

aprendtech.com >> [blog](#) >> [this post](#)

If you have trouble viewing this, try [the pdf of this post](#). You can [download the code](#) used to produce the figures in this post.

Multivariate normal random variables

In my [last post](#), I showed that the multivariate normal, abbreviated multinormal, is a good model for the noise w in a linearized x-ray system model. In this post, I will discuss some of the properties of the multinormal distribution. I will show a rationale for its expression using vectors and matrices. This will lead me to discuss matrix calculus. I will describe diagonalizing and whitening transformations and derive the moment generating functions of the uninormal and multinormal to show that linear combinations of multinormals are also multinormal. This post will provide math background for my discussions of detection and maximum likelihood estimation with the linearized x-ray model.

Matrix expression for the multinormal distribution

Matrices and vectors are natural ways to arrange and keep track of multivariate data so they are widely used in statistical signal processing but most references simply state the matrix expression for the multinormal distribution. I will give a rationale for it and use it to derive some basic properties of the distribution. For this, I start with the univariate normal (uninormal) distribution discussed in elementary probability books (see for example Sec. 5.4 of Ross[1])

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \left(\frac{x-m}{\sigma} \right)^2 \right]. \quad (1)$$

where m is the expected value and σ^2 is the variance. Suppose we have a set of n independent uninormal random variables, their joint density function is

$$\begin{aligned} f(\mathbf{x}) &= f_1(x_1)f_2(x_2)\dots f_n(x_n) \\ &= \frac{1}{(2\pi)^{n/2}\sigma_1\sigma_2\dots\sigma_n} \exp \left[-\frac{1}{2} \sum_{k=1}^n \left(\frac{x_k-m_k}{\sigma_k} \right)^2 \right] \end{aligned} \quad (2)$$

We can summarize the $\{x_k\}$ as the components of a vector \mathbf{x} , the expected values as a vector \mathbf{m} , and the variances as the diagonal covariance matrix

$$\mathbf{C} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \dots & \\ 0 & & \sigma_n^2 \end{bmatrix}.$$

The matrix is diagonal since the covariance of independent random variables is zero. We notice that $\sigma_1\sigma_2\dots\sigma_n$ is the square root of the determinant of the covariance $|\mathbf{C}|^{1/2}$, and that we can express the exponent as the quadratic form $(\mathbf{x}-\mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x}-\mathbf{m})$. This leads to the final expression

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}|^{1/2}} \exp \left[-1/2 (\mathbf{x}-\mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x}-\mathbf{m}) \right]. \quad (3)$$

Eq. 3 was derived for independent variables but I will now show that it makes sense to use it for the general multinormal distribution when \mathbf{C} is any symmetric, positive definite matrix and \mathbf{m} is any vector of real numbers. First, since the exponential function is always

non-negative, and the leading factor in (3) is greater 0, it is clear that $\phi(\mathbf{x}) \geq 0$. Next, I will show that $\int \phi(\mathbf{x})d\mathbf{x} = 1$ by taking a detour through matrix calculus and the principal components of \mathbf{C} . The principal components give us an orthogonal transformation that diagonalizes the matrix. We can also use them to “whiten” the covariance. Applying the transformation allows us to transform $\int \phi(\mathbf{x})d\mathbf{x}$ into a product of integrals of unimodals, each of which is equal to one. Finally, I will derive the moment generating function of the multinormal and use it to show that linear combinations of multinormals are also normal.

Matrix calculus

I will introduce and derive some results from matrix calculus that will be used here and in other posts. I also list some matrix manipulation basics in Eq. 20 at the end of this post. For other results, you can refer to [The Matrix Cookbook](#), which is available free online, and has a huge list of formulas. The book by Harville[2] proves provides proofs for many of the formulas.

The derivative of a matrix whose elements are functions of a scalar is simply the matrix of the derivatives

$$\left[\frac{d\mathbf{A}}{dt} \right]_{jk} = \frac{dA_{jk}}{dt} \quad (4)$$

Similarly the integral of a matrix is the integrals of its components

$$\left[\int \mathbf{A}(t)dt \right]_{jk} = \int A_{jk}(t)dt \quad (5)$$

The matrix formulas are applicable to a vector, which I take to be a column matrix.

