Probabilistic models of population genetics

Étienne Pardoux

February 5, 2009

Contents

Chapter 1

Branching processes

1.1 Bienaymé–Galton–Watson processes

Consider a Bienaymé–Galton–Watson process, i. e. a process $\{Z_n,\ n \geq 0\}$ with values in $\mathbb{N}$ such that

$$Z_{n+1} = \sum_{k=1}^{Z_n} \xi_{n,k},$$

where $\{\xi_{n,k},\ n \geq 0, k \geq 1\}$ are i. i. d. r. v.'s with as joint law that of ξ whose generating function f satisfies

$$\mu := \mathbb{E}[\xi] = f'(1) = 1 + r, \text{ and } 0 < q := f(0) = \mathbb{P}(\xi = 0) < 1.$$

We call f the probability generating function (p. g. f. in short) of the Bienaymé–Galton–Watson process $\{Z_n,\ n \geq 0\}$. In order to exclude trivial situations, we assume that $\mathbb{P}(\xi = 0) = f(0) > 0$, and that $\mathbb{P}(\xi > 1) > 0$. This last condition implies that $s \to f(s)$, which is increasing on $[0, 1]$, is a strictly convex function.

The process is said to be *subcritical* if $\mu < 1$ ($r < 0$), *critical* if $\mu = 1$ ($r = 0$), and *supercritical* if $\mu > 1$ ($r > 0$). We shall essentially be interested in the supercritical case.

First note that the process $\{Z_n,\ n \geq 0\}$ is a Markov process, which has the so–called branching property, which we now formulate. For $x \in \mathbb{N}$, let $\mathbb{P}_x$ denote the law of the Markov process $\{Z_n,\ n \geq 0\}$ starting from $Z_0 = x$. The law of $\{Z_n,\ n \geq 0\}$ under $\mathbb{P}_{x+y}$ is the same as that of the sum of two independent copies of $\{Z_n,\ n \geq 0\}$, one having the law $\mathbb{P}_x$, the other the law $\mathbb{P}_y$.

We next define

$$T = \inf\{k > 0; Z_k = 0\},$$

which is the time of extinction. We first recall the

Proposition 1.1.1. *Assume that $Z_0 = 1$. Then the probability of extinction $\mathbb{P}(T < \infty)$ is one in the subcritical and the critical cases, and it is the unique root $\eta < 1$ of the equation $f(s) = s$ in the supercritical case.*

Proof: Let $f^{\circ n}(s) := f \circ \cdots \circ f(s)$, where f has been composed n times with itself. It is easy to check that $f^{\circ n}$ is the generating function of the r. v. Z_n.

On the other hand, clearly $\{T \leq n\} = \{Z_n = 0\}$. Consequently

$$\begin{aligned}\mathbb{P}(T < \infty) &= \lim_n \mathbb{P}(T \leq n) \\ &= \lim_n \mathbb{P}(Z_n = 0) \\ &= \lim_n f^{\circ n}(0).\end{aligned}$$

Now the function $s \to f(s)$ is continuous, increasing and strictly convex, starts from $q > 0$ at $s = 0$, and ends at 1 at $s = 1$. If $\mu = f'(1) \leq 1$, then $\lim_n f^{\circ n}(0) = 1$. If however $f'(1) = 1 + r > 1$, then there exists a unique $0 < \eta < 1$ such that $f(\eta) = \eta$, and it is easily seen that $\eta = \lim_n f^{\circ n}(0)$. □

Note that the state 0 is absorbing for the Markov chain $\{Z_n,\ n \geq 0\}$, and it is accessible from each state. It is then easy to deduce that all other states are transient, hence either $Z_n \to 0$, or $Z_n \to \infty$, as $n \to \infty$. In other words, the population tends to infinity a. s. on the set $\{T = \infty\}$.

Denote $\sigma^2 = \mathrm{Var}(\xi)$, which is assumed to be finite. We have the

Lemma 1.1.2.

$$\begin{aligned}\mathbb{E}Z_n &= \mu^n \mathbb{E}Z_0 \\ \mathbb{E}[Z_n^2] &= \frac{\mu^{2n} - \mu^n}{\mu^2 - \mu}\sigma^2 \mathbb{E}Z_0 + \mu^{2n}\mathbb{E}(Z_0^2).\end{aligned}$$

PROOF: We have

$$
\begin{aligned}
\mathbb{E}Z_n &= \mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^{Z_{n-1}} \xi_{n-1,k}|Z_{n-1}\right]\right] \\
&= \mu\mathbb{E}Z_{n-1} \\
&= \mu^n\mathbb{E}Z_0,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}[Z_n^2] &= \mathbb{E}\left[\mathbb{E}\left[\left(\sum_{k=1}^{Z_{n-1}} \xi_{n-1,k}\right)^2 |Z_{n-1}\right]\right] \\
&= \mu^2\mathbb{E}[Z_{n-1}(Z_{n-1}-1)] + (\sigma^2+\mu^2)\mathbb{E}Z_{n-1} \\
&= \mu^2\mathbb{E}[Z_{n-1}^2] + \sigma^2\mathbb{E}Z_{n-1} \\
&= \mu^2\mathbb{E}[Z_{n-1}^2] + \sigma^2\mu^{n-1}\mathbb{E}Z_0.
\end{aligned}
$$

Consequently $a_n := \mu^{-2n}\mathbb{E}[Z_n^2]$ satisfies

$$
\begin{aligned}
a_n &= a_{n-1} + \sigma^2\mu^{-(n+1)}\mathbb{E}Z_0 \\
&= a_0 + \sigma^2\mathbb{E}Z_0\sum_{k=1}^{n}\mu^{-(k+1)}.
\end{aligned}
$$

□

Let now Z_n^* denote the number of individuals in generation n with an infinite line of descent. Under $\mathbb{P}_1$, $\{T=\infty\} = \{Z_0^* = 1\}$. ξ denoting a r. v. whose law is that of the number of offsprings of each individual, let $\xi^* \le \xi$ denote the number of those offsprings with an infinite line of descent. Let $\overline{q} := 1-q$. We have the

Proposition 1.1.3. *Assume that $Z_0 = 1$.*

1. *Conditional upon $\{T=\infty\}$, $\{Z_n^*,\ n\ge 0\}$ is again a Bienaymé–Galton–Watson process, whose p. g. f. is given by*

$$f^*(s) = [f(q+\overline{q}s)-q]/\overline{q}.$$

2. *Conditional upon $\{T<\infty\}$, the law of $\{Z_n,\ n\ge 0\}$ is that of a Bienaymé–Galton–Watson process, whose p. g. f. is given by*

$$\tilde{f}(s) = f(qs)/q.$$

3. For all $0 \le s, t \le 1$,

$$\mathbb{E}\left[s^{\xi-\xi^*}t^{\xi^*}\right] = f(qs+\overline{q}t)$$
$$\mathbb{E}\left[s^{Z_n-Z^*_n}t^{Z^*_n}\right] = f^{\circ n}(qs+\overline{q}t).$$

4. Conditional upon $\{T=\infty\}$, *the law of* $\{Z_n,\ n\ge 0\}$ *is that of* $\{Z^*_n,\ n\ge 0\}$ *to which we add individuals with finite line of descent, by attaching to each individual of the tree of the* Z^*_n*'s* N *independent copies of a Bienaymé–Galton–Watson tree with p. g. f.* $\tilde{f}$, *where*

$$\mathbb{E}[s^N|Z^*] = \frac{D^n f(qs)}{D^n f(q)},$$

where $D^n f$ *denotes the* n*–th derivative of* f, *and* n *is the number of sons of the considered individual in the tree* Z^*.

Proof: Let us first prove the first part of *3.* Consider on the same probability space mutually independent r. v.'s $\{\xi, Y_i,\ i\ge 1\}$, where the law of ξ is given as above, and $\mathbb{P}(Y_i=1)=\overline{q}=1-\mathbb{P}(Y_i=0)$, $\forall i\ge 1$. Note that $\overline{q}$ is the probability that any given individual has an infinite line of descent, so that the joint law of $(\xi-\xi^*, \xi^*)$ is that of

$$\left(\sum_{i=1}^{\xi}(1-Y_i), \sum_{i=1}^{\xi} Y_i\right).$$

$$\begin{aligned}
\mathbb{E}\left[s^{\xi-\xi^*}t^{\xi^*}\right] &= \mathbb{E}\left[\mathbb{E}\left[s^{\xi-\xi^*}t^{\xi^*}|\xi\right]\right]\\
&= \mathbb{E}\left[s^{\sum_{i=1}^{\xi}(1-Y_i)}t^{\sum_{i=1}^{\xi}Y_i}\right]\\
&= \mathbb{E}\left[\mathbb{E}[s^{1-Y_1}t^{Y_1}]^{\xi}\right]\\
&= \mathbb{E}[(qs+\overline{q}t)^{\xi}]\\
&= f(qs+\overline{q}t).
\end{aligned}$$

A similar computation yields the second statement in *3.* Indeed

$$\begin{aligned}
\mathbb{E}\left[s^{Z_n-Z^*_n}t^{Z^*_n}\right] &= \mathbb{E}\left[\mathbb{E}\left(s^{Z_n-Z^*_n}t^{Z^*_n}|Z_{n-1}\right)\right]\\
&= \mathbb{E}\left[\left(\mathbb{E}\left[s^{\xi-\xi^*}t^{\xi^*}\right]\right)^{Z_{n-1}}\right]\\
&= f^{\circ(n-1)}(f(qs+\overline{q}t))
\end{aligned}$$

We next prove *1* as follows

$$\begin{aligned}
\mathbb{E}\left(t^{\xi^*}|\xi^*>0\right) &= \frac{\mathbb{E}(1^{\xi-\xi^*}t^{\xi^*};\xi^*>0)}{\mathbb{P}(\xi^*>0)} \\
&= \frac{\mathbb{E}(1^{\xi-\xi^*}t^{\xi^*})-\mathbb{E}(1^{\xi-\xi^*}t^{\xi^*};\xi^*=0)}{\mathbb{P}(\xi^*>0)} \\
&= \frac{f(q+\overline{q}t)-f(q)}{\overline{q}} \\
&= \frac{f(q+\overline{q}t)-q}{\overline{q}}.
\end{aligned}$$

We now prove *2*. It suffices to compute

$$\begin{aligned}
\mathbb{E}\left(s^{\xi}|\xi^*=0\right) &= \mathbb{E}\left(s^{\xi-\xi^*}|\xi^*=0\right) \\
&= \frac{f(sq+0\overline{q})}{q}.
\end{aligned}$$

Finally we prove *4*. All we have to show is that

$$\mathbb{E}[s^{\xi-\xi^*}|\xi^*=n] = \frac{D^n f(qs)}{D^n f(q)}.$$

This follows from the two following identities

$$\begin{aligned}
n!\mathbb{E}[s^{\xi-\xi^*};\xi^*=n] &= \overline{q}^n D^n f(qs+\overline{q}t)|_{t=0} \\
&= \overline{q}^n D^n f(qs), \\
n!\mathbb{P}(\xi^*=n) &= \overline{q}^n D^n f(qs+\overline{q}t)|_{s=1,t=0} \\
&= \overline{q}^n D^n f(q).
\end{aligned}$$

□

1.2 A continuous time Bienaymé–Galton–Watson process

Consider a continuous time $\mathbb{N}$–valued branching process $Z=\{Z_t^k,\ t\geq 0, k\in\mathbb{N}\}$, where t denotes time, and k is the number of ancestors at time 0. Such a process is a Bienaymé–Galton–Watson process in which to each individual is attached a random vector describing its lifetime and its numbers of offsprings. We assume that those random vectors are i. i. d.. The rate of reproduction is governed by a finite measure μ on $\mathbb{N}$, satisfying $\mu(1) = 0$. More precisely, each individual lives for an exponential time with parameter $\mu(\mathbb{N})$, and is replaced by a random number of children according to the probability $\mu(\mathbb{N})^{-1}\mu$. Hence the dynamics of the continuous time jump Markov process Z is entirely characterized by the measure μ. We have the

Proposition 1.2.1. *The generating function of the process Z is given by*

$$\mathbb{E}\left(s^{Z_t^k}\right) = \psi_t(s)^k, \ s \in [0,1], \ k \in \mathbb{N},$$

where

$$\frac{\partial \psi_t(s)}{\partial t} = \Phi(\psi_t(s)), \quad \psi_0(s) = s,$$

and the funcion Φ is defined by

$$\Phi(s) = \sum_{n=0}^{\infty} (s^n - s)\mu(n), \ s \in [0,1].$$

PROOF: Note that the process Z is a continuous time $\mathbb{N}$–valued jump Markov process, whose infinitesimal generator is given by

$$Q_{n,m} = \begin{cases} 0, & \text{if } m < n-1, \\ n\mu(m+1-n), & \text{if } m \geq n-1 \text{ and } m \neq n, \\ -n\mu(\mathbb{N}), & \text{if } m = n. \end{cases}$$

Define $f : \mathbb{N} \to [0,1]$ by $f(k) = s^k$, $s \in [0,1]$. Then $\psi_t(s) = P_t f(1)$. It follows from the backward Kolmogorov equation for the process Z (see e. g. Theorem 7.6 in [21]) that

$$\begin{aligned} \frac{dP_t f(1)}{dt} &= (QP_t f)(1) \\ \frac{\partial \psi_t(s)}{\partial t} &= \sum_{k=0}^{\infty} Q_{1,k}\psi_t(s)^k \\ &= \sum_{k=0}^{\infty} \mu(k)\psi_t(s)^k - \psi_t(s)\sum_{k=0}^{\infty}\mu(k) \\ &= \Phi(\psi_t(s)). \end{aligned}$$

□

The branching process Z is called immortal if $\mu(0) = 0$.

