MTH 9831, FALL 2015. LECTURE 1

ELENA KOSYGINA

Abstract. Brownian motion is probably the most important continuous time stochastic process. The goal of the next two lectures is to give a brief summary of some of its properties.

1. Multivariate normal (Gaussian) random variables.
2. Equivalent definitions of Brownian motion.
3. Brownian motion as a scaling limit of random walks. Donsker's invariance principle.
4. Lévy's construction of Brownian motion.
5. Brownian motion as a Markov process. Strong Markov property. Reflection principle.

1. Multivariate Gaussian random variables.

Recall that X is said to have a (one-dimensional) normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ if X has density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

To include the degenerate case $\sigma = 0$ we shall say that X has a degenerate normal distribution if $X = \mu$ a.s.. The degenerate distribution, clearly, has no density, its distribution μ_X is δ_μ.

Definition 1.1. *A random vector $X = (X_1, X_2, \dots, X_n)$ is said to be Gaussian (equivalently, normal) is for every $a \in \mathbb{R}^n$ the random variable $\langle a, X\rangle := \sum_{i=1}^n a_i X_i$ has a (one-dimensional, possibly degenerate) normal distribution.*

Definition 1.2. *The characteristic function of a random variable $X = (X_1, X_2, \dots, X_n)$ is*

$$\varphi_X(t) = E(e^{i\langle t, X\rangle}), \quad t \in \mathbb{R}^n.$$

Example 1.3. Let X be a normal random variable with mean μ and variance σ^2. Then

$$\varphi_X(t) = E(e^{itX}) = e^{it\mu - \sigma^2 t^2/2}.$$

Indeed, write $X = \mu + \sigma Z$ where Z is a standard normal random variable. Then $\varphi_X(t) = E(e^{it(\mu+\sigma Z)}) = e^{it\mu}E(e^{it\sigma Z})$. To compute the last expression we write

$$\begin{aligned} E(e^{it\sigma Z}) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{it\sigma z - z^2/2}\, dz &= \frac{e^{-\sigma^2 t^2/2}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(z-it\sigma)^2/2}\, dz \\ &= \frac{e^{-\sigma^2 t^2/2}}{\sqrt{2\pi}} \int_{-\infty - it\sigma}^{\infty - it\sigma} e^{-u^2/2}\, du = e^{-\sigma^2 t^2/2}. \end{aligned}$$

In the last step we used the (unjustified) fact that the integral over the line $\{u + it\sigma, u \in \mathbb{R}\}$ is equal to the integral over the real line.

Fact: the characteristic function uniquely characterizes the distribution of a random variable, i.e. if two random variables have the same characteristic function then they have the same distribution.

Theorem 1.4. *$X = (X_1, X_2, \dots, X_n)$ is a Gaussian vector iff its characteristic function has the form*

$$\varphi_X(t) = \exp\left(i\langle t, \mu\rangle - \frac{1}{2}\langle t, Ct\rangle \right),\ t \in \mathbb{R}^n, \tag{1.1}$$

where $\mu \in \mathbb{R}^n$ and C is $n \times n$ symmetric non-negative semi-definite matrix. C is then the covariance matrix of X and μ is the mean of X.

Proof. Suppose (1.1) holds. Let $a \in \mathbb{R}^n$ and consider $Y = \langle a, X\rangle$. Then for $t \in \mathbb{R}$

$$\varphi_Y(t) = \varphi_X(at) = e^{it\langle \mu, a\rangle - t^2\langle a, Ca\rangle/2},$$

which means that $\varphi_Y(t)$ is the characteristic function of a normal random variable with mean $\langle \mu, a\rangle$ and variance $\langle a, Ca\rangle$. Therefore, Y is normal. Since a was arbitrary, we conclude that X is Gaussian.

Now suppose that X is Gaussian. Then we know that for every $a \in \mathbb{R}$ the random variable $Y = \langle a, X\rangle$ is (one-dimensional) normal. We can compute the mean and variance of Y: $E(Y) = \langle a, \mu\rangle$,

$$\begin{aligned}\mathrm{Var}(Y) &= E((Y - E(Y))^2) = E((\langle a, X\rangle - \langle a, \mu\rangle)^2) = E((\langle a, X - \mu\rangle)^2)\\ &= E((a^T(X-\mu))^2) = E((a^T(X-\mu)(X-\mu)^T a)) = a^T C a = \langle a, Ca\rangle,\end{aligned}$$

where $C = E((X-\mu)(X-\mu)^T)$. Since Y is normal,

$$\varphi_Y(t) = E(e^{it\langle a,\mu\rangle - t^2\langle a, Ca\rangle/2}).$$

Then

$$\varphi_Y(1) = E(e^{i\langle a, X\rangle}) = \varphi_X(a) = e^{i\langle a,\mu\rangle - \langle a, Ca\rangle/2},$$

and we conclude that the characteristic function of X has the claimed form (replace a with t in the last line of formulas. □

