BROWNIAN MOTION

1. Expectations and Covariances of Random Vectors

A *random vector*, or more precisely, a *random n-vector* is a column vector $\vec{\boldsymbol{X}} \in \mathbb{R}^n$ whose coordinates are jointly defined random variables:

$$\vec{\boldsymbol{X}} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} .$$

Detailed probability work with random vectors requires knowledge of the joint distribution of the coordinates $X_1, \ldots, X_n$, often based on a joint probability density function. But here we will not worry about the general theory, and will only concern ourselves with the joint distribution in one special case, namely Gaussian random vectors, to be introduced later. For now, we will focus only on expected values and covariances for random vectors.

1.1. **Mean vectors.** We take the expected value of a random vector $\vec{\boldsymbol{X}}$ in the following natural way:

$$E(\vec{\boldsymbol{X}}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} = \vec{\boldsymbol{\mu}} ,$$

where $\mu_i = E(X_i)$ for each i. We call the vector $\vec{\boldsymbol{\mu}}$ the *mean vector* of $\vec{\boldsymbol{X}}$. Note that the mean vector is simply a vector of numbers. These numbers are the means (expected values) of the random variables X_i.

Exercise 1. Let $\vec{\boldsymbol{X}}$ be a random 2-vector. Calculate its mean vector if $\vec{\boldsymbol{X}}$ is a point randomly selected from within

(i) The unit square (that is, the square with vertices $(0,0), (1,0), (0,1), (1,1)$).
(ii) The unit circle centered at the origin.
(iii) The triangle with vertices $(0,0), (1,0), (0,1)$.
(iv) The region above the parabola $y = x^2$ and inside the unit square.

Not surprisingly, linear algebra plays an important role with random vectors. For example, it will be common to multiply a random n-vector $\vec{\boldsymbol{X}}$ by a non-random $m \times n$ matrix A, giving us the random m-vector $A\vec{\boldsymbol{X}}$. We can also multiply random vectors by scalars, and add random vectors to other vectors (random or non-random). Suppose $\vec{\boldsymbol{X}}$ and $\vec{\boldsymbol{Y}}$ are random vectors, and let A and B be non-random matrices. Then it is not hard to prove that

$$E(A\vec{\boldsymbol{X}} + B\vec{\boldsymbol{Y}}) = AE(\vec{\boldsymbol{X}}) + BE(\vec{\boldsymbol{Y}})\,, \tag{1}$$

assuming that the sizes of the vectors $\vec{\boldsymbol{X}}, \vec{\boldsymbol{Y}}$ and the matrices A, B are chosen so that the matrix products $A\vec{\boldsymbol{X}}$ and $B\vec{\boldsymbol{Y}}$ both make sense, and also so that the vector addition $A\vec{\boldsymbol{X}} + B\vec{\boldsymbol{Y}}$ makes sense.

An important special case of (1) is the following: suppose $\vec{\boldsymbol{X}}$ is a random n-vector, A is a non-random $m \times n$ matrix, and $\vec{\boldsymbol{b}}$ is a non-random m-vector (column). Then

$$E(A\vec{\boldsymbol{X}} + \vec{\boldsymbol{b}}) = A\vec{\boldsymbol{\mu}} + \vec{\boldsymbol{b}}\,, \tag{2}$$

where $\vec{\boldsymbol{\mu}} = E(\vec{\boldsymbol{X}})$.

Exercise 2. Explain why Equation (2) is a special case of Equation (1).

Exercise 3. Prove Equation (2) in the case that $\vec{\boldsymbol{X}}$ is a random 2-vector, A is a non-random 2×2 matrix, and $\vec{\boldsymbol{b}}$ is a non-random 2-vector.

Similarly, we can take transposes of our random vectors to get row vectors, and multiply by matrices on the right:

$$E(\vec{\boldsymbol{X}}^T A + \vec{\boldsymbol{Y}}^T B) = E(\vec{\boldsymbol{X}})^T A + E(\vec{\boldsymbol{Y}})^T B\,, \tag{3}$$

where we again assume that the sizes of the vectors and matrices are appropriate. This means, for example, that if $\vec{\boldsymbol{X}}$ is a random n-vector, then the matrix A must $n \times m$ for some m, rather than $m \times n$. Here is the special case corresponding to our previous Equation (2):

$$E(\vec{\boldsymbol{X}}^T A + \vec{\boldsymbol{b}}^T) = \vec{\boldsymbol{\mu}}^T A + \vec{\boldsymbol{b}}^T\,, \tag{4}$$

where $\vec{\boldsymbol{X}}$ is a random n-vector with mean vector $\vec{\boldsymbol{\mu}}$, A is a non-random $n \times m$ matrix, and $\vec{\boldsymbol{b}}$ is a non-random m-vector (column).

1.2. **Covariance matrices.** We are also interested in defining something that would correspond to the "variance" of the random vector $\vec{\boldsymbol{X}}$. This will be a matrix, consisting of all the possible covariances of pairs of the random variables $X_1, X_2, \ldots, X_n$. Because it contains covariances, we call it the *covariance matrix* of $\vec{\boldsymbol{X}}$:

$$\mathrm{Cov}(\vec{\boldsymbol{X}}) = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) & \ldots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \ldots & \mathrm{Cov}(X_n, X_n) \end{pmatrix}$$

Remember that for any random variable Y, $\mathrm{Cov}(Y, Y) = \mathrm{Var}(Y)$, so the diagonal elements in the covariance matrix are actually variances, and we could write the covariance matrix this way:

$$\mathrm{Cov}(\vec{\boldsymbol{X}}) = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \ldots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \ldots & \mathrm{Var}(X_n) \end{pmatrix}$$

Of course, in all of these definitions and formulas involving mean vectors and covariance matrices, we need to assume that the means and covariances of the underlying random variables $X_1, X_2, \ldots, X_n$ exist. Since we will be primarily dealing with Gaussian random variables, this assumption will typically be satisfied.

Exercise 4. Calculate the covariance matrix of the random vector $\vec{\boldsymbol{X}}$ for each of the parts of Exercise 1.

