ltcc4.tex **Week 4** 12.3.2018

Chapter IV. STOCHASTIC PROCESSES IN CONTINUOUS TIME. BROWNIAN MOTION.

1. Markov Processes.

X is *Markov* if for each t, each $A \in \sigma(X_s : s > t)$ (the 'future') and $B \in \sigma(X_s : s < t)$ (the 'past'),

$$P(A|X_t, B) = P(A|X_t).$$

That is, if you know where you are (at time t), how you got there doesn't matter so far as predicting the future is concerned – equivalently, past and future are conditionally independent given the present.

The same definition applied to Markov processes in discrete time.

If both time and state are discrete, the term *Markov chain* is usually used. We may then label the states as $1, 2, \ldots$ (there may be a finite number of states $1, \ldots, N$, or an infinite number; one then speaks of a finite Markov chain or an infinite one. The process $X = (X_n)$ may then be specified by its *transition probability matrix* $P = (p_{ij})$, where

$$p_{ij} := P(X_{n+1} = j|X_n = i)$$

(we restrict attention to *stationary* Markov chains, where this matrix does not depend on time n).

Markov processes (and chains) have been much studied. They have an extensive and interesting theory, and they provide models for many of the standard situations studied in Applied Probability. See e.g. Norris [N].

A situation is Markov if knowing the present is all that is needed to study the future. Roughly speaking, non-Markovian situations, in which one needs to know not only the present but also how one got there, are much harder, and are usually intractable. Again roughly speaking, the two main kinds of dependence where one can get useful results are mgs and Markov processes.

X is said to be *strong Markov* if the Markov property holds with the *fixed* time t replaced by a *stopping time* T (a random variable). This is a real restriction of the Markov property in the continuous-time case (though not in discrete time).

Example. If we take T an exponentially distributed random variable, and define a stochastic process X by

$$X(t) = 0 \qquad (t \leq T), \qquad t - T \qquad (t \geq T),$$

then the Markov property holds at any fixed time t, but not at T.

Another standard example of a process which is Markov but not strong Markov is provided by the left-continuous Poisson process (i.e., a (right-continuous) Poisson process made left-continuous at its jumps).

1a. Diffusions.

A diffusion is a path-continuous strong-Markov process such that for each time t and state x the following limits exist:

$$\mu(t,x) := \lim_{h\downarrow 0}\frac{1}{h}E[(X_{t+h}-X_t)|X_t = x], \quad \sigma^2(t,x) := \lim_{h\downarrow 0}\frac{1}{h}E[(X_{t+h}-X_t)^2|X_t = x].$$

Then $\mu(t,x)$ is called the *drift*, $\sigma^2(t,x)$ the *diffusion coefficient*.

Diffusions are closely linked to Brownian motion $B = (B_t)$ (below), and to martingales. In Week 5, we introduce the *Itô integral*, which allows one to integrate a suitable random integrand $Y = (Y_t)$ with respect to Brownian motion, thus defining a *stochastic integral* $\int_0^t Y(u)dB(u)$, or $\int_0^t Y dB$. One may then study *stochastic differential equations* (SDEs), such as

$$dX_t = \mu(t,X_t)dt + \sigma(t,X_t)dB_t.$$

Under suitable conditions, such an SDE has a solution $X = (X_t)$, which is a diffusion with drift μ and diffusion coefficient σ. All this extends to the multidimensional case. In $\mathbb{R}^d$, X_t, μ are d-vectors, σ a $d\times d$ matrix.

Note. As with ODEs and PDEs, one needs to have existence theorems and uniqueness theorems – and one has more than one sense in which 'solution' can be taken. With SDEs, one needs to discriminate between *weak* and *strong* solutions. For background, see e.g. Øksendal [Ø].

