Generating Random Vectors from the Multivariate Normal Distribution

István T. Hernádvölgyi *

July 27, 1998

Abstract

The *multivariate normal distribution* is often the assumed distribution underlying data samples and it is widely used in pattern recognition and classification [2][3][6][7]. It is undoubtedly of great benefit to be able to generate random values and vectors from the distribution of choice given its sufficient statistics or chosen parameters. We present a detailed account of the theory and algorithms involved in generating random vectors from the multivariate normal distribution given its mean vector $\vec{\mu}$ and covariance matrix Σ using a random number generator from the univariate uniform distribution $U(0,1)$.

1 Road map

First we introduce our notation, characterize the multivariate normal distribution and state some basic definitions. This is followed by the relevant theory needed to understand the algorithm. Then we describe how to put it altogether to generate the random vectors. The stages are:

- generate n random values $x_1, ..., x_n$ (*by separate invocations of the random number generator*), where $x_i \sim U(0,1)$. Then $x = \frac{\sum_{i=1}^{n} x_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \sim N(0,1)$ approximately, according to the *Central Limit Theorem.*
- generate a d dimensional vector $\vec{x}$, where $\vec{x}_i \sim N(0,1)$. The distribution of $\vec{x}$ is $N(\vec{0}, I_d)$, where I is the $d \times d$ identity matrix.

*istvan@csi.uottawa.ca

- using diagonalization and the derivation of the mean and variance of a linear transformation, transform $\vec{x} \rightarrow \vec{y}$ such that $\vec{y} \sim N(\vec{\mu}, \Sigma)$.

2 Notation

$\vec{v}$ denotes a column vector
$\vec{0}$ represents the 0 *vector*
I denotes the *identity matrix*
$\vec{p} \cdot \vec{q}$ is the *dot* product of $\vec{p}$ and $\vec{q}$
$\vec{v}_i$ denotes the i^{th} element of $\vec{v}$
$|A|$ is the *determinant* of the square matrix A
Σ represents a *covariance* matrix of a multivariate distribution
$\vec{\mu}$ is the *mean vector* of a multivariate distribution
μ is the *mean* of a univariate distribution
σ^2 denotes the *variance* of a univariate distribution
A^T *and* $\vec{v}^T$ denote the *transpose* of matrix A and vector $\vec{v}$ respectively
$\|\vec{v}\|$ is the *norm* of $\vec{v}$
$N(\mu, \sigma^2)$ represents the *univariate normal* distribution with *mean* μ and *variance* σ^2
$N(\vec{\mu}, \Sigma)$ represents the *multivariate normal* distribution with *mean vector* $\vec{\mu}$ and *covariance matrix* Σ
$U(a, b)$ represents the *univariate uniform* distribution on $[a, b]$
$X \sim N(0, 1)$ denotes that random variable X is normally distributed
$E[X]$ represents the *expectation* of random variable X
Λ is a diagonal matrix whose diagonal entries are *eigenvalues*
Φ is a matrix, whose columns are normalized *eigenvectors*

3 The Multivariate Normal Distribution

The probability density function of the general form of the multivariate normal distribution is

$$\frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}$$

where $\vec{\mu}$ is the mean or expectation, Σ is the covariance matrix and d is the dimension of the vectors. The covariance matrix measures how dependent the individual dimensions are.

$$\Sigma = E[(X - \vec{\mu})(X - \vec{\mu})^T] = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,d} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,d} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{d,1} & \sigma_{d,2} & \cdots & \sigma_{d,d} \end{pmatrix}$$

where the covariance of dimension i and j is defined as

$$\sigma_{i,j} = E[(\vec{x}_i - \vec{\mu}_i)(\vec{x}_j - \vec{\mu}_j)]$$

Since $\sigma_{i,j} = \sigma_{j,i}$ and $\sigma_{i,j} \geq 0 \ \forall i,j$, Σ is symmetric and positive definite[1]. Figure 1 illustrates the