In connection with the covariance matrix, it will be useful for us to take the expected value of a *random matrix*. As you might expect, this gives us the matrix whose elements are the expected values of the random variables that are the corresponding elements of the random matrix. Thus, if M is a random $m \times n$ matrix whose elements are random variables X_{ij}, then

$$E(M) = \begin{pmatrix} E(X_{11}) & E(X_{12}) & \ldots & E(X_{1n}) \\ E(X_{21}) & E(X_{22}) & \ldots & E(X_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_{m1}) & E(X_{m2}) & \ldots & E(X_{mn}) \end{pmatrix}.$$

It should be easy for you to see that if M and N are random $m \times n$ matrices and if a and b are real numbers, then

$$E(aM + bN) = aE(M) + bE(N). \tag{5}$$

Now we have a fancy and useful way to write the covariance matrix:

$$\text{Cov}(\vec{\boldsymbol{X}}) = E((\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}})(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}})^T) = E(\vec{\boldsymbol{X}}\vec{\boldsymbol{X}}^T) - \vec{\boldsymbol{\mu}}\vec{\boldsymbol{\mu}}^T . \tag{6}$$

Can you see how this expression makes sense? In order to do so, you must recall that if $\vec{\boldsymbol{v}}$ is a column vector in $\mathbb{R}^n$, then the product $\vec{\boldsymbol{v}}\vec{\boldsymbol{v}}^T$ is an $n \times n$ matrix. Thus, $(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}})(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}})^T$ is a random $n \times n$ matrix:

$$(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}})(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}})^T = \begin{pmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \dots & (X_1 - \mu_1)(X_n - \mu_n) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \dots & (X_2 - \mu_2)(X_n - \mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n - \mu_n)(X_1 - \mu_1) & (X_n - \mu_n)(X_2 - \mu_2) & \dots & (X_n - \mu_n)^2 \end{pmatrix}$$

Now you should be able to see that if you take the expected value of this random matrix, you get the covariance matrix of $\vec{\boldsymbol{X}}$.

Exercise 5. Prove the second equality in (6). (Use (5).)

Now we are equipped to get a formula for $\text{Cov}(A\vec{\boldsymbol{X}} + \vec{\boldsymbol{b}})$, where $\vec{\boldsymbol{X}}$ is a random n-vector, A is a non-random $m \times n$ matrix, and $\vec{\boldsymbol{b}}$ is a non-random m-vector. Note that since the vector $\vec{\boldsymbol{b}}$ won't affect any of the covariances, we have

$$\text{Cov}(A\vec{\boldsymbol{X}} + \vec{\boldsymbol{b}}) = \text{Cov}(A\vec{\boldsymbol{X}}) ,$$

so we can ignore $\vec{\boldsymbol{b}}$. But we obviously can't ignore A! Let's use our fancy formula (6) for the covariance matrix, keeping in mind that $E(A\vec{\boldsymbol{X}}) = A\vec{\boldsymbol{\mu}}$, where $\vec{\boldsymbol{\mu}}$ is the mean vector $\vec{\boldsymbol{X}}$.

$$\text{Cov}(A\vec{\boldsymbol{X}}) = E((A\vec{\boldsymbol{X}} - A\vec{\boldsymbol{\mu}})(A\vec{\boldsymbol{X}} - A\vec{\boldsymbol{\mu}})^T) .$$

Now we use rules of matrix multiplication from linear algebra:

$$(A\vec{\boldsymbol{X}} - A\vec{\boldsymbol{\mu}})(A\vec{\boldsymbol{X}} - A\vec{\boldsymbol{\mu}})^T = A(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}})(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}})^T A^T$$

Take the expected value, and then use (2) and (4) to pull A and A^T outside of the expected value, giving us

$$\text{Cov}(A\vec{\boldsymbol{X}} + \vec{\boldsymbol{b}}) = A\,\text{Cov}(\vec{\boldsymbol{X}})A^T . \tag{7}$$

This is an important formula that will serve us well when we work with Gaussian random vectors.

1.3. **Change of coordinates.** Note that $\mathrm{Cov}(\vec{\boldsymbol{X}})$ is a symmetric matrix. In fact, this matrix is nonnegative definite. How do we know this? From the Spectral Theorem, we know there exists an orthogonal matrix O such that $O^T \mathrm{Cov}(\vec{\boldsymbol{X}})O$ is a diagonal matrix D, with the eigenvalues of $\mathrm{Cov}(\vec{\boldsymbol{X}})$ being the diagonal elements, and the columns of O being the corresponding eigenvectors, normalized to have length 1. But by (7), this diagonal matrix D is the covariance matrix of the random vector $O^T\vec{\boldsymbol{X}}$, and the diagonal entries of any covariance matrix are variances, hence nonnegative. This tells us that $\mathrm{Cov}(\vec{\boldsymbol{X}})$ has nonnegative eigenvalues, making it nonnegative definite. Let $\sqrt{D}$ stand for the diagonal matrix whose diagonal elements are the nonnegative square roots of the diagonal elements of D. Then it is easy to see that $\mathrm{Cov}(\vec{\boldsymbol{X}}) = O\sqrt{D}\sqrt{D}^T O^T = \Sigma\Sigma^T$, where $\Sigma = O\sqrt{D}$. Even though the matrix Σ is not uniquely determined, it seems OK to call it the *standard deviation matrix* of $\vec{\boldsymbol{X}}$. So we have

$$D = O^T \mathrm{Cov}(\vec{\boldsymbol{X}})O \quad \text{and} \quad \Sigma = O\sqrt{D}\,,$$

where

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \quad \text{and} \quad O = \begin{pmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \vec{\boldsymbol{v}}_1 & \vec{\boldsymbol{v}}_2 & \ddots & \vec{\boldsymbol{v}}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{pmatrix}$$

and $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of $\mathrm{Cov}(\vec{\boldsymbol{X}})$, with corresponding eigenvectors $\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \dots, \vec{\boldsymbol{v}}_n$. We will always assume that we have put the eigenvalues into decreasing order, so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

Now suppose that $\vec{\boldsymbol{X}}$ is a random vector with covariance matrix $\Sigma\Sigma^T$, where $\Sigma = O\sqrt{D}$ and O and D are as in the previous paragraph. Then clearly $\vec{\boldsymbol{Y}} = O^T\vec{\boldsymbol{X}}$ is a random vector with covariance matrix D. This means that by simply performing a rotation in $\mathbb{R}^n$, we have transformed $\vec{\boldsymbol{X}}$ into a random vector whose coordinates are *uncorrelated*. If the coordinates of $\vec{\boldsymbol{X}}$ are uncorrelated, then we say the $\vec{\boldsymbol{X}}$ itself is uncorrelated. Informally, we have the following statement:

Every random vector is uncorrelated if you look at it from the right direction!

Mathematically, this statement can be written as

$$\mathrm{Cov}\left(O^T\vec{\boldsymbol{X}}\right) = O^T\Sigma\Sigma^T O = D\,.$$

Exercise 6. For each part of Exercise 1, determine the change of coordinates (rotation) that changes $\vec{\boldsymbol{X}}$ into an uncorrelated random vector.

