PfS: HANDOUT (for information only – not examinable)

Supplementary material – relevant to various sections.

1. *III.1. The Gamma function* Γ. Many of the constants in the standard densities in Statistics involve the Gamma function, so we record here what we will need. For complex $z = x + iy$ with $Re\ z = x > 0$, the *Gamma function* $\Gamma(z)$ is defined by

$$\Gamma(z) := \int_0^\infty t^{x-1} e^{-t} dt$$

(note that the integral diverges at 0 for $x = 0$). Integrating by parts,

$$\Gamma(z+1) = z\Gamma(z),$$

the *functional equation for the Gamma function.* Using this, one can extend the domain of definition from $Re\ z > 0$ to $Re\ z > -1$ (the pole at $z = 0$ gives another at $z = -1$). Repeating this indefinitely, we can extend the domain of definition to the whole complex z-plane, and the extended function is regular except at poles $z = 0, -1, \ldots, -n, \ldots$. This process is called Analytic Continuation: for details, see e.g. my website, link to M2P3 Complex Analysis, L22. Also from the functional equation, we see by induction that

$$\Gamma(n+1) = n! \qquad (n = 0, 1, 2, \ldots);$$

thus Gamma provides a *continuous extension to the factorial.*

We have

$$\Gamma(\frac{1}{2}) = \sqrt{\pi};$$

this is equivalent to the standard normal density being a density.

We will need *Stirling's formula* (James STIRLING (1692-1770) in 1730):

$$\Gamma(x) \sim \sqrt{2\pi} e^{-x} x^{x-\frac{1}{2}} \qquad (x \to \infty); \qquad n! \sim \sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}} \qquad (n \to \infty).$$

Beta functions. We will also need the closely related *Beta function*, $B(\alpha, \beta)$ $(\alpha, \beta > 0)$:

$$B(\alpha, \beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx.$$

By *Euler's integral for the Beta function*, this can be expressed in terms of Gammas:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

2. *III.2: Notes on LLN*

The first result of this kind is the WLLN for Bernoulli trials (tossing a coin that falls heads with probability p, tails with probability $q := 1-p$, due to Jakob BERNOULLI (1654-1705); *Ars conjectandi*, 1713, posth.) The general WLLN above, and its strengthening the SLLN below, constitute precise forms of the 'Law of Averages', known to the man in the street. The passage from Bernoulli's theorem of 1713 (perhaps the earliest substantial theorem in Probability Theory) to Kolmogorov's SLLN in 1933, 220 years later, is remarkable both for its length and for the part it played in the 1933 birth of modern measure-theoetic Probability Theory.

The CLT for Bernoulli trials is due to Abraham de MOIVRE (1667-1754), *Doctrine of Chances* 1738 (de Moivre found the normal distribution in 1733), later extended by P. S. de LAPLACE (1749-1827), *Théorie Analytiques des Probabilités*, 1812. The general CLT is due to J. W. LINDEBERG (1876-1932) in 1922 (the name 'central limit theorem' is due to Pólya, also in 1922). The CLT is the precise form of the 'Law of Errors', known to the physicist in the street as saying 'errors are normally distributed about the mean'.

Note. 1. The CLT largely explains why the normal distribution is so ubiquitous in Statistics – basically, this is why Statistics works.

2. The CLT and the normal distribution are static. We shall need their dynamic counterparts. The stochastic process (dynamic counterpart) corresponding to the normal distribution is *Brownian motion* (VI.1); that of the CLT is the Erdös-Kac-Donsker *invariance principle*.

LLN, CLT and LIL complete the trilogy of classical limit theorems in Probability Theory.

The strong law has two main methods of proof:

(i) by Kolmogorov's inequality [see e.g. SA, I.11 L3];

(ii) by Etemadi's method of geometric subsequences [SP, L14]. This has the advantage that it applies with the X_n only *pairwise* independent; Kolmogorov's inequality does not extend to pairwise independence.

The strong law has two important generalisations:

(i) the martingale convergence theorem [see e.g. SA, L5];

(ii) the (Birkhoff-Khinchin) ergodic theorem (G. D. BIRKHOFF (1884-1944) in 1931, Khinchin in 1933).

Note. 1. The term *ergodic* arises in Statistical Mechanics; there one is concerned with whether time averages and phase averages coincide.