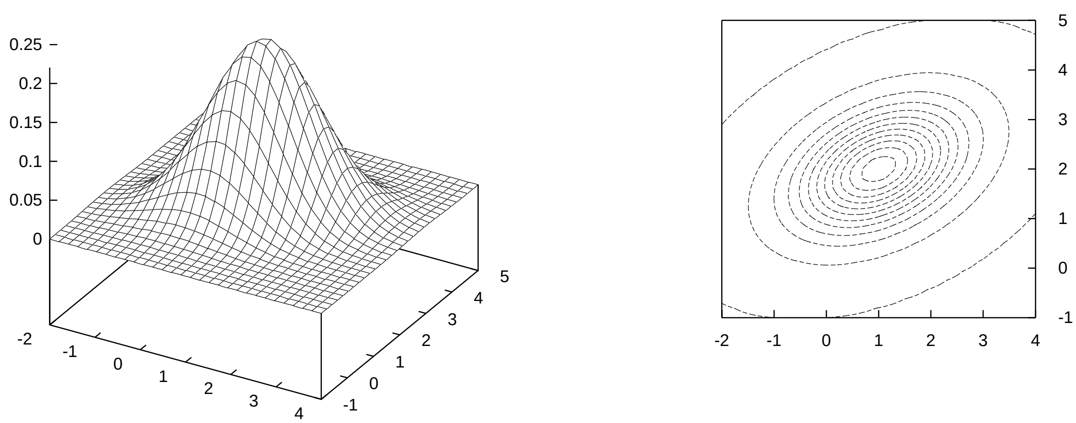


Figure 1: Bivariate Normal with $\vec{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 0.6 \end{pmatrix}$

general shape of the bivariate normal distribution and its level curves[2]. If Σ is diagonal then the principal axes of the ellipses are parallel to the axes of the coordinate system and the dimensions

[1] In theory, Σ is positive *semi-definite*, as values from a particular dimension may have 0 variance. We do not consider such a case, and it is very unlikely to occur in sampling if there is sufficient data available.

[2] Level curves are solutions to $f(\vec{x}) = k$ where k is some constant. Points on the same level curve (*or level surface in higher dimensions*) are *equiprobable*.

are *independent*. In particular, if $\Sigma = \lambda I$ – where I is the identity matrix and λ is a positive constant – then the level curves are circles (*or hyper spheres in higher dimensions*).

As we have already noted, we use a random number generator from the univariate uniform distribution $U(0,1)$. The univariate uniform distribution $U(a,b)$ is usually defined by the two parameters a and b, with probability density function:

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases}$$

with $\mu = \frac{a+b}{2}$ and $\sigma^2 = \frac{(b-a)^2}{12}$. There is ample literature on how to generate pseudo random values from $U(0,1)$ and code is readily available[3].

4 Theory

In this section, we present the theory and the proofs that are needed to understand the mathematics of generating the random vectors. First, we start with the *Central Limit Theorem* which provides the means of generating random values from $N(0,1)$. we investigate the distribution of the random variable $Y = PX$ obtained from X by applying the linear transformation P. Then we show how to find P such that if $X \sim N(\vec{0}, I)$ then $Y = PX \sim N(\vec{\mu}, \Sigma)$.

Theorem: *Central Limit Theorem*

> If $X_1, ..., X_n$ are a random sample from the same distribution with mean μ and variance σ^2, then the distribution of
>
> $$\frac{(\sum_{i=1}^{n} X_i) - n\mu}{\sigma\sqrt{n}}$$
>
> is $N(0,1)$ in the limit as $n \to \infty$.

The *Central Limit Theorem* gives us an algorithm to generate random values distributed $N(0,1)$ from a random sample of values from $U(0,1)$.

$$\lim_{n\to\infty} \frac{(\sum_{i=1}^{n} U(0,1)) - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \sim N(0,1)$$

The question is, how large should n be. If $n = 2$ then the bell curve approximates a triangle, if $n = 3$ then the true distribution is 3 pieces of quadratic curves and in general for $n = k$ k-many

[3] 48 bit pseudo random number generator is part of the core `C` library *random*.

$k-1$ degree polynomial pieces are approximated by $N(0,1)$ [1]. Of course, the larger n is, the better the approximation. In practice[4], $n=12$ yields very good approximation.

The probability density function of $\vec{x}$ where each $x_i \sim N(0,1)$ and the elements are not correlated has level curves of concentric circles. Points with the same distance from the origin are *equiprobable* and $\Sigma = I_d$ where I_d is the identity matrix with dimension d. However we are interested in obtaining random vectors with covariance matrix Σ. Our strategy is to use linear transformations which turn I_d into Σ.

