CHAPTER 2: STOCHASTIC PROCESSES, BROWNIAN MOTION, DIFFUSIONS

This chapter contains background material on stochastic processes in general, and on Brownian motions and other diffusions in particular. Appendix A provides more detail on some of the topics treated here.

1. Random variables, stochastic processes

To define a random variable one starts with a **probability space** $(\Omega, \mathfrak{F}, P)$, where Ω is a set, $\mathfrak{F}$ is a σ-algebra of its subsets, and P is a probability measure on $\mathfrak{F}$. Each $\omega \in \Omega$ is an **outcome**, and each set $E \in \mathfrak{F}$ is an **event**. Given a probability space $(\Omega, \mathfrak{F}, P)$, a **random variable** is a measurable function $x : \Omega \to \mathbf{R}$. For each $\omega \in \Omega$, the real number $x(\omega)$ is the **realization** of the random variable. The probability measure for x is then

$$\mu(A) = P\left\{x^{-1}(A)\right\} = P\left\{\omega \in \Omega : x(\omega) \in A\right\}, \quad A \in \mathfrak{B},$$

where $\mathfrak{B}$ denotes the Borel sets. Since the function x is measurable, each set $x^{-1}(A)$ is in $\mathfrak{F}$, and the probability $\mu(A)$ is well defined. The distribution function for x is

$$G(a) = \mu((-\infty, a]), \quad a \in \mathbf{R}.$$

Probability measures or distribution functions for (measurable) functions of x can be constructed from μ or G.

To define a stochastic process one proceeds in a similar way, except that a time index t must be added. This index can be discrete or continuous, finite or infinite. In most of what follows the time index is continuous, starting at date 0 and having an infinite horizon, attention here will be focussed on the case $t \in [0, \infty)$. Let Ω, $\mathfrak{F}$ and P as before, but in add an increasing family of $\sigma-$algebras $\mathbb{F} \equiv \{\mathfrak{F}_t,\ t \geq 0\}$ contained in $\mathfrak{F}$. That is,

$$\mathfrak{F}_s \subseteq \mathfrak{F}_t, \text{ all } s \leq t, \quad \text{and} \quad \mathfrak{F}_t \subseteq \mathfrak{F}, \text{ all } t,$$

where $\mathfrak{F} = \mathfrak{F}_\infty$ is the smallest $\sigma-$ algebra containing all the $\mathfrak{F}_t$'s. The family $\mathbb{F} \equiv \{\mathfrak{F}_t\}$ is called a **filtration**, and $(\Omega, \mathbb{F}, P)$ is called a **filtered probability space**. The interpretation is that $\mathfrak{F}_t$ is the set of events known at time t.

A stochastic process is a function on a filtered probability space with certain measurability properties. Specifically, let $(\Omega, \mathbb{F}, P)$ be a filtered probability space with time index $t \in [0, \infty) = \mathbf{R}_+$, and let $\mathfrak{B}_+$ denote the Borel subsets of $\mathbf{R}_+$. A **continuous-time stochastic process** is a mapping x: $[0, \infty) \times \Omega \to \mathbf{R}$ that is measurable with respect to $\mathfrak{B}_+ \times \mathfrak{F}$. That is, x is *jointly* measurable in (t, ω). Given a probability space $(\Omega, \mathfrak{F}, P)$ and a filtration $\mathbb{F} = \{\mathfrak{F}_t\}$, the stochastic process x: $[0, \infty) \times \Omega \to \mathbf{R}$ is **adapted** to $\mathbb{F}$ if $x(t, \omega)$ is $\mathfrak{F}_t-$ measurable, all t.

For each fixed $t \in [0, \infty)$, the mapping $x(t, \cdot) : \Omega \to \mathbf{R}$ is an ordinary random variable on the probability space $(\Omega, \mathfrak{F}_t, P_t)$, where P_t is the restriction of P to $\mathfrak{F}_t$. That is, $x(t, \cdot)$ is an $\mathfrak{F}_t$-measurable function of ω. For each fixed $\omega \in \Omega$, the mapping $x(\cdot, \omega) : [0, \infty) \to \mathbf{R}$ is a Borel-measurable function of t. The mapping $x(\cdot, \omega)$ is called a **realization** or **trajectory** or **sample path**.