Suppose we have a scalar function of a matrix $g(\mathbf{x})$. An example is the atmospheric temperature as a function of the 3D position. The derivative is a matrix (actually a vector) and is the familiar gradient with components

$$\left[\frac{\partial g}{\partial \mathbf{x}} \right]_k = [\nabla g]_k = \frac{\partial g}{\partial x_k} \quad (6)$$

From this we can derive the derivative of the dot product with respect to one of the vectors. Since

$$g(\mathbf{x}) = \mathbf{a}^T \mathbf{x} = \mathbf{x}^T \mathbf{a} = a_1 x_1 + \dots + a_n x_n$$

$$\left[\frac{\partial g}{\partial \mathbf{x}} \right]_k = a_k$$

so

$$\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} \quad (7)$$

I will also use the derivative of a quadratic form $g(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$. This can be derived by writing out the products, see Sec. 15.3 of Harville[2].

$$\begin{aligned} \frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} &= (\mathbf{A} + \mathbf{A}^T) \mathbf{x} \\ &= 2\mathbf{A} \mathbf{x} \quad \mathbf{A} \text{ symmetric} \end{aligned} \quad (8)$$

Principal components of the covariance

The principal components expansion of the covariance can be derived by noticing that the exponent of (3) defines a family of hyper-ellipsoids centered on \mathbf{m} . Translating the origin to the mean value so $\mathbf{z} = \mathbf{x} - \mathbf{m}$, the exponent is proportional to

$$\mathbf{z}^T \mathbf{C}^{-1} \mathbf{z} = c \quad (9)$$

which defines a hyperellipsoid for each positive real number c . The first principal component is the line passing through \mathbf{m} with the longest distance to any point \mathbf{x} on the surface of the ellipsoid (see the discussion in Ch. 3 of Morrison[3]). We can find this line by maximizing the distance (squared) $\mathbf{z}^T \mathbf{z}$ subject to (9) by using Lagrange multipliers. The maximum is found by setting the derivative of the distance after adding a term equal to zero.

$$g(\mathbf{z}) = \mathbf{z}^T \mathbf{z} - \lambda [\mathbf{z}^T \mathbf{C}^{-1} \mathbf{z} - c]$$

Applying (7), the derivative of the first term is

$$\frac{\partial \mathbf{z}^T \mathbf{z}}{\partial \mathbf{z}} = 2\mathbf{z}$$

Using (8) for the derivative of a quadratic form and the fact that \mathbf{C} and therefore its inverse is symmetric

$$\frac{\partial g}{\partial \mathbf{z}} = 2(\mathbf{z} - \lambda \mathbf{C}^{-1} \mathbf{z})$$

Setting this equation equal to zero, we see that \mathbf{z} is an eigenvector of the inverse of covariance matrix.

$$\mathbf{C}^{-1} \mathbf{z} = \frac{1}{\lambda} \mathbf{z} \quad (10)$$

Since \mathbf{C} is invertible \mathbf{z} is also an eigenvector of the covariance matrix, $\mathbf{C}\mathbf{z} = \lambda\mathbf{z}$.

The derivative will be zero for any eigenvector \mathbf{z}_k but only one eigenvector corresponds to the maximum difference. Premultiplying (10) by \mathbf{z}_k^T ,

$$\mathbf{z}_k^T \mathbf{C}^{-1} \mathbf{z}_k = \frac{1}{\lambda_k} \mathbf{z}_k^T \mathbf{z}_k = c$$

we see that the distance is $\mathbf{z}_k^T \mathbf{z}_k = \lambda_k c$. Therefore we can maximize the distance from the centroid by choosing the eigenvector corresponding to the largest eigenvalue. This gives us the principal axis of the ellipsoid. The other eigenvectors correspond to the remaining principal components with smaller distances. The other axes can be listed in order of decreasing eigenvalues.