Theorem: *Mean and Variance of a Linear Transformation*

Let X have mean $E[X]$ and covariance matrix $E[(X-E[X])(X-E[X])^T] = \Sigma_X$, both of dimension d. Then for

$$\Theta = \begin{pmatrix} e_{1,1} & e_{1,2} & \dots & e_{1,d} \\ e_{2,1} & e_{2,2} & \dots & e_{2,d} \\ \dots & \dots & \dots & \dots \\ e_{k,1} & e_{k,2} & \dots & e_{k,d} \end{pmatrix} \quad \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_k \end{pmatrix}$$

$Y = \Theta X + \vec{v}$ has mean $E[Y] = \vec{v} + \Theta E[X]$ and covariance matrix $\Sigma_Y = \Theta \Sigma_X \Theta^T$

Proof.

$$E[Y] = E[\Theta X + \vec{v}] = E[\Theta X] + E[\vec{v}] = \Theta E[X] + \vec{v}$$

$$\Sigma_Y = E[(Y-E[Y])(Y-E[Y])^T] =$$
$$E[(\Theta X + \vec{v} - \Theta E[X] - \vec{v})(\Theta X + \vec{v} - \Theta E[X] - \vec{v})^T] =$$
$$E[(\Theta(X-E[X]))(\Theta(X-E[X]))^T] =$$
$$E[\Theta(X-E[X])(X-E[X])^T\Theta^T]$$
$$\Theta E[(X-E[X])(X-E[X])^T]\Theta^T = \Theta\Sigma_X\Theta^T$$

It is clear that $\vec{v} = \vec{\mu}$ is the translation vector. It is less obvious to find Θ such that $\Theta\Sigma\Theta^T = I$.

Definition: *Similar Matrices*

Two matrices are *similar* if there exists an invertible matrix[5] P such that $B = P^{-1}AP$.

[4] see appendix

[5] invertible: $PP^{-1} = I$ or $|P| \neq 0$.

Definition: *Diagonalizable Matrix*

A matrix is *diagonalizable* if it is similar to a diagonal matrix.

We are looking for P such that $P\Sigma P^{-1} = I$ (*which can also be written as* $P^{-1}IP = \Sigma$). First we find Q such that, $Q^{-1}\Sigma Q = \Lambda$ where Λ is a diagonal matrix, and let $P = \Lambda^{\frac{1}{2}}Q$. Once we have Q, and X has covariance matrix I, then $Y = (\Lambda^{\frac{1}{2}}Q)X$ has covariance matrix $(\Lambda^{\frac{1}{2}}Q)I(\Lambda^{\frac{1}{2}}Q)^T = Q\Lambda^{\frac{1}{2}}I\Lambda^{\frac{1}{2}}Q^T = Q\Lambda Q^T$[6]. Later we prove that, if the columns of Q form an *orthonormal basis* of R^d, then $Q^{-1} = Q^T$. Thus $Q\Lambda Q^{-1} = Q\Lambda Q^T = \Sigma$ and for $P = \Lambda^{\frac{1}{2}}Q$ and $X \sim N(\vec{0}, I)$, $Y = PX + \vec{\mu} \sim N(\vec{\mu}, Q\Lambda Q^T) = N(\vec{\mu}, \Sigma)$. Λ being diagonal also eases the computation of $\Lambda^{\frac{1}{2}}$

$$\Lambda^{\frac{1}{2}} = \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & .... & ... & ... \\ 0 & 0 & ... & \lambda_d \end{pmatrix}^{\frac{1}{2}} = \begin{pmatrix} \lambda_1^{\frac{1}{2}} & 0 & ... & 0 \\ 0 & \lambda_2^{\frac{1}{2}} & ... & 0 \\ ... & .... & ... & ... \\ 0 & 0 & ... & \lambda_d^{\frac{1}{2}} \end{pmatrix}$$

To find such a diagonal matrix, we introduce the concept of *eigenvalue* and *eigenvector*.

Definition: *Eigenvalue and Eigenvector*

If A is an $n \times n$ matrix, then a number λ is an *eigenvalue* of A if

$$A\vec{p} = \lambda\vec{p}$$

for some $\vec{p} \neq \vec{0}$. $\vec{p}$ is called an *eigenvector* of A.

