Gaussian Random Vectors

1. The multivariate normal distribution

Let $\boldsymbol{X} := (X_1, \dots, X_n)'$ be a random vector. We say that $\boldsymbol{X}$ is a *Gaussian random vector* if we can write

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z},$$

where $\boldsymbol{\mu} \in \mathbf{R}^n$, $\boldsymbol{A}$ is an $n \times k$ matrix and $\boldsymbol{Z} := (Z_1, \dots, Z_k)'$ is a k-vector of i.i.d. standard normal random variables.

Proposition 1. *Let X be a Gaussian random vector, as above. Then,*

$$\mathrm{E}X = \boldsymbol{\mu}, \ \mathrm{Var}(X) := \boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}', \ \textit{and } M_X(\boldsymbol{t}) = \mathrm{e}^{\boldsymbol{t}'\boldsymbol{\mu}+\frac{1}{2}\|\boldsymbol{A}'\boldsymbol{t}\|^2} = \mathrm{e}^{\boldsymbol{t}'\boldsymbol{\mu}+\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}},$$

for all $\boldsymbol{t} \in \mathbf{R}^n$.

Thanks to the uniqueness theorem of MGF's it follows that the distribution of $\boldsymbol{X}$ is determined by $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and the fact that it is multivariate normal. From now on, we sometimes write $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, when we mean that $M_X(\boldsymbol{t}) = \exp(\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t})$. Interesetingly enough, the choice of $\boldsymbol{A}$ and $\boldsymbol{Z}$ are typically not unique; only $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ influences the distribution of X.

Proof. The expectation of $\boldsymbol{X}$ is $\boldsymbol{\mu}$, since $\mathrm{E}(\boldsymbol{A}\boldsymbol{Z}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{Z}) = \boldsymbol{0}$. Also,

$$\mathrm{E}(\boldsymbol{X}\boldsymbol{X}') = \mathrm{E}\left([\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z}][\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z}]'\right) = \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{A}\mathrm{E}(\boldsymbol{Z}\boldsymbol{Z}')\boldsymbol{A}'.$$

Since $\mathrm{E}(\boldsymbol{Z}\boldsymbol{Z}') = \boldsymbol{I}$, the variance-covariance of $\boldsymbol{X}$ is $\mathrm{E}(\boldsymbol{X}\boldsymbol{X}') - (\mathrm{E}\boldsymbol{X})(\mathrm{E}\boldsymbol{X})' = \mathrm{E}(\boldsymbol{X}\boldsymbol{X}') - \boldsymbol{\mu}\boldsymbol{\mu}' = \boldsymbol{A}\boldsymbol{A}'$, as desired. Finally, note that $M_X(\boldsymbol{t}) = \exp(\boldsymbol{t}'\boldsymbol{\mu}) \cdot$

$M_{\boldsymbol{Z}}(\boldsymbol{A}'\boldsymbol{t})$. This establishes the result on the MGF of $\boldsymbol{X}$, since $M_{\boldsymbol{Z}}(\boldsymbol{s}) = \prod_{j=1}^{n} \exp(s_j^2/2) = \exp(\frac{1}{2}\|\boldsymbol{s}\|^2)$ for all $\boldsymbol{s} \in \mathbf{R}^n$. □

We say that $\boldsymbol{X}$ has the *multivariate normal distribution* with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} := \boldsymbol{A}\boldsymbol{A}'$, and write this as $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{\mu}\,, \boldsymbol{A}\boldsymbol{A}')$.

Theorem 2. *$\boldsymbol{X} := (X_1\,, \ldots, X_n)'$ has a multivariate normal distribution if and only if $\boldsymbol{t}'\boldsymbol{X} = \sum_{i=1}^{n} t_i X_i$ has a normal distribution on the line for every $\boldsymbol{t} \in \mathbf{R}^n$. That is, $X_1, \ldots, X_n$ are jointly normally distributed if and only if all of their linear combinations are normally distributed.*

Note that the distribution of $\boldsymbol{X}$ depends on $\boldsymbol{A}$ only through the positive semidefinite $n \times n$ matrix $\boldsymbol{\Sigma} := \boldsymbol{A}\boldsymbol{A}'$. Sometimes we say also that $X_1, \ldots, X_n$ are *jointly* normal [or Gaussian] when $\boldsymbol{X} := (X_1\,, \ldots, X_n)'$ has a multivariate normal distribution.

