,

$$\begin{aligned} Z = aX_1 + bX_2 &= a(\mu_1 + \sigma_1 Z_1) + b(\mu_1 + \sigma_2 Z_2) \\ &= (a\mu_1 + b\mu_2) + (a\sigma_1 Z_1 + b\sigma_2 Z_2) \end{aligned} \tag{3}$$

Lemma 1 implies that $(a\sigma_1 Z_1 + b\sigma_2 Z_2) \sim N(0, a^2\sigma_1^2 + b^2\sigma_2^2)$, so $Z \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$. $\square$

Lemma 3 (Sum of Independent Normal Random Variables)**.** *The sum of independent normal random variables is still normal. That is, let* $Z = a_1X_1 + a_2X_2 + ... + a_nX_n$ *such that* $X_i \sim N(\mu_i, \sigma_i^2)$ *for* $i = 1, ..., n$*. Then* $Z \sim N(a_1\mu_1 + a_2\mu_2 + ... + a_n\mu_n, a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + ... + a_n^2\sigma_n^2)$*.*

Proof. We will prove this using induction.
Base Cases: $n = 1$ is trivially true. Lemma 2 shows $n = 2$ is true.
Inductive Hypothesis: For some $k > 2$, let $Z' = a_1X_1 + a_2X_2 + ... + a_kX_k$ such that $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, ..., k$. Then $Z' \sim N(a_1\mu_1 + a_2\mu_2 + ... + a_k\mu_k, a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + ... + a_k^2\sigma_k^2)$.
Inductive Step: We have to show that if $Z = a_1X_1 + a_2X_2 + ... + a_{k+1}X_{k+1}$ such that $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, ..., k+1$, then $Z \sim N(a_1\mu_1 + a_2\mu_2 + ... + a_{k+1}\mu_{k+1}, a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + ... + a_{k+1}^2\sigma_{k+1}^2)$. We know that $Z = Z' + a_{k+1}X_{k+1}$. Since both Z' and $a_{k+1}X_{k+1}$ are normally distributed and independent, $Z \sim N(a_1\mu_1 + a_2\mu_2 + ... + a_{k+1}\mu_{k+1}$ by lemma 2 (i.e. the $n = 2$ case). $\square$

3 Jointly Gaussian Random Vector Definitions

There are many equivalent definitions for a jointly Gaussian (JG) random vector:

1. A random vector $Z = (Z_1, ..., Z_k)^T$ is JG if there exists a base random vector $U = (U_1, ..., U_l)$ whose components are independent standard normal random variables, a transition matrix $R \in \mathbb{R}^{k \times l}$, and a mean vector $\mu \in \mathbb{R}^k$, such that $Z = RU + \mu$.

2. A random vector $Z = (Z_1, ..., Z_k)^T$ is JG if $\sum_{i=1}^{k} a_i Z_i$ is normally distributed for every $a_i \in \mathbb{R}$.

3. (Only if Σ is full rank a.k.a non-degenerate case) A random vector $Z = (Z_1, ..., Z_k)^T$ is JG if

$$f_Z(\vec{z}) = \frac{1}{\sqrt{|\Sigma|}} \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(\vec{z}-\mu)^T \Sigma^{-1} (\vec{z}-\mu)} \tag{4}$$

where Σ is the covariance matrix of Z. In other words, $\Sigma = \mathbb{E}[(Z-\mu)(Z-\mu)^T] = \mathbb{E}[(RU)(RU)^T] = R\,\mathbb{E}[UU^T]R^T = RR^T$

4 Proofs of Equivalence

4.1 Proving (1) $\implies$ (2)

We have to prove that $\sum_{i=1}^k a_i Z_i$ is normally distributed given that $Z = RU + \mu$. Let us define $R = (\vec{r}_1, ..., \vec{r}_k)^T$ such that $\vec{r}_i \in \mathbb{R}^l$. Then,

$$\sum_{i=1}^k a_i Z_i = \sum_{i=1}^k a_i(\vec{r}_i^T U + \mu_i) = \sum_{i=1}^k (a_i \vec{r}_i^T U + a_i \mu_i) \tag{5}$$

By lemma 3, $\sum_{i=1}^k a_i Z_i$ must be normally distributed because it equals the sum of multiple standard independent normal random variables (i.e. components of U).