1.3 Convergence to a continuous branching process

To each integer N, we associate a Bienaymé–Galton-Watson process $\{Z_n^N,\ n \geq 0\}$ starting from $Z_0^N = N$. We now define the continuous time process

$$X_t^N := N^{-1} Z_{[Nt]}^N.$$

We shall let the p. g. f. of the Bienaymé–Galton-Watson process depend upon N in such a way that

$$\begin{aligned} \mathbb{E}[\xi_N] &= f_N'(1) = 1 + \frac{\alpha}{N} \\ \mathrm{Var}[\xi_N] &= \beta, \end{aligned}$$

where $\alpha \in \mathbb{R}$, $\beta > 0$, and we assume that the sequence of r. v.'s $\{\xi_N^2,\ N \geq 1\}$ is uniformly integrable. Let $t \in \mathbb{N}/N$ and $\Delta t = N^{-1}$. It is not hard to check that

$$\begin{aligned} \mathbb{E}[X_{t+\Delta t}^N - X_t^N | X_t^N] &= \alpha X_t^N \Delta t, \\ \mathbb{E}[(X_{t+\Delta t}^N - X_t^N)^2 | X_t^N] &= \beta X_t^N \Delta t + \alpha^2 (X_t^N)^2 (\Delta t)^2. \end{aligned}$$

As $N \to \infty$, $X^N \Rightarrow X$, where $\{X_t,\ t \geq 0\}$ solves the SDE

$$dX_t = \alpha X_t dt + \sqrt{\beta X_t} dB_t,\ t \geq 0. \tag{1.3.1}$$

The detailed proof of the convergence will be treated in the next chapter, for a slightly different model.

1.4 The continuous branching process

Denote by $\{X_t(x),\ x > 0, t > 0\}$ the solution of the SDE (1.3.1), starting from x at time $t = 0$, i. e. such that $X_0(x) = x$. For $x > 0$ and $y > 0$ consider $\{X_t(x),\ t > 0\}$ and $\{X_t'(y),\ t > 0\}$, where $\{X_t'(y),\ t > 0\}$ is a copy of $\{X_t(y),\ t > 0\}$ which is independent of $\{X_t(x),\ t > 0\}$. Let $Y_t^{x,y} = X_t(x) + X_t'(y)$. We have

$$\begin{aligned} dY_t^{x,y} &= \alpha(X_t(x) + X_t'(y))dt + \sqrt{\beta X_t(x)} dB_t + \sqrt{\beta X_t'(y)} dB_t' \\ &= \alpha Y_t^{x,y} dt + \sqrt{\beta Y_t^{x,y}} dW_t, \\ Y_0^{x,y} &= x + y \end{aligned}$$

where $\{B_t,\ t \geq 0\}$ and $\{B'_t,\ t \geq 0\}$ are two mutually independent standard Brownian motions, and $\{W_t,\ t \geq 0\}$ is also a standard Brownian motion. Then clearly $\{Y_t^{x,y},\ t \geq 0\}$ and $\{X_t(x+y),\ t \geq 0\}$ have the same law. This shows that $\{X_t(x),\ x > 0, t > 0\}$ possesses the branching property.

This property entails that for all $t,\ \lambda > 0$, there exists $u(t, \lambda)$ such that

$$\mathbb{E}\left[\exp(-\lambda X_t(x))\right] = \exp[-xu(t, \lambda)]. \tag{1.4.1}$$

From the Markov property of the process $t \to X_t(x)$, we deduce readily the semigroup identity

$$u(t+s, \lambda) = u(t, u(s, \lambda)).$$

We seek a formula for $u(t, \lambda)$. Let us first get by a formal argument an ODE satisfied by $u(\cdot, \lambda)$. For $t > 0$ small, we have that

$$X_t(x) \simeq x + \alpha x t + \sqrt{\beta x} B_t,$$

hence

$$\mathbb{E}\left(e^{-\lambda X_t(x)}\right) \simeq \exp\left(-\lambda x[1 + \alpha t - \beta \lambda t/2]\right),$$

and

$$\frac{u(t, \lambda) - \lambda}{t} \simeq \alpha \lambda - \frac{\beta}{2}\lambda^2.$$

Assuming that $t \to u(t, \lambda)$ is differentiable, we deduce that

$$\frac{\partial u}{\partial t}(0, \lambda) = \alpha \lambda - \frac{\beta}{2}\lambda^2.$$

This, combined with the semigroup identity, entails that

$$\frac{\partial u}{\partial t}(t, \lambda) = \alpha u(t, \lambda) - \frac{\beta}{2}u^2(t, \lambda), \quad u(0, \lambda) = \lambda. \tag{1.4.2}$$

It is easy to solve that ODE explicitly, and we now prove rigorously that u is indeed the solution of (1.4.2), without having to go through the trouble of justifying the above argument. Let $\gamma = 2\alpha/\beta$, $\gamma_t = \gamma(1 - e^{-\alpha t})^{-1}$.

Lemma 1.4.1. *The function* $(t, \lambda) \to u(t, \lambda)$ *which appears in* (1.4.1) *is given by the formula*

$$u(t, \lambda) = \frac{\gamma e^{\alpha t}}{e^{\alpha t} - 1 + \gamma/\lambda} = \frac{\lambda \gamma_t}{\lambda + \gamma_t e^{-\alpha t}}, \tag{1.4.3}$$

and it is the unique solution of (1.4.2).

PROOF: It suffices to show that $\{M_s^x,\ 0 \le s \le t\}$ defined by

$$M_s^x = \exp\left(-\frac{\gamma e^{\alpha(t-s)}}{e^{\alpha(t-s)} - 1 + \gamma/\lambda} X_s(x)\right)$$

is a martingale, which follows from Itô's calculus. □

Remark 1.4.2. *In the critical case (i. e. the case $\alpha = 0$),*

$$u(t, \lambda) = \frac{\lambda}{1 + \lambda t \beta/2},$$

which is the limit as $\alpha \to 0$ of (1.4.3)*. This particular formula can also be established by checking that in the case $\alpha = 0$,*

$$N_s^x = \exp\left(-\frac{\lambda}{1 + \lambda\beta(t-s)/2} X_s(x)\right)$$

is a martingale.

The function $\Psi : \mathbb{R}_+ \to \mathbb{R}$ given by

$$\Psi(r) = \frac{\beta}{2} r^2 - \alpha r$$

is called the *branching mechanism* of the continuous branching process X.

For each fixed $t > 0$, $x \to X_t(x)$ has independent and homogeneous increments with values in $\mathbb{R}_+$. We shall consider its right–continuous modification, which then is a subordinator. Its Laplace exponent is the function $\lambda \to u(t, \lambda)$, which can be rewritten (like for any subordinator, see section 7.4 below) as

$$u(t, \lambda) = d(t)\lambda + \int_0^\infty (1 - e^{-\lambda r})\Lambda(t, dr),$$

where $d(t) \ge 0$ and $\int_0^\infty (r \wedge 1)\Lambda(t, dr) < \infty$. Comparing with (1.4.3), we deduce that $d(t) = 0$, and

$$\Lambda(t, dr) = p(t)\exp(-q(t)r)dr,$$
$$\text{where } p(t) = \gamma_t^2 e^{-\alpha t},\ q(t) = \gamma_t e^{-\alpha t}. \tag{1.4.4}$$

We have defined the two parameter process $\{X_t(x);\ x \ge 0, t \ge 0\}$. $X_t(x)$ is the population at time t made of descendants of the initial population of size x at time 0. We may want to introduce three parameters, if we want to discuss the descendants at time t of a population of a given size at time s. The first point, which is technical but in fact rather standard, is that we can construct the collection of those random variables jointly for all $0 \le s < t$, $x \ge 0$, so that all the properties we may reasonably wish for them are satisfied. More precisely, following [3], we have the

Lemma 1.4.3. *On some probability space, there exists a three parameter process*

$$\{X_{s,t}(x),\ 0 \le s \le t,\ x \ge 0\},$$

such that

1. *For every $0 \le s \le t$, $X_{s,t} = \{X_{s,t}(x),\ x \ge 0\}$ is a subordinator with Laplace exponent $u(t-s,\cdot)$.*
2. *For every $n \ge 2$, $0 \le t_1 < t_2 < \cdots < t_n$, the subordinators $X_{t_1,t_2}, \ldots, X_{t_{n-1},t_n}$ are mutually independent, and*
$$X_{t_1,t_n}(x) = X_{t_{n-1},t_n} \circ \cdots \circ X_{t_1,t_2}(x),\ \forall x \ge 0, \quad a.\ s.$$
3. *The processes $\{X_{0,t}(x),\ t \ge 0, x \ge 0\}$ and $\{X_t(x),\ t \ge 0, x \ge 0\}$ have the same finite dimensional distributions.*

Now consider $\{X_{s,t}(x),\ x \ge 0\}$ for fixed $0 \le s \le t$. It is a subordinator with Laplace exponent (the functions p and q are given in (1.4.4))

$$u(t-s,\lambda) = p(t-s)\int_0^\infty (1-e^{-\lambda r})e^{-q(t-s)r}dr.$$

We shall give a probabilistic description of the process $\{X_{s,t}(x),\ x \ge 0\}$ in a further section. For now on, we shall write $X_t(x)$ for $X_{0,t}(x)$.

Let us first study the large time behaviour of the process $X_t(x)$. Consider the extinction event

$$E = \{\exists t > 0,\ \text{s. t. } X_t(x) = 0\}.$$

We define again $\gamma = 2\alpha/\beta$.

Proposition 1.4.4. *If $\alpha \le 0$, $\mathbb{P}_x(E) = 1$ a.s. for all $x > 0$. If $\alpha > 0$, $\mathbb{P}_x(E) = \exp(-x\gamma)$ and on E^c, $X_t(x) \to +\infty$ a. s.*

PROOF: If $\alpha \leq 0$, $\{X_t(x),\ t \geq 0\}$ is a positive supermartingale. Hence it converges a. s. The limit r. v. $X_\infty(x)$ takes values in the set of fixed points of the SDE (1.3.1), which is $\{0, +\infty\}$. But from Fatou and the supermartingale property,

$$\mathbb{E}(\lim_{t\to\infty} X_t(x)) \leq \lim_{t\to\infty} \mathbb{E}(X_t(x)) \leq x.$$

Hence $\mathbb{P}(X_\infty(x) = +\infty) = 0$, and $X_t(x) \to 0$ a. s. as $t \to \infty$.