Let $\vec{v}$ be a vector of dimension d. Then $\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}$ and $\frac{1}{\|\vec{v}\|}\vec{v}$ is a *unit vector* in R^d in the direction of $\vec{v}$ (*by the Pythagoras' Theorem*). Suppose A has n eigenvectors $\vec{p}_1, ..., \vec{p}_n$ associated with eigenvalues $\lambda_1, ..., \lambda_n$. Then for

$$Q = \left(\left(\frac{1}{\|\vec{q}_1\|}\right)\vec{q}_1, ..., \left(\frac{1}{\|\vec{q}_n\|}\right)\vec{q}_n\right) = \begin{pmatrix} \frac{1}{\|\vec{q}_1\|}\vec{q}_{1,1} & \frac{1}{\|\vec{q}_2\|}\vec{q}_{2,1} & \cdots & \frac{1}{\|\vec{q}_n\|}\vec{q}_{n,1} \\ \frac{1}{\|\vec{q}_1\|}\vec{q}_{1,2} & \frac{1}{\|\vec{q}_2\|}\vec{q}_{2,2} & \cdots & \frac{1}{\|\vec{q}_n\|}\vec{q}_{n,2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{\|\vec{q}_1\|}\vec{q}_{1,n} & \frac{1}{\|\vec{q}_2\|}\vec{q}_{2,n} & \cdots & \frac{1}{\|\vec{q}_n\|}\vec{q}_{n,n} \end{pmatrix}$$

$$Q^{-1}AQ = \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \lambda_n \end{pmatrix} = \Lambda$$

[6]A diagonal matrix (Λ) of size $n \times n$ commutes with all $n \times n$ matrices with respect to multiplication. $\Lambda A = A\Lambda$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of A.

Proof.

$$A\frac{1}{||\vec{q}_1||}\vec{q}_1 = \frac{\lambda_1}{||\vec{q}_1||}\vec{q}_1$$

$$A\frac{1}{||\vec{q}_2||}\vec{q}_2 = \frac{\lambda_2}{||\vec{q}_2||}\vec{q}_2$$

$$...$$

$$A\frac{n}{||\vec{q}_n||}\vec{q}_n = \frac{\lambda_n}{||\vec{q}_n||}\vec{q}_n$$

hence

$$AQ = \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \lambda_n \end{pmatrix} Q = \Lambda Q$$

$$Q^{-1}AQ = Q^{-1}\Lambda Q$$

but Λ is diagonal, thus

$$Q^{-1}AQ = \Lambda Q^{-1}Q = \Lambda$$

As the columns of Q form an orthonormal basis of R^n (*they are orthogonal with unit length, and there are n of them by assumption*) $Q^{-1} = Q^T$.

Proof.

The diagonal entries are obtained by $\left(\frac{1}{||\vec{q}_i||}\right)\vec{q}_i^T \cdot \left(\frac{1}{||\vec{q}_i||}\right)\vec{q}_i = 1$, while the off diagonal entries are the dot products of orthogonal vectors, hence $Q^TQ = I$.

For $\Sigma = A$, the linear transformation $P = \Lambda^{\frac{1}{2}}Q$ is such that if $X \sim N(\vec{0}, I)$ then $Y = PX \sim N(\vec{0}, \Sigma)$. All that left to do is to prove that Σ has orthogonal eigenvectors and there is an algorithm to calculate them.

Theorem: *Eigenvectors of Symmetric Matrices*

If A is a symmetric $n \times n$ matrix, then A has orthogonal eigenvectors.

Proof.

If A is symmetric, then $(A\vec{p}) \cdot \vec{q} = \vec{p} \cdot (A\vec{q})$. As $A = A^T$,

$$(A\vec{p}) \cdot \vec{q} = (A\vec{p})^T \vec{q} =$$

$$\vec{p}^T A^T \vec{q} = \vec{p}^T A \vec{q} = \vec{p} \cdot (A\vec{q})$$

Let $A\vec{p}_1 = \lambda_1 \vec{p}_1$ and $A\vec{p}_2 = \lambda_2 \vec{p}_2$ where $\lambda_1 \neq \lambda_2$. Then

$$\lambda_1(\vec{p}_1 \cdot \vec{p}_2) = (A\vec{p}_1) \cdot \vec{p}_2 =$$

$$\vec{p}_1 \cdot (A\vec{p}_2) = \lambda_2(\vec{p}_1 \cdot \vec{p}_2)$$

Hence $\vec{p}_1 \cdot \vec{p}_2 = 0$, which implies that $\vec{p}_1$ and $\vec{p}_2$ are orthogonal.