2. Regarding how a result of 1931 can be a generalisation of a result of 1933: the ergodic theorem generalises the direct part of the LLN (existence of the

mean implies a.s. convergence), but there is no converse.

3. *III.3: Kolmogorov-Smirnov.*

Variants on the problem above include:

1. *The two-sample Kolmogorov-Smirnov test.*

Given two populations, with unknown distributions F, G, we wish to test whether they are the same, on the basis of empiricals F_n, G_m.

2. *Kolmogorov-Smirnov tests with parameters estimated from the data.*

A common case here is *testing for normality*. In one dimension, our hypothesis of interest is whether or not $F \in \{N(\mu, \sigma^2) : \mu \in \mathbf{R}, \sigma > 0\}$. Here (μ, σ) are *nuisance parameters*: they occur in the formulation of the problem, but not in the hypothesis of interest.

Although the Glivenko-Cantelli Theorem is useful, it does not tell us, say, whether or not the law F is absolutely continuous, discrete etc. For, there are discrete G arbitrarily close to an absolutely continuous F (discretise), and absolutely continuous F arbitrarily close to a discrete F (by smooth approximation to F at its jump points). So sampling alone cannot tell us what *type* of law F is – absolutely continuous (with density f, say), discrete, continuous singular, or some mixture of these. So it makes sense for the statistician to *choose* what kind of population distribution he is going to assume. Often (usually), this will be absolutely continuous; again, it makes sense to *assume* what smoothness properties of the density f we will assume. This leads on to the important subject of *density estimation*; see e.g. SMF Day 3.

4. *III.3: Reparametrisation and the Delta Method.*

Suppose we are using parameter θ, but wish to change to some alternative parametrisation, $g(\theta)$, where g is continuously differentiable. A CLT for θ such as

$$\sqrt{n}(T_n - \theta) \to N(0, \sigma(\theta)^2)$$

(as holds above, with T_n the MLE $\hat{\theta}$ based on a sample of size n and $\sigma^2(\theta) = 1/I(\theta)$) transforms into a CLT for $g(\theta)$:

$$\sqrt{n}(g(T_n) - g(\theta)) \to N(0, [g'(\theta)\sigma(\theta)]^2).$$

For,

$$g(T_n) - g(\theta) = (T_n - \theta)(g'(\theta) + \epsilon_n),$$

with ϵ_n a (random) error term – negligible for large n, so

$$g(T_n) - g(\theta) \sim (T_n - \theta)g'(\theta).$$

Since $var(cX) = c^2 \; var \; X$, the result follows.

This is called the *delta method*, and is often useful. It can be extended from random variables to stochastic processes (i.e., from one or finitely many to infinitely many dimensions), and we shall meet it again later.

5. *III.3: Location and scale: type.*

Example: Temperature. In the UK, before entry to the EU (or Common Market as it was then), temperature was measured in degrees Fahrenheit, F (freezing point of water $32^o F$, boiling point $212^o F$; these odd choices are only of historical interest – but dividing the freezing-boiling range into 180 parts rather than 100 is better attuned to homo sapiens being warm-blooded, and most of us having trouble with decimals and fractions!) The natural choice for freezing is 0; 100 parts for the freezing-boiling range is also natural when using the metric system – whence the Centigrade (= Celsius) scale. Back then, one used F for ordinary life, C for science, and the conversion rules

$$C = \frac{5}{9}(F - 32), \qquad F = \frac{9}{5}C + 32$$

were part of the lives of all schoolchildren (and the mechanism by which many of them grasped the four operations of arithmetic!)

Pivotal quantities.

A *pivotal quantity*, or *pivot*, is one whose distribution is independent of parameters. Pivots are very useful in forming *confidence intervals.*

Defn. A *location family* is one where, for some reference density f, the density has the form

$$f(x - \mu);$$

here μ is a *location parameter*. A *scale family* (usually for $x \geq 0$) is of the form

$$f(x/\sigma);$$

here σ is a *scale parameter*. A *location-scale family* is of the form

$$f(\frac{x - \mu)}{\sigma}).$$

Pivots here are

$$\bar{X} - \mu \quad \text{(location)}; \quad \bar{X}/\sigma \quad \text{(scale)}; \quad \frac{\bar{X} - \mu}{\sigma} \quad \text{(location-scale)}.$$

Examples. The normal family $N(\mu, \sigma^2)$ is a location-scale family.
The *Cauchy location family* is

$$f(x-\mu) = \frac{1}{\pi[1+(x-\mu)^2]}.$$

In higher dimensions, the location parameter is the mean μ (now a *vector*); the scale parameter is now the *covariance matrix*

$$\Sigma = (\sigma_{ij}), \qquad \sigma_{ij} := cov(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)].$$

Because the choice of location and scale are often arbitrary (as with the temperature example above), we often work 'modulo location and scale', and look at the 'type' of a distribution.