1.4. **Standardization.** We can go a little bit further. As stated earlier, we select the matrix O so that the nonzero diagonal elements of D, which are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, are in decreasing order. This means that the nonzero eigenvalues come before the eigenvalues that equal 0. Let $m \leq n$ be the number of nonzero diagonal eigenvalues. This is the same as saying that m is the rank of the covariance matrix. Then for $i > m$, the coordinates Y_i of the vector $\vec{\boldsymbol{Y}} = O^T(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}})$ (note that we centered $\vec{\boldsymbol{X}}$ by subtracting its mean vector) have expected value 0 and variance 0, implying that $Y_i = 0$ for $i > m$. Let

$$\vec{\boldsymbol{Z}} = \begin{pmatrix} Y_1/\sqrt{\lambda_1} \\ \\ Y_2/\sqrt{\lambda_2} \\ \vdots \\ Y_m/\sqrt{\lambda_m} \end{pmatrix} .$$

This is the same as letting $\vec{\boldsymbol{Z}} = B\vec{\boldsymbol{Y}}$, where B is the $m \times n$ matrix

$$B = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & & & \\ 0 & 0 & \ldots & \frac{1}{\sqrt{\lambda_m}} & 0 & \ldots & 0 \end{pmatrix}$$

Then $\vec{\boldsymbol{Z}}$ is a random m-vector with mean vector $\vec{\boldsymbol{0}}$ and covariance matrix equal to the $m \times m$ identity matrix I_m. We say that $\vec{\boldsymbol{X}}$ has been *centered and scaled*, or in other words, *standardized*. Informally speaking, we center $\vec{\boldsymbol{X}}$ by subtracting its mean vector, and we scaled it by "dividing" by its standard deviation matrix (after eliminating the extra dimensions that correspond to 0 eigenvalues). To summarize, the standardized version of $\vec{\boldsymbol{X}}$ is $\vec{\boldsymbol{Z}}$, where

$$\vec{\boldsymbol{Z}} = BO^T(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}}) \,. \tag{8}$$

In many cases, all of the diagonal elements of D are nonzero, and $B = \sqrt{D^{-1}}$. This last condition is equivalent to saying that Σ is invertible, and then we have $\Sigma^{-1} = \sqrt{D^{-1}}O^T$.

Theorem 1. *Let $\vec{\boldsymbol{X}}$ be a random n-vector with mean vector $\vec{\boldsymbol{\mu}}$ and standard deviation matrix Σ. If Σ has rank n (or equivalently, is invertible), then the standardized random vector*

$$\vec{\boldsymbol{Z}} = \Sigma^{-1}(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}}) \tag{9}$$

is a random n-vector with mean vector $\vec{\mathbf{0}}$ and covariance matrix I_n, where I_n is the $n \times n$ identity matrix. If Σ has rank $m < n$, then there is an m-dimensional subspace $V \in \mathbb{R}^n$ such that

$$P((\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}}) \in V) = 1\,,$$

and the random m-vector $\vec{\boldsymbol{Z}}$ defined in (8) *has mean vector $\vec{\mathbf{0}}$ and covariance matrix I_m.*

Note that if the covariance matrix of a random vector is the identity matrix, then its standard deviation matrix is also the identity matrix.

Exercise 7. Standardize $\vec{\boldsymbol{X}}$ for each part of Exercise 1.

Exercise 8. Let V_1 and V_2 be the outcomes (0 or 1) of independent tosses of a fair coin. Let

$$\vec{\boldsymbol{X}} = \begin{pmatrix} V_1 + V_2 \\ \\ V_1 \\ \\ V_1 - V_2 \end{pmatrix}\,.$$

Calculate the mean vector and covariance matrix of $\vec{\boldsymbol{X}}$, and then standardize $\vec{\boldsymbol{X}}$.

2. Gaussian Random Vectors

Let $\vec{\boldsymbol{X}}$ be a random n-vector. We say that $\vec{\boldsymbol{X}}$ is a *Gaussian random n-vector* if for all non-random n-vectors $\vec{\boldsymbol{a}}$, the random variable $\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{X}}$ is normally distributed. Note that

$$\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{X}} = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n\,,$$

so this condition simply says that $\vec{\boldsymbol{X}}$ is a Gaussian random vector if any linear combination of its coordinates is normally distributed. In particular, the individual coordinates $X_1, X_2, \ldots, X_n$ must all be normally distributed, but that alone is not enough for $\vec{\boldsymbol{X}}$ to be Gaussian.

Here is how you get Gaussian random vectors. Start with the random n-vector $\vec{\boldsymbol{Z}}$ whose coordinates are independent random variables, all of which have the standard normal distribution. Then it is easy to see that $\vec{\boldsymbol{Z}}$ has mean vector $\vec{\boldsymbol{0}}$ and covariance matrix I_n. We know from the text that any sum of independent normal random variables is itself normal, so $\vec{\boldsymbol{Z}}$ satisfies the definition of a Gaussian random vector. This is the *standard normal (or Gaussian) random n-vector.*

Let Σ be a non-random $m \times n$ matrix and let $\vec{\boldsymbol{\mu}}$ be a non-random m-vector. Then the random m-vector

$$\vec{\boldsymbol{X}} = \Sigma\vec{\boldsymbol{Z}} + \vec{\boldsymbol{\mu}}$$

has mean vector μ and covariance matrix $\Sigma\Sigma^T$. Let $\vec{\boldsymbol{a}}$ be a non-random m-vector. Then

$$\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{X}} = (\vec{\boldsymbol{a}}\Sigma) \cdot \vec{\boldsymbol{Z}} + \vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{\mu}}\,.$$

Since $\vec{\boldsymbol{a}}\Sigma$ is a non-random n-vector, we know that $(\vec{\boldsymbol{a}}\Sigma) \cdot \vec{\boldsymbol{Z}}$ must be normally distributed. And since $\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{\mu}}$ is just a constant, it follows that $\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{X}}$ is normally distributed. Therefore, by definition, $\vec{\boldsymbol{X}}$ is a Gaussian random m-vector. This procedure shows us how we can get a Gaussian random vector with any desired mean vector and covariance matrix.

The preceding argument also works to prove the following fact: if $\vec{\boldsymbol{X}}$ is a Gaussian random n-vector with mean vector $\vec{\boldsymbol{\mu}}$ and covariance matrix Γ, and if A is a non-random $m \times n$ matrix and $\vec{\boldsymbol{b}}$ is a non-random m-vector, then the random m-vector

$$Y = A\vec{\boldsymbol{X}} + \vec{\boldsymbol{b}}$$

is a Gaussian random vector with mean vector $A\vec{\boldsymbol{\mu}} + \vec{\boldsymbol{b}}$ and covariance matrix $A\Gamma A^T$.