We also need a guarantee that Σ of dimension d has d eigenvectors. As far as Σ has linearly independent rows (*or columns*), this is guaranteed. On the other hand, if $|\Sigma| = 0$, or it has linearly dependent rows, then the general form of the normal density is not defined ($|\Sigma|$ *is in the denominator*). Often the parameters (*including* Σ) are obtained by estimation from a random sample. If there is sufficient data available, it is next to impossible to encounter the problem of $|\Sigma| = 0$ unless some dimensions differ so little that arithmetic resolution may render them zero. If that is the case, a particular dimension or dimensions can be scaled by the linear transformation

$$\Theta = \begin{pmatrix} c_1 & 0 & ... & 0 \\ 0 & c_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & c_d \end{pmatrix}$$

where c_i scales dimension i. If $X \sim N(\vec{\mu}, \Sigma)$, then $Y = \Theta(X - \vec{\mu}) + \vec{\mu} \sim N(\vec{\mu}, \Theta\Sigma\Theta^T)$. If we can generate random vectors $\vec{y} \sim Y$, then $\vec{x} = \Theta^{-1}(\vec{y} - \vec{\mu}) + \vec{\mu} \sim N(\vec{\mu}, \Sigma)$.

Obtaining the eigenvectors and eigenvalues of a matrix is not trivial. By definition, the eigenvalues of A are the solutions of the *characteristic polynomial* $p_A(x) = |xI - A|$, where I is the identity matrix. For $n = 2$ and $n = 3$, it is trivial to calculate the eigenvalues, as closed form formulae exist for quadratic and cubic polynomials. For larger n, finding the roots of the polynomial is impractical. Instead a version of the iterative QR algorithm [4][5][8][9] is used.

5 Algorithm

All what is left is to put the theory together to generate the random vectors.

Objective: generate random vectors from $N(\vec{\mu}, \Sigma)$ given $\vec{\mu}$ and Σ ($\vec{\mu}$ and Σ are of dimension d).

1. Generate $\vec{x}$, such that $\vec{x}_i = 12U(0,1) - 6$, where $U(0,1)$ denotes one random value from $U(0,1)$ obtained by an *independent* invocation of a pseudo random number generator from $U(0,1)$. According to the *Central Limit Theorem* $\vec{x}$ is *approximately* from $N(\vec{0}, I)$.

2. Let Φ be the $d \times d$ matrix whose columns are the normalized eigenvectors of Σ and let Λ be the diagonal matrix whose diagonal entries are the eigenvalues of Σ in the order corresponding to the columns of Φ. Let $Q = \Lambda^{\frac{1}{2}}\Phi$. According to the derivation of the mean and variance of a linear transformation $\vec{y} = Q\vec{x} + \vec{\mu}$ is from $N(\vec{\mu}, \Sigma)$.

6 Appendix

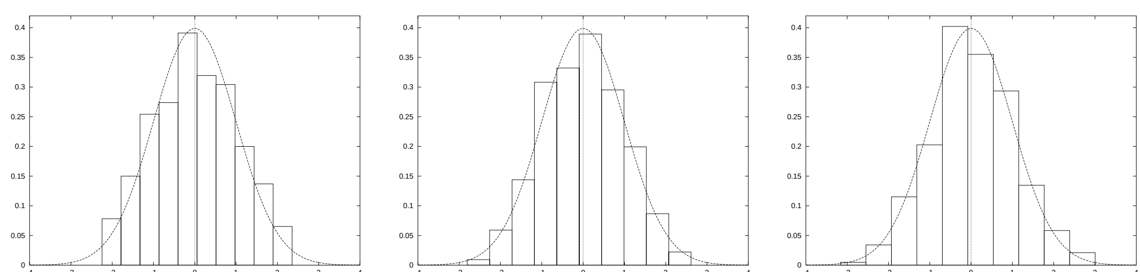
6.1 Random values from $N(0,1)$

The following experiments use the formula

$$\frac{\sum_{i=1}^{n} U(0,1) - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$

for $n = 2$, $n = 3$ and $n = 12$ to generate random values from $N(0,1)$.

Figure 2 shows the histograms of 1000 randomly generated points overfitted by the real density function of $N(0,1)$, $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.



$n = 2$ $n = 3$ $n = 12$

Figure 2: 1000 pseudo random values from $N(0,1)$.

6.2 A Detailed Example

Let us walk through the steps *visually* with the bivariate normal. Our objective is to obtain random vectors from $N(\vec{\mu}, \Sigma)$ where $\vec{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 0.6 \end{pmatrix}$.