Theorem 1.5. *Let $X = (X_1, X_2, \ldots, X_n)$ be a Gaussian vector. The components $X_1, X_2, \ldots, X_n$ are independent iff they are uncorrelated (i.e. C is diagonal).*

The proof is is based on the fact that random variables $X_1,\ X_2, \ldots,\ X_n$ are independent if the characteristic function of $(X_1, X_2, \ldots, X_n)$ splits into the product of characteristic functions of $X_1,\ X_2, \ldots,\ X_n$. We omit the details.

Theorem 1.6. *Let X be an n-dimensional Gaussian vector with mean vector μ. Then there exist independent normal random variables $Y_1, Y_2, \ldots, Y_n$, $Y_i \sim \mathrm{Normal}(0, \sigma_i^2)$, $\sigma_i \geq 0$, and an orthogonal[1] $n \times n$ matrix A such that $X = \mu + AY$, where $Y = (Y_1, Y_2, \ldots, Y_n)$.*

Proof. Let C be the covariance matrix of X. Since C is symmetric, non-negative semi-definite, there is an orthogonal matrix A and a diagonal matrix D with non-negative diagonal elements such that $C = ADA^{-1}$. Set $Y = A^T(X - \mu)$. Since X is Gaussian, Y (as an affine transformation of X) is also Gaussian. Moreover, $E(Y) = 0$ and

$$E(YY^T) = E(A^T(X-\mu)(X-\mu)^T A) = A^T C A = A^T(ADA^T)A = D,$$

since $AA^T = A^T A = I_n$. By Theorem 1.5, $Y_1, Y_2, \ldots, Y_n$ are independent and $\mathrm{Var}(Y_i) = D_{ii} =: \sigma_i^2 \geq 0$. □

Theorem 1.7. *An n-dimensional Gaussian vector X has a density iff C is non-degenerate. In this case*

$$f_X(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det C}}\, e^{-\frac{1}{2}\langle (x-\mu, C^{-1}(x-\mu)\rangle}.$$

We shall first recall the following useful fact.

Lemma 1.8. *Let $X = (X_1, X_2, \ldots, X_n)$ be a random vector and have density $f_X(x)$ then for every non-degenerate $n \times n$ matrix A and $\mu \in \mathbb{R}^n$ the random vector $Y = \mu + AX$ has density*

$$f_Y(y) = \frac{1}{|\det A|}\, f_X(A^{-1}(y - \mu)).$$

[1] An $n \times n$ matrix A is said to be orthogonal if $AA^T = A^T A = I_n$, where I_n is the n-dimensional identity matrix.

Proof. Let $B \in \mathcal{B}^n$, $g(x) = \mu + Ax$, and $D = g^{-1}(B)$. Then by the change of variable formula we have

$$P(Y \in B) = P(X \in D) = \int_D f_X(x)\,dx = \int_{g^{-1}(B)} f_X(x)\,dx = \int_B f_X(A^{-1}(y-\mu))|\det A^{-1}|\,dy.$$

This implies that Y has density

$$f_Y(y) = f_X(A^{-1}(y-\mu))|\det A^{-1}| = \frac{1}{|\det A|}\, f_X(A^{-1}(y-\mu)). \qquad \square$$

Proof of Theorem 1.7. Suppose that C is non-degenerate. By Theorem 1.6, $X = \mu + AY$ for some orthogonal matrix A and a vector $Y = (Y_1, Y_2, \ldots, Y_n)$ of independent normals Y_i with strictly positive variances σ_i^2, $i \in \{1, 2, \ldots, n\}$. Since

$$f_Y(y) = \frac{1}{(2\pi)^{n/2}\sigma_1\sigma_2\ldots\sigma_n}\, e^{-\frac{1}{2}\sum_{i=1}^n \left(y_i/\sigma_i^2\right)} = \frac{1}{(2\pi)^{n/2}\sqrt{\det D}}\, e^{-\frac{1}{2}\langle y, D^{-1}y\rangle},$$

by Lemma 1.8 (recall also that $(A^{-1})^T = (A^T)^{-1}$ and $\det C = \det D$) we get

$$\begin{aligned} f_X(x) &= \frac{1}{(2\pi)^{n/2}\sqrt{\det D}}\, e^{-\frac{1}{2}\langle (A^{-1}(x-\mu), D^{-1}A^{-1}(x-\mu)\rangle} = \frac{1}{(2\pi)^{n/2}\sqrt{\det C}}\, e^{-\frac{1}{2}\langle (x-\mu), (A^{-1})^T D^{-1}A^{-1}(x-\mu)\rangle} \\ &= \frac{1}{(2\pi)^{n/2}\sqrt{\det C}}\, e^{-\frac{1}{2}\langle ((x-\mu), (ADA^T)^{-1}(x-\mu)\rangle} = \frac{1}{(2\pi)^{n/2}\sqrt{\det C}}\, e^{-\frac{1}{2}\langle ((x-\mu), C^{-1}(x-\mu)\rangle}. \end{aligned}$$