6. *V.2: Martingales in discrete time*

The word 'martingale' is taken from an article of harness, to control a horse's head. The word also means a system of gambling which consists in doubling the stake when losing in order to recoup oneself (1815).

Thackeray: 'You have not played as yet? Do not do so; above all avoid a martingale if you do.'

Examples.

1. Mean zero random walk: $S_n = \sum X_i$, with X_i independent with $E(X_i) = 0$ is a mg (submg: positive mean; supermg: negative mean).
2. Stock prices: $S_n = S_0 \zeta_1 \cdots \zeta_n$ with ζ_i independent positive r.vs with finite first moment.
3. Accumulating data about a random variable ([W], pp. 96, 166–167). If $\xi \in L_1(\Omega, \mathcal{F}, \mathcal{P})$, $M_n := E(\xi|\mathcal{F}_n)$ (so M_n represents our best estimate of ξ based on knowledge at time n), then using iterated conditional expectations

$$E[M_n|\mathcal{F}_{n-1}] = E[E(\xi|\mathcal{F}_n)|\mathcal{F}_{n-1}] = E[\xi|\mathcal{F}_{n-1}] = M_{n-1},$$

so (M_n) is a martingale – indeed, a 'nice' mg; see below.

Stopping Times and Optional Stopping

Recall that τ taking values in $\{0, 1, 2, \ldots; +\infty\}$ is a *stopping time* if

$$\{\tau \le n\} = \{\omega : \tau(\omega) \le n\} \in \mathcal{F}_n \qquad \forall\, n \le \infty.$$

From $\{\tau = n\} = \{\tau \le n\} \setminus \{\tau \le n-1\}$ and $\{\tau \le n\} = \bigcup_{k \le n} \{\tau = k\}$, we see the equivalent characterization

$$\{\tau = n\} \in \mathcal{F}_n \qquad \forall\, n \le \infty.$$

Call a stopping time τ *bounded* if there is a constant K such that $P(\tau \leq K) = 1$. (Since $\tau(\omega) \leq K$ for some constant K and all $\omega \in \Omega \setminus N$ with $P(N) = 0$ all identities hold true except on a null set, i.e. a.s.)
Example. Suppose (X_n) is an adapted process and we are interested in the time of first entry of X into a Borel set B (e.g. $B = [c, \infty)$):

$$\tau = \inf\{n \geq 0 : X_n \in B\}.$$

Now $\{\tau \leq n\} = \bigcup_{k \leq n}\{X_k \in B\} \in \mathcal{F}_n$ and $\tau = \infty$ if X never enters B. Thus τ is a stopping time. Intuitively, think of τ as a time at which you decide to quit a gambling game: whether or not you quit at time n depends only on the history up to and including time n – NOT the future. Thus stopping times model gambling and other situations where there is no foreknowledge, or prescience of the future; in particular, in the financial context, where there is no insider trading. Furthermore since a gambler cannot cheat the system the expectation of his hypothetical fortune (playing with unit stake) should equal his initial fortune.

Theorem (Doob's Stopping-Time Principle, STP). Let τ be a bounded stopping time and $X = (X_n)$ a martingale. Then X_τ is integrable, and

$$E(X_\tau) = E(X_0).$$

Proof. Assume $\tau(\omega) \leq K$ for all ω (K integer), and write

$$X_{\tau(\omega)}(\omega) = \sum_{k=0}^{\infty} X_k(\omega) I(\tau(\omega) = k) = \sum_{k=0}^{K} X_k(\omega) I(\tau(\omega) = k).$$