If now $\alpha > 0$, it follows from It's formula that

$$e^{-\gamma X_t(x)} = e^{-\gamma x} - \gamma \int_0^t e^{-\gamma X_s(x)} \sqrt{\beta X_s(x)} dB_s,$$

hence $\{M_t = e^{-\gamma X_t(x)},\ t \geq 0\}$ is a martingale with values in $[0, 1]$, hence it converges a. s. as $t \to \infty$. Consequently $X_t(x) = -\gamma \log(M_t)$ converges a. s., and as above its limit belongs to the set $\{0, +\infty\}$. Moreover

$$\begin{aligned}
\mathbb{P}(E) &= \lim_{t\to\infty} \mathbb{P}(X_t(x) = 0) \\
&= \lim_{t\to\infty} \mathbb{E}[\exp\{-xu(t, \infty)\}] \\
&= \lim_{t\to\infty} \exp\left\{-x \frac{\gamma e^{\alpha t}}{e^{\alpha t} - 1}\right\} \\
&= \exp\{-x\gamma\}.
\end{aligned}$$

It remains to prove that

$$\mathbb{P}(E^c \cap \{X_t \to 0\}) = 0. \tag{1.4.5}$$

Define the stopping times

$$\begin{aligned}
\tau_1 &= \inf\{t > 0,\ X_t(x) \leq 1\},\ \text{and for } n \geq 2, \\
\tau_n &= \inf\{t > \tau_{n-1} + 1,\ X_t(x) \leq 1\}.
\end{aligned}$$

On the set $\{X_t(x) \to 0,\ \text{as } t \to \infty\}$, $\tau_n < \infty$, $\forall n$. Define for $n \geq 1$

$$A_n = \{\tau_{n+1} < \infty, X_{\tau_{n+1}}(x) > 0\}.$$

For all $N > 0$,

$$\begin{aligned}
\mathbb{P}(E^c \cap \{X_t \to 0\}) &\leq \mathbb{P}(\cap_{n=1}^N A_n) \\
&\leq \mathbb{E}\left(\prod_{n=1}^N \mathbb{P}(A_n | \mathcal{F}_{\tau_n})\right) \\
&\leq (\mathbb{P}(X_1(1) > 0))^N \\
&\to 0,\ \text{as } N \to \infty,
\end{aligned}$$

where we have used the strong Markov property, and the fact that

$$\mathbb{P}(A_n|X_{\tau_n}) \leq \mathbb{P}(X_1(1) > 0).$$

□

1.5 Back to Bienaymé–Galton–Watson

1.5.1 The individuals with an infinite line of descent

Let us go back to the discrete model, indexed by N. For each $t \geq 0$, let Y_t^N denote the individuals in the population $Z_{[Nt]}^N$ with an infinite line of descent. Let us describe the law of Y_0^N. Each of the N individuals living at time $t = 0$ has the probability $1 - q_N$ of having an infinte line of descent. It then follows from the branching property that the law of Y_0^N is the binomial law $B(N, 1 - q_N)$. It remains to evaluate q_N, the unique solution in the interval $(0, 1)$ of the equation $f_N(x) = x$. Note that

$$f_N''(1) = \mathbb{E}[\xi_N(\xi_N - 1)] = \beta - \frac{\alpha}{N} + \left(\frac{\alpha}{N}\right)^2.$$

We deduce from a Taylor expansion of f near $x = 1$ that

$$q_N = 1 - \frac{2\alpha}{N\beta} + \circ\left(\frac{1}{N}\right),\ 1 - q_N = \frac{2\alpha}{N\beta} + \circ\left(\frac{1}{N}\right).$$

Consequently, Y_0^N converges in law, as $N \to \infty$, towards a Poisson distribution with parameter $\gamma = 2\alpha/\beta$.

1.5.2 The individuals whose progeny survives during tN generations

The result of the last section indicates that if we consider only the *prolific individuals*, i. e. those with an infinite line of descent, in the limit $N \to \infty$, we should not divide by N, also $Z_{[Nt]}^N \to +\infty$, as $N \to \infty$, for all $t \geq 0$. If now we consider those individuals whose progeny is still alive at time tN (i. e. those whose progeny contributes to the population at time $t > 0$ in the limit as $N \to \infty$), then again we should not divide by N. Indeed, we have the (we use again the notation $\gamma = 2\alpha/\beta$)

Theorem 1.5.1. *Under the assumptions from the beginning of section 1.3, with the notation*

$$\gamma_t = \gamma \left(1 - e^{-\alpha t}\right)^{-1},$$

1. *for N large,*

$$\mathbb{P}_1(Z^N_{[Nt]} > 0) = \frac{\gamma_t}{N} + \circ\left(\frac{1}{N}\right),$$

   and

2. *as $N \to \infty$,*

$$\mathbb{E}_1\left(\exp[-\lambda Z^N_{[Nt]}/N] | Z^N_{[Nt]} > 0\right) \to \frac{\gamma_t e^{-\alpha t}}{\lambda + \gamma_t e^{-\alpha t}}.$$

Proof of 1 : It follows from the branching property that

$$\begin{aligned}\mathbb{P}_1(Z^N_{[Nt]} > 0) &= 1 - \mathbb{P}_1(Z^N_{[Nt]} = 0)\\ &= 1 - \mathbb{P}_N(Z^N_{[Nt]} = 0)^{1/N}\\ &= 1 - \mathbb{P}_1(X^N_t = 0)^{1/N}.\end{aligned}$$

But

$$\begin{aligned}\log\left[\mathbb{P}_1(X^N_t = 0)^{1/N}\right] &= \frac{1}{N}\log \mathbb{P}_1(X^N_t = 0)\\ &= \frac{1}{N}\log \mathbb{P}_1(X_t = 0) + \circ\left(\frac{1}{N}\right).\end{aligned}$$

From (1.4.1) and (1.4.3), we deduce that

$$\begin{aligned}\mathbb{P}_1(X_t = 0) &= \lim_{\lambda\to\infty} \exp[-u(t,\lambda)]\\ &= \exp(-\gamma_t).\end{aligned}$$

We then conclude that

$$\begin{aligned}\mathbb{P}_1(Z^N_{[Nt]} > 0) &= 1 - \exp\left[-\frac{\gamma_t}{N} + \circ\left(\frac{1}{N}\right)\right]\\ &= \frac{\gamma_t}{N} + \circ\left(\frac{1}{N}\right).\end{aligned}$$

PROOF OF 2 :

$$\begin{aligned}
\mathbb{E}_1 \exp[-\lambda Z^N_{[Nt]}/N] &= \left(\mathbb{E}_N \exp[-\lambda Z^N_{[Nt]}/N]\right)^{1/N} \\
&\simeq (\mathbb{E}_1 \exp[-\lambda X_t])^{1/N} \\
&= \exp\left(-\frac{\lambda\gamma_t}{N(\lambda+\gamma_t e^{-\alpha t})}\right),
\end{aligned}$$

since again from (1.4.1) and (1.4.3),

$$\mathbb{E}_1 \exp(-\lambda X_t) = \exp\left(-\frac{\lambda\gamma_t}{\lambda+\gamma_t e^{-\alpha t}}\right).$$

But

$$\begin{aligned}
\mathbb{E}_1\left(\exp[-\lambda Z^N_{[Nt]}/N] | Z^N_{[Nt]} > 0\right) &= \frac{\mathbb{E}_1\left(\exp[-\lambda Z^N_{[Nt]}/N]; Z^N_{[Nt]} > 0\right)}{\mathbb{P}_1(Z^N_{[Nt]} > 0)} \\
&= \frac{\mathbb{E}_1\left(\exp[-\lambda Z^N_{[Nt]}/N]\right) - 1 + \mathbb{P}_1(Z^N_{[Nt]} > 0)}{\mathbb{P}_1(Z^N_{[Nt]} > 0)} \\
&= 1 + \frac{\mathbb{E}_1\left(\exp[-\lambda Z^N_{[Nt]}/N]\right) - 1}{\mathbb{P}_1(Z^N_{[Nt]} > 0)} \\
&\simeq 1 - \frac{\lambda}{\lambda+\gamma_t e^{-\alpha t}},
\end{aligned}$$

from which the result follows. □

1.6 Back to the continuous branching process

Note that the continuous limit $\{X_t\}$ has been obtained after a division by N, so that X_t no longer represents a number of individuals, but a sort of density. The point is that there are constantly infinitely many births and deaths, most individuals having a very short live. If we consider only those individuals at time 0 whose progeny is still alive at some time $t > 0$, that number is finite. We now explain how this follows from the last Theorem, and show how it provides a probabilistic description of the subordinator which appeared at the end of section 1.3.

The first part of the theorem tells us that for large N, each of the N individuals from the generation 0 has a progeny at the generation $[Nt]$ with probability $\gamma_t/N + \circ(1/N)$, independently of the others. Hence the number of those individuals tends to the Poisson law with parameter γ_t. The second statement says that those individuals contribute to X_t a quantity which follows an exponential random variable with parameter $\gamma_t e^{-\alpha t}$. This means that

$$X_{0,t}(x) = \sum_{i=1}^{Z_x} Y_i,$$

where Z_x, $Y_1, Y_2, \dots$ are mutually independent, the law of Z_x being Poisson with parameter $x\gamma_t$, and the law of each Y_i exponential with parameter $\gamma_t e^{-\alpha t}$.

Taking into account the branching property, we have more precisely that $\{X_{0,t}(x),\ x \geq 0\}$ is a compound Poisson process, the set of jump locations being a Poisson process with intensity γ_t, the jumps being i. i. d., exponential with parameter $\gamma_t e^{-\alpha t}$. We can recover from this description the formula for the Laplace exponent of $X_t(x)$. Indeed

$$\begin{aligned} \mathbb{E} \exp\left(-\lambda \sum_{i=1}^{Z_x} Y_i\right) &= \sum_{k=0}^{\infty} \left(\mathbb{E} e^{-\lambda Y_1}\right)^k \mathbb{P}(Z_x = k) \\ &= \exp\left(-x \frac{\lambda \gamma_t}{\lambda + \gamma_t e^{-\alpha t}}\right). \end{aligned}$$

We can now describe the genealogy of the population whose total mass follows the SDE (1.3.1).

Suppose that Z ancestors from $t = 0$ contribute respectively $Y_1, Y_2, \dots, Y_Z$ to $X_{0,t}(x)$. Consider now $X_{0,t+s}(x) = X_{t,t+s}(X_{0,t}(x))$. From the Y_1 mass at time t, a finite number Z_1 of individuals, which follows a Poisson law with parameter $Y_1\gamma_s$, has a progeny at time $t+s$, each one contributing an exponential r. v. with parameter $\gamma_s e^{-\alpha s}$ to $X_{0,t+s}(x)$.

Fo any $y, z \geq 0$, $0 \leq s < t$, we say that the individual z in the population at time t is a descendant of the individual y from the population at time s if y is a jump location of the subordinator $x \to X_{s,t}(x)$, and moreover

$$X_{s,t}(y^-) < z < X_{s,t}(y).$$

Note that $\Delta X_{s,t}(y) = X_{s,t}(y) - X_{s,t}(y^-)$ is the contribution to the population at time t of the progeny of the individual y from the population at time s.

1.7 The prolific individuals

We want to consider again the individuals with an infinite line of descent, but directly in the continuous model. Those could be defined as the individuals such that $\Delta X_{0,t}(y) > 0$, for all $t > 0$. However, it should be clear from Proposition 1.4.4 that an a. s. equivalent definition is the following

Definition 1.7.1. *The individual y from the population at time s is said to be prolific if $\Delta X_{s,t}(y) \to \infty$, as $t \to \infty$.*

For any $s \geq 0$, $x > 0$, let

$$\begin{aligned}\mathcal{P}_s(x) &= \{y \in [0, X_s(x)];\ \Delta X_{s,t}(y) \to \infty,\ \text{as } t \to \infty\},\\ P_s(x) &= \text{card}(\mathcal{P}_s(x)).\end{aligned}$$

Define the conditional probability, given extinction

$$\begin{aligned}\mathbb{P}_e &= \mathbb{P}(\cdot|E)\\ &= e^{x\gamma}\mathbb{P}(\cdot \cap E)\end{aligned}$$

It follows from Theorem 4.8.1 below

Proposition 1.7.2. *Under $\mathbb{P}_e$, there exists a standard Brownian motion $\{B^e_t,\ t \geq 0\}$ such that $X_\cdot(x)$ solves the SDE*

$$X_t(x) = x - \alpha \int_0^t X_s(x)ds + \int_0^t \sqrt{\beta X_s(x)}\ dB^e_s.$$

The branching mechanism of X under $\mathbb{P}_e$ is given by

$$\Psi_e(r) = \frac{\beta}{2}r^2 + \alpha r = \Psi(\gamma + r).$$

Next we identify the conditional law of $X_t(x)$, given that $P_t(x) = n$, for $n \geq 0$.