For some purposes it is convenient to view a sample path $x(\cdot, \omega)$: $[0, +\infty) \to \mathbf{R}$ as a point in an appropriate space of functions. Here the focus will be on stochastic processes that have continuous sample paths. Let $C = C[0, \infty)$ denote the space of continuous functions x: $[0, \infty) \to \mathbf{R}$. A stochastic process x is **continuous** if $x(\cdot, \omega) \in C$, a.e. $\omega \in \Omega$. Stated a little differently, a continuous stochastic process is a mapping $x\colon \Omega \to C[0,\infty)$.

2. Independence

Let $(\Omega, \mathfrak{F}, P)$ be a probability space. Two events $D, E \in \mathfrak{F}$ are **independent** if

$$P(D \cap E) = P(D)P(E).$$

Two families of sets $\mathfrak{D}, \mathfrak{E} \subset \mathfrak{F}$ are **independent** if every pair of events chosen from the two families are independent, i.e., if

$$P(D \cap E) = P(D)P(E), \quad \text{all } D \in \mathfrak{D},\ E \in \mathfrak{E}.$$

The random variables $x_1, x_2, ..., x_n$ (a finite collection) on $(\Omega, \mathfrak{F}, P)$ are **mutually independent** if for any Borel sets $A_1, A_2, ..., A_n$,

$$P(x_i \in A_i,\ i = 1, ..., n) = \prod_{i=1}^{n} P(x_i \in A_i),$$

where

$$P(x_i \in A_i) \equiv P(\omega \in \Omega : x_i(\omega) \in A_i).$$

The random variables $x_1, x_2, x_3, ...$ (an infinite collection) are mutually independent if any finite collection of them are mutually independent.

Exercise 2.1. Let $(\Omega, \mathfrak{F}, P)$ be a probability space, let $\mathfrak{D}, \mathfrak{E} \subset \mathfrak{F}$ be $\sigma-$algebras, and let x, y be random variables that are $\mathfrak{D}-$measurable and $\mathfrak{E}-$measurable respectively. Show that if $\mathfrak{D}$ and $\mathfrak{E}$ are independent, then x and y are independent. Show that if x and y are independent, then the $\sigma-$algebras $\mathfrak{D}_x$ and $\mathfrak{E}_y$ generated by x and y are independent.

3. Wiener processes, Brownian motion

A **Wiener process** (or **standard Brownian motion**) is a stochastic process W having

i. continuous sample paths

ii. stationary independent increments

iii. $W(t) \sim N(0, t)$, all t.

If $W(t)$ is a Wiener process, then over any time interval Δt, the corresponding (random) change is normally distributed with mean zero and variance Δt. That is,

$$\Delta W = \epsilon_t \sqrt{\Delta t}, \quad \text{where} \quad \epsilon_t \sim N(0, 1).$$

As Δt becomes infinitesimally small, write $dW = \epsilon_t \sqrt{dt}$, so

$$\begin{aligned} \mathrm{E}[dW] &= \mathrm{E}\left[\epsilon_t \sqrt{dt}\right] = 0, \\ \mathrm{E}\left[(dW)^2\right] &= \mathrm{E}\left[\epsilon_t^2 dt\right] = dt. \end{aligned}$$

A stochastic process X is a **Brownian motion** with **drift** μ and **variance** σ^2 if

$$X(t) = X(0) + \mu t + \sigma W(t), \quad \text{all } t, \tag{1}$$

where W is a Wiener process. Clearly, the state space for any Brownian motion, including a Wiener process, is all of $\mathbf{R}$. Since $\mathrm{E}[W(t)] = 0$ and $\mathrm{Var}[W(t)] = t$, it follows that

$$\begin{aligned} \mathrm{E}[X(t) - X(0)] &= \mu t, \\ \mathrm{Var}[X(t) - X(0)] &= \sigma^2 t, \quad \text{all } t. \end{aligned}$$

Notice that the both the *drift* and *variance* of a Brownian motion increase *linearly* with the time interval.

