3 Brownian motion (1 lecture)

We mainly follow Oksendal's book: Stochastic Differential Equations: an Introduction with Applications (6th edition) in this section. The computer code in the first subsection is from D. J. Higham, An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations, SIAM Review, 43 (2001) 525–546. See the appendix of the book of Choe for an introduction to Matlab.

3.1 Computer experiments

```
randn('state',100)            % set the state of randn
T = 1; N = 500; dt = T/N;
dW = zeros(1,N);              % preallocate arrays for efficiency
W = zeros(1,N);
dW(1) = sqrt(dt)*randn;       % prepare the iteration W(j) = W(j-1) + dW(j)
W(1) = dW(1);
for j = 2:N                   % start the iteration
    dW(j) = sqrt(dt)*randn;   % general increment
    W(j) = W(j-1) + dW(j);
end
plot([0:dt:T],[0,W],'r-')     % plot W against t
xlabel('t','FontSize',16)
ylabel('W(t)','FontSize',16,'Rotation',0)
```

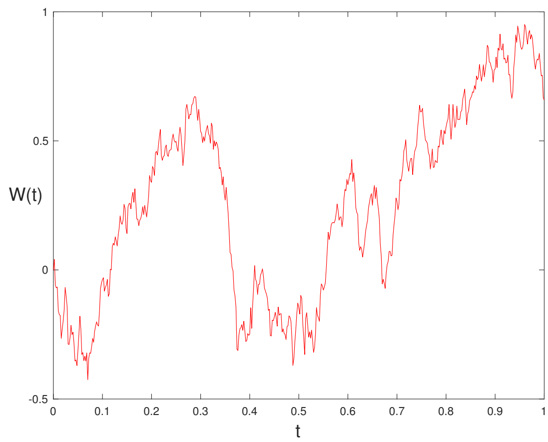


Figure 3.1: Discretized Brownian path

3.2 Probability space

Probability theory studies random experiments. The set of all possible outcomes of an experiment is called the sample space and is denoted by Ω. Any subset of Ω is called an event. The most basic operations on sets are union and intersection. Correspondingly, in practice, we need to study the intersection or union of different events. These needs motivate the introduction of σ-algebra (also called tribe) $\mathcal{F}$ that we have introduced before in Definition 2.5.

Definition 3.1 *The pair $(\Omega, \mathcal{F})$ we mentioned above is called a measurable space. A probability measure $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that*

i) $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$.

ii) whenever a sequence of sets $A_1, A_2, \cdots$ belongs to $\mathcal{F}$ and $\{A_i\}_{i=1}^{\infty}$ is disjoint (i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$), then

$$\mathbb{P}\left(\cup_{n=1}^{\infty} A_n\right) \overset{def}{=} \mathbb{P}\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.

Remark: Note that a measure or a probability measure is a mapping that maps a set to a number. You can compare that with the definition of a scalar function which is a mapping that maps a number to a number. A scalar random variable X is a mapping that maps an element in Ω to a number.

For practical reason, we often need to study $\mathbb{P}(X \in [a, b])$ or $\mathbb{P}(X \in (a, b])$ or $\mathbb{P}(X \in [a, b))$. Let $I = [a, b]$ or $(a, b]$ or $[a, b)$. I is the so called Borel set. To study its probability, the event

$$\{\omega \in \Omega : X(\omega) \in I\} \overset{\text{def}}{=} X^{-1}(I)$$

must belongs to $\mathcal{F}$ because only set that belongs to $\mathcal{F}$ has a $\mathbb{P}$ value. By Definition 2.8, this means X is $\mathcal{F}$-measurable. So, finally, we **define** a scalar random variable Y as an $\mathcal{F}$-measurable function $Y : \Omega \to \mathbb{R}$.

If there is a function $\rho_X(x)$ so that

$$\mathbb{P}(X \in I) = \int_I \rho_X(x)dx \tag{3.1}$$

for every Borel set $I \subset \mathbb{R}$ (you can think of I as an interval), we say $\rho_X(x)$ is the probability density function (pdf) of X and X is then called a continuous random variable. For example, if

$$\mathbb{P}(X \leq a) = \int_{-\infty}^{a} \underbrace{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}_{\text{pdf } \rho_X(x)} dx = \underbrace{F_X(a)}_{cdf}, \tag{3.2}$$

then we say that X is normally distributed with mean μ and standard deviation σ which can be written as $X \sim N(\mu, \sigma^2)$.

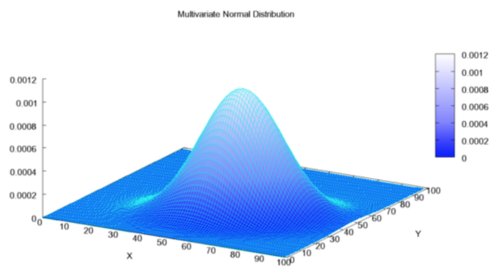
X can also be a vector in $\mathbb{R}^n$. For example, the prices of n different stocks form a n-dimensional vector. A random variable is called a continuous random variable, if we can find ρ_X, called the joint probability density function of X, such that

$$\mathbb{P}(X \in B) = \int_B \rho_X(x)dx \tag{3.3}$$

for any Borel set $B \subset \mathbb{R}^n$. You can safely think of B as a rectangular box. Note that $dx = dx_1 dx_2 \cdots dx_n$ and $\rho_X(x)$ maps $x \in \mathbb{R}^n$ to a value in $\mathbb{R}$. For example, if an n-dimensional vector X satisfies (3.3) with

$$\rho_X(x) = \frac{1}{(2\pi)^{n/2}|\det\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right), \tag{3.4}$$

then we say X is a n-dimensional normally distributed random variable, and write $X \sim N(\mu, \Sigma)$. Its components X_1, X_2, $\cdots$, X_n are called jointly normally distributed. In (3.4), $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$ are column vectors and $\Sigma \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix. Σ^{-1} is the inverse of Σ. The superscript $^\top$ denotes the transpose of a column vector or a matrix.



Let us briefly review some basic properties of normally distributed random variables. First, by the standard change of variable formula in multivariable integration, we can prove that if $X \sim N(\mu, \Sigma)$, then

$$Y = AX + \beta \sim N(A\mu + \beta, A\Sigma A^\top) \tag{3.5}$$

where A is an invertible $n \times n$ matrix and β is an n-dimensional column vector. In particular, Y, and also each component of it, are still normally distributed. Hence, by letting the first row of B be $(1, 1, ..., 1)$, we see that the summation of jointly normally distributed random variables are still normally distributed.

Using (3.5), one can manage to prove that

$$\mathbb{E}[(X-\mu)(X-\mu)^\top] = \Sigma, \quad i.e., \quad \mathbb{E}[(X_i-\mu_i)(X_j-\mu_j)] = \Sigma_{ij}, \tag{3.6}$$

where Σ_{ij} is the (i,j) entry of Σ. Hence Σ is the covariance matrix. The leads to a nice property about multivariate normally distributed random variable: once we know its mean and its variance, we know its pdf and hence we know everything about it.

One can prove that for discrete random variable Y, the expectation of $g(Y)$ (defined in Definition 2.19) can be computed in another two ways

$$\mathbb{E}[g(Y)] \stackrel{\text{def}}{=} \sum_{\omega\in\Omega} g(Y(\omega))\mathbb{P}(\omega) \stackrel{(a)}{=} \sum_{y_i} g(y_i)\mathbb{P}(Y=y_i) \stackrel{(b)}{=} \sum_{g_j} g_j\mathbb{P}(g(Y)=g_j), \tag{3.7}$$

with ω running through all possible outcomes in Ω, y_i running through all possible values of Y, g_j running through all possible values of g. The equalities are obviously since they are just three different ways to count when taking a weighted average of all the values $g(Y(\omega))$: (def) for different ω, (a) for different $Y(\omega)$, (b) for different $g(Y(\omega))$. (See for example, Ross, First course in Probability, 9th edition, Pages 167, 133.)

Example 3.1 *Suppose that two independent flips of a coin that comes up heads with probability p are made, and let Y denote the number of heads obtained. Let $g(y)=y^2$. We can compute $\mathbb{E}[g(Y)]$ in three different ways.*

$$\begin{aligned}
&\sum_{\omega\in\Omega} g(Y(\omega))\mathbb{P}(\omega)\\
=&(Y(HH))^2\mathbb{P}(HH)+(Y(HT))^2\mathbb{P}(HT)+(Y(TH))^2\mathbb{P}(TH)+(Y(TT))^2\mathbb{P}(TT)\\
=&2^2p^2+1^2p(1-p)+1^2(1-p)p+0(1-p)^2=2p^2+2p.\\
&\sum_{y_i\in\{0,1,2\}} g(y_i)\mathbb{P}(Y=y_i)\\
=&0^2\mathbb{P}(Y=0)+1^2\mathbb{P}(Y=1)+2^2\mathbb{P}(Y=2)\\
=&0(1-p)^2+2p(1-p)+4p^2=2p^2+2p.\\
&\sum_{g_j\in\{0,1,4\}} g_j\mathbb{P}(g(Y)=g_j)\\
=&0\mathbb{P}(g(Y)=0)+1\mathbb{P}(g(Y)=1)+4\mathbb{P}(g(Y)=4)\\
=&0(1-p)^2+2p(1-p)+4p^2=2p^2+2p.
\end{aligned}$$

Normally, in an undergraduate textbook on probability, for continuous random variable Y with pdf $\rho_Y(y)$, the expectation $\mathbb{E}[g(Y)]$ can be defined as (See for example, Ross, First course in Probability, 9th edition, Pages 196)

$$\mathbb{E}[g(Y)] = \int_{\mathbb{R}} g(y)\rho_Y(y)dy. \tag{3.8}$$

As the right hand side is the limit of the Riemann sum $\sum_{i=1}^N g(y_i)\underbrace{\rho_Y(y_i)(y_{i+1}-y_i)}_{\approx\mathbb{P}(y_i\le Y<y_{i+1})}$, (3.8) is a generalization of $\mathbb{E}[g(Y)]=\sum_{y_i} g(y_i)\mathbb{P}(Y=y_i)$ in (3.7). To use (3.8), one needs to know

$\rho_Y(y)$. You may have wondered whether we can generalize $\mathbb{E}[g(Y)] = \sum_{\omega\in\Omega} g(Y(\omega))\mathbb{P}(\omega)$ in (3.7) to the continuous random variable case so that we do not need to know precisely the pdf of Y when computing $\mathbb{E}[g(Y)]$. The answer is the following definition

Definition 3.2 *If $\int_\Omega |g(Y(\omega))|d\mathbb{P}(\Omega) < \infty$, then the number*

$$\mathbb{E}[g(Y)] = \int_\Omega g(Y(\omega))d\mathbb{P}(\omega) \tag{3.9}$$

is called the <u>*expectation of $g(Y)$ w.r.t. $\mathbb{P}$.*</u> *See Section 1.1.3 of Shreve II for the precise definition of $\int_\Omega g(Y(\omega))d\mathbb{P}(\omega)$.*

One can prove that (Theorem 1.5.2 of Shreve II)

$$\mathbb{E}[g(Y)] = \underbrace{\int_\Omega g(Y(\omega))d\mathbb{P}(\omega)}_{\text{integrate/sum on }\omega} = \underbrace{\int_{\mathbb{R}^n} g(y)\rho_Y(y)dy}_{\text{integrate/sum on }y}. \tag{3.10}$$

From (3.9), one immediately obtains

$$\mathbb{E}[f(X) + g(Y)] = \mathbb{E}[f(X)] + \mathbb{E}[g(Y)] \tag{3.11}$$

for any random variables X and Y.