Proposition 1.7.3. *For any Borel measurable $f : \mathbb{R} \to \mathbb{R}_+$,*

$$\mathbb{E}[f(X_t(x))|P_t(x) = n] = \frac{\mathbb{E}_e[f(X_t(x))(X_t(x))^n]}{\mathbb{E}_e[(X_t(x))^n]}.$$

PROOF: Recall that the law of $P_0(x)$ is the Poisson distribution with parameter $x\gamma$. Clearly from the Markov property of $X_\cdot(x)$, the conditional law of $P_t(x)$, given $X_t(x)$, is the Poisson law with parameter $X_t(x)\gamma$. Consequently for $\lambda > 0$, $0 \leq s \leq 1$,

$$\begin{aligned}\mathbb{E}\left(\exp[-\lambda X_t(x)]s^{P_t(x)}\right) &= \mathbb{E}\left(\exp[-\lambda X_t(x)]\exp[-\gamma(1-s)X_t(x)]\right)\\ &= \mathbb{E}\left(\exp[-(\lambda+\gamma)X_t(x)]\exp[\gamma s X_t(x)]\right)\\ &= \sum_{n=0}^{\infty}\frac{(s\gamma)^n}{n!}\mathbb{E}\left(\exp[-(\lambda+\gamma)X_t(x)](X_t(x))^n\right).\end{aligned}$$

Now define

$$h(t,\lambda,x,n) = \mathbb{E}\left(\exp[-\lambda X_t(x)]|P_t(x)=n\right).$$

Note that

$$\begin{aligned}\mathbb{P}(P_t(x)=n) &= \mathbb{E}\left[\mathbb{P}(P_t(x)=n|X(t,x))\right]\\ &= \frac{\gamma^n}{n!}\mathbb{E}\left(e^{-\gamma X_t(x)}(X_t(x))^n\right).\end{aligned}$$

Consequently, conditioning first upon the value of $P_t(x)$, and then using the last identity, we deduce that

$$\mathbb{E}\left(\exp[-\lambda X_t(x)]s^{P_t(x)}\right) = \sum_{n=0}^{\infty}\frac{(s\gamma)^n}{n!}h(t,\lambda,x,n)\mathbb{E}\left(\exp[-\gamma X_t(x)](X_t(x))^n\right).$$

Comparing the two series, and using the fact that, on $\mathcal{F}_t$, $\mathbb{P}_e$ is absolutely continuous with respect to $\mathbb{P}$, with density $e^{x\gamma}\exp[-\gamma X_t(x)]$, we deduce that for all $n \geq 0$,

$$\begin{aligned}h(t,\lambda,x,n) &= \frac{\mathbb{E}\left(\exp[-(\lambda+\gamma)X_t(x)](X_t(x))^n\right)}{\mathbb{E}\left(\exp[-\gamma X_t(x)](X_t(x))^n\right)}\\ &= \frac{\mathbb{E}_e\left(\exp[-\lambda X_t(x)](X_t(x))^n\right)}{\mathbb{E}_e\left[(X_t(x))^n\right]}.\end{aligned}$$

□

To any probability law ν on $\mathbb{R}_+$ with finite mean c, we associate the so–called law of its size–biased picking as the law on $\mathbb{R}_+$ $c^{-1}y\nu(dy)$. We note that the conditional law of $X_t(x)$, given that $P_t(x) = n+1$ is obtained from the conditional law of $X_t(x)$, given that $P_t(x) = n$ by sized–biased picking.

We now describe the law of $\{P_t(x),\ t \geq 0\}$, for fixed $x > 0$. Clearly this is a continuous time B–G–W process as considered in section 1.2 above. We have the

Theorem 1.7.4. *For every $x > 0$, the process $\{P_t(x),\ t \geq 0\}$ is an $\mathbb{N}$–valued immortal Branching process in continuous time, with initial distribution the Poisson law with parameter $x\gamma$, and reproduction measure μ_P given by*

$$\mu_P(n) = \begin{cases} \alpha, & \text{if } n = 2, \\ 0, & \text{if } n \neq 2. \end{cases}$$

In other words, $\{P_t(x),\ t \geq 0\}$ is a Yule tree with the intensity α.

Remark 1.7.5. *If we call Φ_P the Φ–function (with the notations of section 1.2) associated to the measure μ_P, we have in terms of the branching mechanism Ψ of X*

$$\Phi_P(s) = \alpha(s^2 - s) = \frac{1}{\gamma}\Psi(\gamma(1 - s)).$$

Note that Ψ_e describes the branching process X, conditioned upon extinction, while Φ_P describes the immortal part of X. Φ_P depends upon the values $\Phi(r)$, $0 \leq r \leq \gamma$, while Ψ_e depends upon the values $\Phi(r)$, $\gamma \leq r \leq 1$. The mapping $\Psi \to (\Psi_e, \Phi_P)$ should be compared with the mapping $f \to (\tilde{f}, f^*)$ from Proposition 1.1.3.

PROOF: The process P inherits its branching property from that of X. The immortal character is obvious. $P_0(x)$ is the number of individuals from the population at time 0, whose progeny survives at time t, for all $t > 0$. Hence it is the limit as $t \to \infty$ of the law of the number of jumps of $\{X_t(y),\ 0 \leq y \leq x\}$, which is the Poisson distribution with parameter $x\gamma$. This coincides with the result in the subsection 1.5.1, as expected.

Now from the Markov property of X, the conditional law of $P_t(x)$, given $X_t(x)$, is the Poisson law with parameter $X_t(x)\gamma$. Consequently

$$\begin{aligned} \mathbb{E}\left(s^{P_t(x)}\right) &= \mathbb{E}\left(\exp[-(1 - s)\gamma X_t(x)]\right) \\ &= \exp[-xu(t, (1 - s)\gamma)]. \end{aligned}$$

Moreover, if we call $\psi_t(s)$ the generating function of the continuous time B–G–W process $\{P_t(x),\ t \geq 0\}$, we have that

$$\begin{aligned} \mathbb{E}\left(s^{P_t(x)}\right) &= \mathbb{E}\left(\psi_t(s)^{P_0(x)}\right) \\ &= \exp[-x\gamma(1 - \psi_t(s))]. \end{aligned}$$

Comparing those two formulas, we deduce that

$$1 - \psi_t(s) = \frac{1}{\gamma} u(t, (1-s)\gamma).$$

Taking the derivative with respect to the time variable t, we deduce from the differential equations satisfied by $\psi_t(\cdot)$ and by $u(t,\cdot)$ the identity

$$\Phi_P(\psi_t(s)) = \frac{1}{\gamma}\Psi(u(t,(1-s)\gamma)) = \frac{1}{\gamma}\Psi(\gamma(1-\psi_t(s))).$$

Consequently

$$\Phi_P(r) = \frac{1}{\gamma}\Psi(\gamma(1-r)).$$

The measure μ_P is then recovered easily from Φ_P. □

We next note that the pair $(X_t(x), P_t(x))$, which we now write $(X_t(x), P_t(x))$, enjoys the Branching property, in the following sense. For every $x > 0$, $n \in \mathbb{N}$, denote by $(X_\cdot(x,n), P_\cdot(x,n))$ a version of the process $\{(X_t(x), P_t(x)),\ t \geq 0\}$, conditioned upon $P_0(x) = n$. What we mean here by the branching property is the fact that for all $x, x' > 0$, $n, n' \in \mathbb{N}$,

$$(X_\cdot(x+x', n+n'), P_\cdot(x+x', n+n'))$$

has the same law as

$$(X_\cdot(x,n), P_\cdot(x,n)) + (X'_\cdot(x',n'), P'_\cdot(x',n')),$$

where the two processes $(X_\cdot(x,n), P_\cdot(x,n))$ and $(X'_\cdot(x',n'), P'_\cdot(x',n'))$ are mutually independent.

We now characterize the joint law of $(X_t(x,n), P_t(x,n))$.

Proposition 1.7.6. *For any $\lambda \geq 0$, $s \in [0,1]$, $t \geq 0$, $x > 0$, $n \in \mathbb{N}$,*

$$\begin{aligned}&\mathbb{E}\left(\exp[-\lambda X_t(x,n)] s^{P_t(x,n)}\right)\\ &= \exp[-x(u(t,\lambda+\gamma)-\gamma)]\left(\frac{u(t,\lambda+\gamma) - u(t,\lambda+\gamma(1-s))}{\gamma}\right)^n.\end{aligned}$$

Proof: First consider the case $n = 0$. We note that $X_\cdot(x,0)$ is a version of the continuous branching process conditioned upon extinction, i. e. with branching mechanism $\Psi_e(r) = \Psi(\gamma + r)$, while $P_t(x,0) \equiv 0$. Hence

$$\mathbb{E}\left(\exp[-\lambda X_t(x,0)] s^{P_t(x,0)}\right) = \exp[-x(u(t,\lambda+\gamma)-\gamma)]. \tag{1.7.1}$$

Going back to the computation in the beginning of the proof of Proposition 1.7.3, we have

$$\begin{aligned}\mathbb{E}\left(\exp[-\lambda X_t(x)]s^{P_t(x)}\right) &= \mathbb{E}\left(\exp[-(\lambda+\gamma(1-s))X_t(x)]\right)\\ &= \exp[-xu(t,\lambda+\gamma(1-s))].\end{aligned}$$

Since the law of $P_0(x)$ is Poisson with parameter $x\gamma$,

$$\mathbb{E}\left(\exp[-\lambda X_t(x)]s^{P_t(x)}\right) = \sum_{n=0}^{\infty} e^{-x\gamma}\frac{(x\gamma)^n}{n!}\mathbb{E}\left(\exp[-\lambda X_t(x,n)]s^{P_t(x,n)}\right).$$

From the branching property of (X,P),

$$\begin{aligned}\mathbb{E}\left(\exp[-\lambda X_t(x,n)]s^{P_t(x,n)}\right) &= \mathbb{E}\left(\exp[-\lambda X_t(x,0)]s^{P_t(x,0)}\right)\\ &\quad\times\left[\mathbb{E}\left(\exp[-\lambda X_t(0,1)]s^{P_t(0,1)}\right)\right]^n.\end{aligned}\tag{1.7.2}$$

Combining the four above identities, we obtain

$$\begin{aligned}&\sum_{n=0}^{\infty}\frac{(x\gamma)^n}{n!}\left[\mathbb{E}\left(\exp[-\lambda X_t(0,1)]s^{P_t(0,1)}\right)\right]^n\\ &\qquad= \exp\left\{x\left[u(t,\lambda+\gamma)-u(t,\lambda+\gamma(1-s))\right]\right\}\\ &\qquad= \sum_{n=0}^{\infty}\frac{\left\{x\left[u(t,\lambda+\gamma)-u(t,\lambda+\gamma(1-s))\right]\right\}^n}{n!}.\end{aligned}$$

Identifying the coefficients of x in the two series yields

$$\mathbb{E}\left(\exp[-\lambda X_t(0,1)]s^{P_t(0,1)}\right) = \frac{u(t,\lambda+\gamma)-u(t,\lambda+\gamma(1-s))}{\gamma}.$$

The result follows from this, (1.7.1) and (1.7.2). □

1.8 Bibliographical comments

We have essentially followed the treatment from [18] in section 1.1. Section 1.2 is inspired from [16]. Section 1.4 owes much to [16], [3] and [17]. The subsection 1.5.1 is taken from [19], 1.5.2 from [20]. Section 1.7 is a translation of the results in [2] to our particular case.

Chapter 2

Genealogical models for fixed–size populations

Consider a population of fixed size N, which evolves in discrete generations. Assume that each individual can be of two different types (a and A, say).

2.1 The simplest Wright–Fisher model

Consider first the case where those are neutral, i. e. there is no selective advantage attached to either of those two types, and there is no mutation.

Reproduction is random (and asexual). More precisely, we assume that each individual picks his parent uniformly from the previous generation (with replacement), and copy his type. Denote

$$Y_k^N := \text{ number of type } A \text{ individuals in generation } k.$$

Clearly

$$\mathbb{P}(Y_{k+1}^N = i | Y_k^N = j) = C_N^i \left(\frac{j}{N}\right)^i \left(1 - \frac{j}{N}\right)^{N-i}.$$

From this, we see that $\{Y_k^N,\ k \geq 0\}$ is both a finite state Markov chain, and a bounded martingale. Note that the two states 0 and N are absorbing, and all other states are transient. Consequently

$$Y_\infty^N = \lim_{k\to\infty} Y_k^N \in \{0, N\}.$$

Moreover

$$j = \mathbb{E}[Y_\infty^N | Y_0^N = j] = N\mathbb{P}(Y_\infty^N = N),$$

hence the probability of fixation of type A is its initial frequency j/N.

The next question is what can we say about the time we have to wait until the population is homogeneous (i. e. $Y_k^N = 0$ or N) ?

If we want to study this and other questions for large N, we should understand the behaviour of the proportions of both alleles in a population of infinite size. Define the following continuous time process :

$$X_t^N = N^{-1} Y_{[Nt]}^N, \ t \geq 0.$$

This means that we consider the fraction of type A–individuals, and the time is a number of generations divided by the size of the population.

Let $t \in \mathbb{N}/N$ and $\Delta t = N^{-1}$. It is not hard to check that

$$\mathbb{E}[X_{t+\Delta t}^N - X_t^N | X_t^N] = 0, \ \mathbb{E}[(X_{t+\Delta t}^N - X_t^N)^2 | X_t^N] = X_t^N(1 - X_t^N)\Delta t.$$

We now want to let $N \to \infty$.

Theorem 2.1.1. *Suppose that $X_0^N \Rightarrow X_0$, as $N \to \infty$. Then $X^N \Rightarrow X$ in $D(\mathbb{R}_+; [0,1])$, where $\{X_t, \ t \geq 0\}$ solves the SDE*

$$dX_t = \sqrt{X_t(1 - X_t)} dB_t, \ t \geq 0.$$

PROOF: The idea is to prove that $\forall f \in C^3([0,1])$, the process

$$M_t^f := f(X_t) - f(X_0) - \frac{1}{2} \int_0^t X_s(1 - X_s) f''(X_s) ds, \ t \geq 0 \qquad (2.1.1)$$

is a martingale (with respect to its own filtration).

It is known that this martingale problem has a unique solution (the SDE has a unique strong solution, see next chapter). Hence the theorem follows from the two following statements

1. the sequence $\{X^N, \ N = 1, 2, \ldots\}$ is tight;
2. any weak limit of a sub–sequence solves the above martingale problem.