First we obtain Λ and Φ.

$$p_\Sigma(\lambda) = |\lambda I - \Sigma| = \lambda^2 - 1.6\lambda + 0.51$$

$$\lambda_1 = 0.8 + 0.05\sqrt{52}$$

$$\lambda_2 = 0.8 - 0.05\sqrt{52}$$

hence

$$\Lambda = \begin{pmatrix} 0.8 + 0.05\sqrt{52} & 0 \\ 0 & 0.8 - 0.05\sqrt{52} \end{pmatrix} = \begin{pmatrix} 1.1606 & 0 \\ 0 & 0.4394 \end{pmatrix}$$

Now we calculate Φ

$$\begin{pmatrix} 1 & 0.3 \\ 0.3 & 0.6 \end{pmatrix} \begin{pmatrix} e_{1,1} \\ e_{1,2} \end{pmatrix} = 0.8 + 0.05\sqrt{52} \begin{pmatrix} e_{1,1} \\ e_{1,2} \end{pmatrix}$$

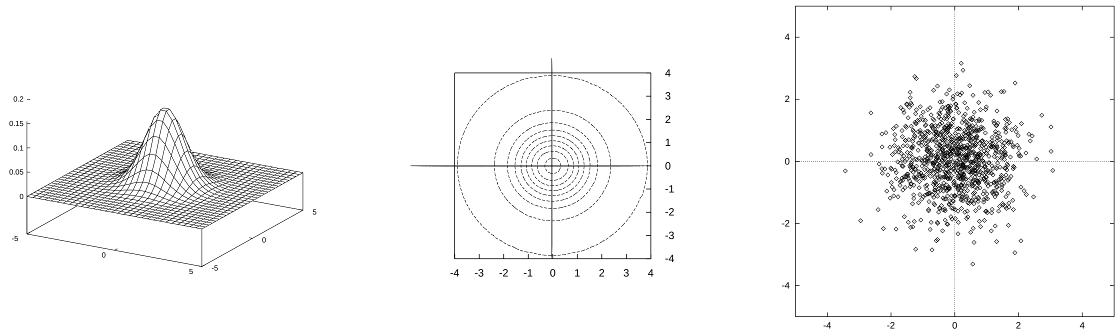
$$\begin{pmatrix} 0.2 - 0.05\sqrt{(52)} & 0.3 \\ 0.3 & -0.2 - 0.05\sqrt{52} \end{pmatrix} \to \begin{pmatrix} 1 & \frac{0.3}{0.2-0.05\sqrt{52}} \\ 0 & 0 \end{pmatrix} \to e_1 = \begin{pmatrix} \frac{1}{\sqrt{1^2+\left(\frac{0.2-0.05\sqrt{52}}{-0.3}\right)^2}} \\ \frac{\frac{0.2-0.05\sqrt{52}}{-0.3}}{\sqrt{1^2+\left(\frac{0.2-0.05\sqrt{52}}{-0.3}\right)^2}} \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0.3 \\ 0.3 & 0.6 \end{pmatrix} \begin{pmatrix} e_{2,1} \\ e_{2,2} \end{pmatrix} = 0.8 - 0.05\sqrt{52} \begin{pmatrix} e_{2,1} \\ e_{2,2} \end{pmatrix}$$

$$\begin{pmatrix} 0.2 + 0.05\sqrt{(52)} & 0.3 \\ 0.3 & -0.2 + 0.05\sqrt{52} \end{pmatrix} \to \begin{pmatrix} 1 & \frac{0.3}{0.2+0.05\sqrt{52}} \\ 0 & 0 \end{pmatrix} \to e_2 = \begin{pmatrix} \frac{1}{\sqrt{1^2+\left(\frac{0.2+0.05\sqrt{52}}{-0.3}\right)^2}} \\ \frac{\frac{0.2+0.05\sqrt{52}}{-0.3}}{\sqrt{1^2+\left(\frac{0.2+0.05\sqrt{52}}{-0.3}\right)^2}} \end{pmatrix}$$