We omit the proof of the converse statement. It can be found in J. Jacod and P. Protter, Probability essentials, Chapter 16. $\square$

Theorem 1.9. *Let X be an n-dimensional Gaussian vector and Y be an m-dimensional Gaussian vector. If X and Y are independent then (X, Y) is an $n+m$-dimensional Gaussian vector.*

Proof. Let $Z = (X, Y)$. By independence of X and Y,

$$\varphi_Z(t) = \varphi_X(u)\varphi_Y(v), \quad t = (u, v),\ u \in \mathbb{R}^n,\ v \in \mathbb{R}^m.$$

Therefore,

$$\varphi_Z(t) = e^{i\langle u, \mu_x\rangle - \frac{1}{2}\langle u, C_X u\rangle} e^{i\langle v, \mu_y\rangle - \frac{1}{2}\langle v, C_Y v\rangle} = e^{-\langle (u,v), (\mu_X, \mu_Y)\rangle - \frac{1}{2}\langle t, Ct\rangle},$$

where

$$C = \begin{bmatrix} C_X & 0 \\ 0 & C_Y \end{bmatrix}$$

By Theorem 1.4, we conclude that Z is $n+m$-dimensional Gaussian vector. $\square$

2. Equivalent definitions of Brownian motion.

Definition 2.1 (Brownian motion)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Suppose that for each $\omega \in \Omega$ there is a continuous function $B(t)$, $t \geq 0$, such that $B(0) = 0$. Then $B := (B(t)_{t\geq 0}$ is said to be a Brownian motion if*

1. *for each $m \in \mathbb{N}$ and $t_0 = 0 < t_1 < \cdots < t_m$ the random variables*
$$B(t_1) - B(t_0),\ B(t_2) - B(t_1),\ \ldots,\ B(t_m) - B(t_{m-1})$$
are independent;
2. *for all $s > 0,\ t \geq 0$ the increment $B(s+t) - B(t)$ has a normal distribution with mean 0 and variance s.*

There are many situations when it is convenient to have alternative definitions of Brownian motion. Several of them are collected in the next theorem.

Theorem 2.2 (Alternative characterizations of Brownian motion)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Suppose that for each $\omega \in \Omega$ there is a continuous function $B(t)$, $t \geq 0$, such that $B(0) = 0$. Each of the following four statements is equivalent to properties (1) and (2) of Definition 2.1.*

(i) *the process B is a Gaussian process (i.e. for each $m \in \mathbb{N}$ and $t_0 = 0 < t_1 < \cdots < t_m$ the random variables $B(t_1)$, $B(t_2)$, ..., $B(t_m)$ are jointly normally distributed) with mean $E(B(t)) = 0$ for all $t \geq 0$ and the covariance function $C(t,s) := E(B(t)B(s)) = s \wedge t$ for all $s, t \geq 0$;*

(ii) *for each $m \in \mathbb{N}$ and $t_0 = 0 < t_1 < \cdots < t_m$ the joint moment generating function of random variables $B(t_1)$, $B(t_2)$, ..., $B(t_m)$ is given by*

$$M(u) = e^{\frac{1}{2}\langle u, Cu\rangle}, \text{ where } u \in \mathbb{R}^m, \ C = \|C_{ij}\|, \ C_{ij} = C(t_i, t_j), \ i, j \in \{1, 2, \ldots, m\}.$$

(iii) *for each $m \in \mathbb{N}$ and $t_0 = 0 < t_1 < \cdots < t_m$ the joint characteristic function of random variables $B(t_1)$, $B(t_2)$, ..., $B(t_m)$ is given by*

$$\varphi(u) = e^{-\frac{1}{2}\langle u, Cu\rangle}, \text{ where } u \in \mathbb{R}^m, \ C = \|C_{ij}\|, \ C_{ij} = C(t_i, t_j), \ i, j \in \{1, 2, \ldots, m\}.$$