Example 3.2 *Later on, we will learn Brownian motion $W_t \sim N(x_0, t)$,*

$$\mathbb{E}[g(W_t)] = \int_\Omega g(W_t(\omega))d\mathbb{P}(\omega) = \int_{\mathbb{R}} g(y)\rho_{W_t}(y)dy = \int_{\mathbb{R}} g(y)dF_{W_t}(y),$$

where $\rho_{W_t}(y) = \frac{1}{\sqrt{2\pi t}}e^{-(y-x_0)^2/(2t)}$, $F_{W_t}(y) = \mathbb{P}(W_t \le y) = \int_{-\infty}^y \rho_{W_t}(s)ds$, $dF_{W_t}(y) = \frac{dF_{W_t}(y)}{dy}dy = \rho_{W_t}(y)dy$. (see (3.2)).

Example 3.3 *Let r and σ be constants. Define*

$$S(t) = S_0 e^{(r-\frac{1}{2}\sigma^2)t+\sigma W(t)} \tag{3.12}$$

with $W(t) \sim N(0,t)$. Let K be a positive constant. Show that

$$\mathbb{E}\left[e^{-rT}(S(T)-K)^+\right] = S_0 N(d_+) - Ke^{-rT}N(d_-) \tag{3.13}$$

with

$$d_\pm = \frac{1}{\sigma\sqrt{T}}\left[\log\frac{S_0}{K} + \left(r \pm \frac{\sigma^2}{2}\right)T\right], \tag{3.14}$$

$$N(y) = \frac{1}{\sqrt{2\pi}}\int_{-y}^{\infty} e^{-\frac{1}{2}z^2}dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}z^2}dz. \tag{3.15}$$

The above formula has appeared before in (2.11).

Proof: *$W(T) \sim N(0,T)$ and its pdf is $\frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}$. By (3.10),*

$$\begin{aligned}&\mathbb{E}\left[e^{-rT}(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma W(T)} - K)^+\right]\\ =&e^{-rT}\int_{-\infty}^{\infty}\left(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma x} - K\right)^+ \frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}dx.\end{aligned}$$

The integrand is zero if $S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma x} - K \le 0$. $S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma x} - K > 0$ if and only if $x > \frac{1}{\sigma}\left(\log\frac{K}{S_0} - (r-\frac{1}{2}\sigma^2)T\right) = \frac{1}{\sigma}\left(-\log\frac{S_0}{K} - (r-\frac{1}{2}\sigma^2)T\right) = -\sqrt{T}d_-$. Hence

$$\begin{aligned}\mathbb{E}\left[e^{-rT}(S(T)-K)^+\right] =&e^{-rT}\int_{-\sqrt{T}d_-}^{\infty}\left(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma x} - K\right)\frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}dx\\ \overset{y=x/\sqrt{T}}{=}&e^{-rT}\int_{-d_-}^{\infty}\left(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma\sqrt{T}y} - K\right)\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}dy.\end{aligned}$$

In the last step, we have used the change of variable $y = x/\sqrt{T}$. After removing the bracket, we get two terms

$$\begin{aligned}&\mathbb{E}\left[e^{-rT}(S(T)-K)^+\right]\\ =&S_0\int_{-d_-}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\sigma^2 T+\sigma\sqrt{T}y-\frac{1}{2}y^2}dy - Ke^{-rT}\int_{-d_-}^{\infty}\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}dy\\ =&S_0\int_{-d_-}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-\sigma\sqrt{T})^2}dy - Ke^{-rT}N\left(d_-\right).\end{aligned}$$

In the last step, we have used the function N introduced in (3.15). To handle the first term, we do a change of variable $z = y - \sigma\sqrt{T}$ and notice $-d_- - \sigma\sqrt{T} = -d_+$:

$$\begin{aligned}\mathbb{E}\left[e^{-rT}(S(T)-K)^+\right] =&S_0\int_{-d_- - \sigma\sqrt{T}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz - Ke^{-rT}N(d_-)\\ =&S_0 N(d_+) - Ke^{-rT}N(d_-). \qquad \square\end{aligned}$$

Remark: By the same calculation as above[21], one can show that if $S = e^X$ with $X \sim N(\mu, \gamma^2)$, then

$$\mathbb{E}[(S-K)^+] = e^{\mu+\frac{1}{2}\gamma^2}N(d_+) - KN(d_-) \tag{3.16}$$

with $d_- = \frac{\log\frac{1}{K}+\mu}{\gamma}$, $d_+ = d_- + \gamma$.

We say that two random variable X and Y are independent if for any intervals A and B on $\mathbb{R}$,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \tag{3.17}$$

[21]Set $S_0 = 1$, $\sigma\sqrt{T} = \gamma$, $rT - \frac{\sigma^2 T}{2} = \mu$. Then $rT + \frac{\sigma^2 T}{2} = \mu + \gamma^2$, $rT = \mu + \frac{1}{2}\gamma^2$.

If a continuous 2-dimensional random variable (X, Y) has joint pdf $\rho_{X,Y}(x, y)$, one can introduce the marginal pdf

$$\rho_X(x) = \int_{-\infty}^{\infty} \rho_{X,Y}(x, y) dy \quad \text{and} \quad \rho_Y(y) = \int_{-\infty}^{\infty} \rho_{X,Y}(x, y) dx. \tag{3.18}$$

Then one can manage to show that X and Y are independent if and only if

$$\rho_{X,Y}(x, y) = \rho_X(x)\rho_Y(y). \tag{3.19}$$

Indeed, as long as $\rho_{X,Y}(x, y)$ is separable, namely, there are functions $f(x)$ and $g(y)$ so that

$$\rho_{X,Y}(x, y) = f(x)g(y), \tag{3.20}$$

then X and Y are independent. Moreover, there is a constant C so that $f(x) = C\rho_X(x)$ and $g(y) = \frac{1}{C}\rho_Y(y)$.

If X and Y are independent, for any function f and g

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}^2} f(x)g(y)\rho_{X,Y}(x, y)dxdy \\ &= \int_{\mathbb{R}} f(x)\rho_X(x)dx \times \int_{\mathbb{R}} g(y)\rho_Y(y)dy = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]. \end{aligned} \tag{3.21}$$

Criteria (3.20) for independence can be generalized to n-dimensional random variables. For example, If $(X_1, \cdots, X_n) \stackrel{\text{def}}{=} X \sim N(\mu, \Sigma)$ and Σ is a diagonal matrix, i.e, $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2)$, it is easy to see that $\det\Sigma = \Pi_{i=1}^n \sigma_i^2$, $\Sigma^{-1} = \text{diag}(\sigma_1^{-2}, \sigma_2^{-2}, \cdots, \sigma_n^{-2})$. Then

$$\rho_X(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}|\det\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) = \Pi_{i=1}^n \rho_{X_i}(x_i), \tag{3.22}$$

with $\rho_{X_i} = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(x_i-\mu_i)^2/(2\sigma_i^2)}$. We then conclude that X_i's are independent of each other.

Definition 3.3 *(Definition 2.1.4 of Oksendal) A <u>stochastic process</u> is a parameterized collection of random variables*

$$\{X_t\}_{t\in\mathcal{T}}$$

defined on a prabability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assuming values in $\mathbb{R}^n$. [22]

The parameter space $\mathcal{T}$ is usually the half space $[0, \infty)$ or an interval $[0, T]$. Note that for each $t \in \mathcal{T}$ fixed, we have a random variable

$$\omega \to X_t(\omega), \qquad \omega \in \Omega.$$

[22]In this lecture notes, X_t and $X(t)$ denote the same thing. $X_t(\omega)$ can also be written as $X(t, \omega)$.

On the other hand, fixing $\omega \in \Omega$, we can consider the function

$$t \to X_t(\omega), \qquad t \in \mathcal{T},$$

which is called a path of X_t.

It may be useful for the intuition to think of t as "time" and each ω as an individual "particle" [23] or "the result of the experiment which randomly selects a particle". With this picture, $X_t(\omega)$ would represent the position (or result) at time t of the particle (experiment) ω.

Given an m-dimensional stochastic process $\{X_t\}_{t\in T}$, we can introduce the finite-dimensional distributio of $\{X_t\}_{t\in T}$ as the collection of the measures $\mu_{t_1,\cdots,t_n}$ defined on $\underbrace{\mathbb{R}^m \times \mathbb{R}^m \cdots \times \mathbb{R}^m}_{n} = \mathbb{R}^{mn}$, $n = 1, 2, \cdots$, such that

$$\mu_{t_1,\cdots,t_n}(\Gamma_1 \times \Gamma_2 \cdots \times \Gamma_n) = \mathbb{P}\left(X_{t_1} \in \Gamma_1, \cdots, X_{t_n} \in \Gamma_n\right), \qquad t_i \in \mathcal{T}. \tag{3.23}$$

Here $\Gamma_1, \cdots, \Gamma_n$ denote Borel sets (think of them as intervals when $m = 1$ or boxes when $m > 1$) in $\mathbb{R}^m$.