Then

$$\begin{aligned}
E(X_\tau) &= E[\sum_{k=0}^{K} X_k I(\tau = k)] && \text{(by the decomposition above)} \\
&= \sum_{k=0}^{K} E[X_k I(\tau = k)] && \text{(linearity of } E\text{)} \\
&= \textstyle\sum_{k=0}^{K} E[E(X_K|\mathcal{F}_k) I(\tau = k)] && (X \text{ a mg, } \{\tau = k\} \in \mathcal{F}_k\ ) \\
&= \sum_{k=0}^{K} E[X_K I(\tau = k)] && \text{(defn. of conditional expectation)} \\
&= E[X_K \sum_{k=0}^{K} I(\tau = k)] && \text{(linearity of } E\text{)}
\end{aligned}$$

$$
\begin{aligned}
&= E[X_K] \qquad \text{(the indicators sum to 1)} \\
&= E[X_0] \qquad (X \text{ a mg}) \qquad //.
\end{aligned}
$$

If $X = (X_n)$ is a supermg, one obtains similarly

$$EX_\tau \le EX_0.$$

Also, alternative conditions such as
(i) $X = (X_n)$ is bounded ($|X_n|(\omega) \le L$ for some L and all n, ω);
(ii) $E\tau < \infty$ and $(X_n - X_{n-1})$ is bounded;
suffice for the proof of the stopping time principle – which is important in many areas of statistics, such as sequential analysis.

We now wish to create the concept of the σ-algebra of events observable up to a stopping time τ, in analogy to the σ-algebra $\mathcal{F}_n$ which represents the events observable up to time n.
Definition. For τ a stopping time, the *stopping time* $\sigma-$*algebra* $\mathcal{F}_\tau$ is

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \le n\} \in \mathcal{F}_n, \text{ for all } n\}.$$

Proposition. For τ a stopping time, $\mathcal{F}_\tau$ is a $\sigma-$algebra.

Proof. Clearly $\Omega, \emptyset$ are in $\mathcal{F}_\tau$. Also for $A \in \mathcal{F}_\tau$ we find

$$A^c \cap \{\tau \le n\} = \{\tau \le n\} \setminus (A \cap \{\tau \le n\}) \in \mathcal{F}_n,$$

thus $A^c \in \mathcal{F}_\tau$. Finally, for a family $A_i \in \mathcal{F}_\tau,\ i = 1, 2, \ldots$ we have

$$\left(\bigcup_{i=1}^{\infty} A_i\right) \cap \{\tau \le n\} = \bigcup_{i=1}^{\infty} (A_i \cap \{\tau \le n\}) \in \mathcal{F}_n,$$

showing $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_\tau$. //

One can show similarly that for σ, τ stopping times with $\sigma \le \tau$, $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$. Similarly, for any adapted sequence of random variables $X = (X_n)$ and a.s. finite stopping time τ, define

$$X_\tau := \sum_{n=0}^{\infty} X_n I(\tau = n).$$

Then X_τ is $\mathcal{F}_\tau$-measurable.

We now give an important extension of the Stopping-Time Principle.

Theorem (Doob's Optional-Sampling Theorem, OST). Let $X = (X_n)$ be a mg and σ, τ be bounded stopping times with $\sigma \le \tau$. Then

$$E[X_\tau | \mathcal{F}_\sigma] = X_\sigma, \qquad \text{and so} \qquad E(X_\tau) = E(X_\sigma).$$

Proof. First observe that X_τ and X_σ are integrable (use the sum representation and the fact that τ is bounded by an integer K) and X_σ is $\mathcal{F}_\sigma$-measurable by above. So it only remains to prove that

$$E(I_A X_\tau) = E(I_A X_\sigma) \qquad \forall A \in \mathcal{F}_\sigma.$$

For any such fixed $A \in \mathcal{F}_\sigma$, define ρ by

$$\rho(\omega) = \sigma(\omega) I_A(\omega) + \tau(\omega) I_{A^c}(\omega).$$

Since

$$\{\rho \le n\} = (A \cap \{\sigma \le n\}) \cup (A^c \cap \{\tau \le n\}) \in \mathcal{F}_n$$

ρ is a stopping time, and from $\rho \le \tau$ we see that ρ is bounded. So the STP implies $E(X_\rho) = E(X_0) = E(X_\tau)$. But

$$E(X_\rho) = E\left(X_\sigma I_A + X_\tau I_{A^c}\right), \qquad E(X_\tau) = E\left(X_\tau I_A + X_\tau I_{A^c}\right).$$