The next result says that any continuous stochastic process $\{X(t),\ t \geq 0\}$ that has stationary independent increments is a Brownian motion.

Theorem 2.1: If the stochastic process $\{X(t),\ t \geq 0\}$ has continuous sample paths with stationary, independent, and identically distributed increments, then it is a Brownian motion. That is, there exists (μ, σ^2) such that for each $t \geq 0$, the random variable $[X(t) - X(0)]$ has a normal distribution with parameters $(\mu t, \sigma^2 t)$.

See Breiman (1968, Prop. 12.4) for a proof. The idea is to use the Central Limit Theorem.

Theorem 2.1 says that in order to work in continuous time with a stochastic process that has continuous sample paths and i.i.d. increments, one must accept normality. The parameters (μ, σ^2) of the Brownian motion in (1) can be chosen, but nothing more. (Thus, part (iii) in the definition of a Wiener process is redundant.) In section 7 a broader class of continuous stochastic processes will be defined by dropping the requirement that the increments be identically distributed.

Since the mean of a Brownian motion grows like t and the standard deviation like $\sqrt{t}$, the standard deviation dictates the overall nature of the path in the short run and the drift—unless it is zero—dominates in the long run. Figure 2.1 displays the expected value and three confidence three bands, at 66%, 95%, and 99%, for Brownian motions $X(t)$ with and without drift.

4. Random walk approximation of Brownian motion

A Brownian motion can be viewed as the limit of discrete-time random walks as the time interval and the step size shrink together in a certain way. Suppose that over each time increment Δt the process X increases by h with probability p and decreases by h with probability $(1-p)$. Let $\Delta X = X(t + \Delta t) - X(t)$ denote the increment in X. Then

$$\begin{aligned} \mathrm{E}[\Delta X] &= ph - (1-p)\,h \\ &= (2p-1)\,h, \end{aligned}$$

$$\begin{aligned}\mathrm{Var}\left[\Delta X\right] &= \mathrm{E}\left[(\Delta X)^2\right] - \left[\mathrm{E}\left(\Delta X\right)\right]^2 \\ &= \left[1-(2p-1)^2\right]h^2.\end{aligned}$$

Thus, to approximate the drift and variance of a Brownian motion with parameters (μ, σ^2), p and h must satisfy

$$\begin{aligned}\mu\Delta t &= (2p-1)\,h, \\ \sigma^2\Delta t &= 4p\,(1-p)\,h^2.\end{aligned} \tag{2}$$

Eliminating h gives a quadratic in p,

$$p^2 - p + \frac{\sigma^2}{4\left(\sigma^2+\mu^2\Delta t\right)} = 0.$$

Hence

$$\begin{aligned}p &= \frac{1}{2}\left(1+\sqrt{1-\frac{\sigma^2}{\sigma^2+\mu^2\Delta t}}\right) \\ &= \frac{1}{2}\left(1+\frac{\mu\sqrt{\Delta t}}{\sqrt{\sigma^2+\mu^2\Delta t}}\right) \\ &\approx \frac{1}{2}\left(1+\frac{\mu}{\sigma}\sqrt{\Delta t}\right),\end{aligned} \tag{3}$$

where the roots are allocated so that $p \gtreqless 1/2$ as $\mu \gtreqless 0$, and the last line is a good approximation if Δt is small relative to σ^2/μ^2. For the step size h, use (2) to find that

$$h = \frac{\mu\Delta t}{2p-1} = \sigma\sqrt{\Delta t}. \tag{4}$$

The idea is that since the drift and variance both have order Δt, if Δt is small relative to σ^2/μ^2, then $\left[\mathrm{E}\left(\Delta X\right)\right]^2 = \mu^2\left(\Delta t\right)^2$ is negligible relative to the variance $\mathrm{Var}[\Delta X] = \sigma^2\Delta t$. Hence $\mathrm{Var}[\Delta X] \approx \mathrm{E}\left[(\Delta X)^2\right]$ or $\sigma^2\Delta t \approx h^2$, as in (4).