The family of all finite-dimensional distributions determine many (but not all) important properties of the process $\{X_t\}$. On the other hand, given a family $\{\nu_{t_1,\cdots,t_n} : n \in \mathbb{Z}_+, t_i \in \mathcal{T}\}$ of probability measures on $\mathbb{R}^{mn}$, can we construct a stochastic process $\{Y_t\}_{t\in\mathcal{T}}$ having $\{\nu_{t_1,\cdots,t_n} : n \in \mathbb{Z}_+, t_i \in \mathcal{T}\}$ as its finite-dimensional distribution? Here $\mathbb{Z}_+ = \{1, 2, 3, 4, \cdots\}$. One of Kolmogorov's famous theorems states that this can be done provided $\{\nu_{t_1,\cdots,t_n} : n \in \mathbb{Z}_+, t_i \in \mathcal{T}\}$ satisfies two natural consistency conditions:

Theorem 3.1 *(Kolmogorov's extension theorem. Theorem 2.1.5 of Oksendal's book) Suppose we have a collection of probability measures $\{\nu_{t_1,\cdots,t_n}\}$ that satisfies for any $n \in \mathbb{Z}_+$, for any $t_1, \cdots, t_n \in \mathcal{T}$, $\nu_{t_1,\cdots,t_n}$ is a probability measure on $\mathbb{R}^{mn}$ such that*

$$\nu_{t_{\sigma(1)},\cdots,t_{\sigma(n)}}(F_1 \times \cdots \times F_n) = \nu_{t_1,\cdots,t_n}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(n)}) \tag{3.24}$$

for all permutation σ on $\{1, 2, \cdots, n\}$ [24] and

$$\nu_{t_1,\cdots,t_n}(F_1 \times \cdots \times F_n) = \nu_{t_1,\cdots,t_n,t_{n+1},\cdots,t_{n+k}}(F_1 \times \cdots \times F_n \times \underbrace{\mathbb{R}^m \times \cdots \times \mathbb{R}^m}_{k}) \tag{3.25}$$

for all $k \in \mathbb{Z}_+$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an m-dimensional (meaning $X_t \in \mathbb{R}^m$) stochastic process $\{X_t\}$ on Ω such that

$$\nu_{t_1,\cdots,t_n}(F_1 \times F_2 \cdots \times F_n) = \mathbb{P}\left(X_{t_1} \in F_1, \cdots, X_{t_n} \in F_n\right), \tag{3.26}$$

for all $t_i \in \mathcal{T}$, $n \in \mathbb{Z}_+$ and all Borel sets $F_i \subset \mathbb{R}^m$.

[23] think about Brownian motion as random motion of particles suspended in a fluid.

[24] Note that we do not require $t_1 < t_2 < t_3 < \cdots$. (3.24) is a natural requirement. For example, $n = 3$, $\sigma : (1, 2, 3) \mapsto (2, 3, 1)$, which means $\sigma(1) = 2$, $\sigma(2) = 3$, $\sigma(3) = 1$. $\nu_{t_{\sigma(1)},t_{\sigma(2)},t_{\sigma(3)}}(F_1, F_2, F_3) \overset{\text{def of } \sigma(i)}{=} \nu_{t_2,t_3,t_1}(F_1, F_2, F_3) \overset{\text{since we want (3.26)}}{=} \mathbb{P}(X_{t_2} \in F_1, X_{t_3} \in F_2, X_{t_1} \in F_3) = \mathbb{P}(X_{t_1} \in F_3, X_{t_2} \in F_1, X_{t_3} \in F_2) = \nu_{t_1,t_2,t_3}(F_3, F_1, F_2) = \nu_{t_1,t_2,t_3}(F_{\sigma^{-1}(1)}, F_{\sigma^{-1}(2)}, F_{\sigma^{-1}(3)},)$.

3.3 Brownian motion in $\mathbb{R}^m$

In 1828 the Scottish botanist Robert Brown observed the pollen grains suspended in liquid performed an irregular motion. We will introduce a stochastic process $\{W_t(\omega)\}$, interpreted as the position at time t of the pollen grain ω. Since all our later studies are based on this $\{W_t\}$ which has some nice but sometimes quite weird properties, we'd better confirm that such a nice (or weird) thing does exist.

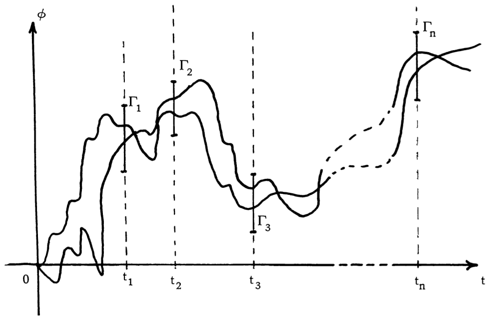


Figure 3.2: Picture taken from "Functional integration and partial differential equations" by Mark Freidlin. $x_0 = 0$. $m = 1$.

To construct $\{W_t\}_{t\geq 0}$, it suffices, by the Kolmogorov's extension theorem (Theorem 3.1) , to specify a family $\nu_{t_1,\cdots,t_n}$ of probability measures satisfying (3.24) and (3.25). Let $W = (W_{t_1}, W_{t_2}, \cdots, W_{t_n})$. If we are able to do so, we obtain (3.26), which says

$$\begin{aligned}\nu_{t_1,\cdots,t_n}(F_1 \times F_2 \cdots \times F_n) &= \mathbb{P}\left(W_{t_1} \in F_1, \cdots, W_{t_n} \in F_n\right) \\ &= \int_{F_1\times F_2\times\cdots\times F_n} \rho_W(w_1, w_2, \cdots, w_n) dw_1 dw_2 \cdots dw_n.\end{aligned}$$

In the last step, we have used (3.3). So, when we specify $\{\nu_{t_1,\cdots,t_n}\}$, what we are really doing is indeed specifing ρ_W, the joint pdf of W.

What kind of joint pdf of W do we want to obtain? We take $m = 1$ and $k = 2$ as an example.

(1) We want W_{t_1} to be normally distributed with mean x_0 (starting point) and variance t_1. So the pdf of W_{t_1} is $\rho_{W_{t_1}}(x_1) = \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{(x_1-x_0)^2}{2t_1}}$.

(2) Once we know $W_{t_1} = x_1$, we hope W_{t_2} is normally distributed around x_1, with variance $= t_2 - t_1$.

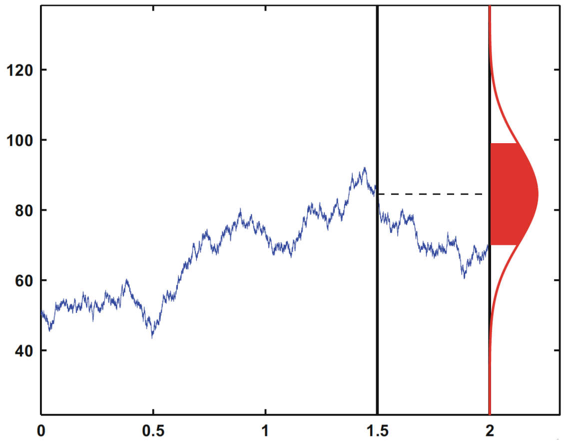


Figure 3.3: $t_1 = 1.5$. $t_2 = 2$. Given the value of W_{t_1}, W_{t_2} is normally distributed around the value of W_{t_1}.

Hence the pdf of W_{t_2}, given $W_{t_1} = x_1$, is $\rho_{W_{t_2}|W_{t_1}}(x_2|x_1) = \frac{1}{\sqrt{2\pi(t_2-t_1)}} e^{-\frac{(x_2-x_1)^2}{2(t_2-t_1)}}$. See Question 8 of Homework III for the notation.

Combining (1) and (2) and using the standard formula for conditional pdf

$$\rho_{W_{t_2}|W_{t_1}}(x_2|x_1) = \frac{\rho_{W_{t_1},W_{t_2}}(x_1, x_2)}{\rho_{W_{t_1}}(x_1)}$$

(see (3.52) in Question 8 of Homework III or Page 270 of A First Course in Probability by Ross), we get

$$\rho_{W_{t_1},W_{t_2}}(x_1, x_2) = \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{(x_1-x_0)^2}{2t_1}} \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{(x_2-x_1)^2}{2(t_2-t_1)}}. \tag{3.27}$$

So, if we introduce (now, consider general $m \in \mathbb{Z}_+$ and I_m is $m \times m$ identity matrix)

$$p(t, x, y) = (2\pi t)^{-m/2} e^{-\frac{|y-x|^2}{2t}} \sim \text{pdf of } N(x, tI_m), \qquad \text{for } x, y \in \mathbb{R}^m, \quad t > 0, \tag{3.28}$$

we see that $\rho_{W_{t_1},W_{t_2}}(x_1, x_2) = p(t_1, x_0, x_1)p(t_2 - t_1, x_1, x_2)$.

Now, we are ready to propose the $\nu_{t_1,\cdots,t_n}$ that we want: If $0 \le t_1 \le t_2 \le \cdots \le t_n$, define a measure $\nu_{t_1,\cdots,t_n}$ on $\mathbb{R}^{mn}$ by (Here Γ_i is any Borel set or m-dimensional box in $\mathbb{R}^m$. $x_0 \in \mathbb{R}^m$ is the starting point. $x_1, \cdots, x_n \in \mathbb{R}^m$ are the integrating/dummy variables.)

$$\begin{aligned} &\nu_{t_1,\cdots,t_n}(\Gamma_1 \times \cdots \times \Gamma_n) \\ &= \int_{\Gamma_1\times\cdots\times\Gamma_n} p(t_1, x_0, x_1)p(t_2 - t_1, x_1, x_2) \cdots p(t_n - t_{n-1}, x_{n-1}, x_n)dx_1 \cdots dx_n. \end{aligned} \tag{3.29}$$

Extend this definition to all finite sequences of t_i's by using (3.24). Since $\int_{\mathbb{R}^n} p(t,x,y)dy = 1$ for all $t > 0$, (3.25) holds, so by Theorem 3.1 there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $\{W_t\}_{t\geq 0}$ on Ω such that the finite-dimensional distributions of W_t are given by (3.29), i.e.,

$$\mathbb{P}(W_{t_1} \in \Gamma_1, \cdots, W_{t_n} \in \Gamma_n) = \int_{\Gamma_1\times\cdots\times\Gamma_n} p(t_1, x_0, x_1)\cdots p(t_n - t_{n-1}, x_{n-1}, x_n)dx_1\cdots dx_n. \tag{3.30}$$

Definition 3.4 *(Definition 2.2.1 of Oksendal) W_t is called Brownian motion starting at x_0.*