Generators. Write $D = d/dx$ for the differentiation operator in one dimension, $D_i = \partial/\partial x_i$ in $\mathbb{R}^d$; thus $D^2 = d^2/dx^2$, $D_{ij} = \partial^2/\partial x_i \partial x_j$. Write

$$L_t := \frac{1}{2}\sigma(t,.)D^2 + \mu(t,.)D, \qquad \text{or} \qquad \frac{1}{2}\sum_{i,j=1}^{d}\sigma_{ij}(t,.)D_{ij} + \sum_{i=1}^{d}\mu_i(t,.)D_i;$$

then L is an elliptic differential operator (linear, second-order, partial if $d > 1$). Under suitable conditions, the parabolic PDE

$$L_t f + \partial f/\partial t = 0 \qquad (PPDE)$$

has as solutions the transition prob. density function for the diffusion X.

Example: Brownian motion. The prototype here is Brownian motion (below), where $\mu = 0$, $\sigma = 1$ (or I in higher dimensions), $L = \frac{1}{2}D^2$ (or $\frac{1}{2}\Delta$ in higher dimensions, with Δ the Laplacian) and $(PPDE)$ is the *heat equation.*

In one dimension, the usual treatment of diffusions uses the *scale function* and *speed measure*; see e.g. Breiman [Bre], Ch. 16, Rogers & Williams [R-W2], V.46, 47. Here one uses the total ordering of the real line (so this is specific to one dimension). In higher dimensions, one uses the Stroock-Varadhan approach via *martingale problems*; see [SV].

2. Gaussian Processes.

Recall the multivariate normal distribution $N(\mu, \mathbf{\Sigma})$ in n dimensions. If $\mu \in \mathbb{R}^n$, $\mathbf{\Sigma}$ is a non-negative definite $n \times n$ matrix, $\mathbf{X}$ has distribution $N(\mu, \mathbf{\Sigma})$ if it has characteristic function

$$\phi_{\mathbf{X}}(\mathbf{t}) := E[\exp\{i\mathbf{t}^T.\mathbf{X}\}] = \exp\{i\mathbf{t}^T.\mu - \frac{1}{2}\mathbf{t}^T\mathbf{\Sigma}\mathbf{t}\} \qquad (\mathbf{t} \in \mathbb{R}^n).$$

If further $\mathbf{\Sigma}$ is positive definite (so non-singular), $\mathbf{X}$ has density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{1}{2}n}|\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\mathbf{x} - \mu)^T\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu)\}$$

(*Edgeworth's Theorem*, 1893).

A process $X = (X_t)_{t \geq 0}$ is *Gaussian* if all its finite-dimensional distributions are Gaussian. Such a process can be specified by:
(i) a measurable function $\mu = \mu(t)$ with $EX_t = \mu(t)$,
(ii) a non-negative definite function $\sigma(s,t)$ with $\sigma(s,t) = cov(X_s, X_t)$.

3. Brownian Motion.

The Scottish botanist Robert Brown observed pollen particles in suspension under a microscope in 1828 and 1829 (though Leeuwenhoek had observed the phenomenon before him – indeed, so had Lucretius in antiquuity, in *De rerum natura* – The Nature of Things), and observed that they were in constant irregular motion.

In 1900 L. Bachelier considered Brownian motion a possible model for stock-market prices (for a recent translation with commentary, see [Bach]) - the first time Brownian motion had been used to model financial or economic phenomena, and before a mathematical theory had been developed.

In 1905 Albert Einstein considered Brownian motion as a model of particles in suspension, and used it to estimate *Avogadro's number* ($N \sim 6 \times 10^{23}$), based on the diffusion coefficient D in the *Einstein relation*

$$var X_t = Dt \qquad (t > 0).$$

Definition. **Brownian motion** (BM) on $\mathbb{R}$ is the process $B = (B_t : t \geq 0)$ such that:
(i) $B_0 = 0$;
(ii) B has stationary independent increments (so B is a Lévy process);
(iii) B has Gaussian increments: for $s, t \geq 0$, $B_{t+s} - B_s \sim N(0, t)$;
(iv) B has continuous paths: $t \mapsto B_t$ is continuous ($t \mapsto B(t, \omega)$ is continuous for all $\omega \in \Omega$).
[The path-continuity in (iv) can be relaxed by assuming it only a.s.; we can then get continuity by excluding some null-set from our probability space.]