PROOF OF 1. Also the sequence is a sequence in the space $D(\mathbb{R}_+; [0,1])$ of discontinuous processes, since the limit is continuous, a tightness criteria in

$C(\mathbb{R}_+;[0,1])$ is good enough. It will prove at the same time that any limit point has continuous paths. For $0 \le i < j$, let $s = i/N$, t=j/N. We have

$$\begin{aligned}
\mathbb{E}\left[|X_t^N - X_s^N|^4\right] &= N^{-4}\mathbb{E}\left(\left[\sum_{k=i}^{j-1}(Y_{k+1}^N - Y_k^N)\right]^4\right) \\
&= N^{-4}\mathbb{E}\sum_{k_1,\dots,k_4=i}^{j-1}\prod_{\ell=1}^{4}(Y_{k_\ell+1}^N - Y_{k_\ell}^N) \\
&= N^{-4}\Big(\mathbb{E}\sum_{k=i}^{j-1}(Y_{k+1}^N - Y_k^N)^4 + 2\mathbb{E}\sum_{i\le k_1<k_2\le j-1}(Y_{k_1+1}^N - Y_{k_1}^N)^2(Y_{k_2+1}^N - Y_{k_2}^N)^2 \\
&\quad + \mathbb{E}\sum_{i\le k_1<k_2\le j-1}(Y_{k_1+1}^N - Y_{k_1}^N)(Y_{k_2+1}^N - Y_{k_2}^N)^3 \\
&\quad + 2\mathbb{E}\sum_{i\le k_1<k_2<k_3\le j-1}(Y_{k_1+1}^N - Y_{k_1}^N)(Y_{k_2+1}^N - Y_{k_2}^N)(Y_{k_3+1}^N - Y_{k_3}^N)^2\Big) \\
&\le C\left(\frac{j-i}{N}\right)^2 = C(t-s)^2.
\end{aligned}$$

Indeed, we first note that

$$\begin{aligned}
\mathbb{E}\left[(Y_{k+1}^N - Y_k^N)^2\right] &= \mathbb{E}\left\{\mathbb{E}\left[(Y_{k+1}^N - Y_k^N)^2|Y_k^N\right]\right\} \\
&\le N/4,
\end{aligned}$$

from which it follows that the second term above has the right size. Concerning the first term, we note that

$$\mathbb{E}\left[(Y_{k+1}^N - Y_k^N)^4\right] = \mathbb{E}\left\{\mathbb{E}\left[(Y_{k+1}^N - Y_k^N)^4|Y_k^N\right]\right\}$$

Conditionally upon $Y_k^N = y$, Y_{k+1}^N follows the binomial law $B(N,p)$ where $p = y/N$. But if $Z_1,\dots,Z_n$ are Bernoulli with $\mathbb{P}(Z_i = 1) = p$, then

$$\begin{aligned}
\mathbb{E}\left(\left[\sum_{i=1}^{N}(Z_i - p)\right]^4\right) &= \mathbb{E}\sum_{i=1}^{N}(Z_i - p)^4 + 4\mathbb{E}\sum_{1\le i<j\le N}(Z_i - p)^2(Z_j - p)^2 \\
&\le 2N^2
\end{aligned}$$

Consequently

$$\mathbb{E}\left[(Y_{k+1}^N - Y_k^N)^4\right] \le 2N^2,$$

and the first term is bounded by $cN^{-1}(t-s) \le c(t-s)^2$, since $1 \le N(t-s)$. Moreover we have

$$\begin{aligned}\mathbb{E}\left[(Y_{k+1}^N - Y_k^N)^3 | Y_k^N\right] &= Y_k^N\left(1-\frac{Y_k^N}{N}\right)\left(1-2\frac{Y_k^N}{N}\right),\\ \left|\mathbb{E}\left[(Y_{k+1}^N - Y_k^N)^3 | Y_k^N\right]\right| &\le N,\\ \left|\mathbb{E}\left[(Y_{k_1+1}^N - Y_{k_1}^N)(Y_{k_2+1}^N - Y_{k_2}^N)^3\right]\right| &\le N^2.\end{aligned}$$

Consequently the third term is estimated exactly as the second one. It remains to consider the last term, which is bounded above by

$$\begin{aligned}N^{-4}\mathbb{E}\sum_{i<k<j}(Y_k^N - Y_i^N)^2(Y_{k+1}^N - Y_k^N)^2 &\le N^{-3}\sum_{i<k<j}\mathbb{E}[(Y_k^N - Y_i^N)^2]\\ &= N^{-3}\sum_{i<k<j}\sum_{i\le \ell<k}\mathbb{E}[(Y_{\ell+1}^N - Y_\ell^N)^2]\\ &\le N^{-2}\sum_{i<k<j}(k-i)\\ &\le (t-s)^2.\end{aligned}$$

The wished estimate of $\mathbb{E}[|X_t^N - X_s^N|^4]$ is established. Keeping in mind that our process takes values in the compact set $[0,1]$, it remains to apply Theorem 12.3 page 95 from Billingsley [4] (see also theorem 7.2.1 below), in conjunction with Proposition 3.10.4 page 149 from Ethier, Kurtz [11].

Proof of 2. With the same notations as above, in particular $\Delta t = N^{-1}$,

$$\begin{aligned}G^N f(x) &:= \mathbb{E}\left[f(X_{t+\Delta t}^N) - f(X_t^N) | X_t^N = x\right]\\ &= \mathbb{E}\left[f\left(N^{-1}\sum_{i=1}^N Z_i\right) - f(x)\right]\\ &= \frac{1}{2}\mathbb{E}\left[\left(N^{-1}\sum_{i=1}^N Z_i - x\right)^2\right]f''(x) + \frac{1}{6}\mathbb{E}\left[\left(N^{-1}\sum_{i=1}^N Z_i - x\right)^3 f'''(\xi)\right]\\ &= \frac{1}{2N}x(1-x)f''(x) + r_N(x),\end{aligned}$$

where $r_N(x) = O(N^{-3/2})$, since

$$\begin{aligned}\frac{1}{6}\mathbb{E}\left|\left(N^{-1}\sum_{i=1}^{N} Z_i - x\right)^3 f'''(\xi)\right| &\leq \frac{1}{6}\|f'''\|_\infty \sup_{0\leq x\leq 1} \mathbb{E}\left(\left|N^{-1}\sum_{i=1}^{N} Z_i - x\right|^3\right) \\ &\leq \frac{1}{6}\|f'''\|_\infty \sup_{0\leq x\leq 1} \left(\mathbb{E}\left(\left|N^{-1}\sum_{i=1}^{N} Z_i - x\right|^4\right)\right)^{3/4} \\ &= O(N^{-3/2}).\end{aligned}$$

Now it is easily seen that

$$\begin{aligned}M_t^N &:= f(X_t^N) - f(X_0^N) - \sum_{i=0}^{[Nt]-1} G^N f(X_{i/N}^N) \\ &= f(X_t^N) - f(X_0^N) - \frac{1}{2}\int_0^{[Nt]/N} X_s^N(1-X_s^N)f''(X_s^N)ds - \int_0^{[Nt]/N} N r_N(X_s^N)ds\end{aligned}$$

is a bounded martingale (with respect to the natural filtration generated by the process X^N).

This means that for any $0 \leq s < t$ and any bounded and continuous function $\varphi : D([0,s];[0,1]) \to \mathbb{R}$,

$$\mathbb{E}\left[M_t^N \varphi\left((X_r^N)_{0\leq r\leq s}\right)\right] = \mathbb{E}\left[M_s^N \varphi\left((X_r^N)_{0\leq r\leq s}\right)\right].$$

Taking the limit along any converging subsequence, we get that any limit point X satisfies (2.1.1). □

2.2 Wright–Fisher model with mutations

Assume that mutation converts at birth an A–type to an a–type with probability α_1, and converts an a–type to an A–type with probability α_0. Here

$$\mathbb{P}(Y_{k+1}^N = i | Y_k^N = j) = C_N^i p_j^i \left(1 - p_j\right)^{N-i},$$

where

$$p_j = \frac{j(1-\alpha_1) + (N-j)\alpha_0}{N}.$$

We now want to let $N \to \infty$. Assume that α_0 and α_1 are of the form

$$\alpha_1 = \gamma_1/N,\ \alpha_0 = \gamma_0/N, \quad \text{where } \gamma_1 > 0,\ \gamma_0 > 0 \text{ are fixed,}$$

and define again the continuous time process

$$X_t^N = N^{-1} Y_{[Nt]}^N,\ t \geq 0.$$

$$\begin{aligned} \mathbb{E}[X_{t+\Delta t}^N - X_t^N | X_t^N] &= [-\gamma_1 X_t^N + \gamma_0 (1 - X_t^N)]\Delta t, \\ \mathbb{E}[(X_{t+\Delta t}^N - X_t^N)^2 | X_t^N] &= X_t^N (1 - X_t^N)\Delta t + O(\Delta t^2). \end{aligned}$$

As $N \to \infty$, $X^N \Rightarrow X$, where $\{X_t,\ t \geq 0\}$ solves the SDE

$$dX_t = \gamma_0 (1 - X_t) dt - \gamma_1 X_t dt + \sqrt{X_t(1 - X_t)} dB_t,\ t \geq 0.$$

2.3 Wright–Fisher model with selection

Assume that type A is selectively superior to type a. Then

$$\mathbb{P}(Y_{k+1}^N = i | Y_k^N = j) = C_N^i p_j^i (1 - p_j)^{N-i},$$

where

$$p_j = \frac{j(1+s)}{j(1+s) + N - j}.$$

If we want to combine mutations and selection, we choose

$$p_j = \frac{(1+s)[j(1-\alpha_1) + (N-j)\alpha_2]}{(1+s)[j(1-\alpha_1) + (N-j)\alpha_2] + j\alpha_1 + (N-j)(1-\beta)}.$$

We again want to let $N \to \infty$. Let $\alpha_1 = 0$, $\alpha_0 = 0$, and $s = \beta/N$, with $\beta > 0$. We define the continuous time process

$$X_t^N = N^{-1} Y_{[Nt]}^N,\ t \geq 0.$$

$$\begin{aligned} \mathbb{E}[X_{t+\Delta t}^N - X_t^N | X_t^N] &= \beta X_t^N (1 - X_t^N)]\Delta t + 0(\Delta t^2), \\ \mathbb{E}[(X_{t+\Delta t}^N - X_t^N)^2 | X_t^N] &= X_t^N (1 - X_t^N)\Delta t + 0(\Delta t^2). \end{aligned}$$

As $N \to \infty$, $X^N \Rightarrow X$, where $\{X_t,\ t \geq 0\}$ solves the SDE

$$dX_t = \beta X_t(1 - X_t)dt + \sqrt{X_t(1 - X_t)}dB_t,\ t \geq 0.$$

Suppose now that $\alpha_0 = \alpha_1 = \gamma/N$. Then

$$\begin{aligned} \mathbb{E}[X^N_{t+\Delta t} - X^N_t | X^N_t] &= \beta X^N_t(1 - X^N_t)]\Delta t + \gamma(1 - 2X^N_t)]\Delta t + 0(\Delta t^2), \\ \mathbb{E}[(X^N_{t+\Delta t} - X^N_t)^2 | X^N_t] &= X^N_t(1 - X^N_t)\Delta t + 0(\Delta t^2). \end{aligned}$$

Then as $N \to \infty$, $X^N \Rightarrow X$, where $\{X_t,\ t \geq 0\}$ solves the SDE

$$dX_t = [\beta X_t(1 - X_t) + \gamma(1 - 2X_t)]dt + \sqrt{X_t(1 - X_t)}dB_t,\ t \geq 0.$$

2.4 Bibliographical comments

Section 2.1 is essentially borrowed from [6], while the next sections are inspired by [15].

Chapter 3

Connections between branching and fixed–size population models

3.1 Discrete models

Consider a B–G–W process $\{Z_n,\ n \geq 0\}$ with initial condition $Z_0 = N$, and Poisson offspring distribution with mean λ.

Proposition 3.1.1. *For any $T \geq 1$, the law of $\{Z_n,\ 0 \leq n \leq T\}$, conditioned upon $\{Z_n = N,\ 0 \leq n \leq T\}$, is the law of the Wright–Fisher model.*

PROOF: Let ξ_i denote the number of offsprings in generation 1 of individual i from the 0 generation. For any $0 \leq k_1, \ldots, k_N \leq N$ satisfying $k_1 + \cdots + k_N = N$, since the law of the sum of N independent Poisson (λ) r. v.'s is Poisson ($N\lambda$),

$$\begin{aligned}
\mathbb{P}(\xi_1 = k_1, \ldots, \xi_N = k_N | \xi_1 + \cdots + \xi_N = N) &= \frac{\mathbb{P}(\xi_1 = k_1) \times \cdots \times \mathbb{P}(\xi_N = k_N)}{\mathbb{P}(Z_1 = N | Z_0 = N)} \\
&= \frac{e^{-\lambda} \frac{\lambda^{k_1}}{k_1!} \times \cdots \times e^{-\lambda} \frac{\lambda^{k_N}}{k_N!}}{e^{-N\lambda} \frac{(N\lambda)^N}{N!}} \\
&= \frac{N!}{k_1! \times \cdots \times k_N!} \left(\frac{1}{N}\right)^N.
\end{aligned}$$

We recognize the multinomial distribution. □

Consider now the two–types Wright–Fisher model where the initial population contains ℓ type A individuals and $N-\ell$ type a individuals, and there is no selection and no mutation. Denote by Y_k^N the number of type A individuals in generation k.