Proof. If $\boldsymbol{X} \in \mathrm{N}_n(\boldsymbol{\mu}\,, \boldsymbol{A}\boldsymbol{A}')$ then we can write it as $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z}$, we as before. In that case, $\boldsymbol{t}'\boldsymbol{X} = \boldsymbol{t}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{A}\boldsymbol{Z}$ is a linear combination of $Z_1, \ldots, Z_k$, whence has a normal distribution with mean $t_1\mu_1 + \cdots + t_n\mu_n$ and variance $\boldsymbol{t}'\boldsymbol{A}\boldsymbol{A}'\boldsymbol{t} = \|\boldsymbol{A}'\boldsymbol{t}\|^2$.

For the converse, suppose that $\boldsymbol{t}'\boldsymbol{X}$ has a normal distribution for every $\boldsymbol{t} \in \mathbf{R}^n$. Let $\boldsymbol{\mu} := \mathrm{E}\boldsymbol{X}$ and $\boldsymbol{\Sigma} := \mathrm{Var}(\boldsymbol{X})$, and observe that $\boldsymbol{t}'\boldsymbol{X}$ has mean vector $\boldsymbol{t}'\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}$. Therefore, the MGF of the univariate normal $\boldsymbol{t}'\boldsymbol{X}$ is $M_{\boldsymbol{t}'\boldsymbol{X}}(s) = \exp(s\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}s^2\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t})$ for all $s \in \mathbf{R}$. Note that $M_{\boldsymbol{t}'\boldsymbol{X}}(s) = \mathrm{E}\exp(s\boldsymbol{t}'\boldsymbol{X})$. Therefore, apply this with $s := 1$ to see that $M_{\boldsymbol{t}'\boldsymbol{X}}(1) = M_{\boldsymbol{X}}(\boldsymbol{t})$ is the MGF of a multivariate normal. The uniqueness theorem for MGF's (Theorem 1, p. 27) implies the result. □

2. The nondegenerate case

Suppose $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{\mu}\,, \boldsymbol{\Sigma})$, and recall that $\boldsymbol{\Sigma}$ is always positive semidefinite. We say that $\boldsymbol{X}$ is *nondegenerate* when Σ is positive definite (equivalently, invertible).

Take, in particular, $X \sim \mathrm{N}_1(\mu\,, \Sigma)$; μ can be any real number and Σ is a positive semidefinite 1×1 matrix; i.e., $\Sigma \geq 0$. The distribution of X is defined via its MGF as

$$M_X(t) = \mathrm{e}^{t\mu + \frac{1}{2}t^2\Sigma}.$$

When X is nondegenerate ($\Sigma > 0$), $X \sim \mathrm{N}(\mu\,, \Sigma)$. If $\Sigma = 0$, then $M_X(t) = \exp(t\mu)$; therefore by the uniqueness theorem of MGFs, $\mathrm{P}\{X = \mu\} = 1$. Therefore, $\mathrm{N}_1(\mu\,, \sigma^2)$ is the generalization of $\mathrm{N}(\mu\,, \sigma^2)$ in order to include the case that $\sigma = 0$. We will not write $\mathrm{N}_1(\mu\,, \sigma^2)$; instead we always write $\mathrm{N}(\mu\,, \sigma^2)$ as no confusion should arise.

Theorem 3. *$X \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ has a probability density function if and only if it is nondegenerate. In that case, the pdf of X is*

$$f_X(\boldsymbol{a}) = \frac{1}{(2\pi)^{n/2}(\det \boldsymbol{\Sigma})^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{a}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{a}-\boldsymbol{\mu})\right)$$

for all $\boldsymbol{a} \in \mathbf{R}^n$.