To summarize:

(i) If $\vec{\boldsymbol{Z}}$ is a random n-vector with independent coordinates that have the standard normal distribution, then $\vec{\boldsymbol{Z}}$ is a Gaussian random n-vector, called the standard Gaussian random n-vector.

(ii) if $\vec{\boldsymbol{Z}}$ is a standard Gaussian random n-vector, Σ a non-random $m \times n$ matrix and $\vec{\boldsymbol{\mu}}$ a non-random m-vector, then $\vec{\boldsymbol{X}} = \Sigma\vec{\boldsymbol{Z}} + \vec{\boldsymbol{\mu}}$ is a Gaussian random m-vector with mean vector $\vec{\boldsymbol{\mu}}$ and covariance matrix $\Sigma\Sigma^T$.

(iii) if $\vec{\boldsymbol{X}}$ is any Gaussian random n-vector, then for any non-random $m \times n$ matrix A and non-random m-vector $\vec{\boldsymbol{b}}$, the random m-vector $A\vec{\boldsymbol{X}} + \vec{\boldsymbol{b}}$ is a Gaussian random m-vector.

(iv) Theorem 1 shows how to use matrix multiplication and vector addition to turn any Gaussian random vector into a standard Gaussian random vector.

It turns out that this is all you need to know about Gaussian random vectors, because the distribution of a Gaussian random vector is uniquely determined by its mean vector and covariance matrix. This fact requires advanced techniques for its proof, so we won't give the proof here. But we state it as a theorem:

Theorem 2. *The distribution of a Gaussian random n-vector $\vec{\boldsymbol{X}}$ is uniquely determined by its mean vector $\vec{\boldsymbol{\mu}}$ and covariance matrix Γ. If Γ is invertible, the joint probability density function of $\vec{\boldsymbol{X}}$ is*

$$f(x_1, x_2, \ldots, x_n) = f(\vec{\boldsymbol{x}}) = \frac{1}{\sqrt{2\pi}^n \sqrt{\det(\Gamma)}} \exp(-\frac{1}{2}(\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})^T \Gamma^{-1} (\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}}))$$

If Γ is a diagonal matrix, then the coordinates of $\vec{\boldsymbol{X}}$ are independent, and for each i, the i^{th} coordinate X_i has the distribution $\mathcal{N}(\mu_i, \Gamma_{ii})$.

If we combine Theorems 1 and 2, we see that

> **Every Gaussian random vector has independent coordinates if you look at it from the right direction!**

Exercise 9. If Γ is a diagonal matrix with strictly positive diagonal elements, show that the joint density function given in Theorem 2 is indeed the joint density function of n independent random variables $X_1, X_2, \ldots, X_n$ such that for each i, X_i has the distribution $\mathcal{N}(\mu_i, \Gamma_{ii})$.

Exercise 10. (i) Which of the following two matrices is a covariance matrix?

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & \sqrt{2} \\ -1 & \sqrt{2} & 2 \end{pmatrix} \quad \begin{pmatrix} 1 & \sqrt{2} & 0 \\ \sqrt{2} & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

(ii) Using three independent standard normal random variables Z_1, Z_2, Z_3, create a random vector $\vec{\boldsymbol{X}}$ whose mean vector is $(-1, 1, 0)^T$ and whose covariance matrix is from (i).

Example 1. Let

$$\Gamma = \begin{pmatrix} \frac{11}{3} & -\frac{8}{3} & \frac{8}{3} \\ -\frac{8}{3} & \frac{11}{3} & -\frac{8}{3} \\ \frac{8}{3} & -\frac{8}{3} & \frac{11}{3} \end{pmatrix}$$

It turns out that the eigenvalues of Γ are $\lambda_1 = 9$, $\lambda_2 = \lambda_3 = 1$, with corresponding orthogonal eigenvectors

$$\vec{\boldsymbol{v}}_1 = \begin{pmatrix} \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \end{pmatrix} \quad \vec{\boldsymbol{v}}_2 = \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \quad \vec{\boldsymbol{v}}_3 = \begin{pmatrix} -\frac{\sqrt{6}}{3} \\ -\frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{6} \end{pmatrix}$$

Let O be the orthogonal matrix whose columns are these eigenvectors, and let D be the diagonal matrix whose diagonal elements are the eigenvalues $9, 1, 1$. Then $\sqrt{D}$ is the diagonal matrix whose diagonal elements are $3, 1, 1$. You should check that all of these statements are correct, or at least check some of them and know how to check the rest.

Since Γ is positive definite, we can think of Γ as a covariance matrix, with corresponding standard deviation matrix

$$\Sigma = O\sqrt{D} = \begin{pmatrix} \sqrt{3} & 0 & -\frac{\sqrt{6}}{3} \\ -\sqrt{3} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \sqrt{3} & \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} \end{pmatrix}$$

You should check that $\Gamma = \Sigma\Sigma^T$.

Now, if $\vec{\boldsymbol{X}}$ is a random 3-vector with covariance matrix Γ and mean vector, say,

$$\vec{\boldsymbol{\mu}} = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix},$$

then $\vec{\boldsymbol{Y}} = O^T(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}})$ is uncorrelated, with covariance matrix D and mean vector $\vec{\mathbf{0}}$:

$$\vec{\boldsymbol{Y}} = O^T(\vec{\boldsymbol{X}} - \vec{\boldsymbol{\mu}}) = \begin{pmatrix} \frac{\sqrt{3}}{3} & -\frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{6}}{3} & -\frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} \end{pmatrix} \begin{pmatrix} X_1 - 1 \\ X_2 \\ X_3 + 2 \end{pmatrix}$$

Doing the matrix multiplication gives

$$\vec{\boldsymbol{Y}} = \begin{pmatrix} \frac{\sqrt{3}}{3}(X_1 - X_2 + X_3 + 1) \\ \frac{\sqrt{2}}{2}(X_2 + X_3 + 2) \\ \frac{\sqrt{6}}{6}(-2X_1 - X_2 + X_3 + 4) \end{pmatrix}$$