Proposition 3.1.2. *If $Y_0^N \to Z_0$ as $N \to \infty$, then $Y^N \Rightarrow Z$ as $N \to \infty$, where $\{Z_k,\ k \geq 0\}$ is a B–G–W process with Poisson(1) offspring distribution.*

Proof: It suffices to prove the weak convergence of finite dimensional distributions. But the conditional law of Y_{k+1}^N, given that $Y_k^N = r$ is the binomial law with parameters $(N, \frac{r}{N})$, which, as $N \to \infty$, converges to the Poisson(r) distribution. □

3.2 Diffusion models

Consider the Wright–Fisher model with selection

$$dX_t = \beta X_t(1-X_t)dt + \sqrt{X_t(1-X_t)}dB_t.$$

Define the increasing process

$$A_t = \int_0^t (1-X_s)ds,$$

and its inverse

$$\sigma_t = \inf\{s;\ A_s > t\},$$

which is defined for all $0 \leq t \leq A_\infty$, where

$$A_\infty = \int_0^\tau (1-X_s)ds, \quad \tau = \inf\{s;\ X_s = 1\}.$$

Note that $\{\tau = \infty\} = \{A_\infty = \infty\}$. Define the process $Y_t = X_{\sigma_t}$ for $0 \leq t < A_\infty$. $\{Y_t;\ 0 \leq t < A_\infty\}$ solves the SDE

$$dY_t = \beta Y_t dt + \sqrt{Y_t}dB_t,$$

which is a continuous branching diffusion.

3.3 Bibliographical comments

Section 3.1 is borrowed from [16].

Chapter 4

One–dimensional diffusions

4.1 A uniqueness theorem of Yamada–Watanabe

We consider the one–dimensional SDE of the form

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dB_t, \ t \geq 0; \\ \ X_0 = x; \end{cases} \tag{4.1.1}$$

where $x \in \mathbb{R}$, $\{B_t, \ t \geq 0\}$ is a one dimensional standard Brownian motion, $b : \mathbb{R} \to \mathbb{R}$ is uniformly Lipschitz continuous and $\sigma : \mathbb{R} \to \mathbb{R}$ satisfies

$$\text{there exists a function } \rho : \mathbb{R}_+ \to \mathbb{R}_+ \text{ such that} \\ \begin{array}{ll} (i) & \rho \text{ is strictly increasing, } \rho(0) = 0, \\ (ii) & \displaystyle\int_{0+} \frac{dr}{\rho^2(r)} = +\infty, \\ (iii) & |\sigma(x) - \sigma(y)| \leq \rho(|x-y|), \ x, y \in \mathbb{R}. \end{array} \tag{4.1.2}$$

Note that $\rho(r) = r^{1/2}$ satisfies (4.1.2) (i) and (ii). We have the

Theorem 4.1.1. *Under the above assumptions, if moreover* $|\sigma(x)| \leq K(1 + |x|)$ $\forall x \in \mathbb{R}$, *then the equation* (4.1.1) *has at most one solution.*

PROOF: Let $1 = a_0 > a_1 > a_2 > \cdots > 0$ be such that

$$\forall n \geq 1, \quad \int_{a_n}^{a_{n-1}} \frac{dr}{\rho^2(r)} = n.$$

Clearly, $a_n \to 0$, as $n \to \infty$. For each $n \geq 1$, let ψ_n be a continuous function with support in (a_n, a_{n-1}) such that

$$0 \leq \psi_n(r) \leq \frac{2}{n\rho^2(r)} \quad \text{and} \quad \int_{a_n}^{a_{n-1}} \psi_n(r)dr = 1.$$

Set

$$\varphi_n(x) = \int_0^{|x|} dy \int_0^y \psi_n(r)dr, \quad x \in \mathbb{R}.$$

It is easily seen that $\varphi_n \in C^2(\mathbb{R})$, $|\varphi_n'(x)| \leq 1$, and $\varphi_n(x) \uparrow |x|$ as $n \to \infty$.

Let X^1 and X^2 be two solutions of (4.1.1).Then from It's formula

$$\begin{aligned}\mathbb{E}[\varphi_n(X_t^1 - X_t^2)] &= \mathbb{E}\int_0^t \varphi_n'(X_s^1 - X_s^2)[b(X_s^1) - b(X_s^2)]ds + \\ &\quad + \frac{1}{2}\int_0^t \varphi_n''(X_s^1 - X_s^2)[\sigma(X_s^1) - \sigma(X_s^2)]^2 ds \\ &\leq K\int_0^t \mathbb{E}[|X_s^1 - X_s^2|]ds + t/n,\end{aligned}$$

where we have used (4.1.2) (iii) and $\psi_n(r) \leq 2/n\rho^2(r)$ for the last inequality. Taking the limit as $n \to \infty$ in this inequality yields

$$\mathbb{E}[|X_t^1 - X_t^2|] \leq K\int_0^t \mathbb{E}[|X_s^1 - X_s^2|]ds.$$

The result now follows from Gronwall's Lemma. □

Remark 4.1.2. *The Lipschitz condition on b can be replaced by the more general condition*

$$|b(x) - b(y)| \leq \kappa(|x - y|),$$

with $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ concave and increasing, satisfying $\kappa(0) = 0$ and $\int_{0+} \kappa^{-1}(r)dr = \infty$ if we use Bihari's generalization of Gronwall's Lemma.

We now have the

Corollary 4.1.3. *Under the assumptions of Theorem 4.1.1, the SDE* (4.1.1) *has a unique weak solution, i. e. there exists a probability space with a filtration $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ on which one can define a progressively measurable process $\{(X_t, B_t),\ t \geq 0\}$ such that B is a standard Brownian motion, and* (4.1.1) *is satisfied, and moreover the law of $\{X_t,\ t \geq 0\}$ on $C(\mathbb{R}_+)$ is uniquely determined by that statement.*

PROOF: Since the coefficients b and σ are continuous, it is not hard to construct a solution of the martingale problem by weak convergence of an approximating Markov chain, as in Theorem 2.1.1. the existence of a solution in the sense of the statement then follows from the martingale representation Theorem 7.3.1 below.

It remains to prove the uniqueness of the law of $\{X_t,\ t \geq 0\}$. Let $\mathcal{C}_x$ denote the space of continuous functions from $\mathbb{R}_+$ into $\mathbb{R}$ which start from x at time 0. Let $\{(X_t, B_t),\ t \geq 0\}$ and $\{(X'_t, B'_t),\ t \geq 0\}$ denote two solutions of (4.1.1) (possibly defined on different probability spaces), and let $\mathbb{P}_x$ (resp. $\mathbb{P}'_x$) be its law on $\mathcal{C}_x \times \mathcal{C}_0$. If we denote by Π the projection in $\mathcal{C}_x \times \mathcal{C}_0$ onto its second coordinate, then we have that

$$\Pi(\mathbb{P}_x) = \Pi(\mathbb{P}'_x) = W,$$

where W denotes the Wiener maesure (i. e. the law of the standard Brownian motion) on $\mathcal{C}_0$. Let $Q^{w_2}(dw_1)$ [resp. $Q'^{w_2}(dw_1)$] denote a regular conditional probability distribution of w_1, given w_2 under $\mathbb{P}_x$ [resp. under $\mathbb{P}'_x$]. This means that

1. $\forall w_2 \in \mathcal{C}_0$, Q^{w_2} and Q'^{w_2} are two probability measures on $(\mathcal{C}_x, \mathcal{B}(\mathcal{C}_x))$.
2. $\forall A \in \mathcal{B}(\mathcal{C}_x)$, $w_2 \longrightarrow Q^{w_2}(A)$ and $w_2 \longrightarrow Q'^{w_2}(A)$ are $\mathcal{B}(\mathcal{C}_0$–measurable.
3. $\forall A \in \mathcal{B}(\mathcal{C}_x)$, $B \in \mathcal{B}(\mathcal{C}_0)$,

$$\mathbb{P}_x(A \times B) = \int_B Q^{w_2}(A) W(dw_2),$$
$$\mathbb{P}'_x(A \times B) = \int_B Q'^{w_2}(A) W(dw_2).$$

Consider the measure

$$Q(dw_1, dw_2, dw_3) = Q^{w_3}(dw_1) Q'^{w_3}(dw_2) W(dw_3)$$

on the space $(\Omega, \mathcal{F})$, where $\Omega = \mathcal{C}_x \times \mathcal{C}_x \times \mathcal{C}_0$ and $\mathcal{F}$ is its Borel σ–field, completed with the class $\mathcal{N}$ of Q–null sets. We equip this probability space with the filtration

$$\mathcal{F}_t = \cap_{\varepsilon > 0} \left(\mathcal{B}_{t+\varepsilon} \vee \mathcal{N} \right),$$

where $\mathcal{B}_t = \mathcal{B}_t(\mathcal{C}_x) \times \mathcal{B}_t(\mathcal{C}_x) \times \mathcal{B}_t(\mathcal{C}_0)$.

Then the law of (w_1, w_3) [resp. of (w_2, w_3)] under Q is $\mathbb{P}_x$ [resp. P'_x]. But (w_1, w_3) and (w_2, w_3) are two solutions of (4.1.1), defined on the same probability space. Hence from Theorem 4.1.1, $w_1 = w_2$ Q a. s. But since for W almost all w_3, w_1 and w_2 are independent, this implies that there exists a measurable mapping $F : \mathbb{R} \times \mathcal{C}_0$ such that W a. s.

$$Q^{w_3}(dw) = Q'^{w_3}(dw) = \Delta_{F(x,w_3)}(dw).$$

Hence certainly for all $A \in \mathcal{B}(\mathcal{C}_x)$,

$$\mathbb{P}(X \in A) = \int_{\mathcal{C}_0} Q^w(A)W(dw) = \int_{\mathcal{C}_0} Q'^w(A)W(dw) = \mathbb{P}(X' \in A),$$

which proves the uniqueness in law. □

Remark 4.1.4. *With a little more effort, one can deduce from the above argument that given the Brownian motion* $\{B_t,\ t \geq 0\}$ *defined on a probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$*, one can construct the solution* $\{X_t,\ t \geq 0\}$ *of* (4.1.1) *driven by that Brownian motion as* $X = F(x, B)$ *(with the notation from the preceding proof) which is then unique from Theorem 4.1.1. The point is that weak existence and pathwise uniqueness implies existence of a unique strong solution, see [13] for details.*

4.2 The local time of a one dimensional diffusion

4.2.1 Local time of the Brownian motion

Note that Brownian motion spends zero time at any point $x \in \mathbb{R}$. We shall now define a process, called the *local time*, which in a sense measures the time spent by the Brownian motion near any point $x \in \mathbb{R}$. The local time at x of the Brownian motion up to time t can be intuitively defined as

$$L_t^x = \int_0^t \delta_x(B_s)ds,$$

where δ_x is the Dirac measure at the point x. It can be rigorously defined through the Itô–Tanaka formula, which says that for all $x \in \mathbb{R}$, $t \geq 0$,

$$(B_t - x)_+ = (B_0 - x)_+ + \int_0^t \mathbf{1}_{\{B_s > x\}} dB_s + \frac{1}{2} L_t^x. \tag{4.2.1}$$

Since all terms except the last one in this formula are well defined, that formula provides a definition of the process $\{L_t^x,\ t \geq 0\}$. Note that the intuitive definition given above is related to the fact that, if we take any reasonable approximation of the Dirac measure at x by a sequence of functions f_n (for instance with compact support), and apply Itô's formula to a function whose second derivative if f_n, then we obtain the Itô–Tanaka formula in the limit.

It is easily seen from the definition that for each $x \in \mathbb{R}$, $t \to L_t^x$ is continuous and increasing. We have moreover the

Proposition 4.2.1. *There exists a version of the two–parameter process* $\{L_t^x,\ x \in \mathbb{R},\ t \geq 0\}$ *whose trajectories are jointly continuous with respect to* (t, x).