So, $p(t_1, x_0, x_1)\cdots p(t_n - t_{n-1}, x_{n-1}, x_n)$, is the joint pdf of the n-dimensional random variable $Y = (W_{t_1}, \cdots, W_{t_n})$. If we introduce $t_0 = 0$ and let $\delta t_i = t_i - t_{i-1}$, then

$$p(t_1, x_0, x_1)\cdots p(t_n - t_{n-1}, x_{n-1}, x_n) = (2^n\pi^n\delta t_1\cdots\delta t_n)^{-m/2}e^{-\sum_{i=1}^{n}\frac{1}{2\delta t_i}|x_i - x_{i-1}|^2}. \tag{3.31}$$

Let's summarize it as a theorem,

Theorem 3.2 *If $t_0 = 0$, given the initial position $W_{t_0} = x_0 \in \mathbb{R}^m$, the joint probability density function of $(W_{t_1}, \cdots, W_{t_n})$ is*

$$\frac{1}{(2\pi)^{mn/2}[(t_1 - t_0)(t_2 - t_1)\cdots(t_n - t_{n-1})]^{m/2}}\exp\left(-\frac{|x_1 - x_0|^2}{2(t_1 - t_0)} - \frac{|x_2 - x_1|^2}{2(t_2 - t_1)} - \cdots\frac{|x_n - x_{n-1}|^2}{2(t_n - t_{n-1})}\right). \tag{3.32}$$

So, the joint pdf indeed takes a rather simple form. In particular, it can be written as the product of pdfs of the increments. This reminds us of the criteria (3.20) for independence and motivates the following change of variable from X to Y:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} W_{t_1} \\ W_{t_2} - W_{t_1} \\ W_{t_3} - W_{t_2} \\ \vdots \\ W_{t_n} - W_{t_{n-1}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} W_{t_1} \\ W_{t_2} \\ W_{t_3} \\ \vdots \\ W_{t_n} \end{pmatrix} \stackrel{\text{def}}{=} AX \tag{3.33}$$

In fact, using the change of variable formula in multivariate calculus/probability [25], we see that

$$\begin{aligned}\rho_Y(y_1, y_2, \cdots, y_n) =&(2^n\pi^n\delta t_1\cdots\delta t_n)^{-m/2}e^{-\sum_{i=1}^n \frac{1}{2\delta t_i}|x_i-x_{i-1}|^2}\frac{1}{|\mathrm{det}A|}\\ =&(2^n\pi^n\delta t_1\cdots\delta t_n)^{-m/2}e^{-\sum_{i=1}^n \frac{1}{2\delta t_i}|y_i|^2}.\end{aligned}$$

We have used the fact that $\det A = 1$. $(2\pi\delta t_i)^{-m/2}e^{-\frac{1}{2\delta t_i}|y_i|^2}$ is the pdf of a $N(0, \delta t_i)$ random variable. ρ_Y is the production of them. By (3.3), we recognize that

Theorem 3.3 *The increments, $W_{t_i} - W_{t_{i-1}}$ and $W_{t_j} - W_{t_{j-1}}$, are independent as long as $i \neq j$. $W_{t_i} - W_{t_{i-1}} \sim N(0, t_i - t_{i-1})$.*

The proof that we have just presented is FYI only, and won't be tested. I just want to show you how the independence of the increments is hidden in our construction at the beginning. You should be able to apply the theorem, for example, to perform the calculation in (3.36) later on.

From the increments, we can recover W_t by the telescoping sum

$$W_t = (W_t - W_{t_{n-1}}) + (W_{t_{n-1}} - W_{t_{n-2}}) + (W_{t_{n-2}} - W_{t_{n-3}}) + \cdots + (W_{t_1} - W_{t_0}) + x_0 \quad (3.35)$$

with $W_{t_0} = x_0$ being the initial position. This is precisely the inverse of the mapping (3.33) with

$$A^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}.$$

(See also Question 16 of Homework III for another interesting construction.)

[25](See for example Page 283 of A First Course in Probability by Ross) If $(X_1, X_2, \ldots, X_n)$ is an n-dimensional continuous random variable with probability density function $\rho_{X_1,X_2,\ldots,X_n}$, and

$$Y_i = g_i(X_1, X_2, \ldots, X_n), \quad 1 \leq i \leq n$$

$$J(x_1, x_2, \ldots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix} \neq 0,$$

then

$$\rho_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = \rho_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)\,|J(x_1, x_2, \ldots, x_n)|^{-1}. \quad (3.34)$$

Remark: In particular, by (3.5), the increments are normal implies $W_{t_1}, W_{t_2}, \cdots, W_{t_n}$ are normal, even though the later can also be derived directly from (3.32). To see how things are connected, let us perform some direct calculation to verify it:

Example 3.4 *(The matrix-vector multiplication way to write the joint pdf) Let $t_2 > t_1$. If (W_{t_1}, W_{t_2}) is multivariate normal distribution with pdf*

$$f(x_1, x_2) = \frac{1}{(2\pi)(det\ \Sigma)^{1/2}} e^{-\frac{1}{2}(x_1-x_0,x_2-x_0)\Sigma^{-1}(x_1-x_0,x_2-x_0)^\top},$$

where $\Sigma = \begin{pmatrix} t_1 & t_1 \\ t_1 & t_2 \end{pmatrix}$, *then* $\Sigma^{-1} = \frac{1}{t_1t_2-t_1^2}\begin{pmatrix} t_2 & -t_1 \\ -t_1 & t_1 \end{pmatrix}$ *and*

$$f(x_1, x_2) = \frac{1}{(2\pi)(t_1t_2 - t_1^2)^{1/2}} e^{-\frac{1}{2}\frac{(x_1-x_0,x_2-x_0)\begin{pmatrix} t_2 & -t_1 \\ -t_1 & t_1 \end{pmatrix}\begin{pmatrix} x_1-x_0 \\ x_2-x_0 \end{pmatrix}}{t_1(t_2-t_1)}}.$$

After the matrix vector multiplication, we obtain

$$\begin{aligned} f(x_1, x_2) =& \frac{1}{(2\pi t_1)^{1/2}} \frac{1}{(2\pi(t_2-t_1))^{1/2}} e^{-\frac{t_2(x_1-x_0)^2+t_1(x_2-x_0)^2-2t_1(x_1-x_0)(x_2-x_0)}{2t_1(t_2-t_1)}} \\ =& \frac{1}{(2\pi t_1)^{1/2}} \frac{1}{(2\pi(t_2-t_1))^{1/2}} e^{-\frac{(t_2-t_1)(x_1-x_0)^2+t_1(x_2-x_1)^2}{2t_1(t_2-t_1)}} \\ =& \frac{1}{(2\pi t_1)^{1/2}} e^{-\frac{(x_1-x_0)^2}{2t_1}} \frac{1}{(2\pi(t_2-t_1))^{1/2}} e^{-\frac{(x_2-x_1)^2}{2(t_2-t_1)}}. \end{aligned}$$

Hence

$$\mathbb{P}(W_{t_1} \in \Gamma_1, W_{t_2} \in \Gamma_2) = \mathbb{P}((W_{t_1}, W_{t_2}) \in \Gamma_1 \times \Gamma_2) = \int_{\Gamma_1\times\Gamma_2} f(x_1, x_2) dx_1 dx_2$$
$$= \int_{\Gamma_1\times\Gamma_2} \frac{1}{(2\pi t_1)^{1/2}} e^{-\frac{(x_1-x_0)^2}{2(t_1-t_0)}} \frac{1}{(2\pi(t_2-t_1))^{1/2}} e^{-\frac{(x_2-x_1)^2}{2(t_2-t_1)}} dx_1 dx_2.$$

The right hand side is exactly (3.29) with the integrand being (3.31).

Remark: By (3.6),

$$\begin{aligned} \begin{pmatrix} t_1 & t_1 \\ t_1 & t_2 \end{pmatrix} = \Sigma = \mathbb{E}\left[\begin{pmatrix} W_{t_1} - x_0 \\ W_{t_2} - x_0 \end{pmatrix} \begin{pmatrix} W_{t_1} - x_0 & W_{t_2} - x_0 \end{pmatrix}\right] \\ = \mathbb{E}\left[\begin{pmatrix} (W_{t_1} - x_0)^2 & (W_{t_1} - x_0)(W_{t_2} - x_0) \\ (W_{t_1} - x_0)(W_{t_2} - x_0) & (W_{t_2} - x_0)^2 \end{pmatrix}\right]. \end{aligned}$$

We hence obtain $\mathbb{E}[(W_{t_i} - x_0)^2] = t_i$ (for $i = 1, 2$) and $\mathbb{E}[(W_{t_1} - x_0)(W_{t_2} - x_0)] = t_1 = \min(t_2, t_1)$ as by-products of the above calculation.

To see how things are connected, let us also use the independence and mean zero properties of the increments to compute:

$$
\begin{aligned}
&\mathbb{E}[(W_{t_1}-x_0)(W_{t_2}-x_0)]\\
=&\mathbb{E}[(W_{t_1}-x_0)(W_{t_2}-W_{t_1}+W_{t_1}-x_0)]\\
=&\mathbb{E}[(W_{t_1}-x_0)(W_{t_1}-x_0)]+\mathbb{E}[(W_{t_1}-W_{t_0})(W_{t_2}-W_{t_1})]\\
\overset{(3.21)}{=}&\mathrm{Var}[W_{t_1}]+\mathbb{E}[(W_{t_1}-W_{t_0})]\mathbb{E}[(W_{t_2}-W_{t_1})]=t_1+0=t_1.
\end{aligned}
\tag{3.36}
$$

Then using (3.6) and the fact that $(W_{t_1}, W_{t_2}, \cdots, W_{t_n})$ is multi-dimensional normal (Remark after Theorem 3.3), we obtain $\Sigma_{1,2}=\Sigma_{2,1}=t_1$.

Indeed, by simple calculations like in (3.36), we can obtain

$$
\mathbb{E}[(X-M)(X-M)^\top]=\begin{pmatrix}
t_1 & t_1 & t_1 & \cdots & t_1\\
t_1 & t_2 & t_2 & \cdots & t_2\\
t_1 & t_2 & t_3 & \cdots & t_3\\
\vdots & \vdots & \vdots & & \vdots\\
t_1 & t_2 & t_3 & \cdots & t_n
\end{pmatrix}. \tag{3.37}
$$

where $X=(W_{t_1},\cdots,W_{t_n})^\top\in\mathbb{R}^{n\times 1}$ and $M=\mathbb{E}[X]=(x_1,\cdots,x_0)^\top$. Again, we have taken $m=1$ for simplicity and required $t_1<t_2<\cdots<t_n$. Note that the (i,j) entry of the above matrix is $\min(t_i,t_j)=t_{\min(i,j)}$.