You should be able to check directly that this is correct, and also that this random vector is uncorrelated and has mean vector $\vec{\mathbf{0}}$. In order to fully standardize the random vector $\vec{\boldsymbol{X}}$, we need to multiply $\vec{\boldsymbol{Y}}$ by $\sqrt{D^{-1}}$ to get

$$\vec{\boldsymbol{Z}} = \sqrt{D^{-1}}\vec{\boldsymbol{Y}} = \begin{pmatrix} \frac{\sqrt{3}}{9}(X_1 - X_2 + X_3 + 1) \\ \frac{\sqrt{2}}{2}(X_2 + X_3 + 2) \\ \frac{\sqrt{6}}{6}(-2X_1 - X_2 + X_3 + 4) \end{pmatrix}$$

This random vector $\vec{\boldsymbol{Z}}$ has mean vector $\vec{\mathbf{0}}$ (that's easy to check) and covariance matrix I_3. Let's do a little checking of the covariance matrix. According to the rules we learned about covariances of ordinary random variables,

$$\begin{aligned} \operatorname{Var}(Z_1) = \frac{1}{27}[&\operatorname{Var}(X_1) + \operatorname{Var}(X_2) + \operatorname{Var}(X_3) \\ &- 2\operatorname{Cov}(X_1, X_2) + 2\operatorname{Cov}(X_1, X_3) - 2\operatorname{Cov}(X_2, X_3)]\,. \end{aligned}$$

We can obtain these variances and covariances from the covariance matrix Γ, so

$$\operatorname{Var}(Z_1) = \frac{1}{27}\left[\frac{11}{3} + \frac{11}{3} + \frac{11}{3} + \frac{16}{3} + \frac{16}{3} + \frac{16}{3}\right] = 1\,.$$

That's what we expected. Let's try one of the covariances:

$$\begin{aligned}\operatorname{Cov}(Z_2, Z_3) = \frac{\sqrt{12}}{12}[&-2\operatorname{Cov}(X_2, X_1) - \operatorname{Var}(X_2) + \operatorname{Cov}(X_2, X_3) \\ &- 2\operatorname{Cov}(X_3, X_1) + \operatorname{Cov}(X_3, X_2) + \operatorname{Var}(X_3)]\end{aligned}$$

Filling in the values of the variances and covariances from Γ, you can easily check that this equals 0, as expected.

Let's do one more thing. Suppose we started with a standard Gaussian random 3-vector $\vec{\boldsymbol{Z}}$. How can we use it to create a random Gaussian 3-vector $\vec{\boldsymbol{X}}$ with covariance matrix Γ and mean vector $\vec{\boldsymbol{\mu}}$? We simply multiply $\vec{\boldsymbol{Z}}$ by the matrix Σ and then add $\vec{\boldsymbol{\mu}}$:

$$\vec{\boldsymbol{X}} = \Sigma\vec{\boldsymbol{Z}} + \vec{\boldsymbol{\mu}} = \begin{pmatrix} \sqrt{3} & 0 & -\frac{\sqrt{6}}{3} \\ -\sqrt{3} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} \\ \sqrt{3} & \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}.$$

You can carry out the indicated operations to find that, for example, $X_1 = \sqrt{3}Z_1 - \sqrt{6}Z_3/3 + 1$.

3. Random Walk

Let $X_1, X_2, \ldots$ be independent random variables, all having the same distribution, each having mean μ and variance σ^2. For $n = 1, 2, 3, \ldots$, define

$$W_n^{(1)} = X_1 + X_2 + \cdots + X_n\,.$$

Let $\vec{\boldsymbol{W}}$ be the following random vector (with infinitely many coordinates):

$$\vec{\boldsymbol{W}}^{(1)} = (W_1^{(1)}, W_2^{(1)}, W_3^{(1)}, \ldots)\,.$$

This is the *random walk* with *steps* $X_1, X_2, X_3, \ldots$. (The reason superscript (1) will be clear later.) We often think of the subscript as time, so that $W_n^{(1)}$ is the *position* of the random walk $\vec{\boldsymbol{W}}^{(1)}$ at *time* n. That is, we think of $\vec{\boldsymbol{W}}^{(1)}$ as a (random) dynamical system in discrete time, and at each time n, the system takes the step X_n to move from its previous position $W_{n-1}^{(1)}$ to its current position $W_n^{(1)}$. In those cases where it is desirable to talk about time 0, then we define $W_0^{(1)} = 0$. But we typically don't include this value as one of the coordinates of $\vec{\boldsymbol{W}}^{(1)}$.

It is easy to calculate the mean vector $\vec{\boldsymbol{\mu}}$ and covariance matrix of $\vec{\boldsymbol{W}}^{(\mathbf{1})}$:

$$\vec{\boldsymbol{\mu}} = \begin{pmatrix} \mu \\ 2\mu \\ 3\mu \\ \vdots \end{pmatrix} \quad \text{and} \quad \Gamma = \begin{pmatrix} \sigma^2 & \sigma^2 & \sigma^2 & \dots \\ \sigma^2 & 2\sigma^2 & 2\sigma^2 & \dots \\ \sigma^2 & 2\sigma^2 & 3\sigma^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Thus, $\vec{\boldsymbol{\mu}}_n = n\mu$ and $\Gamma_{ij} = \min(i,j)\sigma^2$.

Exercise 11. Verify these formulas for the mean vector and covariance matrix of $\vec{\boldsymbol{W}}^{(\mathbf{1})}$.

Exercise 12. The *simple symmetric random walk* has steps X_i such that $P(X_i = 1) = P(X_i = -1) = 1/2$. Find the mean vector and covariance matrix of this random walk.

In the next section we will want to make the transition from random walks in discrete time to Brownian motion in continuous time. The key to this transition is to chop discrete time up into smaller and smaller pieces. Fix a value $\Delta > 0$. For $n = 1, 2, 3, \dots$ and $t = n\Delta$, define

$$W_t^{(\Delta)} = Y_1 + Y_2 + \cdots + Y_n \,,$$

where for $k = 1, 2, 3, \dots,$

$$Y_k = \sqrt{\Delta}(X_k - \mu) + \Delta\mu \,.$$

Note that the random walk $\vec{\boldsymbol{W}}^{(\mathbf{1})}$ defined earlier agrees with the case $\Delta = 1$.

We think of the random variables Y_k as the steps taken by a random walk at times t that are multiples of Δ. Since the random variables X_k are assumed to be independent, and have the same distribution, with mean μ and variance σ^2, the random variables Y_k are also independent, and they have the same distribution as each other, but their common mean is $\Delta\mu$ and their common variance is $\Delta\sigma^2$. Thus, if $t = n\Delta$, the random variable $W_t^{(\Delta)}$ has mean $t\mu$ and variance $t\sigma^2$. And if $s = m\Delta$, then $\mathrm{Cov}(W_s^{(\Delta)}, W_t^{(\Delta)}) = \min(s,t)\sigma^2$.