Remark. *Note that we can construct cases were the summation turns out to be a constant. However, we still consider these point means to be normal with variance 0.*

4.2 Proving (2) $\implies$ (1)

Lemma 4. *$Cov(AX) = ACov(X)A^T$ where X is an n-dimensional random vector and $A \in \mathbb{R}^{n\times n}$.*

Proof.

$$\begin{aligned} Cov(AX) &= \mathbb{E}[(AX - \mathbb{E}[AX])(AX - \mathbb{E}[AX])^T] \\ &= \mathbb{E}[(AX - A\mathbb{E}[X])(AX - A\mathbb{E}[X])^T] \\ &= \mathbb{E}[A(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T A^T] \\ &= A\mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]A^T \\ &= ACov(X)A^T \end{aligned} \tag{6}$$

□

With lemma 4, we can now show that if $\sum_{i=1}^k a_i Z_i$ is normally distributed, then $Z = RU + \mu$.
Define $\mu = \mathbb{E}[Z]$ and $\Sigma = Cov(Z) = \mathbb{E}[(Z-\mu)(Z-\mu)^T]$. Since Σ is symmetric and positive semidefinite, we can write $\Sigma = RR^T$ where R is the same dimensionality as Σ. Now consider the new random variable $U = R^{-1}(Z-\mu)$.

We know that U is normal since we have assumed that $\sum_{i=1}^{k} a_i Z_i$ is normally distributed (notice that U is just a linear combination of the components of Z). We also know that $\mathbb{E}[U] = \mathbb{E}[R^{-1}(Z - \mu)] = R^{-1}\mathbb{E}[Z - \mu] = R^{-1}(0) = 0$ and $Cov[U] = Cov[R^{-1}(Z - \mu)] = R^{-1}Cov(Z - \mu)(R^{-1})^T = R^{-1}RR^T(R^{-1})^T = I$. In other words, we know that U consists of independent standard normal random variables such that $Z = RU + \mu$.

Remark. *Notice that within the proof, we have assumed that R is invertible since we took R^{-1}. However, R does not have to be invertible. If R is not invertible, that simply means that Z only needs to be written as a linear combination of a subset of all the standard normal components in U (i.e. If R is not invertible, $\Sigma = RR^T$ is not invertible which indicates that one of the random variables Z_i can be written as a linear combination of the others).*
So, when R is not invertible, instead of using R^{-1} for our proof, we take the pseudoinverse of R with singular value decomposition (SVD) instead. That is, by SVD, $R = U\Sigma_R V^T$ where U and V are orthonormal and Σ_R is a diagonal matrix with non-negative numbers on its diagonal. We define the pseudoinverse to be $R^+ = V\Sigma_R^{-1}U^T$ where the zeros on the diagonal of Σ_R^{-1} remain as zeros. Thus, we can show that $R^+R = RR^+ = \begin{bmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}$ where r is the rank of R by simply multiplying their decompositions together.
With this, we can see that $\mathbb{E}[U] = \mathbb{E}[R^+(Z-\mu)] = 0$ and $Cov[U] = Cov[R^+(Z-\mu)] = R^+RR^T(R^+)^T = \begin{bmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}$. We reach the same conclusion.

4.3 Proving (1) $\implies$ (3)

First, we need to show that $f_U(\vec{u}) = |R|f_Z(\vec{z})$ where $\vec{z} = R\vec{u} + \mu$.
Let R' represent an arbitrary region in the U space. Then,

$$\mathbb{P}[U \subseteq R'] = \int ... \int \int_{R'} f_U(\vec{u})d\vec{u} \tag{7}$$

Let us the transformation T as transforming a vector $\vec{u} \in U$ to $R\vec{u} + \vec{\mu} \in Z$. Then,

$$\begin{aligned} \mathbb{P}[Z \subseteq T(R')] &= \int ... \int \int_{T(R')} f_Z(\vec{z})d\vec{z} \\ &= \int ... \int \int_{R'} f_Z(\vec{z})|\frac{\partial \vec{v}}{\partial \vec{u}}|d\vec{u} \\ &= \int ... \int \int_{R'} f_Z(\vec{z})|R|d\vec{z} \end{aligned} \tag{8}$$