Diagonalizing transform

We can construct a coordinate transformation matrix Φ that diagonalizes the covariance matrix from the eigenvectors. First, we normalize the eigenvectors to have unit length by dividing by their length i.e. their norm. We can always do this because they are guaranteed to have non-zero norm. We then form a matrix with the eigenvectors as columns. To avoid confusion, I will denote the eigenvectors as ϕ_k

$$\Phi = \begin{bmatrix} \phi_1 & \vdots & \phi_n \end{bmatrix}$$

With this definition

$$\mathbf{C}\Phi = \Phi \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = \Phi \mathbf{D} \quad (11)$$

We can show that if ϕ_j and ϕ_k are two eigenvectors with different eigenvalues λ_j and λ_k , then they are orthogonal. The eigen equations for these components are

$$\begin{aligned} \mathbf{C}\phi_j &= \lambda_j \phi_j \\ \mathbf{C}\phi_k &= \lambda_k \phi_k \end{aligned} \quad (12)$$

Premultiplying the top equation in (12) by ϕ_k^T and taking the transpose of the bottom equation and postmultiplying by ϕ_j

$$\begin{aligned} \phi_k^T \mathbf{C}\phi_j &= \lambda_j \phi_k^T \phi_j \\ \phi_k^T \mathbf{C}^T \phi_j &= \lambda_k \phi_k^T \phi_j \end{aligned}$$

Since \mathbf{C} is symmetric, the left hand sides of the equations are equal, so subtracting them

$$(\lambda_k - \lambda_j) \phi_k^T \phi_j = 0$$

The eigenvalues are different so $\phi_k^T \phi_j = 0$ and the eigenvectors corresponding to different eigenvalues are orthogonal. This implies that the eigenvector matrix Φ is orthogonal so its transpose is equal to its inverse, $\Phi^T = \Phi^{-1}$.

The covariance of the transformed coordinates $\mathbf{z}' = \Phi^T \mathbf{z}$ is

$$\begin{aligned} \mathbf{C}' &= \langle \mathbf{z}' \mathbf{z}'^T \rangle \\ &= \langle \Phi^T \mathbf{z} \mathbf{z}^T \Phi \rangle \\ &= \Phi^T \mathbf{C} \Phi \\ &= \Phi^T \Phi \mathbf{D} \\ &= \mathbf{D} \end{aligned} \quad (13)$$

Whitening transform

The covariance of the Φ transformed coordinates is diagonal but the variances are different. In some cases, we want them to be equal i.e. whitened. Studying Eq. 13, the whitening transform is $\Phi_w = \Phi \mathbf{D}^{-1/2}$. Since \mathbf{D} is diagonal, $\mathbf{D}^{-1/2} = \text{diagonal}[1/\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_n}]$. The covariance of the transformed coordinates $\mathbf{z}' = \Phi_w^T \mathbf{z}$ is

$$\begin{aligned} \mathbf{C}'_w &= \langle \mathbf{z}' \mathbf{z}'^T \rangle \\ &= \langle \mathbf{D}^{-1/2} \Phi^T \mathbf{z} \mathbf{z}^T \Phi \mathbf{D}^{-1/2} \rangle \\ &= \mathbf{D}^{-1/2} \Phi^T \mathbf{C} \Phi \mathbf{D}^{-1/2} \\ &= \mathbf{D}^{-1/2} \Phi^T \Phi \mathbf{D} \mathbf{D}^{-1/2} \\ &= \mathbf{D}^{-1/2} \mathbf{D}^{1/2} \\ &= \mathbf{I} \end{aligned}$$

We can use the whitening transform to define a useful factorization of the covariance. Defining $\mathbf{V} = \Phi_w^T$

$$\begin{aligned} \mathbf{V}^T \mathbf{V} &= \Phi_w \Phi_w^T \\ &= \Phi \mathbf{D}^{-1/2} \mathbf{D}^{-1/2} \Phi^T \\ &= \Phi (\Phi \mathbf{D})^{-1} && \text{the transpose of an orthogonal is also the inverse} \\ &= \Phi (\mathbf{C}\Phi)^{-1} \\ &= \Phi \Phi^{-1} \mathbf{C}^{-1} \\ &= \mathbf{C}^{-1} \end{aligned}$$

Proof that $\int f(\mathbf{x})d\mathbf{x} = 1$ for multinormal

In this section, I will show that the density function for the general multinormal Eq. 3) has the correct normalizing constant. For this, I need to compute the integral. Defining an unnormalized density function as

$$d(\mathbf{x}) = \exp \left[-1/2 (\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}) \right]$$

we need to compute $\int d(\mathbf{x}) d\mathbf{x}$. Translating the origin to \mathbf{m} so $\mathbf{z} = \mathbf{x} - \mathbf{m}$, and transforming by Φ so $\mathbf{z} = \Phi \mathbf{u}$, the exponent in the multinormal distribution is

$$(\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}) = \mathbf{z}^T \mathbf{C}^{-1} \mathbf{z} = (\Phi \mathbf{u})^T \mathbf{C}^{-1} (\Phi \mathbf{u}) \quad (14)$$