(iv) *for each $m \in \mathbb{N}$ and $t_0 = 0 < t_1 < \cdots < t_m$ the joint density function of random variables $B(t_1)$, $B(t_2)$, ..., $B(t_m)$ is given by (we set $y_0 = 0$)*

$$\begin{aligned} f(y_1, y_2, \ldots, y_m) &= \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi(t_j - t_{j-1})}} e^{-\frac{(y_j - y_{j-1})^2}{2(t_j - t_{j-1})}} \\ &= \prod_{j=1}^{m} p(t_j - t_{j-1}, y_{j-1}, y_j), \text{ where } p(t,x,y) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}. \end{aligned}$$

Sketch of the proof. Suppose that $B(t)$, $t \geq 0$ is a Brownian motion according to Definition 2.1. We shall show that (i) is satisfied. To compute the joint distribution of $B(t_1)$, $B(t_2)$, ..., $B(t_m)$ we notice that

$$\begin{bmatrix} B(t_1) \\ B(t_2) \\ B(t_3) \\ \cdots \\ B(t_m) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 1 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & 1 & \ldots & 1 \end{bmatrix} \begin{bmatrix} B(t_1) \\ B(t_2) - B(t_1) \\ B(t_3) - B(t_2) \\ \cdots \\ B(t_m) - B(t_{m-1}) \end{bmatrix}. \tag{2.1}$$

This equality can be recorded as $Y = AX$, where Y and X are m-dimensional random vectors and A is an $m \times m$ matrix with zeroes above the diagonal and ones on and below the diagonal. We are given that X is Gaussian and we know its parameters from Definition 2.1. We want to show that Y is Gaussian and find its parameters. Take an arbitrary $a \in \mathbb{R}^m$ and consider

$$\langle a, Y\rangle = \langle a, AX\rangle = \langle A^t a, X\rangle.$$

Since X is Gaussian and $A^t a \in \mathbb{R}^m$, $\langle A^t a, X\rangle = \langle a, Y\rangle$ is also Gaussian, and, thus, Y is Gaussian. Now let us compute the covariance function (the mean is zero). For $0 \leq s \leq t$ we have

$$C(t,s) = E(B(t)B(s)) = E((B(t) - B(s))B(s)) + E((B(s)^2) = E(B(t) - B(s))E(B(s)) + s = s = s \wedge t.$$

The case $0 \leq t \leq s$ is similar.

Let us assume (i) and prove (ii). We need to compute $M_Y(u) = E(e^{\langle u, Y\rangle})$, where Y is the vector in the left-hand side of (2.1) and $u \in \mathbb{R}^m$. We know that Y is a Gaussian vector and, thus, its moment generating function is equal to $e^{\frac{1}{2}\langle u, C_Y u\rangle}$, where C_Y is the covariance matrix of Y (see p. 96 of Shreve, Vol. 2). But the covariance matrix of Y is equal to

$$(C_Y)_{ij} = E(Y_j Y_j) = E((B(t_i)B(t_j)) = t_i \wedge t_j = C(t_i, t_j), \text{ that is } C_Y = \begin{bmatrix} t_1 & t_1 & t_1 & \ldots & t_1 \\ t_1 & t_2 & t_2 & \ldots & t_2 \\ t_1 & t_2 & t_3 & \ldots & t_3 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ t_1 & t_2 & t_3 & \ldots & t_m \end{bmatrix}.$$

The proof of (ii) is complete.

Now assume (ii) and show (iii). We shall use the fact that the joint moment generating function determines the distribution of the random vector Y. From the form of the moment generating function

we conclude that Y is Gaussian with mean 0 and covariance matrix C_Y shown above. Now for the Gaussian random variable Y with these parameters the characteristic function is given by the claimed formula. This gives us (iii).

Assume (iii) and prove (iv). Again, from the form of the characteristic function of Y we know that Y is a Gaussian vector with mean vector 0 and covariance matrix C_Y as above. To deduce that Y has a density we need to know that C_Y is non-degenerate. Notice that $C_Y = AC_XA^t$, where A was defined above and C_X is the covariance matrix of X. But

$$X = \begin{bmatrix} B(t_1) \\ B(t_2) - B(t_1) \\ B(t_3) - B(t_2) \\ \dots \\ B(t_m) - B(t_{m-1}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix} \begin{bmatrix} B(t_1) \\ B(t_2) \\ B(t_3) \\ \dots \\ B(t_m) \end{bmatrix} = A^{-1}Y,$$

and is Gaussian (since Y is Gaussian) and $C_X = A^{-1}C_Y(A^{-1})^t$, so that

$$C_X = \begin{bmatrix} t_1 & 0 & 0 & \dots & 0 \\ 0 & t_2 - t_1 & 0 & \dots & 0 \\ 0 & 0 & t_3 - t_2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & t_m - t_{m-1} \end{bmatrix}$$

and C_Y as a product of non-degenerate matrices is non-degenerate. Therefore, Y has a density and it is equal to

$$f(y_1, y_2, \dots, y_m) = \frac{1}{(2\pi)^{m/2}\sqrt{\det C_Y}}\, e^{-\frac{1}{2}\langle y, C_Y^{-1} y\rangle}.$$