The fact that Brownian motion so defined *exists* is quite deep, and was first proved by Norbert Wiener (1894-1964) in 1923. In honour of this, Brownian motion is also known as the *Wiener process*, and the probability measure generating it - the measure W on $C[0,1]$ (one can extend to $C[0,\infty)$) by

$$W(A) = P(B_. \in A) = P(\{t \mapsto B_t(\omega)\} \in A)$$

for all Borel sets $A \in C[0,1]$ is called *Wiener measure.*
Covariance. Before addressing existence, we first find the covariance function. For $s \leq t$, $B_t = B_s + (B_t - B_s)$, so as $E[B_t] = 0$,

$$cov(B_s, B_t) = E[B_s B_t] = E[B_s^2] + E[B_s(B_t - B_s)].$$

The last term is $E[B_s]E[B_t - B_s]$ by independent increments, $= 0$, so

$$cov(B_s, B_t) = E[B_s^2] = s \qquad (s \leq t): \qquad cov(B_s, B_t) = \min(s, t).$$

A Gaussian process (one whose finite-dimensional distributions are Gaussian) is specified by its mean function and its covariance function, so among centred (zero-mean) Gaussian processes, the covariance function $\min(s, t)$ serves as the signature of Brownian motion.
Finite-Dimensional Distributions.

For $0 \leq t_1 < \ldots < t_n$, the joint law of $X(t_1), X(t_2), \ldots, X(t_n)$ can be obtained from that of $X(t_1), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$. These are jointly Gaussian, hence so are $X(t_1), \ldots, X(t_n)$: the finite-dimensional distributions are *multivariate normal.* Recall that the multivariate normal law in n dimensions, $N_n(\mu, \Sigma)$ is specified by the mean vector μ and the covariance matrix Σ (non-negative definite) by its CF:

$$E[\exp\{i\mathbf{u}^T\mathbf{X}\}] = \exp\{i\mathbf{u}^T\mathbf{X} - \frac{1}{2}\mathbf{u}^T\Sigma\mathbf{u}\},$$

and when Σ is positive definite (so non-singular), the joint density is given by Edgeworth's theorem. So to check the finite-dimensional distributions of BM - stationary independent increments with $B_t \sim N(0,t)$ - it suffices to show that they are multivariate normal with mean zero and covariance $cov(B_s, B_t) = \min(s,t)$ as above.
Construction of BM.

It suffices to construct BM for $t \in [0,1]$). This gives $t \in [0,n]$ by dilation, and $t \in [0,\infty)$ by letting $n \to \infty$.

First, take $L^2[0,1]$, and any complete orthonormal system (cons) (ϕ_n) on it. Now L^2 is a Hilbert space, under the inner product

$$\langle f, g\rangle = \int_0^1 f(x)g(x)dx \text{ (or } \textstyle\int fg),$$

so norm $\|f\| := (\int f^2)^{1/2})$. By Parseval's identity,

$$\int_0^1 fg = \sum_{n=0}^{\infty} \langle f, \phi_n\rangle \langle g, \phi_n\rangle$$

(where convergence of the series on the right is in L^2, or in mean square: $\|f - \sum_0^n \langle f, \phi_k\rangle \phi_k\|$
$\to 0$ as $n \to \infty$). Now take, for $s, t \in [0,1]$,

$$f(x) = I_{[0,s]}(x), \qquad g(x) = I_{[0,t]}(x).$$

Parseval's identity becomes

$$\min(s,t) = \sum_{n=0}^{\infty} \int_0^s \phi_n dx \int_0^t \phi_n(x)dx.$$

Now take (Z_n) independent and identically distributed $N(0,1)$, and write

$$B_t = \sum_{n=0}^{\infty} Z_n \int_0^t \phi_n(x)dx.$$

This is a sum of independent random variables. Kolmogorov's theorem on random series ('three-series theorem' – see e.g. [Brei] §3.4, [G-S], 7.11.35) says that it converges a.s. if the sum of the variances converges. This is $\sum_{n=0}^{\infty} (\int_0^t \phi_n(x)dx)^2, = t$ by above. So the series above converges a.s., and by

excluding the exceptional null set from our prob. (as we may), everywhere.
The Haar System. Define the 'mother wavelet' H by

$$H(t) = 1 \text{ on } [0, \tfrac{1}{2}), \qquad -1 \text{ on } [\tfrac{1}{2}, 1], \qquad 0 \text{ else.}$$