The Brownian motion defined in Defintion 3.4 is not unique and we indeed can ask for more about it: By the following Kolmogorov's continuity theorem and Question 2 of Homework III, for almost all $\omega\in\Omega$, we can have a continuous path $t\to W_t(\omega)$ from $[0,\infty)$ to $\mathbb{R}^n$ which satisfies (3.30).

Theorem 3.4 *(Kolmogorov's continuity theorem) Suppose that the process $X=\{X_t\}_{t\geq 0}$ satisfies the following condition: For all $T>0$, there exist positive constants α,β,D such that*

$$
E[|X_t-X_s|^\alpha]\leq D|t-s|^{1+\beta},\qquad 0\leq t,s\leq T. \tag{3.38}
$$

Then there exists a process $\tilde{X}=\{\tilde{X}_t\}_{t\geq 0}$ that is continuous with respect to t such that

$$
\mathbb{P}(\{\omega: X_t(\omega)=\tilde{X}_t(\omega)\})=1\qquad\forall\, t. \tag{3.39}
$$

To summarize, we have proved the following three basic properties of Brownian motion. They sometimes are used as the definition of Brownian motion [26] and it has already been used in the Matlab code in Section 3.1 .

[26] See for example Chapter 7.1 of "Probability: Theory and Examples" by Durrett. Indeed, Durrett then needs to answer "Is there a process with these properties". The way to find such a process is by defining measure (3.29) and then use Kolmogorov's Theorems 3.1 and 3.4.

1) $t \to W_t(\omega)$ is continuous and $\mathbb{P}(\{\omega : W_0(\omega) = x_0\}) = 1$.

2) For $0 \le s < t \le T$, the random variable given by the increment $W(t) - W(s)$ is normally distributed with mean zero and variance $t - s$; equivalently, $W(t) - W(s) \sim \sqrt{t-s}N(0,1)$ where $N(0,1)$ denotes a normally distributed random variable with zero mean and unit variance.

3) For $0 \le t_1 < t_2 < \cdots < t_k$,

$$W_{t_1}, W_{t_2} - W_{t_1}, \cdots, W_{t_k} - W_{t_{k-1}} \text{ are independent.} \tag{3.40}$$

In particular, $W_{t_k} - W_{t_{k-1}}$ is independent of $W_{t_{k-1}}$ since the latter is $(W_{t_{k-1}} - W_{t_{k-2}}) + \cdots + (W_{t_2} - W_{t_1}) + W_{t_1}$

A third equivalent definition for $\{W_t\}_{t\ge 0}$ to be a n-dimensional Brownian motion is the following (see for example the book "Functional integration and partial differential equations" by Freidlin):

1) $t \to W_t(\omega)$ is continuous and $\mathbb{P}(\{\omega : W_0(\omega) = x_0\}) = 1$.

2) W_t is a Gaussian process, which means for all $0 \le t_1 \le \cdots \le t_k$, the random variable $Z = (W_{t_1}, \cdots, W_{t_k})$ has a multi normal distribution.

3) $\mathbb{E}W_s = x_0$ and $\mathbb{E}[(W_s - x_0)(W_t - x_0)^\top] = \min(s,t)I_n \stackrel{\text{def}}{=} (s \wedge t)I_n$.

Please note that our previous Example 3.4 indeed says that the third definition implies Theorem 3.2 which is the construction behind Defintion 3.4, which is the first definition of Brownian motion.

Even though we have not yet introduced "continuous-time martingale" [27], we'd like to document a fourth way to tell if a stochastic process is a Brownian motion: In Question 6 of Homework III, we have seen that $[W,W](t) \stackrel{\text{def}}{=} \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} |M_{t_{j+1}} - M_{t_j}|^2 = t$ if W is a Brownian motion. The opposite is also true:

Theorem 3.5 *(Levy) Let M_t, $t \ge 0$, be a martingale with respect to a filtration [28] $\mathcal{F}_t$, $t \ge 0$. Assume that $M_0 = 0$, M_t has continuous path, and*

$$[M,M](t) \stackrel{def}{=} \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} |M_{t_{j+1}} - M_{t_j}|^2 = t \quad \forall t \ge 0$$

where $0 = t_0 < t_1 \cdots < t_n = t$ is a partition of the interval $[0,t]$. Then M_t is a Brownian motion.

[27] We have already introduced discrete-time martingale in Defintion 2.4

[28] Definition 2.6

See Theorem 4.6.4 of Shreve II for the proof.

Example 3.5 *(Paley-Wiener representation)* [29] *This is FYI only and won't be tested. It is the example from Page 21 of the book "Diffusion processes and their sample paths" by Ito and McKean. See also Page 29 of the book "Functional integration and partial differential equations" by Freidlin. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of i.i.d. random variables $\xi_i \sim N(0,1)$. Define*

$$\phi_t(\omega) = t\xi_0 + \frac{\sqrt{2}}{\pi}\sum_{k=1}^{\infty}\frac{\sin(k\pi t)}{k}\xi_k, \qquad 0 \le t \le 1. \tag{3.41}$$

One can check that this series converges uniformly on $[0,1]$ with probability 1. Since the sum of independent Gaussian (i.e. normally distributed) random variables is also a Gaussian random variable , and since the limit of Gaussian random variables is also a Gaussian random variable, we conclude that the process $\phi_t(\omega)$ is also Gaussian. From (3.41), we know

$$\mathbb{E}\phi_s\phi_t = ts + \sum_{k=1}^{\infty}\frac{2}{\pi^2}\frac{\sin(k\pi s)\sin(k\pi t)}{k^2} = s \wedge t, \qquad s,t \in [0,1].$$

Hence $\phi_t(\omega)$ is a 1-dimensional Brownian motion. The last identity can be proved using the identity $\sum_{k=1}^{\infty}\frac{\cos(kx)}{k^2} = \frac{\pi^2}{6} - \frac{\pi x}{2} + \frac{x^2}{4}$ for $0 \le x \le 2\pi$ [30]*. More general construction can be found in Section 1.5 of Itô and McKean's book.*

Example 3.6 *Let $\{W_t, t \ge 0\}$ be Brownian motion with $W_0 = 0$. What is the distribution of $W_1 + W_3$?*
Solution: *Recall that the sum of jointly normally distributed random variables is still normally distributed. $\mathbb{E}[W_1 + W_3] = 0 + 0 = 0$. To determine the variance, note that*

$$W_1 + W_3 = W_3 - W_1 + 2W_1.$$

$W_3 - W_1$ and $2W_1$ are independent. $W_3 - W_1 \sim N(0,2)$. $Var(W_3 - W_1 + 2W_1) = Var(W_3 - W_1) + Var(2W_1) = 2 + 4\,Var(W_1) = 2 + 4 = 6$. Hence $W_1 + W_3 \sim N(0,6)$.

Example 3.7 *Let $\{W_t, t \ge 0\}$ be Brownian motion with $W_0 = 0$. Recall we have learned in (3.37) that $\mathbb{E}[W_tW_s] = \min(t,s)$. Find $Cov(W_2 + W_4, W_1 + W_3)$ where $Cov(X,Y) \stackrel{def}{=} \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$. Then show that $W_2 + W_4 - \frac{7}{10}(W_1 + W_3)$ is independent of $W_1 + W_3$.*

[29] If the summation in (3.41) is finite $\sum_{k=1}^{N}$ instead of $\sum_{k=1}^{\infty}$, then you can take any times derivative of $\phi_t(\omega)$ with respect to t. But once we let $N \to \infty$, this $\phi_t(\omega)$ becomes not differentiable in t.

[30] See for example Page 47 of "Table of Integrals, Series, and Products" 7th edition by Gradshteyn and Ryzhik.

Solution*:*

$$\begin{aligned} Cov(W_2+W_4, W_1+W_3) =& \mathbb{E}[(W_2+W_4-\mathbb{E}[W_2+W_4])(W_1+W_3-\mathbb{E}[W_1+W_3])] \\ =& \mathbb{E}[(W_2+W_4)(W_1+W_3)] \\ =& \mathbb{E}[W_2W_1]+\mathbb{E}[W_2W_3]+\mathbb{E}[W_4W_1]+\mathbb{E}[W_4W_3] \\ =& 1+2+1+3=7. \end{aligned}$$

$$\begin{aligned} & Cov(W_2+W_4-\frac{7}{6}(W_1+W_3), W_1+W_3) \\ =& Cov(W_2+W_4, W_1+W_3)-\frac{7}{6}Cov(W_1+W_3, W_1+W_3) \\ =& 7-\frac{7}{6}6=0. \end{aligned}$$

Let $X = W_2+W_4-\frac{7}{6}(W_1+W_3)$ and $Y = W_1+W_3$. Then X and Y are normally distributed random variables with zero covariance. Hence X and Y are independent.

Example 3.8 *Let W_t be a 2-dimensional Brownian motion which means that its pdf is $\frac{1}{2\pi t}e^{-\frac{x_1^2+x_2^2}{2t}}$. Let $D_\rho = \{x \in \mathbb{R}^2 : \|x\| = \sqrt{x_1^2+x_2^2} < \rho\}$. Compute*

$$\mathbb{P}^0[W_t \in D_\rho].$$

Solution*:*

$$\mathbb{P}^0(W_t \in D_\rho) = \iint_{D_\rho} \frac{1}{2\pi t}e^{-\frac{x_1^2+x_2^2}{2t}}dx_1dx_2.$$

Let $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$ with $r \in [0,\rho]$, $\theta \in [0, 2\pi]$.

$$det\begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{pmatrix} = r.$$

By the change of variable formula from multivariable calculus or simply the polar coordinates,

$$\mathbb{P}^0(W_t \in D_\rho) = \frac{1}{2\pi t}\int_0^{2\pi}\int_0^\rho e^{-\frac{r^2}{2t}}rdrd\theta = \frac{1}{2\pi}2\pi\left(-e^{-\frac{r^2}{2t}}\right)\Big|_{r=0}^{r=\rho} = 1-e^{-\frac{\rho^2}{2t}}.$$

3.4 Homework III

(Only submit solutions to Questions 1,4,6,9,14. Unless otherwise specified, a Brownian motion is assume to be 1-dimensional with zero initial value.)