Proof. First of all let us consider the case that X is degenerate. In that case $\boldsymbol{\Sigma}$ has some number $k < n$ of strictly-positive eigenvalues. The proof of Theorem 2 tells us that we can write $X = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\mu}$, where $\boldsymbol{Z}$ is a k-dimensional vector of i.i.d. standard normals and $\boldsymbol{A}$ is an $n \times k$ matrix. Consider the k-dimensional space

$$\boldsymbol{E} := \left\{\boldsymbol{x} \in \mathbf{R}^n : \ \boldsymbol{x} = \boldsymbol{A}\boldsymbol{z} + \boldsymbol{\mu} \text{ for some } \boldsymbol{z} \in \mathbf{R}^k\right\}.$$

Because $\mathrm{P}\{\boldsymbol{Z} \in \mathbf{R}^k\} = 1$, it follows that $\mathrm{P}\{X \in \boldsymbol{E}\} = 1$. If X had a pdf f_X, then

$$1 = \mathrm{P}\{X \in \boldsymbol{E}\} = \int_{\boldsymbol{E}} f_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

But the n-dimensional volume of $\boldsymbol{E}$ is zero since the dimension of $\boldsymbol{E}$ is $k < n$. This creates a contradiction [unless X did not have a pdf, that is].

If X is nondegenerate, then we can write $X = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\mu}$, where $\boldsymbol{Z}$ is an n-vector of i.i.d. standard normals and $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}'$ is invertible; see the proof of Theorem 2. Recall that the choice of $\boldsymbol{A}$ is not unique; in this case, we can always choose $\boldsymbol{A} := \boldsymbol{\Sigma}^{1/2}$ because $\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z} + \boldsymbol{\mu} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In other words,

$$X_i = \sum_{j=1}^n A_{i,j}Z_j + \mu_i = \sum_{j=1}^n \Sigma_{i,j}^{1/2} Z_j + \mu_i := g_i(Z_1, \ldots, Z_n) \qquad (1 \le i \le n).$$

If $\boldsymbol{a} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{z} + \boldsymbol{\mu}$, then $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{a} - \boldsymbol{\mu})$. Therefore, the change of variables formula of elementary probability implies that

$$f_X(\boldsymbol{a}) = \frac{f_Z\left(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{a}-\boldsymbol{\mu})\right)}{|\det J|},$$

as long as $\det J \neq 0$, where

$$J := \begin{pmatrix} \frac{\partial g_1}{\partial z_1} & \cdots & \frac{\partial g_1}{\partial z_n} \\ \vdots & & \vdots \\ \frac{\partial g_n}{\partial z_1} & \cdots & \frac{\partial g_n}{\partial z_n} \end{pmatrix} = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \cdots & A_{n,n} \end{pmatrix} = \boldsymbol{A}.$$

Because $\det(\boldsymbol{\Sigma}) = \det(\boldsymbol{A}\boldsymbol{A}') = (\det \boldsymbol{A})^2$, it follows that $\det \boldsymbol{A} = (\det \boldsymbol{\Sigma})^{1/2}$, and hence

$$f_X(\boldsymbol{a}) = \frac{1}{(\det \boldsymbol{\Sigma})^{1/2}} f_Z\left(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{a}-\boldsymbol{\mu})\right).$$

Because of the independence of the Z_j's,

$$f_Z(\boldsymbol{z}) = \prod_{j=1}^n \frac{e^{-z_j^2/2}}{\sqrt{2\pi}} = \frac{1}{(2\pi)^{n/2}} e^{-\boldsymbol{z}'\boldsymbol{z}/2}$$

for all $\boldsymbol{z} \in \mathbf{R}^n$. Therefore,

$$f_Z\left(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{a}-\boldsymbol{\mu})\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(\boldsymbol{a}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{a}-\boldsymbol{\mu})\right),$$

and the result follows. □

3. The bivariate normal distribution

A *bivariate normal distribution* has the form $\mathrm{N}_2(\boldsymbol{\mu}\,,\boldsymbol{\Sigma})$, where $\mu_1 = \mathrm{E}X_1$, $\mu_2 = \mathrm{E}X_2$, $\Sigma_{1,1} = \mathrm{Var}(X_1) := \sigma_1^2 > 0$, $\Sigma_{2,2} = \mathrm{Var}(X_2) := \sigma_2^2 > 0$, and $\Sigma_{1,2} = \Sigma_{2,1} = \mathrm{Cov}(X_1\,,X_2)$. Let

$$\rho := \mathrm{Corr}(X_1\,,X_2) := \frac{\mathrm{Cov}(X_1\,,X_2)}{\sqrt{\mathrm{Var}(X_1)\cdot\mathrm{Var}(X_2)}}$$

denote the correlation between X_1 and X_2, and recall that $-1 \le \rho \le 1$. Then, $\Sigma_{1,2} = \Sigma_{2,1} = \rho\sigma_1\sigma_2$, whence