Write $H_0(t) \equiv 1$, and for $n \geq 1$, express n in dyadic form as $n = 2^j + k$ for a unique $j = 0, 1, \ldots$ and $k = 0, 1, \ldots, 2_j - 1$. Using this notation for n, j, k throughout, define the 'daughter wavelets' by

$$H_n(t) := 2^{j/2} H(2^j t - k)$$

(so H_n has support $[k/2^j, (k+1)/2^j]$). So if m, n have the same j, $H_m H_n \equiv 0$, while if m, n have different js, one can check that $H_m H_n$ is $2^{(j_1+j_2)/2}$ on half its support, $-2^{(j_1+j_2)/2}$ on the other half, so $\int H_m H_n = 0$. Also H_n^2 is 2^j on $[k/2^j, (k+1)/2^j]$, so $\int H_n^2 = 1$. Combining:

$$\int H_m H_n = \delta_{mn},$$

and (H_n) form an orthonormal system, called the *Haar system*. For completeness: the indicator of any dyadic interval $[k/2^j, (k+1)/2^j]$ is in the linear span of the H_n (difference two consecutive H_ns and scale). Linear combinations of such indicators are dense in $L^2[0, 1]$. Combining: the Haar system (H_n) is a cons in $L^2[0, 1]$.
The Schauder System.

We obtain the *Schauder system* by integrating the Haar system. Consider the triangular function (or 'tent function') as mother wavelet:

$$\Delta(t) := 2t \qquad (0 \leq t \leq \frac{1}{2}), \qquad 2(1-t) \qquad (\frac{1}{2} \leq t \leq 1), \qquad 0 \qquad \text{else.}$$

Write $\Delta_0(t) := t$, $\Delta_1(t) := \Delta(t)$, and define the nth *Schauder function* Δ_n (daughter wavelets) by

$$\Delta_n(t) := \Delta(2^j t - k) \qquad (n = 2^j + k \geq 1).$$

Note that Δ_n has support $[k/2^j, (k+1)/2^j]$ (so is 'localized' on this dyadic interval, which is small for n, j large). We see that

$$\int_0^t H(u)du = \frac{1}{2}\Delta(t), \qquad \int_0^t H_n(u)du = \lambda_n \Delta_n(t),$$

$$\lambda_0 = 1, \qquad \lambda_n = \frac{1}{2}.2^{-j/2} \qquad (n = 2^j + k \geq 1).$$

THEOREM (Paley-Wiener-Zygmund, 1933). For $(Z_n)_0^\infty$ independent $N(0,1)$ random variables, λ_n, Δ_n as above,

$$B_t := \sum_{n=0}^{\infty} \lambda_n Z_n \Delta_n(t)$$

converges uniformly on $[0,1]$, a.s. The process $B = (B_t : t \in [0,1])$ is Brownian motion.

Lemma. For Z_n independent $N(0,1)$,

$$|Z_n| \leq C\sqrt{\log n} \qquad \forall n \geq 2,$$

for some random variable $C < \infty$ a.s.

Proof. For $x > 1$,

$$P(|Z_n| \geq x) = \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}u^2} du \leq \sqrt{2/\pi} \int_x^\infty u e^{-\frac{1}{2}u^2} du = \sqrt{2/\pi} e^{-\frac{1}{2}x^2}.$$

So for any $a > 1$,

$$P(|Z_n| > \sqrt{2a \log n}) \leq \sqrt{2/\pi} \exp(-a \log n) = \sqrt{2/\pi}.n^{-a}.$$

Since $\sum n^{-a} < \infty$ for $a > 1$, the Borel-Cantelli lemma (see e.g. [Brei] §3.3, or [G-S] §7.3 Th. 10) gives

$$P(|Z_n| > \sqrt{2a \log n} \text{ for infinitely many } n) = 0 : \qquad C := \sup_{n \geq 2} \frac{|Z_n|}{\sqrt{\log n}} < \infty \qquad a.s.$$