PROOF: □

A very important result is the *occupation times formula* :

Theorem 4.2.2. *If* $f : \mathbb{R} \to \mathbb{R}$ *is locally integrable, then for all* $t \geq 0$, *a. s.,*

$$\int_0^t f(B_s)ds = \int_{\mathbb{R}} f(x)L_t^x dx.$$

PROOF: It suffices to proves the result in the case f has a compact support. For such an f, let

$$F(x) := \int_{-\infty}^x f(y)dy, \quad \mathcal{F}(x) := \int_{-\infty}^x F(y)dy.$$

The following two identities are easy to check :

$$F(B_t) = \int_{\mathbb{R}} f(x)\mathbf{1}_{\{B_t > x\}}dx, \quad \mathcal{F}(B_t) = \int_{\mathbb{R}} f(x)(B_t - x)_+dx.$$

The result now follows by multiplying the Itô–Tanaka formula by f(x), integrating over $\mathbb{R}$ with respect to Lebesgue's measure, and interchanging the Lebesgue integral with respect to dx and the Itô integral with respect to dB_t. The justification of that point is left to the reader. □

4.2.2 Local time of a one–dimensional diffusion

If $\{X_t,\ t \geq 0\}$ is the solution of a one–dimensional SDE, it is in particular a continuous semimartingale. We can now state the

Proposition 4.2.3. *For any real number x, there exists a continuous increasing process L^x called the local time of X at x, such that*

$$|X_t - x| = |X_0 - x| + \int_0^t sign(X_s - x)dX_s + L_t^x, \tag{4.2.2}$$

where

$$sign(x) = \begin{cases} -1 & , \text{ if } x \leq 0; \\ +1 & , \text{ if } x > 0. \end{cases}$$

Proof: For any $\varepsilon > 0$, let $y \to \varphi_\varepsilon(y)$ be defined by

$$\varphi_\varepsilon(y) = \begin{cases} \frac{|y-x|^2}{2\varepsilon}, & \text{if } y \in [x-\varepsilon, x+\varepsilon]; \\ |y-x| - \varepsilon/2, & \text{if } y \notin [x-\varepsilon, x+\varepsilon]. \end{cases}$$

Note that

$$\varphi'_\varepsilon(y) = \left(\frac{y-x}{\varepsilon} \wedge 1\right) \vee (-1), \quad \varphi''_\varepsilon(y) = \varepsilon^{-1}\mathbf{1}_{[x-\varepsilon,x+\varepsilon]}(y), \quad y \in \mathbb{R}.$$

Itô's formula (taking the limit along a regularizing sequence made of C^2 approximations of φ_ε whose second derivative increases to that of φ_ε) yields

$$\varphi_\varepsilon(X_t) = \varphi_\varepsilon(X_0) + \int_0^t \varphi'_\varepsilon(X_s)dX_s + L_t^{x,\varepsilon}, \tag{4.2.3}$$

where

$$L_t^{x,\varepsilon} = \frac{1}{2\varepsilon}\int_0^t \mathbf{1}_{[x-\varepsilon,x+\varepsilon]}(X_s)d\langle X, X\rangle_s.$$

As $\varepsilon \to 0$, the three first terms in the identity (4.2.3) converge to the corresponding terms of (4.2.2), hence $L_t^{x,\varepsilon} \to L_t^x$ in probability as $\varepsilon \to 0$. The a. s. continuity follows from that of the other terms in (4.2.2). The fact that $t \to L_t^x$ is a. s. increasing follows from the smae property for $L_t^{x,\varepsilon}$. □

Exercise 4.2.4. *Justify the argument leading to formula* (4.2.3).

Let us prove the

Proposition 4.2.5. *(Occupation times formula) Outside a $\mathbb{P}$–negligible subset of Ω, for all $t > 0$ and Borel measurable functions φ from $\mathbb{R}$ into $\mathbb{R}_+$,*

$$\int_0^t \varphi(X_s)d\langle X, X\rangle_s = \int_{\mathbb{R}} \varphi(x)L_t^x dx.$$

PROOF: It suffices to prove the result for φ continuous with compact support. With such a φ, define

$$f(x) = \frac{1}{2}\int_{\mathbb{R}} |x-y|\varphi(y)dy.$$

Then $f \in C^2(\mathbb{R})$,

$$f'(x) = \frac{1}{2}\int_{\mathbb{R}} \text{sign}(x-y)\varphi(y)dy, \quad f''(x) = \varphi(x).$$

It follows from Itô's formula that a. s. for all $t \geq 0$,

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dX_s + \frac{1}{2}\int_0^t f''(X_s)d\langle X, X\rangle_s.$$

On the other hand, if we multiply equation (4.2.2) by $\varphi(x)/2$ and integrate over $\mathbb{R}$ with respect to dx, we deduce that a. s. for all $t \geq 0$,

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dX_s + \frac{1}{2}\int_{\mathbb{R}} L_t^x \varphi(x)dx. \tag{4.2.4}$$

The result follows from the comparison of the two last identities. □

Exercise 4.2.6. *Justify the argument leading to the identity* (4.2.4)*.*

4.3 A comparison theorem

Consider two one–dimensional SDEs

$$X_t^{(1)} = x_1 + \int_0^t b_1(X_s^{(1)})ds + \int_0^t \sigma(X_s^{(1)})dB_s, \tag{4.3.1}$$

$$X_t^{(2)} = x_2 + \int_0^t b_2(X_s^{(2)})ds + \int_0^t \sigma(X_s^{(2)})dB_s. \tag{4.3.2}$$

The aim of this section is to prove the

Theorem 4.3.1. *Let $\{X_t^{(1)},\ t \geq 0\}$ (resp. $\{X_t^{(2)},\ t \geq 0\}$) be a solution of equation (4.3.1) (resp. (4.3.2)). Assume that σ satisfies the same condition as in Theorem 4.1.1, that either b_1 or b_2 is globally Lipschitz. If moreover $x_1 \leq x_2$, $b_1(x) \leq b_2(x)$, $\forall x \in \mathbb{R}$, then $X_t^{(1)} \leq X_t^{(2)}$ for all $t \geq 0$, a. s.*

Proof:

4.4 Classification of boundary points

As we will see, all we need to understand is the behaviour at 0 of Bessel square processes. Let $\{W_t,\ t \geq 0\}$ be a d–dimensional standard Brownian motion, starting from $y \neq 0$. Define $X_t := |W_t|^2$, $x = |y|^2$. From Itô's formula,

$$X_t = x + d \times t + 2 \int_0^t < W_s, dW_s > .$$

Note that the continuous martingale $M_t = \int_0^t < W_s, dW_s >$ has its quadratic variation given by $< M >_t = \int_0^t X_s ds$. Hence there exists a one–dimensional Brownian motion $\{B_t,\ t \geq 0\}$ such that

$$M_t = \int_0^t \sqrt{X_s} dB_s,$$

and $\{X_t,\ t \geq 0\}$ solves the SDE

$$X_t = x + d \times t + 2 \int_0^t \sqrt{X_s} dB_s.$$

We now consider for any $\delta \geq 0$, $x > 0$, the SDE

$$\begin{cases} dX_t = \delta \times dt + 2\sqrt{X_t} dB_t, \\ X_0 = x. \end{cases} \tag{4.4.1}$$

From a well–known result of Yamada–Watanabe (see e. g. [22] Theorem 3.5 page 371), the SDE (4.4.1) has a unique strong solution. It is not hard to see that the solution remains non negative for all $t > 0$. The only delicate question is whether or not the point 0 can be reached in finite time (whether it is *accessible* or *inaccessible*).

Theorem 4.4.1. *The left endpoint 0 is accessible whenever $\delta < 2$, in particular it is an exit point (absorbing) if $\delta = 0$, instantaneously reflecting if $0 < \delta < 2$; it is inaccessible (and an entrance point) if $\delta \geq 2$.*

PROOF: For $\delta = 1$, X_t is the absolute value of a one–dimensional Brownian motion, hence 0 is accessible. We deduce by comparison that the same is true for $0 \leq \delta \leq 1$. Now the point 0 is polar for the two–dimensional Brownian motion, hence 0 is inaccessible in case $\delta = 2$, hence also by comparison if $\delta \geq 2$. For the other cases, we refer for the proof to [22] p. 423.
AN ALTERNATIVE PROOF OF THE THEOREM. We consider the diffusion $\{Y_t;\ t \geq 0\}$, solution of the SDE

$$Y_t = x + \delta \int_0^t Y_s ds + 2 \int_0^t Y_s dB_s.$$

That SDE has the explicit solution

$$Y_t = x\ \exp\left[(\delta - 2)t + 2B_t\right],\ t \geq 0.$$

We note that

$$\limsup_{t\to\infty} Y_t = \begin{cases} +\infty, & \text{if } \delta \geq 2; \\ 0, & \text{if } \delta < 2. \end{cases}$$

THE CASE $\delta \geq 2$ In that case,

$$A(t) := \int_0^t Y_s ds \to \infty \text{ a. s., as } t \to \infty.$$

Consequently

$$\sigma(t) = \inf\{s;\ A(s) > t\} = A^{-1}(t)$$

is defined for all $t > 0$. Let $X_t := Y_{\sigma(t)}$. It is not hard to see that

$$X_t = x + \delta t + 2 \int_0^t \sqrt{X_s} dB_s,\ X_0 = x. \tag{4.4.2}$$

Indeed, we have

$$Y_t = x + \delta A(t) + 2M_t, \quad \text{with } < M >_t = \int_0^t Y_s^2 ds.$$

Now $X_t = x + \delta t + 2N_t$, where $N_t = M_{\sigma(t)}$ is a martingale from Doob's optional stopping theorem, since the $\sigma(t)$'s are stopping times, and

$$
\begin{aligned}
< N >_t &=< M >_{\sigma(t)} \\
&= \int_0^{\sigma(t)} Y_s^2 ds \\
&= \int_0^t Y_{\sigma(r)}^2 \sigma'(r) dr \\
&= \int_0^t Y_{\sigma(r)} dr \\
&= \int_0^t X_s ds,
\end{aligned}
$$

where we have used the fact that $\sigma'(r) = 1/A'(\sigma(r)) = 1/Y_{\sigma(r)}$. It remains to make use of the martingale representation theorem 7.3.1 in order to conclude (4.4.2).

But since $Y_t > 0$, $\forall t \geq 0$, the same is true for X. Hence 0 is inaccessible. The case $\delta < 2$ Since $B_t/t \to 0$ a. s., as $t \to \infty$, $T(\omega) > 0$ such that $\mathbb{P}$ a. s.,

$$Y_t(\omega) \leq \exp\left[(\delta - 2)t/2\right], \ \forall t \geq T(\omega).$$

As a consequence, $A(t) \to A(\infty)$, with $A(\infty) < \infty$ a. s., as $t \to \infty$. Now $\sigma(t) < \infty$, $\forall t < A(\infty)$, and $\sigma(t) \to \infty$, as $t \to A(\infty)$. Consequently, $X_t = Y_{\sigma(t)} \to 0$, as $t \to A(\infty)$. Consequently $X_{A(\infty)} = 0$, and 0 is accessible. □

4.5 Application to the Wright–Fisher diffusion

Let us first translate the last result for the diffusion

$$
\begin{cases}
dX_t = ct + \sqrt{X_t} dB_t, \\
X_0 = x.
\end{cases}
\tag{4.5.1}
$$

We can transform (4.5.1) into (4.4.1) either by time change, or by a change of spatial scale. At any rate, the result translates to the fact that the solution of

(4.5.1) never hits zero iff $c \geq 1/2$. Consider first the Wright–Fisher diffusion without mutation nor selection

$$dX_t = \sqrt{X_t(1-X_t)}dB_t.$$

Clearly it behaves near 0 as the solution of (4.5.1) with $c = 0$, and similarly near 1, i. e. both end points are accessible and absorbing. The Wright–Fisher diffusion with selection

$$dX_t = \beta X_t(1-X_t)dt + \sqrt{X_t(1-X_t)}dB_t$$

has the same behaviour, since it is obtained from the previous one by an absolute continuous Girsanov transformation (one can also use a comparison argument). The fact that the two end points are absorbing follows easily from pathwise uniqueness, which is a consequence a Theorem 4.1.1. Consider finally the Wright–Fisher diffusion with mutation

$$dX_t = \gamma_0(1-X_t)dt - \gamma_1 X_t dt + \sqrt{X_t(1-X_t)}dB_t, \ t \geq 0.$$

0 is inaccessible iff $\gamma_0 \geq 1/2$, 1 is inaccessible iff $\gamma_1 \geq 1/2$. Neither point is absorbing, unless the corresponding γ vanishes.

4.6 Probability of fixation and time to fixation

We consider here the Wright–Fisher diffusion without mutations, with or without selection.