Premultiplying (11) by Φ^{-1} , $\mathbf{C} = \Phi \mathbf{D} \Phi^{-1}$. Therefore $\mathbf{C}^{-1} = \Phi \mathbf{D}^{-1} \Phi^{-1}$. Substituting in (14),

$$\mathbf{z}^T \mathbf{C}^{-1} \mathbf{z} = \mathbf{u}^T \Phi^T \Phi \mathbf{D}^{-1} \Phi^{-1} \Phi \mathbf{u} = \mathbf{u}^T \mathbf{D}^{-1} \mathbf{u}$$

The last step follows because the transpose of an orthogonal matrix is its inverse, $\Phi^T = \Phi^{-1}$. Since \mathbf{D} is diagonal

$$\mathbf{D}^{-1} = \begin{bmatrix} 1/\lambda_1 & & 0 \\ & \ddots & \\ 0 & & 1/\lambda_n \end{bmatrix}$$

The exponent is then

$$\mathbf{u}^T \mathbf{D}^{-1} \mathbf{u} = \sum_{k=1}^n \frac{u_k^2}{\lambda_k}$$

and

$$\int d(\mathbf{x}) d\mathbf{x} = \int \exp \left[-\frac{1}{2} \sum_{k=1}^n \frac{u_k^2}{\lambda_k} \right] |\Phi| d\mathbf{u}$$

For an orthogonal transformation the determinant is one, $|\Phi| = 1$ and we can write the transformed integrand as

$$\exp \left[-\frac{1}{2} \sum_{k=1}^n \frac{u_k^2}{\lambda_k} \right] = g_1(u_1) g_2(u_2) \dots g_n(u_n)$$

where

$$g_k(u_k) = \exp \left[-\frac{1}{2} \frac{u_k^2}{\lambda_k} \right]$$

We know from the univariate distribution that $\int \exp \left[-\frac{1}{2} \frac{u_k^2}{\lambda_k} \right] du_k = \sqrt{2\pi\lambda_k}$. Therefore, the integral of $d(\mathbf{x})$ is the product of the univariate integrals and $\int d(\mathbf{x})d\mathbf{x} = (2\pi)^{n/2} \sqrt{\lambda_1\lambda_2 \dots \lambda_n}$. The product of the eigenvalues is the determinant of the \mathbf{D} matrix defined in (11) and since $\mathbf{C} = \Phi \mathbf{D} \Phi^{-1}$ and $|\Phi|$ is orthogonal $|\mathbf{C}| = |\mathbf{D}|$. The integral is therefore

$$\int d(\mathbf{x}) d\mathbf{x} = (2\pi)^{n/2} |\mathbf{C}|^{1/2}$$

so the proposed multinormal density (3) has the correct normalizing constant.

Linear combinations of multinormal random variables–moment generating functions

An important property of multinormal variables is that linear combinations are also normal. I will prove this by deriving the moment generating function, which is also useful for other purposes. Let's start with the moment generating function of the “standard” unnormal random variable $\mathcal{N}(0, 1)$ with 0 mean and variance equal to 1. By definition this is

$$\begin{aligned}
 M_{\mathcal{N}}(t) &= \langle e^{tX} \rangle \\
 &= \frac{1}{\sqrt{2\pi}} \int \exp(tx) \exp(-x^2/2) dx \\
 &= \frac{1}{\sqrt{2\pi}} \int \exp\left(-\frac{x^2-2tx+t^2}{2} + \frac{t^2}{2}\right) dx \\
 &= \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int \exp\left(-\frac{(x-t)^2}{2}\right) dx \\
 &= e^{t^2/2}
 \end{aligned} \tag{15}$$

The general unnormal random variable $\mathcal{N}(m, \sigma^2)$ is a linear combination of a constant and the “standard” unnormal $Y = m + \sigma\mathcal{N}$. Its moment generating function is

$$\begin{aligned}
 M_Y(t) &= \langle e^{t(m+\sigma\mathcal{N})} \rangle \\
 &= e^{mt} \langle e^{\sigma t\mathcal{N}} \rangle \\
 &= e^{mt} M_{\mathcal{N}}(\sigma t) \\
 &= \exp\left(mt + \frac{\sigma^2 t^2}{2}\right)
 \end{aligned} \tag{16}$$