1. a) Let $X \sim N(0,t)$ which means its pdf is $f_X = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$. Define its moment generating function $\varphi(u) = \mathbb{E}[e^{uX}] = \int_{\mathbb{R}} e^{ux} f_X(x)dx$. Show that $\varphi(u) = e^{\frac{1}{2}u^2 t}$.

   b) If $X \sim N(\mu, \sigma^2)$, show that the so called log-normally distributed random variable $S = e^X$ has mean $e^{\mu + \frac{1}{2}\sigma^2}$. [Hint: To simplify the calculation, consider $\mathbb{E}[e^X] = e^{\mu}\mathbb{E}[e^{X-\mu}]$.]

   c) What is the variance of e^X if $X \sim N(\mu, \sigma^2)$?

2. The kurtosis of a random variable is defined to be the ratio of its fourth central moment to the square of its variance. Prove that for a normal random variable $X \sim N(\mu, \sigma^2)$, the kurtosis $\frac{\mathbb{E}[(X-\mu)^4]}{(\mathbb{E}[(X-\mu)^2])^2} = 3$. (Hint: If $\mu = 0$, we can use the moment generating function $\varphi(u) = \mathbb{E}[e^{uX}]$ defined in Question 1. Taking 2 derivatives with respect to u, we get $\varphi''(u) = \mathbb{E}[X^2 e^{\mu X}]$, which is $(t + tu)e^{\frac{1}{2}u^2 t}$ since $\varphi(u) = e^{\frac{1}{2}u^2 t}$. By setting $u = 0$, we obtain $\mathbb{E}[X^2] = t$. By taking 4 derivatives, one can calculate $\mathbb{E}[X^4]$.) Then show that

$$E[|W_t - W_s|^4] = 3|t-s|^2, \qquad 0 \le t, s \le T.$$

   Hence (3.38) is true with $D = 3$, $\alpha = 4$, $\beta = 1$ and "$\le$" is indeed "$=$" for the Brownian motion case.

   Solution: Define $\varphi(u) = \mathbb{E}[e^{u(X-\mu)}]$. Then $\varphi(u) = e^{\frac{1}{2}u^2\sigma^2}$ by Question 1.

$$\varphi^{(4)}(u) = \mathbb{E}[(X-\mu)^4 e^{u(X-\mu)}] = \left(3\sigma^4 + 6u^2\sigma^6 + u^4\sigma^8\right) e^{\frac{1}{2}u^2\sigma^2}.$$

   Letting $u = 0$, we get

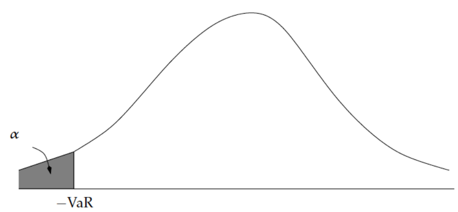
$$\mathbb{E}[(X-\mu)^4] = 3\sigma^4.$$

   Since $W_t - W_s \sim N(0, t-s)$, $E[|W_t - W_s|^4] = 3|t-s|^2$.

3. Value-at-risk (VaR) denotes, within a confidence level, the maximum loss a portfolio could suffer. To be more precise, denote by X the change in the market value of a portfolio during a given time period. Then the VaR with confidence level $1-\alpha$ is defined to be the value VaR in

$$\mathbb{P}(X \le -\mathrm{VaR}) = \alpha. \tag{3.42}$$

   In other words, with probability $1-\alpha$, the maximum loss will not exceed VaR.



Show that if X is assumed to be normally distributed with mean μ and variance σ^2, the VaR with confidence level $1-\alpha$ is given by

$$\mathrm{VaR} = z_\alpha \sigma - \mu \tag{3.43}$$

where z_α satisfies $N(-z_\alpha) = \alpha$ with function N defined in (3.15). For example, $z_{0.1} = 1.2816$, $z_{0.05} = 1.6449$, $z_{0.01} = 2.3263$, $z_{0.005} = 2.5758$.

Proof: $X \sim N(\mu, \sigma^2)$ implies that $\frac{X-\mu}{\sigma} \sim N(0,1)$. Hence

$$\begin{aligned} N(-z_\alpha) = \alpha = \mathbb{P}(X \le -\mathrm{VaR}) &= \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{-\mathrm{VaR}-\mu}{\sigma}\right) \\ &= \int_{-\infty}^{-\frac{\mathrm{VaR}+\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = N\left(-\frac{\mathrm{VaR}+\mu}{\sigma}\right). \end{aligned}$$

So $z_\alpha = \frac{\mathrm{VaR}+\mu}{\sigma}$ which proved (3.43). $\square$

4. A binary call option with maturity T pays one dollar when the stock price at time T is at or above a certain level K and pays nothing otherwise. The payoff can be written in the form of an indicator function

$$X = 1_{\{S_T \ge K\}}.$$

Suppose in the risk neutral world $\log S_T$ is normally distributed with mean $\log S_0 + (r - \frac{\sigma^2}{2})T$ and variance $\sigma^2 T$. Show that the discounted expected payoff $e^{-rT}\mathbb{E}X$ is

$$e^{-rT}\mathbb{E}X = e^{-rT} N(d_-), \tag{3.44}$$

where the function N is defined in (3.15) and the d_- is defined in (3.14). [Hint: Note that $\mathbb{E}X = \mathbb{P}(\log S_T \ge \log K)$.]

5. Let $f(t)$ be a function defined for $0 \le t \le T$. The quadratic variation of f up to time T is defined to be

$$[f,f](T) = \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|^2 \tag{3.45}$$

where $0 = t_0 < t_1 < \cdots < t_n = T$ is a partition of the interval $[0,T]$. Recall that the mean value theorem says that if f has continuous derivative, for any $[t_j, t_{j+1}]$, there is a $t_j^* \in (t_j, t_{j+1})$ so that $f(t_{j+1}) - f(t_j) = f'(t_j^*)(t_{j+1} - t_j)$.

Prove that if f has continuous derivative on $[0,T]$ with $T < \infty$, then

$$[f,f](T) = 0.$$

Here you can use the fact that f has continuous derivative on $[0,T]$ implies that $M \stackrel{\text{def}}{=} \sup_{s\in[0,T]} |f'(s)| < \infty$.

Proof:

$$
\begin{aligned}
0 \leq [f,f](T) &= \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|^2 \\
&= \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} |f'(t_j^*)|^2 |t_{j+1} - t_j|^2 \\
&\leq \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} \left(M^2 \max_j |t_{j+1} - t_j| \right) |t_{j+1} - t_j| \\
&= \lim_{\max_j |t_{j+1}-t_j| \to 0} \left(M^2 \max_j |t_{j+1} - t_j| \right) \sum_{j=0}^{n-1} |t_{j+1} - t_j| \\
&= \lim_{\max_j |t_{j+1}-t_j| \to 0} \left(M^2 \max_j |t_{j+1} - t_j| \right) T = 0.
\end{aligned}
$$

6. By definition, we say that $X_n(\omega) \to X(\omega)$ in mean square sense if $\mathbb{E}[(X_n - X)^2] \to 0$ as $n \to \infty$. Given a Brownian motion W_t, define a random variable $[W,W]_{\Pi,T}$ by

$$[W,W]_{\Pi,T}(\omega) = \sum_{j=0}^{n-1} |W_{t_{j+1}}(\omega) - W_{t_j}(\omega)|^2$$

where $\Pi = \{0 = t_0, t_1, \cdots, t_n = T\}$ is the partition of $[0,T]$. This question asks you to show that the random variable $[W,W]_{\Pi,T} \to T$ in the mean square sense when $\|\Pi\| \stackrel{\text{def}}{=} \max_j |t_{j+1} - t_j| \to 0$ [31] . The proof is split into two steps:

a) Show that

$$\mathbb{E}[[W,W]_{\Pi,T}] = T. \tag{3.47}$$

b) By (3.47) and $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$, the variance of $[W,W]_{\Pi,T}$ is $\text{Var}[[W,W]_{\Pi,T}] = \mathbb{E}[([W,W]_{\Pi,T} - T)^2]$. Show that

$$\text{Var}[[W,W]_{\Pi,T}] = \sum_{j=0}^{n-1} \text{Var}[|W_{t_{j+1}} - W_{t_j}|^2] \to 0$$

as $\|\Pi\| \to 0$.

7. We now continue with Example 2.11 of Chapter 2 to introduce Markov property:

[31] Theorem 3.4.3 of Shreve II says that

$$[W,W]_{\Pi,T} \to T \text{ almost surely when } \max_j |t_{j+1} - t_j| \to 0. \tag{3.46}$$

Note that we say that a statement A that depends on ω is true almost surely if $\mathbb{P}(\{\omega : A(\omega) \text{ is true}\}) = 1$. Hence what we will prove in this question is that the convergence in (3.46) happens in mean square sense.

Definition 3.5 *(Definition 2.5.1 of Shreve I) Consider the binomial asset-pricing model. Let X_0, X_1, ... , X_N be a sequence of random variables, with each X_n depending only on the first n coin tosses (and X_0 constant). If, for every n between 0 and $N-1$ and for every function $f(x)$, there is another function $g(x)$ (depending on f and n) such that*

$$\tilde{\mathbb{E}}_n[f(X_{n+1})] = g(X_n), \tag{3.48}$$

we say that $X_0, X_1, \cdots, X_n$ is a Markov process.

The Markov property says that the dependence of $\tilde{\mathbb{E}}_n[f(X_{n+1})]$ on the first n coin tosses occurs through X_n (i.e., the information about the coin tosses one needs in order to evaluate $\tilde{\mathbb{E}}[f(X_{n+1})]$ is summarized by X_n).

It also implies the "two-step ahead" property:

$$\tilde{\mathbb{E}}_n[h(X_{n+2})] \overset{(2.31)}{=} \tilde{\mathbb{E}}_n[\tilde{\mathbb{E}}_{n+1}[h(X_{n+2})]] = \tilde{\mathbb{E}}_n[f(X_{n+1})] = g(X_n), \tag{3.49}$$

or "multi-step ahead" property: For any function f and any $0 \le n \le m \le N$, there is a function g so that

$$\tilde{\mathbb{E}}_n[h(X_m)] = g(X_n). \tag{3.50}$$

Now comes the question: Consider the stock price model

$$S_{n+1}(\omega_1 \cdots \omega_n \omega_{n+1}) = \begin{cases} uS_n(\omega_1 \cdots \omega_n) & \text{if } \omega_{n+1} = H, \\ dS_n(\omega_1 \cdots \omega_n) & \text{if } \omega_{n+1} = T. \end{cases}$$

Find the function g so that

$$\tilde{\mathbb{E}}_n[f(S_{n+1})](\omega_1 \cdots \omega_n) = g(S_n).$$

Solution: $g(x) = q_u f(ux) + q_d f(dx)$.