Computing (we set $y_0 = 0$)

$$\langle y, C_Y^{-1} y\rangle = \langle y, (A^t)^{-1}C_X^{-1}A^{-1}y\rangle = \langle A^{-1}y, C_X^{-1}(A^{-1}y)\rangle = \sum_{j=1}^{m} \frac{(y_j - y_{j-1})^2}{t_j - t_{j-1}};$$

$$\det C_Y = (\det A)(\det C_X)(\det A^t) = \det C_X = \prod_{j=1}^{m}(t_j - t_{j-1})$$

we obtain (iv).

Finally, assume (iv) and derive conditions (1) and (2) of Definition 2.1. We know that Y is a Gaussian vector and we know its density. We need information about $X = A^{-1}Y$. Since X is a linear transformation of a Gaussian vector it is Gaussian and its covariance matrix is $C_X = A^{-1}C_Y(A^{-1})^t$ which was computed above. Since C_X is diagonal, we know that the coordinates of X are independent. This gives (1). Condition (2) follows immediately if we set $m = 2$, $t_1 = t$, $t_2 = s + t$, and look at the parameters of the resulting Gaussian vector X. □

3. Brownian motion as a scaling limit of random walks. Donsker's invariance principle.

In this section we obtain Brownian motion on $[0, 1]$ as a limit of scaled random walks.

A realization of a Brownian motion $B(t),\ t \in [0, 1]$, is a continuous path on $[0, 1]$. Therefore, our sample space Ω can be taken to be $C([0, 1])$. To make it into a probability space we need a σ-algebra and a probability measure.

Definition 3.1. *Let $\mathcal{F}$ be the smallest σ-algebra on $C([0, 1])$ which contains all subsets of $C([0, 1])$ of the form*

$$\{B(t_1) \le x_1, B(t_2) \le x_2, \dots, B(t_m) \le x_m\},$$

where $m \in \mathbb{N}$, $0 \le t_1 < \dots < t_m \le 1$, and $x_1, x_2, \dots, x_m \in \mathbb{R}$. $\mathcal{F}$ is called a cylindrical σ-algebra *on $C([0, 1])$.*

We set $\Omega = C([0,1])$. The next step is to define a probability measure on $(\Omega, \mathcal{F})$ in such a way that all the properties of Definition 2.1 (or one of the equivalent definitions) are satisfied. One way to do this is to define this measure as a (weak) limit of simpler measures on the same space. And these simpler measures are those which come from a scaled random walk.

Let $(\Omega_0, \mathcal{F}_0, Q)$ be a probability space and $(X_j)_{j\in\mathbb{N}}$ be an i.i.d. sequence of random variables on $(\Omega_0, \mathcal{F}_0, Q)$ with mean zero and variance 1. Let $W_0 = 0$, $W_n = \sum_{j=1}^n X_j$, $n \in \mathbb{N}$. For each $\omega \in \Omega_0$ define a piece-wise linear path $B^{(n)} = \{B^{(n)}(t),\ t \in [0,1]\}$, by

$$B^{(n)}(t) = \frac{W_{[nt]}}{\sqrt{n}} + \frac{(nt - [nt])}{\sqrt{n}}(W_{[nt]+1} - W_{[nt]}), \quad t \in [0,1].$$

This gives us for each $n \in \mathbb{N}$ a measurable mapping from $(\Omega_0, \mathcal{F}_0)$ to $(\Omega, \mathcal{F})$. Now we can define a probability measure P_n on $(\Omega, \mathcal{F})$ by setting

$$P_n(A) = Q(B^{(n)} \in A)$$

for all $A \in \mathcal{F}$. The next theorem which is due to M. Donsker says that the sequence of probability measures P_n converges weakly to a probability measure P on $(C([0,1]), \mathcal{F})$ under which random paths have all the properties required by the definition of Brownian motion.

Theorem 3.2 (Donsker's invariance principle)**.** *Let P_n be defined as above. Then for every continuous*[2] *bounded function g on $\Omega := C([0,1])$*

$$\lim_{n\to\infty} \int_\Omega g\ dP_n = \int_\Omega g\, dP,$$

where P is a probability measure on $(\Omega, \mathcal{F})$ such that this measure satisfies all conditions of Definition 2.1. In short, $B^{(n)}(\cdot) \Rightarrow B(\cdot)$ as $n \to \infty$, where B is a Brownian motion.