In summary, to approximate a Brownian motion with parameters (μ, σ^2), choose Δt small relative to σ^2/μ^2, and choose h and p so that

$$(2p-1)\,h = \mu\Delta t, \qquad \text{and} \qquad h^2 = \sigma^2\Delta t,$$

as in (3) and (4). Notice that as the time increment Δt shrinks to zero, the step size h also shrinks to zero.

The following sequence of approximations, with decreasing time increment Δ_n, illustrates the idea behind Theorem 2.1 and also suggests methods for numerically approximating Brownian motions to any desired degree of accuracy. Fix a time interval $T > 0$, and define

$$\begin{aligned} \Delta_n &\equiv T/n, \qquad h_n \equiv \sigma\sqrt{T/n}, \qquad p_n \equiv \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{T/n}\right), \\ \mathbf{X}_n &\equiv \{-nh_n, ..., 0, ..., +nh_n\} \\ &= \left\{-\sigma\sqrt{nT}, ..., 0, ..., +\sigma\sqrt{nT}\right\}, \qquad\qquad n = 1, 2, ... \end{aligned}$$

For each n, let $\{\xi_i^n\}_{i=1}^n$ be a sequence of i.i.d. random variables taking values $\pm h_n$ with probabilities p_n and $(1 - p_n)$, and let

$$X_n = \xi_1^n + \xi_2^n + ... + \xi_n^n, \quad n = 1, 2, ...,$$

be their sum. Each of the random variables ξ_i^n has mean $\mu T/n$ and variance $\sigma^2 T/n$, so each of the random variables X_n has mean μT and variance $\sigma^2 T$. Moreover, each X_n takes values in the finite set $\mathbf{X}_n$ consisting of a grid of $2n+1$ evenly spaced points centered around zero. The grid is on an interval of length $2\sigma\sqrt{nT}$ and the step size is $\sigma\sqrt{T/n}$. Thus, the number of grid points grows like n, with the interval length and the number of grid points on any subinterval both increasing like $\sqrt{n}$. Hence as $n \to \infty$ the number of points on any subinterval grows without bound and the state space expands to cover all of $\mathbf{R}$. Moreover, by the Central Limit Theorem the sequence $\{X_n\}$ converges in distribution to a normal random variable.

Exercise 2.2. Consider a (μ, σ^2) Brownian motion with initial state $x(0) = 0$. Fix $T > 0$ and $a > 0$. Use a sequence of discrete time approximations to calculate

$$\Pr\left\{\max_{0 \le t \le T} x(t) \ge a\right\}.$$

5. Stopping times

A **stopping time** on a filtered probability space $(\Omega, \mathbb{F}, P)$ is a measurable function $T\colon \Omega \to [0, +\infty) \cup +\infty$ with the property that

$$\{\omega \in \Omega : T(\omega) \leq t\} \in \mathfrak{F}_t, \quad \text{all } t \geq 0. \tag{5}$$

The idea is that $T(\omega)$ is the (random) date when something happens, with $T(\omega) = +\infty$ interpreted as meaning it never happens. The measurability condition in (5) says that it must be possible to tell at date t, given the available information $\mathfrak{F}_t$, whether or not it has happened already. Notice that the filtration $\mathbb{F}$ is crucial for deciding whether a function T is a stopping time. If T takes values in $\mathbf{R}_+$ with probability one, then T is a random variable. Here $T < +\infty$ will be used as a shorthand notation for $P\{T < +\infty\} = 1$.