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Since $\det\boldsymbol{\Sigma} = \sigma_1^2\sigma_2^2(1-\rho^2)$, it follows immediately that our bivariate normal distribution is non-degenerate if and only if $-1 < \rho < 1$, in which case

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \dfrac{1}{\sigma_1^2(1-\rho^2)} & -\dfrac{\rho}{1-\rho^2}\cdot\dfrac{1}{\sigma_1\sigma_2} \\ -\dfrac{\rho}{1-\rho^2}\cdot\dfrac{1}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2(1-\rho^2)} \end{pmatrix}.$$

Because

$$\boldsymbol{z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z} = \left(\frac{z_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{z_1}{\sigma_1}\right)\left(\frac{z_2}{\sigma_2}\right) + \left(\frac{z_2}{\sigma_2}\right)^2$$

for all $\boldsymbol{z} \in \mathbf{R}^n$, the pdf of $\boldsymbol{X} = (X_1\,,X_2)'$—in the non-degenerate case where there *is* a pdf—is

$$f_X(x_1\,,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right).$$

But of course non-degenerate cases are also possible. For instance, suppose $Z \sim \mathrm{N}(0\,,1)$ and define $X := (Z\,,-Z)$. Then $X = \boldsymbol{A}Z$ where $\boldsymbol{A} := (1\,,-1)'$, whence

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}' = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

is singular. In general, if $X \sim \mathrm{N}_n(\boldsymbol{\mu}\,,\boldsymbol{\Sigma})$ and the rank of $\boldsymbol{\Sigma}$ is $k < n$, then X depends only on k [and not n] i.i.d. $\mathrm{N}(0\,,1)$'s. This can be gleaned from the proof of Theorem 2.

4. A few important properties of multivariate normal distributions

Proposition 4. *Let $X \sim \mathrm{N}_n(\boldsymbol{\mu}\,,\boldsymbol{\Sigma})$. If $\boldsymbol{C}$ is an $m \times n$ matrix and $\boldsymbol{d}$ is an m-vector, then $\boldsymbol{C}X + \boldsymbol{d} \sim \mathrm{N}_m(\boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{d}\,,\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')$. In general, $\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'$ is positive semidefinite; it is positive definite if and only if it has full rank m.*

In particular, if $\boldsymbol{a}$ is a nonrandom n-vector, then $\boldsymbol{a}'X \sim \mathrm{N}(\boldsymbol{a}'\boldsymbol{\mu}\,,\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a})$.

Proof. We compute the MGF of $\boldsymbol{C}X + \boldsymbol{d}$ as follows:

$$M_{\boldsymbol{C}X+\boldsymbol{d}}(\boldsymbol{t}) = \mathrm{E}\exp\left(\boldsymbol{t}'\left[\boldsymbol{C}X + \boldsymbol{d}\right]\right) = \mathrm{e}^{\boldsymbol{t}'\boldsymbol{d}} M_X(\boldsymbol{s}),$$

where $\boldsymbol{s} := \boldsymbol{C}'\boldsymbol{t}$. Therefore,

$$M_{\boldsymbol{C}X+\boldsymbol{d}}(\boldsymbol{t}) = \exp\left(\boldsymbol{t}'\boldsymbol{d} + \boldsymbol{s}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{s}'\boldsymbol{\Sigma}\boldsymbol{s}\right) = \exp\left(\boldsymbol{t}'\boldsymbol{\nu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{Q}\boldsymbol{t}\right),$$

where $\boldsymbol{\nu} := \boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{d}$ and $\boldsymbol{Q} := \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'$. Finally, a general fact about symmetric matrices (Corollary 13, p. 17) implies that the symmetric $m \times m$ matrix $\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'$ is nonsingular if and only if it has full rank m. □

Proposition 5. *If $X \in \mathrm{N}_n(\boldsymbol{\mu}\,,\boldsymbol{\Sigma})$, for a nonsingular variance-covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{C}_{m\times n}$ and $\boldsymbol{d}_{n\times 1}$ are nonrandom, then $\boldsymbol{C}X + \boldsymbol{d}$ is nonsingular if and only if $\mathrm{rank}(\boldsymbol{C}) = m$.*