What is the difference between the statement $B^{(n)}(t) \Rightarrow B(t)$ for each $t \in [0,1]$ and the statement of Donsker's theorem? Donsker's theorem implies not only the convergence of all finite-dimensional distributions of vectors of $(B^{(1)}, \dots, B^{(n)}_m)$ to the one of $(B(t_1), \dots, B(t_m))$ for all $m \in \mathbb{N}$ and $t_1, \dots, t_m \in [0,1]$, but also the convergence of a function of the whole random walk paths to the same function of the Brownian motion. For example, for each $t \in [0,1]$

$$\max_{0\le s\le t} B^{(n)}(s) \Rightarrow \max_{0\le s\le t} B(s).$$

In other words, any continuous function of the Brownian motion path can be approximated by taking the same function of the scaled random walk for large n. This gives us a way to simulate Brownian motion and functions of Brownian motion and compute many quantities (probabilities and expectations) needed for applications.

4. Lévy's construction of Brownian motion.

This construction gives us another way to simulate a Brownian motion. It is also known as a Brownian bridge construction and has some advantages over the random walk construction of the previous section (see P. Glasserman, Monte Carlo Methods in Financial Engineering, pp. 82-86). This part of the lecture follows A. Etheridge, A Course in Financial Calculus, pp. 56-59.

We start with a simple lemma which is the main tool of this construction.

Lemma 4.1. *Let $X = B(t/2)$ and $Y = B(t)$, where $t > 0$ is fixed. Then*

$$f_{X|Y}(x|y) = \sqrt{\frac{2}{\pi t}} \exp\left(-\frac{(x - y/2)^2}{2(t/4)}\right).$$

In other words, the conditional distribution of X given that $Y = y$ is normal with mean $y/2$ and variance $t/4$.

[2]Function $g : C([0,1]) \to \mathbb{R}$ is said to be continuous if for every $\varepsilon > 0$ there is a $\delta > 0$ such that $|g(x(\cdot)) - g(y(\cdot))| < \varepsilon$ for all $x(\cdot), y(\cdot) \in C([0,1])$ such that $\max_{t\in[0,1]} |x(t) - y(t)| < \delta$.

Exercise 1. Prove the above lemma.

We shall construct Brownian motion on $[0,1]$. Take all dyadic points in $(0,1]$: $k2^{-n}$, $n \in \mathbb{N}$, $k \in \{1,2,3,\dots,2^n\}$. There are countable many such points. Generate an i.i.d. sequence of standard normal random variables indexed by dyadic points: $\{\xi(k2^{-n})\}$, $n \in \mathbb{N}$, $k \in \{1,2,3,\dots,2^n\}$.

Step 0. Set $X_0(t) = t\xi(1)$, $t \in [0,1]$. Then $X_0(1)$ is a standard normal random variable and $X_0(t)$ is just a straight line connecting the origin and the point $(1, X_0(1))$ in the (t,x) plane.

Step 1. Set $X_1(0) = 0$, $X_1(1) = X_0(1)$, and then let $X_1(1/2) \sim N(X_0(1)/2, 1/4)$. By Lemma 4.1 the distribution of $X_1(1/2)$ is the same as that of $B(1/2)$ given that $B(1) = X_0(1)$. We can construct $X_1(1/2)$ by setting

$$X_1(1/2) = \frac{1}{2}\, X_0(1) + \frac{1}{2}\, \xi(2^{-1}).$$

Connect the tree points $(0,0)$, $(1/2, X_1(1/2))$, and $(1, X_1(1))$ by line segments to get $X_1(t)$, $t \in [0,1]$.

Then we repeat the procedure by choosing a middle point on each of $[0,1/2]$ and $[1/2,1]$ and constructing $X_2(1/4)$ and $X_2(3/4)$ by the same algorithm and then "connecting the dots". More generally, suppose that we have constructed $X_n(t)$, $t \in [0,1]$.