The following are some standard examples of stopping times.

a. Any fixed date $T = \hat{T}$ is a stopping time.

b. Suppose $\{X(t)\}$ is a stochastic process on $(\Omega, \mathbb{F}, P)$, and let A be any Borel set. Then the first date t for which $X(t) \in A$ is a stopping time. The k^{th} date for which $X(t) \in A$ is also a stopping time, for any $k = 2, 3, 4, \ldots$..

c. If S and T are stopping times, then $S + T$, $S \wedge T$, and $S \vee T$ are also stopping times, where $a \wedge b \equiv \min\{a, b\}$, and $a \vee b \equiv \max\{a, b\}$.

Exercise 2.3. Let $\{X(t)\}$ be a stochastic process on $(\Omega, \mathbb{F}, P)$. Explain briefly why each of the following is or is not a stopping time:

a. $T - S$, where S and T are stopping times with $S \leq T$.

b. Let $\{a_j\}_{j=1}^n$ be a sequence of real numbers, and consider the first time X reaches a_n, after first reaching a_{n-1}, after first reaching a_{n-2}, ..., after first reaching a_1.

c. Let $\{A_j\}_{j=1}^n$ be a sequence of measurable sets, and consider the first time X

reaches A_n, after first reaching A_{n-1}, after first reaching A_{n-2}, ..., after first reaching A_1.

d. $T - \Delta$, where T is any one of the stopping times above and $0 < \Delta$ is a constant.

6. Strong Markov property

Let $\{X(t, \omega)\}$ be a stationary stochastic process on the filtered space $(\Omega, \mathbb{F}, P)$. Then X has the **strong Markov property** if, for any sequence of stopping times $T_0 < T_1 < ... < T_n$, any $s > 0$, any $x_0, x_1, ..., x_n$, and any measurable set A,

$$\Pr\{X(T_n + s) \subset A \mid X(T_0) = x_0,\ X(T_1) = x_1,\ ...,\ X(T_n) = x_n\} \tag{6}$$

$$= \Pr\{X(T_n + s) \subset A \mid X(T_n) = x_n\}.$$

That is, given the state at stopping time T_n, outcomes at earlier stopping times are not useful for predicting outcomes at later dates. Stated a little differently, (6) says that the evolution of the system after any stopping time T_n, given $X(T_n)$, is conditionally independent of the earlier outcomes. If the process is not stationary the condition is more complex but similar in spirit.

The next theorem says that if a Brownian motion X is reinitialized at a stopping time T, the resulting process is also a Brownian motion, with the same mean and variance as the original. That is, Brownian motions have the strong Markov property.

THEOREM 2.2: Let X be a (μ, σ^2) Brownian motion and $T < \infty$ a stopping time on the filtered space $(\Omega, \mathbb{F}, P)$, and let

$$X^*(t, \omega) = X(T + t, \omega) - X(T, \omega), \quad \text{all } t \geq 0, \text{ all } \omega \in \Omega.$$

Then X^* is a (μ, σ^2) Brownian motion with initial value zero, and for any $t > 0$, the random variables $X^*(t)$ and T are independent.

See Billingsley (1995, Theorem 37.5) for a proof. The idea is that since X has continuous sample paths and increments that are i.i.d. (and normally distributed), X^* also has these properties. Hence $\{X^*(t)\}$ satisfies the hypotheses of Theorem 2.1.

Since a constant value $T(\omega) = \hat{T}$ is a valid stopping time, the theorem implies that a Brownian motion that is reinitialized at a fixed date is itself a Brownian motion, with the same drift and variance.

Notice, however, that the filtration $\mathbb{F}$ is not appropriate for the process X^*. One that is can be constructed in the usual way. For each $t \geq 0$, let $\mathfrak{F}_t^*$ be the smallest $\sigma-$algebra for which the random variables $X^*(s, \omega)$, $0 \leq s \leq t$ are measurable; let $\mathfrak{F}^* = \mathfrak{F}_\infty^*$ be the smallest $\sigma-$algebra containing all of the $\mathfrak{F}_t^*$'s; and define the filtration $\mathbb{F}^* = \{\mathfrak{F}_t^*,\ t \geq 0\}$. Then $(\Omega, \mathbb{F}^*, P)$ is a filtered space and, by construction, X^* is adapted to $\mathbb{F}^*$.