Thus, for any $\Delta > 0$, we have created a random walk $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ that takes steps at times that are multiples of Δ. The mean vector $\vec{\boldsymbol{\mu}}$ of $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ has coordinates $\vec{\boldsymbol{\mu}}_t = t\mu$ covariance matrix Γ with elements $\Gamma_{s,t} = \min(s,t)\sigma^2$, for all times s, t that are multiples of Δ. Note that these do not depend on Δ, so we might expect something good to happen as we try to make the transition to continuous time by letting

$\Delta \to 0$. In the next section, we will see exactly what it is that does happen.

Exercise 13. The following numbers are the results of 32 rolls of a standard 6-sided die:

$$1, 5, 3, 4, 6, 2, 1, 1, 6, 1, 2, 4, 3, 5, 4, 5, 5, 1, 5, 2, 5, 2, 3, 1, 1, 1, 5, 6, 1, 2, 2, 1$$

Let $X_1, X_2, \ldots, X_{32}$ be these numbers. For $\Delta = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$, draw graphs of $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ for appropriate values of t between 0 and 4.

4. The Central Limit Theorem for Random Walks

We start by considering a special collection of random walks, the *Gaussian random walks*. These are defined as in the previous section, with the additional assumption that the steps X_i are normally distributed, so that for each Δ, $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ is a Gaussian random vector (make sure you understand why this is true). If each step has mean μ and variance σ^2, then $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ is the *Gaussian random walk* with *drift* μ, *diffusion coefficient* σ and *time step size* Δ.

If the drift is 0 and the diffusion coefficient is 1, then the steps of $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ all have mean 0 and variance Δ, and then $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ is called the *standard Gaussian random walk* with time step size Δ. Note that for a fixed time step size Δ, if $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ is a Gaussian random walk with drift μ and diffusion coefficient σ, then the random walk whose position at time t is

$$\frac{1}{\sigma}(W_t^{(\Delta)} - \mu t)$$

is the standard Gaussian random walk with time step size Δ (for times t that are multiples of Δ).

On the other hand, if $\vec{\boldsymbol{W}}^{(\boldsymbol{1})}$ is a standard Gaussian random walk with time step size Δ, then we get a Gaussian random walk with drift μ and diffusion coefficient σ by going the other way. That is, the random walk whose position at time t is

$$\sigma W_t^{(\Delta)} + \mu t$$

is a Gaussian random walk with drift μ and diffusion coefficient σ. So just as we can easily convert back and forth between a standard normally distributed random variable and a normally distributed random variable with mean μ and variance σ^2, we can also easily convert back and forth between a standard Gaussian random walk and a Gaussian random walk with drift μ and diffusion coefficient σ.

What do we know about Gaussian random walks? In the following, the times (such as s and t) are all multiples of Δ. If $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ is a Gaussian random walk with drift μ, diffusion coefficient σ and time step size Δ, then:

(i) For each $t > s \geq 0$, the *increment* $W_t^{(\Delta)} - W_s^{(\Delta)}$ has the distribution $\mathcal{N}(\mu(t-s), \sigma^2(t-s))$.
(ii) If $(s_1, t_1), (s_2, t_2), \ldots, (s_n, t_n)$ are disjoint open intervals in $(0, \infty)$, then the increments
$$W_{t_1}^{(\Delta)} - W_{s_1}^{(\Delta)}, W_{t_2}^{(\Delta)} - W_{s_2}^{(\Delta)}, \ldots, W_{t_n}^{(\Delta)} - W_{s_n}^{(\Delta)}$$
are independent.

Since the distribution of an increment of a Gaussian random walk depends only on the time difference $t - s$, we say that the increments are *stationary*. We summarize the two properties together by saying that a Gaussian random walk has *stationary, independent increments*.

Note that the distributions of the increments do not depend on Δ, except for the assumption that the times involved must be multiples of Δ. It seems reasonable to guess that if we let $\Delta \to 0$, we will get a dynamical system $\mathbf{W} = \{W_t, t \geq 0\}$ that is defined for all nonnegative times t, and this *stochastic process* will have stationary independent normally distributed increments. Furthermore, it seems reasonable to guess that W_t, the position at time t, depends continuously on t. That is, there is a stochastic process $\mathbf{W}$ with the following properties:

(i) The position W_t depends continuously on t for $t \geq 0$.
(ii) For each $t > s \geq 0$, the increment $W_t - W_s$ has the distribution $\mathcal{N}(\mu(t-s), \sigma^2(t-s))$.
(iii) If $(s_1, t_1), (s_2, t_2), \ldots, (s_n, t_n)$ are disjoint open intervals in $(0, \infty)$, then the increments
$$W_{t_1} - W_{s_1}, W_{t_2} - W_{s_2}, \ldots, W_{t_n} - W_{s_n}$$
are independent.

Einstein first talked about this in a meaningful mathematical way, and Norbert Wiener was the first to establish the existence of $\mathbf{W}$ with some rigor. The stochastic process $\mathbf{W}$ is called *Brownian motion* with drift μ and diffusion coefficient σ. When $\mu = 0$ and $\sigma = 1$, it is called *standard Brownian motion*, or the *Wiener process*.

Here is a summary of what we have so far: if we fix the drift μ and the diffusion coefficient σ, then as the time step size Δ goes to 0, the corresponding sequence of Gaussian random walks converges (in a rigorous mathematical sense that will not be made explicit here) to Brownian motion with drift μ and diffusion coefficient σ.

Here is a wonderful fact: the statement in the previous paragraph remains true, even if we drop the Gaussian assumption. It doesn't matter what the original distribution of the steps $X_1, X_2, X_3, \ldots$ is, when Δ goes to 0, we get Brownian motion. This is the "Central Limit Theorem" for random walks.

Why is this result true? The first point is that for any random walk, we know no matter what the time step size Δ equals, the increments of the random walk $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ for non-overlapping time intervals are independent, and we know that the means, variances, and covariances of increments do not depend on Δ. The second point is that for any fixed $t > s \geq 0$, the ordinary Central Limit Theorem tells us that the increment $W_t^{(\Delta)} - W_s^{(\Delta)}$ is approximately normal when Δ is small, because this increment is the sum of a lot of independent random variables that all have the same distribution. This means that for small enough $\Delta > 0$, it is hard to distinguish a non-Gaussian random walk from a Gaussian random walk. In the limit as $\Delta \to 0$, non-Gaussian and Gaussian random walks become indistinguishable.