Proof. Recall that the nonsingularity of $\boldsymbol{\Sigma}$ is equivalent to it being positive definite. Now $\boldsymbol{C}X + \boldsymbol{d}$ is multivariate normal by the preceding result. It is nondegenerate if and only if $\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'$ is positive definite. But $\boldsymbol{x}'\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'\boldsymbol{x} = (\boldsymbol{C}'\boldsymbol{x})'\boldsymbol{\Sigma}(\boldsymbol{C}'\boldsymbol{x}) > 0$ if and only if $(\boldsymbol{C}'\boldsymbol{x}) \neq \boldsymbol{0}$, since $\boldsymbol{\Sigma}$ is positive definite. Therefore, $\boldsymbol{C}X + \boldsymbol{d}$ is nondegenerate if and only if $\boldsymbol{C}'\boldsymbol{x} \neq \boldsymbol{0}$ whenever $\boldsymbol{x} \neq \boldsymbol{0}$. This is equivalent to $\boldsymbol{x}'\boldsymbol{C} \neq \boldsymbol{0}$ for all nonzero vectors $\boldsymbol{x}$; that is, $\boldsymbol{C}$ has row rank—hence rank—m. □

The following is an easy corollary of the previous proposition, and identifies the "standard multivariate normal" distribution as the distribution of i.i.d. standard univariate normal distributions. It also states that we do not change the distribution of a standard multivariate normal if we apply to it an orthogonal matrix.

Corollary 6. *$\boldsymbol{Z} \sim \mathrm{N}_n(\boldsymbol{0}, \boldsymbol{I})$ if and only if $Z_1, \ldots, Z_n$ are i.i.d. $\mathrm{N}(0, 1)$'s. Moreover, if $\boldsymbol{Z} \sim \mathrm{N}_n(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{A}_{n\times n}$ is orthogonal then $\boldsymbol{AZ} \sim \mathrm{N}_n(\boldsymbol{0}, \boldsymbol{I})$ also.*

Next we state another elementary fact, derived by looking only at the MGF's. It states that a subset of a multivariate normal vector itself is multivariate normal.

Proposition 7. *Suppose $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $1 \le i_1 < i_2 < \cdots < i_k \le n$ is a subsequence of $1, \ldots, n$. Then, $(X_{i_1}, \ldots, X_{i_k})' \sim \mathrm{N}_k(\boldsymbol{\nu}, \boldsymbol{Q})$, where*

$$\boldsymbol{\nu} := \mathrm{E}\begin{pmatrix} X_{i_1} \\ \vdots \\ X_{i_k} \end{pmatrix} = \begin{pmatrix} \mu_{i_1} \\ \vdots \\ \mu_{i_k} \end{pmatrix}, \qquad \boldsymbol{Q} := \mathrm{Var}\begin{pmatrix} X_{i_1} \\ \vdots \\ X_{i_k} \end{pmatrix} = \begin{pmatrix} \Sigma_{i_1,i_1} & \cdots & \Sigma_{i_1,i_k} \\ \vdots & & \vdots \\ \Sigma_{i_k,i_1} & \cdots & \Sigma_{i_k,i_k} \end{pmatrix}.$$

Proposition 8. *Suppose $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and assume that we can divide the X_i's into two groups: $(X_i)_{i\in G}$ and $(X_j)_{j\notin G}$, where G is a subset of the index set $\{1, \ldots, n\}$. Suppose in addition that $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \in G$ and $j \notin G$. Then, $(X_i)_{i\in G}$ is independent from $(X_j)_{j\notin G}$.*

Thus, for example, if $(X_1, X_2, X_3)'$ has a trivariate normal distribution and X_1 is uncorrelated from X_2 and X_3, then X_1 is independent of (X_2, X_3). For a second example suppose that (X_1, X_2, X_3, X_4) has a multivariate normal distribution and: $\mathrm{E}(X_1X_2) = \mathrm{E}(X_1)\mathrm{E}(X_2)$, $\mathrm{E}(X_1X_3) = \mathrm{E}(X_1)\mathrm{E}(X_3)$, $\mathrm{E}(X_4X_2) = \mathrm{E}(X_4)\mathrm{E}(X_2)$, and $\mathrm{E}(X_4X_3) = \mathrm{E}(X_4)\mathrm{E}(X_3)$, then (X_1, X_4) and (X_2, X_3) are two independent bivariate normal random vectors.