7. Diffusions

The term **diffusion** will be used here to refer to a continuous-time stochastic process that (a) has continuous sample paths and (b) has the strong Markov property. The state space for a diffusion can be all of $\mathbf{R}$ or any (open, closed or half-open) interval. That is, the state space is an interval of the form (ℓ, r), $[\ell, r)$, $(\ell, r]$, $[\ell, r]$, where $\ell = -\infty$ or $r = +\infty$ are allowed at open endpoints. Clearly, any Brownian motion is a diffusion, and its state space is all of $\mathbf{R}$.

A diffusion is **regular** if, starting from any point in the interior of the state space, any other interior point is, with positive probability, reached in finite time. That is, if T_y denotes the first time the process reaches y, then

$$\Pr\{T_y < \infty \mid X(0) = x\} > 0, \quad \text{all } \ell < x, y < r.$$

This assumption rules out the possibility of non-communicating subsets. All of the diffusions considered here are regular.

Continuity implies that the probability of a large change in the state can be made arbitrarily small by making the time interval sufficiently short. Formally, let $\Delta > 0$ denote a time increment, and let

$$h(t,\Delta) \equiv X(t+\Delta) - X(t)$$

be the change in the level of the process during an interval of length Δ after date t. A diffusion has the property that for any $\varepsilon > 0$

$$\lim_{\Delta \downarrow 0} \Pr \{ |h(t,\Delta)| > \varepsilon \mid X(t) = x \} = 0, \quad \text{all } x, \text{ all } t.$$

That is, the probability of a change of fixed size $\varepsilon > 0$ goes to zero as the time interval Δ gets arbitrarily short.

Any diffusion is characterized by its **infinitesimal parameters** $\mu(t,x)$ and $\sigma^2(t,x)$, defined by

$$\begin{aligned} \mu(t,x) &\equiv \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathrm{E} \left[h(t,\Delta) \mid X(t) = x \right], \\ \sigma^2(t,x) &\equiv \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathrm{E} \left[[h(t,\Delta)]^2 \mid X(t) = x \right], \quad \text{all } x, t, \end{aligned} \tag{7}$$

where $h(t,\Delta)$ is the increment between t and $t+\Delta$, defined above. The function μ is called the **drift** or **infinitesimal mean** and σ^2 is called the **diffusion** parameter or **infinitesimal variance.** The functions μ and σ^2 in (7) will be assumed to be continuous, and the latter to be strictly positive on the interior of the state space. In most of applications of interest here μ and σ are time-invariant, so the argument t does not appear.

All higher moments are usually zero:

$$\lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathrm{E} \left[[h(t,\Delta)]^r \mid X(t) = x \right] = 0, \quad \text{all } x, t, \quad r = 3, 4, ...$$

Although this not required for a diffusion, it holds for all the examples considered here.

Two particular diffusions are widely used in economics.

Example A: In many contexts it is useful to employ a stochastic process that has *relative* increments $\Delta X/X$ that are i.i.d. A stochastic process $X(t)$ is a **geometric Brownian motion** if

$$dX = \mu X dt + \sigma X dW,$$

so (with an obvious abuse of notation) $\mu(t,x) = \mu x$ and $\sigma^2(t,x) = \sigma^2 x$. The state space for a geometric Brownian motion is $\mathbf{R}_+$.

Example B: In other contexts it is useful to employ a stochastic process that is *mean reverting*. A process with this property is the **Ornstein-Uhlenbeck** process, which has infinitesimal parameters $\mu(x) = -\alpha x$, where $\alpha > 0$, and $\sigma^2 > 0$. The state space for this process is all of $\mathbf{R}$, but the process has a central tendency toward zero: the drift is negative if the state exceeds zero and is positive if the state is below zero, with the magnitude of the drift increasing with the distance from zero.