Now, let's consider the vector case. The joint moment generating function is defined to be $M_{\mathbf{X}}(\mathbf{t}) = \langle \exp(\mathbf{t}^T \mathbf{X}) \rangle$, where \mathbf{t} and \mathbf{X} are now vectors of length n . If $\mathcal{N}(0, 1)$ is the “standard” multinormal random vector with independent components, 0 mean and all variances equal to 1

$$\begin{aligned}
 M_{\mathcal{N}}(\mathbf{t}) &= \langle \exp(t_1\mathcal{N}_1 + t_2\mathcal{N}_2 + \dots t_n\mathcal{N}_n) \rangle \\
 &= \langle e^{t_1\mathcal{N}_1} \rangle \dots \langle e^{t_n\mathcal{N}_n} \rangle && \text{components independent} \\
 &= \exp\left[\frac{t_1^2}{2} + \dots \frac{t_n^2}{2}\right] && \text{use Eq. 15} \\
 &= \exp\left(\frac{1}{2}\mathbf{t}^T \mathbf{t}\right)
 \end{aligned} \tag{17}$$

We can use this to derive the moment generating function of the general multinormal distribution. To do this, I will derive the formula for the linear combination of a general multivariate random vector, not necessarily normal. Let $\mathbf{Y} = \mathbf{A} + \mathbf{B}^T \mathbf{X}$ where \mathbf{X} is an $n \times 1$ random vector, \mathbf{A} is an $m \times 1$ constant vector and \mathbf{B} is an $n \times m$ constant matrix. The moment generating function of \mathbf{Y} is

$$\begin{aligned}
 M_Y(t) &= \langle \exp(\mathbf{t}^T \mathbf{A} + \mathbf{t}^T \mathbf{B}^T \mathbf{X}) \rangle \\
 &= \exp(\mathbf{t}^T \mathbf{A}) \langle \exp(\mathbf{t}^T \mathbf{B}^T \mathbf{X}) \rangle \\
 &= \exp(\mathbf{t}^T \mathbf{A}) M_{\mathbf{X}}(\mathbf{B}\mathbf{t})
 \end{aligned} \tag{18}$$

To derive the moment generating function of a general multinormal let $\mathbf{Y} = \mathbf{m} + \mathbf{S}^T \mathcal{N}(0, 1)$ where $\mathcal{N}(0, 1)$ is the standard multinormal (i.e. with zero mean and unit diagonal covariance). Applying the general result for a linear combination (18)

$$\begin{aligned}
 M_Y(\mathbf{t}) &= \exp(\mathbf{t}^T \mathbf{m}) M_{\mathcal{N}}(\mathbf{S}\mathbf{t}) \\
 &= \exp\left[\mathbf{t}^T \mathbf{m} + \frac{1}{2}\mathbf{t}^T \mathbf{S}^T \mathbf{S}\mathbf{t}\right] \\
 &= \exp\left[\mathbf{t}^T \mathbf{m} + \frac{1}{2}\mathbf{t}^T \mathbf{C}\mathbf{t}\right]
 \end{aligned}$$

where the covariance $\mathbf{C} = \mathbf{S}^T \mathbf{S}$.

To show that linear combinations of multinormals are also multinormal, I will use the moment generating function of a linear combination of general multinormal random variable, $\mathbf{Y}_{\text{normal}} = \mathbf{A} + \mathbf{B}^T \mathcal{N}$

$$\begin{aligned} M_{Y_{\text{normal}}}(\mathbf{t}) &= \exp(\mathbf{t}^T \mathbf{A}) M_{\mathcal{N}}(\mathbf{B}\mathbf{t}) && \text{apply (18)} \\ &= \exp(\mathbf{t}^T \mathbf{A}) \exp\left(\mathbf{t}^T \mathbf{B}\mathbf{m} + \frac{1}{2} \mathbf{t}^T \mathbf{B}^T \mathbf{C} \mathbf{B} \mathbf{t}\right) && \text{apply (??)} \\ &= \exp\left(\mathbf{t}^T (\mathbf{A} + \mathbf{B}\mathbf{m}) + \frac{1}{2} \mathbf{t}^T \mathbf{B} \mathbf{C} \mathbf{B}^T \mathbf{t}\right) \end{aligned} \quad (19)$$

This is the moment generating function of a multinormal random vector with expected value $\mathbf{A} + \mathbf{B}\mathbf{m}$ and covariance $\mathbf{B} \mathbf{C} \mathbf{B}^T$.