8. (Conditional distribution) The conditional probability of event A happens under the condition that event B already happens is denoted as $\mathbb{P}(A|B)$ and we have the relation

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)} \tag{3.51}$$

if $\mathbb{P}(B) > 0$. If we have joint density function $\rho_{X,Y}$ for two continuous random variables X and Y, we define marginal density function $\rho_X(x)$ and $\rho_Y(y)$ as in (3.18). Then we can calculate conditional distribution

$$\begin{aligned} \mathbb{P}(X \in A|Y = y) &\overset{\text{def}}{=} \lim_{a \downarrow 0} \mathbb{P}(X \in A|Y \in [y-a, y+a]) \overset{(3.51)}{=} \lim_{a \downarrow 0} \frac{\mathbb{P}(X \in A, Y \in [y-a, y+a])}{\mathbb{P}(Y \in [y-\Delta, y+\Delta])} \\ &= \lim_{a \downarrow 0} \frac{\int_{[y-a,y+a]} \left(\int_A \rho_{X,Y}(x,v)dx\right) dv}{\int_{[y-a,y+a]} \rho_Y(v)dv} \\ &= \lim_{a \downarrow 0} \frac{\left(\int_A \rho_{X,Y}(x,y)dx\right) \times (2a) + \text{ higher order term of } a}{\rho_Y(y) \times (2a) + \text{ higher order term of } a}. \end{aligned}$$

In the last step, we used again the fact that $\int_a^b f(v)dv = f(\frac{a+b}{2})(b-a)+$ higher order term of $(b-a)$. Now, dividing $2a$ from both the numerator and denomonator, we get

$$\begin{aligned}\mathbb{P}(X \in A|Y = y) &= \lim_{a\downarrow 0} \frac{\int_A \rho_{X,Y}(x,y)dx + \frac{\text{higher order term of } a}{2a}}{\rho_Y(y) + \frac{\text{higher order term of } a}{2a}} \\ &= \frac{\int_A \rho_{X,Y}(x,y)dx + \lim_{a\downarrow 0} \frac{\text{higher order term of } a}{2a}}{\rho_Y(y) + \lim_{a\downarrow 0} \frac{\text{higher order term of } a}{2a}}.\end{aligned}$$

In the last step, we have used $\lim_{a\to b} \frac{f(a)}{g(a)} = \frac{\lim_{a\to b} f(a)}{\lim_{a\to b} g(a)}$. Continue, we get

$$\mathbb{P}(X \in A|Y = y) = \frac{\int_A \rho_{X,Y}(x,y)dx + 0}{\rho_Y(y) + 0} = \int_A \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)}dx.$$

Hence by (3.1), the conditional density of X given $Y = y$ is

$$\frac{\rho_{X,Y}(x,y)}{\rho_Y(y)} \stackrel{\text{def}}{=} \rho_{X|Y}(x|y), \tag{3.52}$$

which, by the way, also depends on y. Since we now have the density function, by (3.10), we know

$$\mathbb{E}[h(X)|Y = y] = \int_{\mathbb{R}} h(x)\rho_{X|Y}(x|y)dx \tag{3.53}$$

which is a function of y. By the way, if X and Y are independent, by (3.20), $\rho_{X|Y}(x|y) = \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)} = \rho_X(x)$, then

$$\mathbb{E}[h(X)|Y = y] = \int_{\mathbb{R}} h(x)\rho_{X|Y}(x|y)dx = \int_{\mathbb{R}} h(x)\rho_X(x)dx = \mathbb{E}[h(X)]. \tag{3.54}$$

Go back to (3.53). Note that $\mathbb{E}[h(X)|Y] \stackrel{\text{def}}{=} \mathbb{E}[h(X)|Y = Y]$ is a function of Y which is hence a random variable by itself. As we already know the density function of Y, by (3.10), we can compute

$$\mathbb{E}[\mathbb{E}[h(X)|Y]] = \int_{\mathbb{R}} \mathbb{E}[h(X)|Y = y]\ \rho_Y(y)dy.$$

Here comes the question: Prove the iterated conditioning property

$$\mathbb{E}[\mathbb{E}[h(X)|Y]] = \mathbb{E}[h(X)]. \tag{3.55}$$

Remark: You should compare $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$ with

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}_m[X]] = \mathbb{E}[\mathbb{E}[X|\text{given } \omega_1 \cdots \omega_m]] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_m]], \tag{3.56}$$

which we have learned before in Theorem 2.2 (we used $\tilde{\mathbb{E}}$ there to stress that we were using the risk-neutral probability. But apparently, it is true for any probability). The precise meaning of the last term, $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_m]]$, will be clarified in Theorem 4.3 of the next chapter.

Proof:

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[h(X)|Y]] &\overset{(3.10)}{=} \int_{\mathbb{R}} \mathbb{E}[h(X)|Y=y]\rho_Y(y)dy \\
&\overset{(3.53)}{=} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x)\rho_{X|Y}(x|y)dx \right) \rho_Y(y)dy \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x)\frac{\rho_{X,Y}(x,y)}{\cancel{\rho_Y(y)}}dx \right) \cancel{\rho_Y(y)}dy \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x)\rho_{X,Y}(x,y)dx \right) dy \\
&= \mathbb{E}[h(X)].
\end{aligned}$$

9. (Continue with Question 8.) Let S_1 and S_2 be the prices of two assets. Assume that $X = \log S_1$ and $Y = \log S_2$ have a joint density function

$$\rho_{X,Y} = \frac{\sqrt{3}}{4\pi} e^{-\frac{1}{2}(x^2 - xy + y^2)}, \qquad \text{for } x, y \in \mathbb{R}.$$

Determine $\mathbb{E}[X|Y]$.

10. (This problem is rather difficult and won't be tested. But I hope you can read it so that you have a better understanding of the density function of Brownian motion. This example is taken from Question 2.12 of Hui Wang, "Monte Carlo Simulations with Applications to Finance". Please bear with me for the complicated computations. If your probability/finance intuition already tells you that the conclusion is obvious, you do not need to read the proof.) Given an arbitrary constant θ, let $B = \{B_t : t \geq 0\}$ be a Brownian motion with drift θ, i.e.,

$$B_t = W_t + \theta t, \qquad t \geq 0, \tag{3.57}$$

where $W = \{W_t : t \geq 0\}$ is a Brownian motion. Given $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$, show that the conditional distribution of the $(n-1)$-dimensional random variable $X = (B_{t_1}, \cdots, B_{t_{n-1}})$ given $Y = B_T = y$ does not depend on θ. In particular, letting $\theta = 0$, we conclude that the conditional distribution of $\{B_t : 0 \leq t \leq T\}$ given $B_T = y$ is the same as the conditional distribution of $\{W_t : 0 \leq t \leq T\}$ given $W_T = y$.

Proof: Recall (3.32)

$$\rho_{W_{t_1},\cdots,W_{t_n}}(x_1, \cdots x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\prod_{i=1}^n (t_i - t_{i-1})}} \exp\left(-\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})} \right)$$

and $W_{t_n} \sim N(x_0, t_n)$ which means

$$\rho_{W_{t_n}}(x_n) = \frac{1}{\sqrt{2\pi(t_n - t_0)}} \exp(-\frac{(x_n - x_0)^2}{2t_n}).$$

Since $B_{t_i} = W_{t_i} + \theta t_i \sim N(x_0 + \theta t_i, t_i)$, by the change of variable formula in multivariable calculus or probability, which is (3.34), we get

$$\rho_{B_{t_1},\cdots,B_{t_n}}(x_1, \cdots x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\prod_{i=1}^n (t_i - t_{i-1})}} \exp\left(-\sum_{i=1}^n \frac{[(x_i - \theta t_i) - (x_{i-1} - \theta t_{i-1})]^2}{2(t_i - t_{i-1})}\right)$$

and

$$\rho_{B_{t_n}}(x_n) = \frac{1}{\sqrt{2\pi(t_n - t_0)}} \exp\left(-\frac{(x_n - \theta t_n - x_0)^2}{2t_n}\right).$$

To calculate $\frac{\rho_{B_{t_1},\cdots,B_{t_n}}(x_1,\cdots x_n)}{\rho_{B_{t_n}}(x_n)}$, we need to compute

$$\begin{aligned}
&-\sum_{i=1}^n \frac{[(x_i - x_{i-1}) - \theta(t_i - t_{i-1})]^2}{2(t_i - t_{i-1})} + \frac{[(x_n - x_0) - \theta t_n]^2}{2t_n} \\
&= -\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})} + \cancel{\sum_{i=1}^n \theta(x_i - x_{i-1})} - \frac{\theta^2}{2}\cancel{\sum_{i=1}^n (t_i - t_{i-1})} - \frac{(x_n - x_0)^2}{2t_n} + \cancel{\theta(x_n - x_0)} + \cancel{\frac{\theta^2}{2}t^n} \\
&= -\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})} + \frac{(x_n - x_0)^2}{2t_n}.
\end{aligned}$$

Hence the conditional density function of $(B_{t_1}, \cdots, B_{t_{n-1}})$ given $B_{t_n} = x_n$ is

$$\frac{\rho_{B_{t_1},\cdots,B_{t_n}}(x_1, \cdots x_n)}{\rho_{B_{t_n}}(x_n)} = \frac{\sqrt{t_n - t_0}}{(2\pi)^{(n-1)/2}\sqrt{\prod_{i=1}^n (t_i - t_{i-1})}} \exp\left(-\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})} + \frac{(x_n - x_0)^2}{2t_n}\right). \tag{3.58}$$

The parameter θ has disappeared from the above formula. This proves the result. □

11. (This problem is rather difficult and won't be tested. This example is taken from Page 40 of Hui Wang, "Monte Carlo Simulations with Applications to Finance". It shows an application of conditional expectation. The result is useful for pricing, say, a look back call option whose payoff is $(\max_{0\le t\le T} S_t - K)^+$ and is path-dependent. See Example 2.6 of "Monte Carlo Simulations with Applications to Finance" by Hui Wang for more details.) We say a function is path-dependent if it depends on the sample paths of the relevant process. For example, we can define

$$h(W_{[0,T]}) = \max_{0\le t\le T} W_t - \min_{0\le t\le T} W_t - W_T \tag{3.59}$$

whose value depends on the entire sample path $W_{[0,T]} = \{W_t : 0 \le t \le T\}$.