So subtracting yields the result. //

Write $X^\tau = (X_n^\tau)$ for the sequence $X = (X_n)$ *stopped* at time τ, where we define $X_n^\tau(\omega) := X_{\tau(\omega) \wedge n}(\omega)$. One can show
(i) If τ is a stopping time and X is adapted, then so is X^τ.
(ii) If τ is a stopping time and X is a mg (supermg, submg), then so is X^τ.
Examples and Applications.
1. Simple Random Walk. Recall the simple random walk: $S_n := \sum_{k=1}^n X_k$, where the X_n are independent tosses of a fair coin, taking values ± 1 with equal probability 1/2. Suppose we decide to bet until our net gain is first $+1$, then quit. Let τ be the time we quit; τ is a stopping time. The stopping time τ has been analyzed in detail (see e.g. [GS], 5.3, or Ex. 3.4). From this:
(i) $\tau < \infty$ a.s.: the gambler will certainly achieve a net gain of $+1$ eventually;
(ii) $E\tau = +\infty$: the mean waiting-time for this is infinity. Hence also:
(iii) No bound can be imposed on the gambler's maximum net loss before his net gain first becomes $+1$.

At first sight, this looks like a foolproof way to make money out of nothing: just bet until you get ahead (which happens eventually, by (i)), then quit. But as a gambling strategy, this is hopelessly impractical: because of (ii), you need unlimited time, and because of (iii), you need unlimited capital – both unrealistic.

Notice that the Stopping-time Principle fails here: we start at zero, so $S_0 = 0$, $ES_0 = 0$; but $S_\tau = 1$, so $ES_\tau = 1$. This example shows two things:
1. Conditions are indeed needed here, or the conclusion may fail (none of the conditions in STP or the alternatives given are satisfied in this example).
2. Any practical gambling (or trading) strategy needs to have some integrability or boundedness restrictions to eliminate such theoretically possible but practically ridiculous cases.

7. *VI.1: Construction of Brownian motion; Paley-Wiener-Zygmund theorem*

Covariance. Before addressing existence, we first find the covariance function. For $s \le t$, $W_t = W_s + (W_t - W_s)$, so as $E(W_t) = 0$,

$$cov(W_s, W_t) = E(W_s W_t) = E(W_s^2) + E[W_s(W_t - W_s)].$$

The last term is $E(W_s)E(W_t - W_s)$ by independent increments, and this is zero, so

$$cov(W_s, W_t) = E(W_s^2) = s \qquad (s \le t): \qquad cov(W_s, W_t) = \min(s, t).$$

A Gaussian process is specified by its mean function and its covariance function, so among centered (zero-mean) Gaussian processes, the covariance function $\min(s, t)$ serves as the signature of Brownian motion.

Finite-dimensional Distributions.

For $0 \le t_1 < \ldots < t_n$, 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). So to check the finite-dimensional distributions of BM – stationary independent increments with $W_t \sim N(0, t)$ – it suffices to show that they are multivariate normal with mean zero and covariance $cov(W_s, W_t) = \min(s, t)$ 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, with inner product

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

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

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

Now take (Z_n) independent and identically distributed $N(0,1)$ (we can construct these, indeed from one $X \sim U[0,1]$), and write

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

This is a sum of independent zero-mean random variables. Kolmogorov's theorem on random series says that it converges a.s. if the sum of the variances converges (we quote this). This is $\sum_{n=0}^{\infty} (\int_0^t \phi_n(x)dx)^2$, $= t$ by above. So the series above converges a.s. (wlog, everywhere, excluding a null set).
The Haar System. Define

$$H(t) := 1 \quad \text{on} \quad [0, \frac{1}{2}), \quad -1 \quad \text{on} \quad [\frac{1}{2}, 1), \quad 0 \quad \text{else}.$$

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, write

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

(so H_n has support $[k/2^j, (k+1)/2^j]$). So if m, n $(m \neq n)$ 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 complete orthonormal system in $L^2[0,1]$. One has

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

where $\lambda_0 = 1$ and for $n \geq 1$, $\lambda_n = \frac{1}{2} \times 2^{-j/2}$ $(n = 2^j + k \geq 1)$.
The Schauder System.