One final comment. For any $\Delta > 0$, the relationship between $\vec{\boldsymbol{W}}^{(\boldsymbol{1})}$ and $\vec{W}^{(\Delta)}$ is merely a change of coordinates. If $\mu = 0$, then the change of coordinates is particularly simple: speed up time by the factor $1/\Delta$ and scale space by $\sqrt{\Delta}$.

$$W_t^{(\Delta)} = \sqrt{\Delta}\left(W_{t/\Delta}^{(1)}\right) .$$

This means that if you start with a graph of the original random walk $\vec{\boldsymbol{W}}^{(\boldsymbol{1})}$ and do the proper rescaling, you can get a graph of $\vec{\boldsymbol{W}}^{(\boldsymbol{\Delta})}$ for really small Δ, and this graph will look at lot like a graph of Brownian motion! This works whether your random walk is the simple symmetric random walk, with $P(X_1 = 1) = P(X_1 = -1) = 1/2$, or a standard Gaussian random walk.

5. Geometric Brownian Motion

Let's recall a few facts about the binomial tree model. We assume there is a risk-free interest rate r, and we will let $\Delta > 0$ be the length of a time step. When the stock price goes up, it is multiplied by a factor u, and when it goes down, it is multiplied by a factor d. There is a "probability" p of going up at each time step that comes from the No Arbitrage Theorem, given by the formula:

$$p = \frac{e^{r\Delta} - d}{u - d} .$$

We will assume that

$$u = e^{\sigma\sqrt{\Delta}} \quad \text{and} \quad d = 1/u = e^{-\sigma\sqrt{\Delta}}\,.$$

The idea is that at each time step, the stock price is multiplied by a random quantity $e^{R\sqrt{\Delta}}$, which equals u with probability p and d with probability $1-p$. Therefore, $R = \sigma$ with probability p and $-\sigma$ with probability $1-p$. This makes R like the step of a random walk. In this setting, σ is called the *volatility*.

Let's calculate the approximate mean and variance of R. In order to do this, we will repeatedly use the approximation $e^x \approx 1+x$ when x is small. Then we get $e^{r\Delta} \approx 1 + r\Delta$, $u \approx 1 + \sigma\sqrt{\Delta}$ and $d \approx 1 - \sigma\sqrt{\Delta}$, so

$$p \approx \frac{r\Delta + \sigma\sqrt{\Delta}}{2\sigma\sqrt{\Delta}} = \frac{1}{2} + \frac{r\sqrt{\Delta}}{2\sigma}\,.$$

We know that $E(R) = \sigma(2p-1)$, so we get $E(R) \approx r\sqrt{\Delta}$. We also know that

$$\mathrm{Var}(R) = 4\sigma^2 p(1-p) \approx \sigma^2\,.$$

Now we look at the geometric random walk, which we will call $\vec{\boldsymbol{S}}^{(\boldsymbol{\Delta})}$:

$$S_t^{(\Delta)} = S_0 \exp(\sqrt{\Delta}(R_1 + R_2 + \cdots + R_{t/\Delta})\,,$$

where $R_1, R_2, \ldots$ are the steps of the underlying random walk. When the time step size Δ is small, the expression in the exponent is approximately a Brownian motion $\vec{\boldsymbol{W}}$. Note that

$$E(W_t) = \sqrt{\Delta}E(R)/\Delta \approx r\,,$$

and

$$\mathrm{Var}(W_t) = \Delta(t/\Delta)\,\mathrm{Var}(R) \approx \sigma^2\,.$$

This is exactly what we expect: using the probabilities given by the No Arbitrage Theorem, the approximate average rate of growth should equal the risk-free rate and we have arranged things so that σ^2 is the approximate variance of the growth rate.

It turns out that we have been a little sloppy with our approximations. Brownian motion is wild enough so that the approximation $e^x \approx 1+x$ isn't good enough. Instead, we should use $e^x \approx 1 + x + x^2/2$ in some places. When this is done, the diffusion coefficient of the underlying Brownian motion W_t remains equal to σ, but the drift is changed. The key change occurs in the approximation for p:

$$p \approx \frac{r\Delta + \sigma\sqrt{\Delta} - \frac{1}{2}\sigma^2\Delta}{2\sigma\sqrt{\Delta}} = \frac{1}{2} + \frac{r\sqrt{\Delta}}{2\sigma} - \frac{\sigma\sqrt{\Delta}}{4}\,.$$

Once this change is made, we get $E(R) \approx (r - \sigma^2/2)\sqrt{\Delta}$.

6. Calculations with Brownian motion

Throughout this section, W_t will be the position at time t of a standard Brownian motion, X_t will be the position at time t of a Brownian motion with drift μ and diffusion coefficient σ, and S_t will be the position at time t of a geometric Brownian motion with risk-free interest rate r and volatility σ. We will assume that W_0 and X_0 are always equal to 0, and that the initial stock price S_0 is a positive quantity.

In order to do calculations with a general Brownian motion X_t, it is often best to turn it into a standard Brownian motion:

$$W_t = \frac{X_t - \mu t}{\sigma}\,.$$

Similarly, it can be useful to turn a geometric Brownian motion S_t into a standard Brownian motion. We could do this in two steps. First take the logarithm of S_t and subtract the logarithm of the initial stock price to get a Brownian motion X_t with drift $\mu = r - \sigma^2/2$ and diffusion coefficient σ:

$$X_t = \log(S_t/S_0) = \log(S_t) - \log(S_0)\,.$$

Then we get the standard Brownian motion by subtracting the drift and dividing by the diffusion coefficient:

$$W_t = \frac{X_t - rt + \frac{\sigma^2 t}{2}}{\sigma}\,.$$

The upshot is that for many purposes, if you can do probability calculations with standard Brownian motion, then you can do probability calculations for all Brownian motions, including geometric Brownian motion. The main exception to this is calculations that involve random times in some way. But we will avoid such calculations in this course. You will learn how to do those next year.

For the rest of this section, we will focus on doing some probability calculations with standard Brownian motion W_t. The simplest calculation has to do with the position of the Brownian motion at a specific time. For example, calculate $P(W_4 \geq 3)$. To do this, we use the fact that W_4 has the distribution $\mathcal{N}(0, 4)$. That means that $Z = W_4/2$ has the standard normal distribution, so

$$P(W_4 \geq 3) = P(Z \geq 3/2) = 1 - \Phi(1.5)\,,$$

where Φ is the standard normal cumulative distribution function.