Step $n+1$. Set $X_{n+1}(k2^{-n}) := X_n(k2^{-n})$ for all $k \in \{1,2,3,\dots,2^n\}$. Now we need to construct X_{n+1} at all middle points of $[(k-1)2^{-n}, k2^{-n}]$ which are $(2k-1)2^{-(n+1)}$, $k \in \{1,2,3,\dots,2^n\}$. We need to have

$$X_{n+1}\left(\frac{2k-1}{2^{n+1}}\right) \sim N\left(\frac{1}{2}\left(X_n\left(\frac{k}{2^n}\right) + X_n\left(\frac{k-1}{2^n}\right)\right), \frac{1}{2^{n+2}}\right). \text{ Thus, we set}$$

$$\begin{aligned} X_{n+1}\left(\frac{2k-1}{2^{n+1}}\right) &:= \frac{1}{2}\left(X_n\left(\frac{k}{2^n}\right) + X_n\left(\frac{k-1}{2^n}\right)\right) + \frac{1}{2^{1+n/2}}\,\xi\left(\frac{2k-1}{2^{n+1}}\right) \\ &= X_n\left(\frac{2k-1}{2^{n+1}}\right) + \frac{1}{2^{1+n/2}}\,\xi\left(\frac{2k-1}{2^{n+1}}\right). \end{aligned}$$

Finally we define $X_{n+1}(t)$ to be linear on $[(k-1)2^{-n}, (2k-1)2^{-(n+1)}]$ and on $[(2k-1)2^{-(n+1)}, k2^{-n}]$.

Lemma 4.2. *Define $X_n(t)$, $t \in [0,1]$ as above. Then*

$$P(\lim_{n\to\infty} X_n(t) \text{ exists uniformly in } t \in [0,1]) = 1.$$

Lemma 4.3. *Denote $\lim_{n\to\infty} X_n(t)$ by $X(t)$ if the limit exists uniformly in $t \in [0,1]$ and set $X(t) = 0$ otherwise. Then the process $(X(t))_{t\geq 0}$ satisfies Definition 2.1 (for $t \in [0,1]$ instead of $t \in [0,\infty)$).*

Sketches of proofs of the last two lemmas can be found in A. Etheridge, A Course in Financial Calculus, pp. 58-59.

5. Brownian motion as a Markov process.

To introduce and discuss Markov property of Brownian motion we need the notion of a filtration for a Brownian motion.

Definition 5.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A filtration on $(\Omega, \mathcal{F}, P)$ is a collection of σ-algebras $(\mathcal{F}(t))_{t\geq 0}$ satisfying the following property: $\mathcal{F}(s) \subset \mathcal{F}(t) \subset \mathcal{F}$ for all $0 \leq s < t$.*

Definition 5.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(\mathcal{F}(t))_{t\geq 0}$ be a filtration, and $(X(t))_{t\geq 0}$ be a stochastic process on $(\Omega, \mathcal{F}, P)$. $(X(t))_{t\geq 0}$ is said to be adapted to the filtration $(\mathcal{F}(t))_{t\geq 0}$ if $X(t)$ is $\mathcal{F}(t)$-measurable for each $t \geq 0$.*

Definition 5.3. *Let $(B(t))_{t\geq 0}$ be a Brownian motion on the probability space $(\Omega, \mathcal{F}, P)$. A filtration for the Brownian motion is a filtration $(\mathcal{F}(t))_{t\geq 0}$ such that $(B(t))_{t\geq 0}$ is adapted to it and $B(t+s) - B(t)$ is independent of $\mathcal{F}(t)$ for all $t \geq 0$ and $s > 0$.*

The natural filtration of Brownian motion $(\mathcal{F}(t))_{t\geq 0}$ is the filtration such that each $\mathcal{F}(t)$ is the smallest σ-algebra which contains all sets of the form $\{B(t_1)\leq x_1, B(t_2)\leq x_2, \dots, B(t_m)\leq x_m\}$ for all $m\geq 1$, $0\leq t_1<t_2<\cdots<t_m\leq t$, $x_1, x_2, \dots, x_m\in\mathbb{R}$.

In general, $(\mathcal{F}(t))_{t\geq 0}$ can be generated not only by the Brownian motion but by Brownian motion together with some other process or processes. In such case the filtration $(\mathcal{F}(t))_{t\geq 0}$ can be strictly larger than the natural filtration.

Definition 5.4. *Let $(\Omega,\mathcal{F},P)$ be a probability space, $(\mathcal{F}(t))_{t\geq 0}$ be a filtration, and $(X(t))_{t\geq 0}$ be a stochastic process adapted to the filtration $(\mathcal{F}(t))_{t\geq 0}$. We say that $(X(t))_{t\geq 0}$ is a Markov process if for each $s,t\geq 0$ and for every non-negative Borel function f there is a Borel function g such that*

$$E(f(X(t+s))\,|\,\mathcal{F}(s)) = g(X(s)). \tag{5.1}$$

Remark 5.5. It is more intuitive to state the Markov property as follows: for all $A\in\mathcal{B}$ and all $s,t\geq 0$

$$P(X(t+s)\in A\,|\,\mathcal{F}(s)) = P(X(t+s)\in A\,|\,X(s)). \tag{5.2}$$

Equality (5.1) implies (5.2) if we set $f(x)=1_A(x)$ and $g(x):=E(1_A(X(t+s)\,|\,X(s)=x)$. Therefore, (5.1) seems more demanding. On the other hand, using the standard machine it is possible to show that (5.2) and (5.1) are equivalent. Our definition follows S. Shreve, Stochastic Calculus for Finance, II, p. 74, since it is more convenient for applications.