Line 3 of equation 8 is true because the jacobian matrix $\frac{\partial \vec{v}}{\partial \vec{u}} = \frac{\partial (R\vec{u}+\vec{\mu})}{\partial \vec{u}} = R$. Since $\mathbb{P}[U \subseteq R'] = \mathbb{P}[Z \subseteq T(R')]$, $f_U(\vec{u}) = |R| f_Z(\vec{z})$. Then,

$$
\begin{aligned}
f_Z(\vec{z}) &= \frac{1}{|R|} f_U(R^{-1}(\vec{z}-\vec{\mu})) \\
&= \frac{1}{|R|} \prod_{i=1}^{k} f_{U_i}((R^{-1}(\vec{z}-\vec{\mu}))_i) \\
&= \frac{1}{|R|} \prod_{i=1}^{k} (\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(R^{-1}(\vec{z}-\vec{\mu}))_i^2}) \\
&= \frac{1}{|R|} \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(R^{-1}(\vec{z}-\vec{\mu}))^T (R^{-1}(\vec{z}-\vec{\mu}))} \\
&= \frac{1}{|R|} \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(\vec{z}-\vec{\mu})^T (RR^T)^{-1}(\vec{z}-\vec{\mu}))}
\end{aligned}
\tag{9}
$$

$RR^T = \Sigma$ and $|\Sigma| = |RR^T| = |R|^2$, so

$$
f_Z(\vec{z}) = \frac{1}{\sqrt{|\Sigma|}} \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}(\vec{z}-\vec{\mu})^T \Sigma^{-1}(\vec{z}-\vec{\mu}))} \tag{10}
$$

4.4 Proving (3) $\implies$ (1)

For the proof above, the whole process is reversible. So we can just do what we did in 4.3 but in reverse to proof this implication.

5 Misconceptions

5.1 Marginal Gaussians $\not\Longrightarrow$ JG Vector

Looking at the three definitions for JG random vectors, we might assume that a random vector $Z = (Z_1, ..., Z_k)^T$ is JG if Z_i is normally distributed for all $i = 1, ..., k$. However, this is not true. While JG random vectors have marginal gaussians (i.e. if $\sum_{i=1}^{k} a_i Z_i$ is normally distributed for every $a_i \in \mathbb{R}$, simply plug in combinations of $(a_1, a_2, ..., a_k)$ such that one of the values is one and the rest are zero to see that each individual Z_i is normally distributed), marginal gaussians don't necessarily form a JG random vector (i.e. the converse is not true).

Proof. We will construct a counterexample where a two normal random variables aren't JG when put together in a vector. Consider $X \sim N(0, 1)$ and $Y = X(2B - 1)$ where $B \sim \text{Bernoulli}(1/2)$ (i.e. $Y = \pm X$ with probability 1/2).

We already know that X is normal, but we need to show that Y is normal as well. We know that

$$
\mathbb{P}(Y \leq y) = \frac{1}{2}(\mathbb{P}(Y \leq y | B = 1) + \mathbb{P}(Y \leq y | B = 0)) \tag{11}
$$

We can now break down the two terms on the right:

$$\mathbb{P}(Y \le y | B = 1) = \mathbb{P}(X \le y) = \Phi(y) \tag{12}$$

$$\begin{aligned} \mathbb{P}(Y \le y | B = 0) &= \mathbb{P}(-X \le y) \\ &= \mathbb{P}(X \ge -y) \\ &= 1 - \mathbb{P}(X < -y) \\ &= 1 - \Phi(-y) \\ &= \Phi(y) \end{aligned} \tag{13}$$

Now we can solve equation 11:

$$\mathbb{P}(Y \le y) = \frac{1}{2}(\Phi(y) + \Phi(y)) = \Phi(y) \tag{14}$$

Thus $Y \sim N(0, 1)$. We know that X and Y are both normally distributed. But now, we will show that they are not JG. We will look at the definition 2 of JG, where every linear combination of a JG vector's components must be normal. This case doesn't follow definition 2 since

$$X + Y = \begin{cases} 2X & \text{if } B = 1 \\ 0 & \text{if } B = 0 \end{cases} \tag{15}$$