Proof of the Theorem.
1. Convergence. Choose J and $M \geq 2^J$; then

$$\sum_{n=M}^{\infty} \lambda_n |Z_n| \Delta_n(t) \leq C \sum_M^\infty \lambda_n \sqrt{\log n} \Delta_n(t).$$

The right is majorized by

$$C. \sum_J^\infty \sum_{k=0}^{2^j - 1} \frac{1}{2}.2^{-j/2} \sqrt{j+1} \Delta_{2^j + k}(t)$$

(perhaps including some extra terms at the beginning, using $n = 2^j + k < 2^{j+1}$, $\log n \leq (j+1)\log 2$, and $\Delta_n(.) \geq 0$, so the series is absolutely convergent). In the inner sum, only one term is non-zero (t can belong to only one dyadic interval $[k/2^j, (k+1)/2^j)$), and each $\Delta_n(t) \in [0,1]$. So

$$LHS \leq C \sum_{j=J}^{\infty} \frac{1}{2}.2^{-j/2}\sqrt{j+1} \qquad \forall t \in [0,1],$$

and this tends to 0 as $J \to \infty$, so as $M \to \infty$. So the series $\sum \lambda_n Z_n \Delta_n(t)$ is absolutely and uniformly convergent, a.s. Since continuity is preserved under uniform convergence and each $\Delta_n(t)$ (so each partial sum) is continuous, B_t is continuous in t.

2. Covariance. By absolute convergence, we can interchange integral and expectation (Fubini's theorem):

$$E[B_t] = E[\sum_0^\infty \lambda_n Z_n \Delta_n(t)] = \sum \lambda_n \Delta_n(t).E[Z_n] = \sum 0 = 0.$$

So the covariance is

$$E[B_s B_t] = E[\sum_m Z_m \int_0^s \phi_m . \sum_n Z_n \int_0^t \phi_n] = \sum_m \sum_n E[Z_m Z_n] \int_0^s \phi_m \int_0^t \phi_n,$$

or as $E[Z_m Z_n] = \delta_{mn}$,

$$\sum_n \int_0^s \phi_m \int_0^t \phi_n = \min(s,t),$$

by the Parseval calculation above.

3. Joint Distributions. Take $t_1, \ldots, t_m \in [0,1]$, we have to show that $(B(t_1), \ldots, B(t_n))$ is multivariate normal, with mean vector 0 and covariance matrix $(\min(t_i, t_j))$. The multivariate CF is

$$E[\exp\{i \sum_{j=1}^m u_j B(t_j)\}] = E[\exp\{i \sum_{j=1}^m u_j \sum_{n=0}^\infty \lambda_n Z_n \Delta_n(t)\}],$$

which by independence of the Z_n is

$$\prod_{n=0}^\infty E[\exp\{i\lambda_n Z_n \sum_{j=1}^m u_j \Delta_n(t_j)\}].$$

Since each Z_n is $N(0,1)$, the RHS is

$$\prod_{n=0}^{\infty} \exp\{-\frac{1}{2}\lambda_n^2(\sum_{j=1}^{m} u_j \Delta_n(t_j))^2\} = \exp\{-\frac{1}{2}\sum_{n=0}^{\infty}\lambda_n^2(\sum_{j=1}^{m} u_j \Delta_n(t))^2\}.$$

The sum in the exponent on the right is

$$\sum_{n=0}^{\infty}\lambda_n^2 \sum_{j=1}^{m}\sum_{k=1}^{m} u_j u_k \Delta_n(t_j)\Delta_n(t_k) = \sum_{j=1}^{m}\sum_{k=1}^{m} u_j u_k \sum_{n=0}^{\infty}\int_0^{t_j} H_n(u)du. \int_0^{t_k} H_n(u)du,$$

giving

$$\sum_{j=1}^{m}\sum_{k=1}^{m} u_j u_k \min(t_j, t_k),$$

by the Parseval calculation, as (H_n) are cons. Combining,

$$E[\exp\{i\sum_{j=1}^{m} u_j B(t_j)\}] = \exp\{-\frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{m} u_j u_k \min(t_j, t_k)\}.$$

This says that $(B(t_1), \ldots, B(t_n))$ is multinormal with mean 0 and covariance function $\min(t_j, t_k)$ as required. This completes the construction of BM. //

Wavelets.