With this general results, we can also show that the components of a multinormal random vector are uninormal. In (19), let $\mathbf{A} = 0$ and $\mathbf{B} = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^T$, then $M_Y(\mathbf{t}) = \exp(m_1 + \frac{1}{2} C_{11} t_1^2)$, which is uninormal with mean m_1 and variance C_{11} . This result applies to all the components.

Summary

The material discussed in this post will be used during my discussion of statistical detection and estimation theory applied to x-ray imaging. Eq. 20 summarizes basic matrix manipulations

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^T &= \mathbf{A}^T + \mathbf{B}^T \\ (\mathbf{A}\mathbf{B}\mathbf{C})^T &= \mathbf{C}^T \mathbf{B}^T \mathbf{A}^T \\ (\mathbf{A}\mathbf{B}\mathbf{C})^{-1} &= \mathbf{C}^{-1} \mathbf{B}^{-1} \mathbf{A}^{-1} \quad \mathbf{A}, \mathbf{B}, \mathbf{C} \text{ square, invertible} \\ (\mathbf{A}^T)^{-1} &= (\mathbf{A}^{-1})^T \end{aligned} \quad (20)$$

Eq. 21 shows some basic matrix calculus formulas

$$\begin{aligned} \frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} &= \frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} && \mathbf{a}, \mathbf{b} \text{ equal length column vectors} \\ \frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} &= (\mathbf{A} + \mathbf{A}^T) \mathbf{x} && \mathbf{A} \text{ matrix, } \mathbf{x} \text{ vector} \\ \frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} &= 2\mathbf{A} \mathbf{x} && \mathbf{A} \text{ symmetric} \end{aligned} \quad (21)$$

Eq. 22 has some formulas for normal distributions

$$\begin{aligned} \mathcal{N}(0, 1; x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2\right] && \text{mean} = m, \text{ variance} = \sigma^2 \text{ univariate normal} \\ &\exp\left(mt + \frac{v^2 t^2}{2}\right) && \text{univariate moment generating function} \\ \mathcal{N}(\mathbf{m}, \mathbf{C}; \mathbf{x}) &= \frac{1}{(2\pi)^{n/2} |\mathbf{C}|^{1/2}} \exp\left[-1/2 (\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m})\right] && \text{mean} = \mathbf{m}, \text{ covariance} = \mathbf{C} \text{ multivariate} \\ &\exp\left[\mathbf{t}^T \mathbf{m} + \frac{1}{2} \mathbf{t}^T \mathbf{C} \mathbf{t}\right] && \text{multivariate normal moment generating function} \end{aligned} \quad (22)$$

Eq. 23 lists the diagonalizing and whitening transforms

$$\begin{aligned} \mathbf{C} \phi_k &= \lambda_k \phi_k && \text{eigenvectors of covariance} \\ \mathbf{C} \Phi &= \Phi \mathbf{D} && \text{matrix of eigenvectors} \\ &\mathbf{D} && \text{diagonal matrix of eigenvalues} \\ \mathbf{z}' &= \Phi^T \mathbf{z} && \text{diagonalizing transform} \\ \Phi_w &= \Phi \mathbf{D}^{-1/2}, \mathbf{z}_{\text{white}} = \Phi_w^T \mathbf{z} && \text{whitening transform} \\ \mathbf{V} &= \Phi_w^T && \text{whitening factor} \\ \mathbf{C}^{-1} &= \mathbf{V}^T \mathbf{V} && \text{whitening factorization} \end{aligned} \quad (23)$$

Last edited January 6, 2012

©2012 by Aprend Technology and Robert E. Alvarez

Linking is allowed but reposting or mirroring is expressly forbidden.

References

- [1] S. M. Ross, *A First Course in Probability*, 5th ed., New York: Prentice Hall College Div, 1997.
- [2] D. A. Harville, *Matrix algebra from a statistician's perspective*, Springer, 2008.
- [3] D. F. Morrison, *Multivariate statistical methods*, Thomson/Brooks/Cole, 2005.