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

$$\Delta(t) = 2t \quad \text{on} \quad [0, \frac{1}{2}), \quad 2(1-t) \quad \text{on} \quad [-\frac{1}{2}, 1], \quad 0 \quad \text{else.}$$

Write $\Delta_0(t) := t$, $\Delta_1(t) := \Delta(t)$ ('mother wavelet'), and define the nth *Schauder function* Δ_n ('daughter wavelets') by

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

(Δ_n has support $[k/2^j, (k+1)/2^j]$, so is 'localized' on this dyadic interval, small for n, j large). The $\sum_{n=0}^{2^j}$ below gives the jth stage in the Lévy broken-line construction.

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

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

converges uniformly on $[0,1]$, a.s. The process $W = (W_t : t \in [0,1])$ 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 of the Lemma. For $x > 1$,

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

So for any $a > 1$,

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

As $\sum n^{-a} < \infty$ for $a > 1$, the Borel-Cantelli lemma (see e.g. SP L13) gives

$$P(|Z_n| > \sqrt{2a \log n} \text{ for infinitely many } n) = 0.$$

So

$$C := \sup_{n \geq 2} \frac{|Z_n|}{\sqrt{\log n}} < \infty \qquad a.s.$$

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 (using $n = 2^j + k < 2^{j+1}, \log n \leq (j+1)\log 2$) by

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

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 $\Delta_n(t)$ is continuous, W_t is continuous in t.
2. Covariance. By absolute convergence and Fubini's theorem (see e.g. SP L9),

$$E(W_t) = E\left(\sum_0^\infty \lambda_n Z_n \Delta_n(t)\right) = \sum \lambda_n \Delta_n(t) E(Z_n) = \sum 0 = 0.$$

So the covariance is

$$E(W_s W_t) = E\left[\sum_m Z_m \int_0^s \phi_m \times \sum_n Z_n \int_0^t \phi_n\right] = \sum_{m,n} E[Z_m Z_n] \int_0^s \phi_m \int_0^t \phi_n,$$

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. //

8. *VII.3: Algebraic approach to Markov chains*

We give an algebraic approach to the limit theorem of VII.3. If the e-values are λ_i, write v_i for the right (column) e-vectors,

$$Pv_i = \lambda_i v_i,$$

and u_i for the left (row) e-vectors,

$$u_i P = \lambda_i u_i.$$

Form the matrices U (column of rows u_i), V (row of columns v_i), $\Lambda := diag(\lambda_i)$. Then

$$UP = \Lambda U, \qquad PV = \Lambda V.$$

If the e-values λ_i are *distinct*, the u_i are linearly independent, and similarly the v_i are independent (we quote this from Linear Algebra). So U, V are non-singular, and

$$P = U^{-1} \Lambda U = V \Lambda V^{-1}.$$

For $i \neq j$,

$$u_i P v_j = u_i \lambda_j v_j = \lambda_j u_i v_j,$$

and symmetrically

$$u_i P v_j = \lambda_i u_i v_j.$$

Subtract:

$$(\lambda_i - \lambda_j) u_i v_j = 0 \qquad (i \neq j).$$

As the e-values are assumed *distinct*, this gives

$$u_i v_j = 0 \qquad (i \neq j).$$

So u_i is orthogonal to all v_j for $j \neq i$, so not orthogonal to v_i (the vs span the whole space). So by scaling we can take $u_i v_i = 1$, giving $u_i v_j = \delta_{ij}$, or

$$UV = I : \qquad U = V^{-1}.$$

Write

$$A_i := v_i u_i$$

("column times row = matrix": A_i has rank one, and the general rank-one matrix is of this column-times-row form). Then

$$A_i A_j = v_i u_i v_j u_j = 0 \qquad (i \neq j), \qquad A_i A_i = v_i u_i v_i u_i = v_i u_i = A_i :$$

$$A_i A_j = \delta_{ij} A_i$$

(so each A_i is idempotent). We can now re-write

$$P = U^{-1} \Lambda U = V \Lambda U$$

as

$$P = \sum_i v_i \lambda_i u_i = \sum_i \lambda_i v_i u_i = \sum_i \lambda_i A_i.$$

Then by induction on n,

$$P^n = \sum_i \lambda_i^n A_i.$$

One can now see what will happen as $n \to \infty$. All e-values with modulus < 1 will die out, leaving just those of modulus 1 – the d dth roots of unity in the case of period $d > 1$, and only the PF e-value $\lambda_1 = 1$ in the aperiodic case. So for P aperiodic,

$$P^n \to A_1 = v_1 u_1 \qquad (n \to \infty).$$

As v_1 is a columns of 1s, the matrix $v_1 u_1$ is a column of identical rows u_1, the PF left-e-vector. This re-proves the Theorem – an algebraic proof, under the algebraic assumption that all e-values are distinct.