and thus

$$\Phi = (e_1 e_2) = \begin{pmatrix} \frac{1}{\sqrt{1^2+\left(\frac{0.2-0.05\sqrt{52}}{-0.3}\right)^2}} & \frac{1}{\sqrt{1^2+\left(\frac{0.2+0.05\sqrt{52}}{-0.3}\right)^2}} \\ \frac{\frac{0.2-0.05\sqrt{52}}{-0.3}}{\sqrt{1^2+\left(\frac{0.2-0.05\sqrt{52}}{-0.3}\right)^2}} & \frac{\frac{0.2+0.05\sqrt{52}}{-0.3}}{\sqrt{1^2+\left(\frac{0.2+0.05\sqrt{52}}{-0.3}\right)^2}} \end{pmatrix} = \begin{pmatrix} 0.8817 & 0.4719 \\ 0.4719 & -0.8817 \end{pmatrix}$$



$$\vec{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\hat{\vec{\mu}} = \begin{pmatrix} 0.0149 \\ 0.0392 \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 1.0084 & -0.0251 \\ -0.0251 & 0.9882 \end{pmatrix}$$

Figure 3: Distribution $N(\vec{0}, I)$

Figure 3 shows the real bivariate density $N(\vec{0}, I)$, its level curves and 1000 random points generated from this distribution. $\hat{\vec{\mu}}$ and $\hat{\Sigma}$ are the *most likelihood estimates* [3] of the parameters.

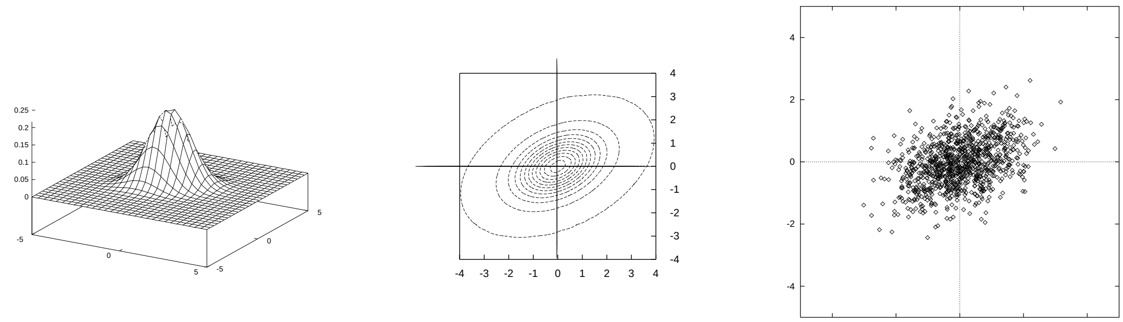
$$\hat{\vec{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \vec{x}_i$$

where $\vec{x}_i$ is the i^{th} vector from the sample.

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} (\vec{x}_i - \hat{\vec{\mu}})(\vec{x}_i - \hat{\vec{\mu}})^T$$

Figure 4 shows the application of the linear transformation $P = \Lambda^{\frac{1}{2}}\Phi$ to the real density as well as to the very same 1000 random points.

Finally, figure 5 represents the distribution $N(\vec{\mu}, \Sigma)$. It is obvious from the figures, that the randomly generated points follow the level curves and density of the underlying distribution. The most likelihood estimates of the parameters are indeed very close to the parameter values supplied.

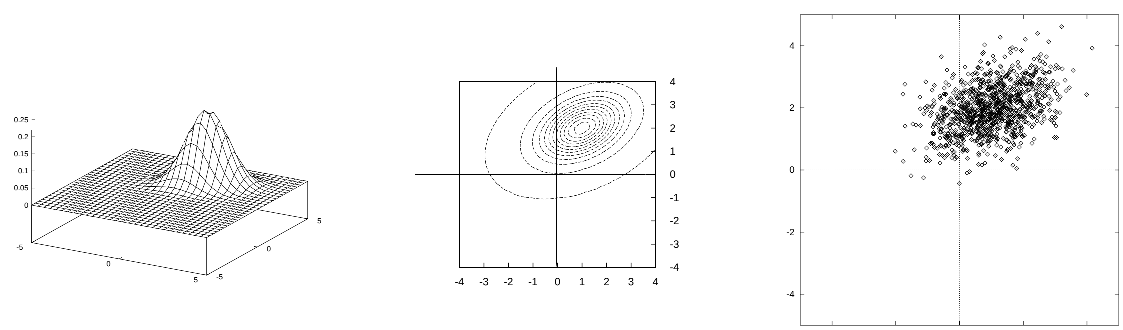


$$\vec{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 0.6 \end{pmatrix}$$

$$\hat{\vec{\mu}} = \begin{pmatrix} -0.0019 \\ -0.0305 \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 1.0213 & 0.2963 \\ 0.2963 & 0.5832 \end{pmatrix}$$