Proof. I will prove the following special case of the proposition; the general case follows from a similar reasoning, but the notation is messier.

Suppose (X_1, X_2) has a bivariate normal distribution and $\mathrm{E}(X_1X_2) = \mathrm{E}(X_1)\mathrm{E}(X_2)$. Then, X_1 and X_2 are independent. In order to prove this we write the MGF of $\boldsymbol{X} := (X_1, X_2)'$:

$$\begin{aligned} M_X(\boldsymbol{t}) &= \mathrm{e}^{\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}} \\ &= \mathrm{e}^{t_1\mu_1 + t_2\mu_2} \cdot \exp\left(\frac{1}{2}(t_1, t_2)\begin{pmatrix} \mathrm{Var}(X_1) & 0 \\ 0 & \mathrm{Var}(X_2) \end{pmatrix}\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}\right) \\ &= \mathrm{e}^{t_1\mu_1 + \frac{1}{2}t_1^2\mathrm{Var}(X_1)} \cdot \mathrm{e}^{t_2\mu_2 + \frac{1}{2}t_2^2\mathrm{Var}(X_2)} \\ &= M_{X_1}(t_1) \cdot M_{X_2}(t_2). \end{aligned}$$

The result follows from the independence theorem for MGF's (Theorem 6, p. 29). $\square$

Remark 9. The previous proposition has generalizations. For instance, suppose we could decompose $\{1, \ldots, n\}$ into k disjoint groups $G_1, \ldots, G_k$ [so $G_i \cap G_j = \varnothing$ if $i \neq j$, and $G_1 \cup \cdots \cup G_k = \{1, \ldots, n\}$] such that $X_{i_1}, \ldots, X_{i_k}$ are [pairwise] uncorrelated for all $i_1 \in G_1, \ldots, i_k \in G_k$. Then, $(X_i)_{i \in G_1}, \ldots, (X_i)_{i \in G_k}$ are independent multivariate normal random vectors. The proof is the same as in the case $k = 2$. $\square$

Remark 10. It is important that X has a multivariate normal distribution. For instance, we can construct two standard-normal random variables X and Y, on the same probability space, such that X and Y are uncorrelated but dependent. Here is one way to do this: Let $Y \sim \mathrm{N}(0, 1)$ and $S = \pm 1$ with probability $1/2$ each. Assume that S and Y are independent, and define $X := S|Y|$. Note that

$$\begin{aligned} \mathrm{P}\{X \leq a\} &= \mathrm{P}\{X \leq a\ ,\ S = 1\} + \mathrm{P}\{X \leq a\ ,\ S = -1\} \\ &= \frac{1}{2}\mathrm{P}\{|Y| \leq a\} + \frac{1}{2}\mathrm{P}\{-|Y| \leq a\}. \end{aligned}$$

If $a \geq 0$, then $\mathrm{P}\{X \leq a\} = \frac{1}{2}\mathrm{P}\{|Y| \leq a\} + \frac{1}{2} = \mathrm{P}\{Y \leq a\}$. Similarly, $\mathrm{P}\{X \leq a\} = \mathrm{P}\{Y \leq a\}$ if $a \leq 0$. Therefore, $X, Y \sim \mathrm{N}(0, 1)$. Furthermore, X and Y are uncorrelated because S has mean zero; here is why: $\mathrm{E}(XY) = \mathrm{E}(SY|Y|) = \mathrm{E}(S)\mathrm{E}(Y|Y|) = 0 = \mathrm{E}(X)\mathrm{E}(Y)$. But X and Y are not independent because $|X| = |Y|$: For instance, $\mathrm{P}\{|X| < 1\} > 0$, but $\mathrm{P}\{|X| < 1 \mid |Y| \geq 2\} = 0$. The problem is [and can only be] that $(X, Y)'$ is *not* bivariate normal. $\square$