In other words, $X + Y$ is a 50/50 mixture model of $N(0, 4)$ and a point mass at 0, which is not normal. ☐

Remark. *$X + Y$ is not normal because $\mathbb{P}[X + Y < 0] = \frac{1}{2}\mathbb{P}[2X < 0] = \frac{1}{2}\mathbb{P}[X < 0] = \Phi(0)/2 = 1/4$. This is impossible for a normal distribution since for any $Z \sim N(0, \sigma^2)$, $\mathbb{P}[Z < 0] = 1/2$.*

Remark. *Note that we had to construct a really artificial example to in order to disprove this implication. This gives us a hint that most normal distributions are probably jointly normal.*

5.2 Uncorrelated Normal Variables $\not\Rightarrow$ Independence

Uncorrelated normal variables are only guaranteed to be independent if they are JG. Note that we can prove this by looking at the JG distribution (for simplicity, let us look at only two uncorrelated normal variables):

$$\begin{aligned} f_Z \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{vmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{vmatrix}^{-\frac{1}{2}} \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}\begin{pmatrix} x - \mu_x & y - \mu_y \end{pmatrix} \begin{bmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{bmatrix}^{-1} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix}} \\ &= \left|\Sigma_{XX}\right|^{-\frac{1}{2}} \frac{1}{(2\pi)^{k_x/2}} e^{-\frac{1}{2}(x-\mu_x)^T \Sigma_{XX}^{-1}(x-\mu_x)} \\ &\quad \left|\Sigma_{YY}\right|^{-\frac{1}{2}} \frac{1}{(2\pi)^{k_y/2}} e^{-\frac{1}{2}(y-\mu_y)^T \Sigma_{YY}^{-1}(y-\mu_y)} \\ &= f_X(x) f_Y(y) \end{aligned} \tag{16}$$

However, if uncorrelated normal variables are not JG, they aren't guaranteed to be independent. We first need to prove a fundamental lemma with expected value before we go into an explanation for why this is true:

Lemma 5 (Expected Value of Two Variables)**.** *Let X and Y be random variables. Then,*

$$\mathbb{E}[g(X,Y)] = \mathbb{E}_X[\mathbb{E}_Y[g(X,Y)|X]] = \mathbb{E}_Y[\mathbb{E}_X[g(X,Y)|Y]] \tag{17}$$

Proof. We can prove this by breaking expected value into integral form. Let $f_{X,Y}(x,y)$ be the joint distribution of X and Y.

$$\begin{aligned}
\mathbb{E}[g(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(y)\frac{f(x,y)}{f(y)}dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(y)f(x|y)dxdy \\
&= \int_{-\infty}^{\infty} f(y)\int_{-\infty}^{\infty} g(x,y)f(x|y)dxdy \\
&= \int_{-\infty}^{\infty} f(y)\mathbb{E}_X[g(X,Y)|Y]dy \\
&= \mathbb{E}_Y[\mathbb{E}_X[g(X,Y)|Y]]
\end{aligned} \tag{18}$$

We can prove for the other form by using $\frac{f(x,y)}{f(x)}$ instead of $\frac{f(x,y)}{f(y)}$ in line 2. $\square$

Now that we have lemma 5, a good counterexample to why two normal variables being uncorrelated doesn't guarantee independence is the example we showed in 5.1 with $X \sim N(0,1)$ and $Y = X(2B-1)$ where $B \sim \text{Bernoulli}(1/2)$. Here,

$$\begin{aligned}
\text{Cov}(X,Y) &= \text{Cov}(X,(2B-1)X) \\
&= \mathbb{E}[(2B-1)X^2] - \mathbb{E}[X]\mathbb{E}[(2B-1)X] \\
&= \mathbb{E}[(2B-1)X^2] - (0)\mathbb{E}[(2B-1)X] \\
&= \mathbb{E}[(2B-1)X^2] \\
&= \mathbb{E}_{(2B-1)}[\mathbb{E}_X[(2B-1)X^2|(2B-1)]] \\
&= \mathbb{E}_{(2B-1)}[2B-1] \\
&= 0
\end{aligned} \tag{19}$$

Here, X and Y are normal and uncorrelated. Yet, by definition they are obviously dependent.