The next simplest calculation has to do with non-overlapping increments. For such calculations, we use the fact that Brownian motion

has stationary independent increments. For example,

$$\begin{aligned} P(|W_1| \le 1 \text{ and } W_3 > W_2 \text{ and } W_5 < W_3 + 1) \\ &= P(-1 \le W_1 \le 1) P(W_3 - W_2 > 0) P(W_5 - W_3 < 1) \\ &= P(-1 \le Z \le 1) P(Z > 0) P(Z < \sqrt{2}/2) \\ &= (2\Phi(1) - 1) \left(\frac{1}{2}\right) \left(\Phi\left(\sqrt{\frac{1}{2}}\right)\right) . \end{aligned}$$

The hardest calculations are for probabilities involving the positions of Brownian motion at several different times. For these calculations, we use the joint distribution function. To keep things from getting too messy, we will restrict our attention to events involving only two different times. For example, suppose we want to calculate

$$P(W_2 < 1 \text{ and } W_5 > -2) .$$

To do this, we use the fact that the random 2-vector $(W_2, W_5)^T$ is a Gaussian random 2-vector with mean vector $\vec{\mathbf{0}}$ and covariance matrix $\begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}$. By Theorem 2, the joint density function of this random 2-vector is

$$f(x, y) = \frac{1}{2\pi\sqrt{6}} \exp\left(-\frac{1}{2} (x \; y) \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} \right)$$

Since

$$\begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{5}{6} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{pmatrix} ,$$

the quadratic expression in the exponential function is

$$-\frac{5}{12} x^2 + \frac{1}{3} xy - \frac{1}{6} y^2 .$$

Therefore

$$P(W_2 < 1 \text{ and } W_5 > -2) = \frac{1}{2\pi\sqrt{6}} \int_{-2}^{\infty} \int_{-\infty}^{1} \exp\left(-\frac{5}{12} x^2 + \frac{1}{3} xy - \frac{1}{6} y^2 \right) dx \, dy .$$

It is best to use computer software to calculate the final numerical answer. Another approach would be to make a rotation that would convert the Gaussian 2-vector into a vector with independent coordinates. This simplifies the joint density function but complicates the region of integration.

7. The fundamental stochastic differential equation

Recall that the basic differential equation for the growth of a bank account with interest rate r (compounded continuously) is

$$\frac{dS}{dt} = rS\,.$$

Let's write this equation in a different way:

$$dS_t = rS_t\,dt\,.$$

This emphasizes the way in which S depends on t, and also makes us think about the approximation to this differential equation in terms of increments:

$$\Delta S_t = S_{t+\Delta t} - S_t \approx rS_t\Delta t\,.$$

There is no randomness in this equation. It describes a "risk-free" situation. Its solution is $S_t = S_0\exp(rt)$.

Here is a way to introduce some randomness:

$$\Delta S_t \approx rS_t\Delta t + \sigma S_t(W_{t+\Delta t} - W_t) = rS_t\Delta t + \sigma S_t\,\Delta W_t\,, \tag{10}$$

where W_t represents the position of a standard Brownian motion at time t. The idea here is that the price of the stock S_t is affected in two ways: first, there is the growth that comes from the risk-free interest rate r, and then there is a random fluctuation which, like the growth due to interest, is expressed as a factor σW_t multiplied by the current price S_t. This factor can be positive or negative. Roughly speaking, one thinks of it as an accumulation of many tiny random factors, so that it should be (approximately) normally distributed. For reasons that go back to the No Arbitrage Theorem, these fluctuations should have mean 0. But the size of the fluctuations varies with the stock, and that is the reason for the "volatility" constant σ. You'll learn more about why this is a reasonable model next year.

As we let Δt go to 0, we get the *stochastic differential equation*

$$dS_t = rS_t\,dt + \sigma S_t\,dW_t\,. \tag{11}$$

This equation does not make sense in the ordinary way, because of the term dW_t. The paths of Brownian motion are so wild that a naive approach to understanding dW_t doesn't work. It took several geniuses (including Norbert Wiener and Kiyosi Ito) to find a way to make sense of this.

We won't go into detail here, but we will see a little bit of what goes on by checking that geometric Brownian motion with growth rate r and volatility σ is a solution to this stochastic differential equation. Thus, let X_t be a Brownian motion with drift $\mu = r - \sigma^2/2$ and diffusion

coefficient σ, and let $S_t = S_0 \exp(X_t)$, where S_0 is the initial price of the stock. To justify our claim that this choice for S_t gives a solution to (11), we check to see if it seems to work in the approximation (10).

$$\begin{aligned}(12)\quad \Delta S_t = S_{t+\Delta t} - S_t &= S_0 \exp(X_t + \Delta X_t) - \exp(X_t) \\ &= S_t(\exp(\Delta X_t) - 1) \approx S_t \left(\Delta X_t + \frac{1}{2}(\Delta X_t)^2 \right) .\end{aligned}$$

In this expression, we used the 2nd degree Taylor expansion of the exponential function to get our approximation of ΔS_t on the right side. Remember that $X_t = \mu t + \sigma W_t$, where W_t is a standard Brownian motion, so

$$\Delta X_t = \left(r - \frac{\sigma^2}{2} \right) \Delta t + \sigma \, \Delta W_t = r \, \Delta t + \sigma \, \Delta W_t - \frac{\sigma^2}{2} \, \Delta t \, .$$

We also have

$$(\Delta X_t)^2 = \sigma^2 (\Delta W_t)^2 + (\text{terms involving } (\Delta t)^2 \text{ and } \Delta t \, \Delta W_t) \, .$$

We saw last semester that the terms involving $(\Delta t)^2$ can be ignored in such approximations. It turns out that the terms involving $\Delta t \, \Delta W_t$ can also be ignored – they are obviously $o(\Delta t)$. But Brownian motion is wild enough so that $(\Delta W_t)^2$ cannot be ignored. Remember:

$$E((\Delta W_t)^2) = E((W_{t+\Delta t} - W_t)^2) = \text{Var}(W_{\Delta t}) = \Delta t \, ,$$

so $(\Delta W_t)^2$ is comparable to Δt. Putting everything into (12) gives

$$\Delta S_t \approx r S_t \, \Delta t + \sigma S_t \, \Delta W_t + \frac{\sigma^2}{2} ((\Delta W_t)^2 - \Delta t) \, .$$

It turns out that the difference $(\Delta W_t)^2 - \Delta t$ can also be ignored – it is a random quantity that is typically small compared to Δt. In other words, the random quantity $(\Delta W_t)^2 - \Delta t$ is typically $o(\Delta t)$. After removing that term, we see that our choice of S_t satisfies (10), as desired.

You can see that there is a lot of fancy mathematics hiding behind all of this. Welcome to the world of stochastic differential equations!