Figure 4: Distribution $N(\vec{0}, \Sigma)$



$$\vec{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 0.6 \end{pmatrix}$$

$$\hat{\vec{\mu}} = \begin{pmatrix} 0.9981 \\ 1.9695 \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 1.0213 & 0.2963 \\ 0.2963 & 0.5832 \end{pmatrix}$$

Figure 5: Distribution $N(\vec{\mu}, \Sigma)$

7 Concluding Remarks

Our focus was to introduce the *theory* of generating random vectors from the multivariate normal distribution, at a level that does not require extensive background in Linear Algebra and Statistics. We did not concern ourselves with the *efficiency* and *complexity* of these algorithms. The reader is encouraged to further investigate and study the particulars of efficiently finding *eigenvalues* and *eigenvectors* and generating pseudo random values from $U(a,b)$ and $N(\mu,\sigma^2)$.

The method we presented can be applied *backwards* to the distribution $X \sim N(\vec{\mu},\Sigma)$. For $P_1 = \Phi^{-1} = \Phi^T$, $Y_1 = P_1(X-\vec{\mu}) + \vec{\mu} \sim N(\vec{\mu},\Lambda)$, where Λ is a diagonal matrix of *eigenvalues*. As the covariance matrix of the distribution of Y_1 is diagonal, the linear transformation P_1 renders the individual dimensions *independent*. Suppose X is a bivariate density that approximates the *weight* and *height* distribution of a particular species. It is reasonable to assume that *height* and *weight* are dependent; the taller the specimen the heavier it is. On the other hand, Y_1 has *independent* or *orthogonal* components, where each dimension is a *linear* expression of *height* and *weight*. To make statistical inferences about the species, these independent measures may be more appropriate. For $P_2 = \Lambda^{-\frac{1}{2}}P_1$, $Y_2 = P_2(X-\vec{\mu}) + \vec{\mu} \sim N(\vec{\mu},I)$, or in other words, not only Y_2 has independent dimensions, but they also have the same variance: 1.

8 Acknowledgement

I would like to thank *Dr. John Oommen* from *Carleton University* for teaching the insipirational course that awoke my interest in Multivariate Statistics and Statistical Pattern Recognition.

References

[1] R. V. Hogg, E. A. Tanis: Probability and Statistical Inference. *Macmillan Publishing Co. 1993*

[2] J. W. Pratt, H. Raiffa, R. Schlaifer: Statistical Decision Theory. *MIT Press. 1995*

[3] R. O. Duda, P. E. Hart: Pattern Classification and Scene Analysis. *John Wiley & Sons. 1973*

[4] W. K. Nicholson: Elementary Linear Algebra with Applications. *PWS-KENT. 1990*

[5] C. G. Cullen: Matrices and Linear Transformations. *Dover Publications. 1990*

[6] S. K. Kachigan: Multivariate Statistical Analysis. *Radius Press. 1991*

[7] H. Chernoff, L. E. Moses: Elementary Decision Theory. *Dover Publications. 1986*

[8] E. Isaacson, H. B. Keller: Analysis of Numerical Methods. *Dover Publications. 1994*

[9] R. L. Burden, J. D. Faires: Numerical Analysis. *PWS-KENT. 1993*

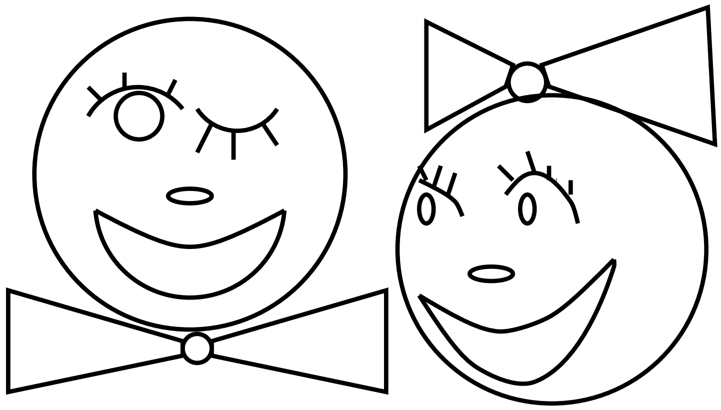


Figure 6: *Muggy & Huggy*