5. Quadratic forms

Given a multivariate-normal random variable $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{0}, \boldsymbol{I})$ and an $n \times n$ positive semidefinite matrix $\boldsymbol{A} := (A_{i,j})$, we can consider the random quadratic form

$$Q_{\boldsymbol{A}}(\boldsymbol{X}) := \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}.$$

We can write $\boldsymbol{A}$, in spectral form, as $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'$, so that

$$Q_{\boldsymbol{A}}(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'\boldsymbol{X}.$$

Since $\boldsymbol{P}$ is orthogonal and $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{Z} := \boldsymbol{P}'\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{0}, \boldsymbol{I})$ as well. Therefore,

$$Q_{\boldsymbol{A}}(\boldsymbol{X}) = \boldsymbol{Z}'\boldsymbol{D}\boldsymbol{Z} = \sum_{i=1}^{n} D_{i,i} Z_i^2.$$

If $\boldsymbol{A}$ is a projection matrix, then all of the $D_{i,i}$'s are ones and zeros. In that case, $Q_{\boldsymbol{A}}(\boldsymbol{X}) \sim \chi^2_r$, where $r :=$ the number of eigenvalues of $\boldsymbol{A}$ that are ones; i.e, $r = \text{rank}(\boldsymbol{A})$. Finally, recall that the rank of a projection matrix is equal to its trace (Corollary 16, p. 16). Let us summarize our findings.

Proposition 11. *If $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{0}\,,\boldsymbol{I})$ and $\boldsymbol{A}$ is a projection matrix, then $\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} \sim \chi^2_{\text{rank}(\boldsymbol{A})} = \chi^2_{\text{tr}(\boldsymbol{A})} = \chi^2_r$, where $r :=$ the total number of nonzero [i.e., one] eigenvalues of $\boldsymbol{A}$.*

Example 12. Let

$$\boldsymbol{A} := \begin{pmatrix} 1 - 1/n & 1/n & -1/n & \cdots & 1/n \\ -1/n & 1 - 1/n & -1/n & \cdots - 1/n & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & -1/n & \cdots & 1 - 1/n \end{pmatrix}.$$

Then we have seen (Example 5, p. 23) that

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \text{for all } \boldsymbol{x} \in \mathbf{R}^n.$$

Now let us observe that $\boldsymbol{A}$ has the form

$$\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{B},$$

where $\boldsymbol{B} := (1/n)\mathbf{1}_{n\times n}$. Note that $\boldsymbol{B}$ is symmetric and $\boldsymbol{B}^2 = \boldsymbol{B}$. Therefore, $\boldsymbol{B}$ is a projection, and hence so is $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{B}$. Clearly, $\text{tr}(\boldsymbol{A}) = n - 1$. Therefore, Proposition 11 implies the familiar fact that if $X_1, \ldots, X_n$ are i.i.d. standard normals, then $\sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2_{n-1}$. □

Example 13. If $\boldsymbol{A}$ is an $n \times n$ projection matrix of rank [or trace] r, then $\boldsymbol{I} - \boldsymbol{A}$ is an $n \times n$ projection matrix of rank [or trace] $n - r$. Therefore, $\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X} \sim \chi^2_{n-r}$, whenever $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{0}\,,\boldsymbol{I})$. □

Example 14. What is the distribution of $\boldsymbol{X}$ is a nonstandard multivariate normal? Suppose $\boldsymbol{X} \sim \mathrm{N}_n(\boldsymbol{\mu}\,,\boldsymbol{\Sigma})$ and $\boldsymbol{A}$ is a projection matrix. If $\boldsymbol{X}$ is nondegenerate, then $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \mathrm{N}_n(\boldsymbol{0}\,,\boldsymbol{I})$. Therefore,

$$(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{A}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_{\text{rank}(\boldsymbol{A})} = \chi^2_{\text{tr}(\boldsymbol{A})},$$

for every $n \times n$ projection matrix $\boldsymbol{A}$. In particular,

$$(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_n,$$

which can be seen by specializing the preceding to the projection matrix $\boldsymbol{A} := \boldsymbol{I}$. Specializing further still, we see that if $X_1, \ldots, X_n$ are independent normal random variables, then we obtain the familiar fact that

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \sim \chi_n^2,$$

where $\mu_i := \mathrm{E}X_i$ and $\sigma_i^2 := \mathrm{Var}(X_i)$. $\square$