In contrast to a Brownian motion or geometric Brownian motion, the Ornstein-Uhlenbeck (OU) process has a stationary distribution. In particular, the stationary distribution is normal with mean zero and variance $\gamma = \sigma^2/2\alpha$. Hence it has the stationary density $\psi(x) = ce^{-\gamma x^2}$, where $c > 0$ is a constant that depends on the parameters α, σ^2.

8. Discrete approximation of an Ornstein-Uhlenbeck process

An Ornstein-Uhlenbeck process can be approximated with a sequence of discrete-time, discrete-state processes in the same way that Brownian motion can be. Fix the parameters (α, σ^2) and a time interval $T > 0$. For each $n = 1, 2, ...$, define the time increment Δ_n as before,

$$\Delta_n \equiv T/n.$$

Suppose, also as before, that during each time increment the process moves up or down one step. Notice that for an OU process both the step size h_n and transition probabilities, p_n and $1 - p_n$, could, in principle, depend on the current state x. It would be extremely awkward if the step size varied with x, however, since a fixed grid could no longer be used: one step up from x followed by one step down from $x + h_n(x)$ would bring the process to $x + h_n(x) - h_n[x + h_n(x)] \neq x$. Fortunately, this does not happen.

As before, the condition for the variance pins down the step size,

$$h_n = \sigma\sqrt{\Delta_n}, \qquad \text{all } n.$$

Hence the step size is independent of x, and the state space

$$\mathbf{X}_n \equiv \{..., -2h_n, -h_n, 0, +h_n, +2h_n, ...\}, \qquad n = 1, 2, ...,$$

can be used. Let

$$p_n(x) = \frac{1}{2} - \varepsilon_n(x), \quad x \in \mathbf{X}_n,$$

be the probability of an upward step. The transition probabilities must satisfy the drift condition,

$$\begin{aligned} -\alpha x \Delta_n &= p_n(x)h_n - [1 - p_n(x)]\, h_n \\ &= \left[\frac{1}{2} - \varepsilon_n(x)\right] h_n - \left[\frac{1}{2} + \varepsilon_n(x)\right] h_n \\ &= -2\varepsilon_n(x)h_n, \end{aligned}$$

so

$$\begin{aligned} \varepsilon_n(x) &= \frac{\alpha x}{2}\frac{\Delta_n}{h_n} \\ &= \frac{\alpha x}{2\sigma}\sqrt{\Delta_n}, \quad \text{all } x \in \mathbf{X}_n, \text{ all } n. \end{aligned}$$

It is straightforward to verify that all higher moments converge to zero as $n \to \infty$.

9. Notes

A good introduction to basic probability theory is Ross (1989). Feller (1968) is a classic treatment at a basic level, and Feller (1971) continues to more advanced material. Both contain many examples and lots of helpful discussion. At an advanced level Breiman (1968), Chung (1974), and Billingsley (1995) are also excellent. Breiman includes a nice introduction to stochastic processes in general and to Brownian motion and Ornstein-Uhlenbeck processes in particular.

For an introduction to the basics of stochastic processes see Cinlar (1975) or Ross (1983). Karlin and Taylor (1975, 1981) provide an outstanding treatment at a more advanced level, with many applications and examples and an emphasis on problem solving. See Chapter 7 for a discussion of many of the properties of Brownian motion. See Chapters 14 and 15 for a more detailed treatment of the strong Markov property and an excellent discussion of diffusions, including many examples and many useful results.

If the stopping times in the definition of the strong Markov property are replaced with fixed dates, then X has the Markov property. For discrete-time processes the two concepts are equivalent, but for continuous-time processes they differ. See Chung (1974, Sect. 9.2) for a further discussion. Note that the definition here is for stationary processes only.