Theorem 5.6. *Let $(B(t))_{t\geq 0}$ be a Brownian motion on the probability space $(\Omega,\mathcal{F},P)$ and $(\mathcal{F}(t))_{t\geq 0}$ be its filtration (in the sense of Definition 5.3). Then $(B(t))_{t\geq 0}$ is a Markov process.*

Idea of a proof. We shall use the Independence lemma (see lecture 4, Lemma 3.1 of the refresher). Fix $s\geq 0,\ t>0$, and a non-negative Borel function f.

$$\begin{aligned} E(f(B(t+s))\,|\,\mathcal{F}(s)) &= E(f(B(t+s)-B(s)+B(s))\,|\,\mathcal{F}(s)) \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} f(y+B(s))\,dy = g(B(s)), \end{aligned}$$

where

$$g(x) := \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} e^{-\frac{y^2}{2t}} f(y+x)\,dy.$$

□

Strong Markov property is the extension of Markov property to stopping times. Let τ be a stopping time relative to filtration $(\mathcal{F}(t))_{t\geq 0}$.[3] Define

$$\mathcal{F}_\tau = \{A\in\mathcal{F}\,:\, A\cap\{\tau\leq t\}\subset\mathcal{F}_t \text{ for all } t\geq 0\}.$$

Here is a version of the strong Markov property for Brownian motion.

Theorem 5.7. *Let $(B(t))_{t\geq 0}$ be a Brownian motion on the probability space $(\Omega,\mathcal{F},P)$, $(\mathcal{F}(t))_{t\geq 0}$ be its filtration (in the sense of Definition 5.3), and τ be a stopping time relative to $(\mathcal{F}(t))_{t\geq 0}$. Then for all $A\in\mathcal{B}$ and $t\geq 0$*

$$P(B(t+\tau)-B(\tau)\in A\,|\,\mathcal{F}_\tau) = p(t,0,A);$$

$$P(B(t+\tau)\in A\,|\,\mathcal{F}_\tau) = p(t,B(\tau),A), \text{ where}$$

$$p(t,x,A) := P(B(t)\in A\,|\,B(0)=x) = \frac{1}{\sqrt{2\pi t}}\int_A e^{-(y-x)^2/(2t)}\,dy.$$

Informally speaking, the strong Markov property says that the process $(B(t+\tau)-B(\tau))_{t\geq 0}$ is again a Brownian motion and is independent from $\mathcal{F}_\tau$ or, in other words, "Brownian motion starts afresh at every stopping time". The proof is omitted. An important consequence of the strong Markov property is the reflection principle.

[3] i.e. $\{\tau\leq t\}\subset\mathcal{F}(t)$ for all $t\geq 0$.

Theorem 5.8 (Reflection principle)**.** *Let $(B(t))_{t\geq 0}$ be a Brownian motion on the probability space $(\Omega, \mathcal{F}, P)$. Then for every $a \geq 0$ and $t > 0$*

$$P(\max_{0\leq s\leq t} B(s) > a) = 2P(B(t) > a).$$

Sketch of a proof. Let $\tau_a = \inf\{t > 0 : B(t) = a\}$. Then τ_a is a stopping time and $B(\tau_a) = a$. By the strong Markov property, $B(t + \tau_a) - B(\tau_a)$ is a Brownian motion which is independent of $\mathcal{F}(\tau_a)$. Notice that $P(\max_{0\leq s\leq t} B(s) > a) = P(\tau_a < t)$. We have

$$\begin{aligned} P(\tau_a < t) &= P(\tau_a < t, B(t) > a) + P(\tau_a < t, B(t) \leq a) \\ &= P(B(t) > a) + P(B(t) \leq B(\tau_a) \,|\, \tau_a < t)P(\tau_a < t) \\ &= P(B(t) > a) + P(B(t) - B(\tau_a) \leq 0 \,|\, \tau_a < t)P(\tau_a < t) = P(B(t) > a) + \frac{1}{2}P(\tau_a < t). \end{aligned}$$

This gives $P(\tau_a < t) = 2P(B(t) > a)$ as claimed. □