9. *VII.1: Mathematical genetics (Fisher-Wright model, and more generally)*
From the Wikipedia entry for Wright:
"*Scientific achievements and credits.*
His papers on inbreeding, mating systems, and genetic drift make him a principal founder of theoretical population genetics, along with R. A. Fisher

and J. B. S. Haldane[1]. Their theoretical work is the origin of the modern evolutionary synthesis or neodarwinian synthesis. Wright was the inventor of the inbreeding coefficient, a standard tool in population genetics. He was the chief developer of the mathematical theory of genetic drift, which is sometimes known as the Sewall Wright effect, cumulative stochastic changes in gene frequencies that arise from random births, deaths, and Mendelian segregations in reproduction. Wright was convinced that the interaction of genetic drift and the other evolutionary forces was important in the process of adaptation. He described the relationship between genotype or phenotype and fitness as fitness surfaces or fitness landscapes. On these landscapes fitness was the height, plotted against horizontal axes representing the allele frequencies or the average phenotypes of the population. Natural selection would lead to a population climbing the nearest peak, while genetic drift would cause random wandering."

In this model, the $2N$ genes in each generation are obtained by sampling with replacement from the genes in the previous one, leading to

$$p_{ij} = \binom{2N}{j} (i/2N)^j (1 - (i/2N))^{2N-j} \qquad (i, j = 0, \ldots, 2N).$$

Note. Mathematical genetics makes extensive use of Markov chain methods and models. For background, see e.g. Ewens [E].
Note: Mathematics and Biology.

The two foundations of modern biology are the Darwinian theory of natural selection (Charles DARWIN (1809-1882): *On the Origin of Species by means of Natural Selection*, 1859 – *The Origin of Species*), and Mendelian genetics (Gregor MENDEL (1822-1884): *Experiments on plant hybridization*, 1866). Mendel's work was largely forgotten, but rediscovered in 1900. It was thought at first that Mendelian genetics and Darwinian natural selection were incompatible, but this is not so; the two were synthesized by Wright, Fisher and Haldane. From the Wikipedia entry for Mendel:
"*Mendel, Darwin and Galton.*
Mendel lived around the same time as the British naturalist Charles Darwin (1809-1882), and many have fantasized about a historical evolutionary synthesis of Darwinian natural selection and Mendelian genetics during their lifetimes. Mendel had read a German translation of Darwin's Origin (as evidenced by underlined passages in the copy in his monastery), after com-

[1] J. B. S. HALDANE (1892-1964)

pleting his experiments but before publishing his paper. Some passages in Mendel's paper are Darwinian in character, evidence that the Origin of the Species influenced Mendel's writing. Darwin did not have a copy of Mendel's paper, but he did have a book by Focke with references to it. The leading expert in heredity at this time was Darwin's cousin Francis Galton[2] who had mathematical skills that Darwin lacked and may have been able to understand the paper had he seen it. In any event, the modern evolutionary synthesis did not start until the 1920s, by which time statistics had become advanced enough to cope with genetics and evolution."
One might add probability and stochastic processes to this last sentence!

The point to note here is that Mathematics and Biology have much to contribute to each other. Within the last century, Biology has developed from being a largely descriptive science to being quantitative. Mathematics – in particular, Statistics and Stochastic Processes – has played a vital role here, and continues to do so, e.g. in the topical and important area of the Human Genome Project. In the other direction, problems from the biological sciences have been important motivation in the development of many mathematical and statistical tools. The point is worth making, for two reasons:
(i) not all students realize that Mathematics can be applied to Biology and the Life Sciences;
(ii) one still encounters a tendency by some in the 'Exact Sciences' (as Physics and Chemistry, and indeed Mathematics, are sometimes called) to look down on Biology and the Life Sciences as 'soft science' rather than 'hard science'. This attitude is now a good half-century out of date.

Another area where mathematical modelling (stochastic and deterministic) in the life sciences is of prime importance is the modelling of infectious diseases such as HIV/AIDS, of Foot and Mouth disease (recall the outbreak of 2001) and of rabies (endemic on the continent, but not in the UK).

[2] Sir Francis GALTON (1822-1911): *Hereditary genius* and regression, in 1869