Introduce B_t and W_t as in Question 10 with $W_0 = 0$. Prove that for any path dependent function h,

$$\mathbb{E}\left[h(B_{[0,T]})\right] = \mathbb{E}\left[e^{\theta W_T - \frac{1}{2}\theta^2 T} h(W_{[0,T]})\right]. \tag{3.60}$$

Proof: Since $W_T \sim N(0,T)$, by (3.55),

$$\begin{aligned}
RHS &= \int_{\mathbb{R}} \mathbb{E}\left[e^{\theta W_T - \frac{1}{2}\theta^2 T} h(W_{[0,T]}) \middle| W_T = x\right] \frac{1}{\sqrt{2\pi T}} e^{-\frac{x^2}{2T}} dx \\
&= \int_{\mathbb{R}} \mathbb{E}\left[e^{\theta x - \frac{1}{2}\theta^2 T} h(W_{[0,T]}) \middle| W_T = x\right] \frac{1}{\sqrt{2\pi T}} e^{-\frac{x^2}{2T}} dx \\
&= \int_{\mathbb{R}} \mathbb{E}\left[h(W_{[0,T]}) \middle| W_T = x\right] \frac{1}{\sqrt{2\pi T}} e^{\theta x - \frac{1}{2}\theta^2 T - \frac{x^2}{2T}} dx \\
&= \int_{\mathbb{R}} \mathbb{E}\left[h(W_{[0,T]}) \middle| W_T = x\right] \frac{1}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(x-\theta T)^2} dx.
\end{aligned}$$

By Question 10, $\mathbb{E}\left[h(W_{[0,T]}) \middle| W_T = x\right] = \mathbb{E}\left[h(B_{[0,T]}) \middle| B_T = x\right]$. Hence

$$RHS = \int_{\mathbb{R}} \mathbb{E}\left[h(B_{[0,T]}) \middle| B_T = x\right] \frac{1}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(x-\theta T)^2} dx$$

and $\frac{1}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(x-\theta T)^2}$ is pdf of B_T. Hence

$$RHS = \int_{\mathbb{R}} \mathbb{E}\left[h(B_{[0,T]}) \middle| B_T = x\right] \rho_{B_T}(x) dx = \mathbb{E}\left[\mathbb{E}\left[h(B_{[0,T]}) \middle| B_T\right]\right] = \mathbb{E}[h(B_{[0,T]})]$$

by (3.55). This finishes the proof.

12. Let $\{W_t, t \geq 0\}$ be a Brownian motion. Find $\mathbb{P}(W_3 \leq 1 | W_2 = \frac{1}{2})$.

    Solution: Note that W_2 and $W_3 - W_2$ are independent.

    $$\mathbb{P}(W_3 \leq 1 | W_2 = \frac{1}{2}) = \mathbb{P}(W_3 - W_2 \leq \frac{1}{2} | W_2 = \frac{1}{2}) = \mathbb{P}(W_3 - W_2 \leq \frac{1}{2})$$

    Since $W_3 - W_2 \sim N(0,1)$, $\mathbb{P}(W_3 - W_2 \leq \frac{1}{2}) = \int_{-\infty}^{1/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \approx 0.6915$.

13. (To compare with Question 12.) Let $\{W_t, t \geq 0\}$ be a Brownian motion with $W_0 = 0$. Find $\mathbb{P}(W_2 \leq 1 | W_3 = \frac{1}{2})$.

    Solution: By (3.52), if $t_1 < t_2$,

    $$\begin{aligned}
    \rho_{W_{t_1}|W_{t_2}}(x_1|x_2) &= \frac{\rho_{W_{t_1},W_{t_2}}(x_1,x_2)}{\rho_{W_{t_2}}(x_2)} = \frac{\sqrt{t_2}}{\sqrt{2\pi}\sqrt{t_1(t_2-t_1)}} e^{-\frac{x_1^2}{2t_1} - \frac{(x_2-x_1)^2}{2(t_2-t_1)} + \frac{x_2^2}{2t_2}} \\
    &= \frac{1}{\sqrt{2\pi \frac{t_1(t_2-t_1)}{t_2}}} e^{-\frac{(t_2 x_1 - t_1 x_2)^2}{2t_1 t_2 (t_2-t_1)}} = \frac{1}{\sqrt{2\pi \frac{t_1(t_2-t_1)}{t_2}}} e^{-\frac{(x_1 - \frac{t_1}{t_2}x_2)^2}{2\frac{t_1(t_2-t_1)}{t_2}}} \sim N(\frac{t_1}{t_2}x_2, \frac{t_1(t_2-t_1)}{t_2}).
    \end{aligned} \tag{3.61}$$

Hence $W_2|W_3 = \frac{1}{2} \sim N(\frac{2}{3}\frac{1}{2}, \frac{2}{3}) = N(\frac{1}{3}, \frac{2}{3})$ with pdf $\frac{1}{\sqrt{2\pi\frac{2}{3}}} e^{-\frac{(x_2-\frac{1}{3})^2}{2\times\frac{2}{3}}}$,

$$\mathbb{P}(W_2 \le 1|W_3 = \frac{1}{2}) = \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi\frac{2}{3}}} e^{-\frac{(x_2-\frac{1}{3})^2}{2\times\frac{2}{3}}} dx_2 \overset{u=\frac{x_2-\frac{1}{3}}{\sqrt{\frac{2}{3}}}}{=} \int_{-\infty}^{\sqrt{2/3}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \approx 0.7929.$$

14. Let $X_t = e^{W_t - \frac{1}{2}t}$ be the price of a stock where W_t is a Brownian motion with $W_0 = 0$. Find $\mathbb{P}(X_3 \le 3)$. [Hint: Event $X_3 \le 3$ and event $\log X_3 \le \log 3$ are equivalent.]

15. Stock prices are sometimes modeled by distributions other than lognormal in order to fit the empirical data more accurately. For instance, soon after his Black-Scholes-Merton work, Merton introduces a jump diffusion model for stock prices. A special case of Merton's model assumes that the underlying stock price S satisfies

$$S = e^Y, \qquad Y = X_1 + \sum_{i=1}^{X_2} Z_i \tag{3.62}$$

where X_1 is $N(\mu, \sigma^2)$, X_2 is Poisson with parameter λ [32], Z_i is $N(0, \nu^2)$, and X_1, X_2, $\{Z_i\}$ are all independent. The evaluation of call options involves expected values such as

$$\mathbb{E}[(S-K)^+], \tag{3.63}$$

where K is some positive constant. Compute this expected value.

Solution: For every $n \ge 0$, we can compute the conditional expected value

$$v_n = \mathbb{E}[(S-K)^+ | X_2 = n].$$

Recall that the sum of jointly normally distributed random variables are still normally distributed. So, conditional on $X_2 = n$, $Y = X_1 + Z_1 + \cdots + Z_n$ is normally distributed as $N(\mu, \sigma^2 + n\nu^2)$ since $\mathbb{E}[Y] = \mu$ and $\mathrm{Var}[Y] = \sigma^2 + n\nu^2$ by the independence assumption. By (3.16), we get

$$v_n = \mathbb{E}[(S-K)^+ | X_2 = n] = e^{\mu + \frac{1}{2}(\sigma^2 + n\nu^2)} N(d_{n,+}) - K N(d_{n,-}) \tag{3.64}$$

with $d_{n,-} = \frac{\log\frac{1}{K} + \mu}{\sqrt{\sigma^2 + n\nu^2}}$ and $d_{n,+} = d_{n,-} + \sqrt{\sigma^2 + n\nu^2}$. By (3.55) of Question 8,

$$\begin{aligned}\mathbb{E}[(S-K)^+] &= \mathbb{E}[\mathbb{E}[(S-K)^+ | X_2]] = \sum_{n=0}^{\infty} \mathbb{E}[(S-K)^+ | X_2 = n] \mathbb{P}(X_2 = n) \\ &= \sum_{n=0}^{\infty} v_n \mathbb{P}(X_2 = n) = e^{-\lambda} \sum_{n=0}^{\infty} v_n \frac{\lambda^n}{n!}.\end{aligned}$$

The last expression can be evaluated numerically.

[32] which means $\mathbb{P}(X_2 = n) = e^{-\lambda}\frac{\lambda^n}{n!}$ for $n = 0, 1, 2, 3, \ldots$.

16. (Cholesky factorization and independent increment) Show that the right hand side of (3.37), called Σ, has a decomposition

$$\Sigma = \mathbb{E}[(Z-M)(Z-M)^\top] = \begin{pmatrix} t_1 & t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & t_2 & \cdots & t_2 \\ t_1 & t_2 & t_3 & \cdots & t_3 \\ \vdots & \vdots & \vdots & & \vdots \\ t_1 & t_2 & t_3 & \cdots & t_n \end{pmatrix} = AA^\top \tag{3.65}$$

where

$$A = \begin{pmatrix} \sqrt{t_1} & 0 & 0 & \cdots & 0 \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & 0 & \cdots & 0 \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \sqrt{t_3 - t_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \sqrt{t_3 - t_2} & \cdots & \sqrt{t_n - t_{n-1}} \end{pmatrix}. \tag{3.66}$$

Let $X = (X_1, \cdots, X_n)^\top \sim N(\mathbf{0}, I_n)$, which means $X_i \in N(0,1)$ and X_i's are independent. **Show** that $AX \sim N(\mathbf{0}, \Sigma)$. By the way, as an application of the result, one can generate $Z = (W_{t_1}, \cdots, W_{t_n})^\top$ as AX. This is precisely the telescoping sum (3.35) and is also the method used in the Matlab code at the beginning of this Chapter:

```
for j = 2:N                    % start the iteration
    dW(j) = sqrt(dt)*randn; % general increment
    W(j) = W(j-1) + dW(j);
end
```

Proof: By direct calculation, one can easily check that $AA^\top = \Sigma$. By (3.5), we know $AX \sim N(\mathbf{0}, AI_nA^\top) = N(\mathbf{0}, AA^\top) = N(\mathbf{0}, \Sigma)$.