The Haar system (H_n) and the Schauder system (Δ_n) are examples of *wavelet systems*. The original function, H or Δ, is a *mother wavelet*, and the 'daughter wavelets' are obtained from it by dilation and translation. The PWZ expansion is the *wavelet expansion of BM* with respect to the Schauder system (Δ_n). For any $f \in C[0,1]$, we can form its wavelet expansion

$$f(t) = \sum_{n=0}^{\infty} c_n \Delta_n(t); \qquad c_n = f(\frac{k+\frac{1}{2}}{2^j}) - \frac{1}{2}[f(\frac{k}{2^j}) + f(\frac{k+1}{2^j})].$$

are the *wavelet coefficients*. This is the form that gives the $\Delta_n(.)$ term its correct triangular influence, localized on the dyadic interval $[k/2^j, (k+1)/2^j]$. Thus for f BM, $c_n = \lambda_n Z_n$, with λ_n, Z_n as above. The wavelet construction of BM above is, in modern language, the classical 'broken-line' construction of BM due to Lévy in his book of 1948. The account above is from [Ste].
Note. 1. We shall see that Brownian motion is a *fractal*, and wavelets are a

useful tool for the analysis of fractals more generally.
2. Wavelets are very useful in *data compression*. This is because many signals with lots of 'local discontinuities' may be accurately summarized by a *sparse* wavelet expansion (one with only a few non-zero coefficients). For example, the FBI digitized its finger-print data bank using wavelets.
Zeros. It can be shown that Brownian motion *oscillates*:

$$\limsup_{t\to\infty} X_t = +\infty, \qquad \liminf_{t\to\infty} X_t = -\infty \qquad a.s.$$

Hence, for every n there are zeros (times t with $X_t = 0$) of X with $t \geq n$ (indeed, infinitely many such zeros). So, denoting the zero-set of $BM(\mathbb{R})$ by

$$Z := \{t \geq 0 : X_t = 0\} :$$

1. Z is an *infinite* set. We quote also:
2. Z is a (Lebesgue) *null* set: Z has Lebesgue measure zero.
3. Z is a *closed* set (contains its limit points – from path-continuity).
Less obvious are the next two properties:
4. Z is a *perfect* set: every point $t \in Z$ is a limit point of points in Z.
So there are *infinitely many* zeros in *every* neighbourhood of *every* zero (so the paths must oscillate amazingly fast!). This shows that *it is impossible to draw a realistic picture of a Brownian path.*
Brownian Scaling. For each $c \in (0, \infty)$, $X(c^2 t)$ is $N(0, c^2 t)$, so $X_c(t) := c^{-1} X(c^2 t)$ is $N(0, t)$. Thus X_c has all the defining properties of a Brownian motion (check). So, X_c **IS** a Brownian motion:

Theorem. If X is $BM(\mathbb{R})$ and $c > 0$, $X_c(t) := c^{-1} X(c^2 t)$, then X_c is again a $BM(\mathbb{R})$.
Corollary. X is *self-similar* (reproduces itself under scaling), so a Brownian path $X(.)$ is a *fractal*. So too is the zero-set Z.

Brownian motion owes part of its importance to belonging to *all* the important classes of stochastic processes: it is (strong) Markov, a (continuous) martingale, Gaussian, a diffusion, a Lévy process (process with stationary independent increments), etc.

Brownian motion is the dynamic counterpart of the standard normal distribution $\Phi = N(0, 1)$, and this owes much of its importance to the Central Limit Theorem (CLT) ('Law of Errors'). The dynamic counterpart of the CLT is *Donsker's Invariance Principle* (see e.g. [Bil]) (Week 5).