It is easy to deduce from the Markov property that for all $0 < x < 1$, $\mathbb{P}_x(\tau_{0,1} < \infty) = 1$. As we have said, both 0 and 1 are absorbing, hence $X_t = X_{t\wedge\tau_{0,1}}$. We are interested in the probability of the event that the allele A gets fixed, i. e. that $\tau_1 < \tau_0 = \infty$. Define

$$u(x) = \mathbb{P}_x(\tau_1 < \infty).$$

Applying the semigroup of the Wright–Fisher diffusion to the function u, we deduce

$$\begin{aligned} P_t u(x) &= \mathbb{E}_x\left[\mathbb{P}_{X_t}(\tau_1 < \infty)\right] \\ &= \mathbb{E}_x\left[\mathbb{P}_x(\tau_1 < \infty|\mathcal{F}_t)\right] \\ &= \mathbb{P}_x(\tau_1 < \infty) \\ &= u(x). \end{aligned}$$

From this we conclude that u belongs to the domain of the generator L of the diffusion X, and that $Lu(x) \equiv 0$.
Consider first the case $\beta = 0$. In that case, $Lu = 0$ implies $u'' = 0$. Since $u(0) = 0$ and $u(1) = 1$, we deduce that $u(x) = x$. Note that the same result can be deduced from the remark that $\{X_t,\ t \geq 0\}$ is a bounded martingale, hence

$$\mathbb{P}_x(\tau_1 < \infty) = \mathbb{E}_x[X_{\tau_{0,1}}] = x.$$

We can also compute the expected time until there is fixation of either allele. Denote by $x \to v(x)$ the solution of the boundary value problem

$$\begin{cases} Lv(x) + 1 = 0,\ 0 < x < 1; \\ \qquad v(0) = v(1) = 0. \end{cases}$$

Applying Itô's formula to develop $v(X_{t\wedge\tau_{0,1}})$, taking the expectation and letting $t \to \infty$, we deduce that $v(x) = \mathbb{E}_x[\tau_{0,1}]$. The unique solution of the above equation is given as follows

$$\begin{aligned} v''(x) &= -\frac{2}{x(1-x)} \\ &= -\frac{2}{1-x} - \frac{2}{x} \\ v'(x) &= 2\log\frac{1}{x} - 2\log\frac{1}{1-x} \\ v(x) &= 2(1-x)\log\frac{1}{1-x} + 2x\log\frac{1}{x}. \end{aligned}$$

Consider now the case $\beta > 0$. The identity $Lu(x) = 0$ implies that

$$\left(e^{2\beta x}u'\right)'(x) = 0,\ 0 < x < 1.$$

This, together with $u(0) = 0$, $u(1) = 1$ implies that

$$\mathbb{P}_x(\tau_1 < \infty) = u(x) = \frac{1 - e^{-2\beta x}}{1 - e^{-2\beta}}.$$

We note that this quantity is increasing both in x and in β, as it should be, and that it converges to x as $\beta \to 0$.

4.7 More on the Wright–Fisher diffusion

Consider the Wright–Fisher diffusion without mutation nor selection, i. e.

$$dX_t = \sqrt{X_t(1-X_t)}dB_t, \quad t \geq 0.$$

All the invariant probability measures of that Markov process are the measures $\{\lambda\delta_0 + (1-\lambda)\delta_1;\ 0 \leq \lambda \leq 1\}$.

Consider now the same process with values in the state space $E = (0,1)$, where the process is considered to be killed when it reaches either 0 or 1. Clearly that process is transient, it cannot possibly possess an invariant probability measure. However the measure

$$\frac{dx}{x(1-x)}, \quad \text{on } (0,1)$$

is an invariant σ–finite measure, and the process is in fact reversible with respect to that invariant σ–finite measure. Indeed, the infinitesimal generator of the Wright–Fisher diffusion on $(0,1)$ is $(\mathcal{D}(L), L)$, where the set

$$\mathcal{A} = \{f \in C^2(0,1) \cap C[0,1], \quad f(0) = f(1) = 0\}$$

is dense is $\mathcal{D}(L)$ and for $f \in \mathcal{A}$,

$$Lf(x) = \frac{1}{2}x(1-x)f''(x).$$

For any $f, g \in \mathcal{A}$, we have

$$\begin{aligned}\int_0^1 f(x)(Lg)(x)\frac{dx}{x(1-x)} &= \frac{1}{2}\int_0^1 f(x)g''(x)dx \\ &= -\int_0^1 f'(x)g'(x)dx \\ &= \int_0^1 g(x)(Lf)(x)\frac{dx}{x(1-x)}\end{aligned}$$

If we call $\mu(dx) = \frac{dx}{x(1-x)}$, we have just shown that the oprerator is a self–adjoint unbounded negative operator on $L^2((0,1);\mu(dx))$.

4.8 Conditioning

We are interested in conditioning the Wright–Fisher diffusion upon fixation of the allele A, that is we want to study $\{X_t,\ t \geq 0\}$, under the law

$$\mathbb{P}_x(\ \cdot\ |\tau_1 < \infty).$$

Recall the notation $u(x) = \mathbb{P}_x(\tau_1 < \infty)$.

Theorem 4.8.1. *The process* $\{u(X_t),\ t \geq 0\}$ *is a positive bounded martingale and*

$$\mathbb{P}_x(A|\tau_1 < \infty) = \mathbb{E}_x\left(\frac{u(X_t)}{u(x)}; A\right),\ x \in (0,1), A \in \mathcal{F}_t.$$

Moreover, the generator L^* *of the process* X *conditonned upon* $\{\tau_1 < \infty\}$, *acts as follows on* $f \in C^2(0,1)$ *:*

$$L^* f(x) = \frac{L(uf)(x)}{u(x)},\ x \in (0,1).$$

Proof: The martingale property of $\{u(X_t),\ t \geq 0\}$ follows readily from the fact that $Lu \equiv 0$. Let $f \in C^2(0,1)$.

$$\begin{aligned} P_t^* f(x) &= \frac{\mathbb{E}_x\left[f(X_t); \{\tau_1 < \infty\}\right]}{\mathbb{P}_x(\tau_1 < \infty)} \\ &= \frac{\mathbb{E}_x\left[f(X_t)u(X_t)\right]}{u(x)} \\ &= \frac{P_t(uf)(x)}{u(x)}. \end{aligned}$$

The first statement follows upon replacing $f(X_t)$ by $\mathbf{1}_A$, with $A \in \mathcal{F}_t$. The second follows by subtracting $f(x)$, dividing by t and letting $t \to 0$. □

Consider now the Wright–Fisher diffusion with selection, i. e.

$$dX_t = \beta X_t(1 - X_t)dt + \sqrt{X_t(1 - X_t)}dB_t,\ X_0 = x.$$

If follows from the formula for $u(x)$ in the last section that the generator of that diffusion, conditioned upon fixation of the advantageous allele, is

$$L^* f(x) = \beta x(1 - x)\coth(\beta x)f'(x) + \frac{1}{2}x(1 - x)f''(x).$$

The corresponding SDE reads

$$dX_t = \beta X_t(1 - X_t)\coth(\beta X_t)dt + \sqrt{X_t(1 - X_t)}dB_t,\ X_0 = x.$$

Since $\lim_{x\to 0}\beta x(1 - x)\coth(\beta x) = 1$, it follows from Theorem 4.4.1 that 0 is an entrance boundary for that SDE. While 0 is an exit point for the unconditioned Wright–Fisher diffusion, it is an entrance point for the Wright–Fisher diffusion conditioned upon fixation.

Note that the limit as $\beta \to 0$ of the last SDE is

$$dX_t = (1 - X_t)dt + \sqrt{X_t(1 - X_t)}dB_t,\ X_0 = x,$$

which coincides with the Wright–Fisher diffusion without selection, conditioned upon fixation.

4.9 Invariant measure

Consider the Wright–Fisher diffusion with selection and mutation, i. e. the solution of the SDE

$$dX_t = [\beta X_t(1 - X_t) + \gamma(1 - 2X_t)]dt + \sqrt{X_t(1 - X_t)}dB_t,\ t \geq 0.$$

This diffusion is irreducible, in the sense that starting from any point $x \in [0, 1]$, any interval $I \subset [0, 1]$ with $|I| > 0$, all $t > 0$, $\mathbb{P}_x(X_t \in I) > 0$. Since moreover the process $\{X_t,\ t \geq 0\}$ is a homogeneous Markov process with values in a compact set, it has a unique invariant probability distribution, which has the density

$$Kx^{\gamma-1}(1 - x)^{\theta-1}e^{\beta x},\ x \in [0, 1],$$

whre K is a normalizing constant.

4.10 Bibliographical comments

Section 4.1 follows mainly the treatment in [13]. Section 4.8 follows [16].

Chapter 5

The coalescent

Consider again the discrete Wright–Fisher model, but this time we consider that each of the N individuals of the initial generation is of a different type. The process describing the types of the individuals of the successive generations (i. e. who was their grand–grand father in the initial generation) is a again a Markov chain, which reaches eventually one of the N absorbing states consisting of all the individuals being the grand–grand children of the same initial ancestor.

Looking backward in time, if we sample n individuals in the present population, we want to describe at which generation any two of those had the same common ancestor, until we reach the most recent common ancestor of the sample.

5.1 Cannings' model

We can generalize the Wright–Fisher model as follows. Suppose at each generation, we label the N individuals randomly. For $r \geq 0$, $1 \leq i \leq N$, let ν_i^r denote the number of offsprings in generation $r+1$ of the i–th individual from generation r. Clearly those r. v.'s must satisfy the requirement that

$$\nu_1^r + \cdots + \nu_N^r = N.$$

Cannings' model stipulates moreover that

$$\nu^r, \; r \geq 0 \;\text{ are i. i. d. copies of } \nu,$$

and that the law of ν is exchangeable, i. e.

$$(\nu_1, \dots, \nu_N) \simeq (\nu_{\pi(1)}, \dots, \nu_{\pi(N)}), \forall \pi \in S_N.$$

The above conditions imply that $\mathbb{E}\nu_1 = 1$. To avoid the trivial case where $\mathbb{P}(\nu_1 = \cdots = \nu_N = 1) = 1$, we assume that $\mathrm{Var}(\nu_1) > 0$. A particular case of Cannings' model is the Wright–Fisher model, in which ν is multinomial.

5.2 Looking backward in time

Consider a population of fixed size N, which has been reproducing for ever according to Cannings' model. We sample $n < N$ individuals from the present generation, and label them $1, 2, \dots, n$. For each $r \geq 0$, we introduce the equivalence relation on the set $\{1, \dots, n\}$: $i \sim_r j$ if the individuals i and j have the same ancestor r generations back in the past. Denote this equivalence relation by $R_r^{N,n}$. For $r \geq 0$, $R_r^{N,n}$ is a random equivalence relation, which can be described by its associated equivalence classes, which is a random partition of $(1, \dots, n)$. Thus $\{R_r^{N,n};\ r \geq 0\}$ is a Markov chains with values in the set $\mathcal{E}_n$ of the partitions of $(1, \dots, n)$, which starts from the trivial *finest* partition $(\{1\}, \dots \{n\})$, and eventually reaches the *coarsest* partition consisting of the set $\{1, \dots, n\}$ alone. We denote by $P_{\xi,\eta}^{N,n}$ the transition matrix of that chain.

The probability that two individuals in today's population have the same ancestor in the previous generation is

$$c_N = \frac{\sum_{i=1}^N \mathbb{E}\left[\binom{\nu_i}{2}\right]}{\binom{N}{2}} = \frac{\sum_{i=1}^N \mathbb{E}[\nu_i(\nu_i - 1)]}{N(N-1)} = \frac{\mathbb{E}[\nu_1(\nu_1 - 1)]}{N-1}.$$

Provided that $c_N \to 0$ as $N \to \infty$, if $r = t/c_N$,

$$\mathbb{P}(1 \not\sim_r 2) = (1 - c_N)^r \approx e^{-t}.$$

This suggests to consider

$$\mathcal{R}_t^{N,n} := R_{[t/c_N]}^{N,n},\ t \geq 0.$$

5.3 Kingman's coalescent

Let $\{\mathcal{R}_t^n;\ t \geq 0\}$ be a continuous time $\mathcal{E}_n$–valued Markov chain with the rate matrix given by (for $\eta \neq \xi$)

$$Q_{\xi\eta} = \begin{cases} 1 & \text{, if } \eta \text{ is obtained from } \xi \text{ by merging exactly two classes,} \\ 0 & \text{, otherwise.} \end{cases} \tag{5.3.1}$$

This is Kingman's n coalescent. In order for $\mathcal{R}^{N,n}$ to converge to Kingman's coalescent, we certainly need that merges of 3 or more lineages are asymptotically negligible. The probability that three individuals in today's population have the same ancestor in the previous generation is

$$d_N := \frac{\sum_{i=1}^N \mathbb{E}\left[\binom{\nu_i}{3}\right]}{\binom{N}{3}} = \frac{\mathbb{E}[\nu_1(\nu_1-1)(\nu_1-2)]}{(N-1)(N-2)}.$$

Theorem 5.3.1. *$\mathcal{R}^{N,n} \Rightarrow \mathcal{R}^n$ in $D(\mathbb{R}_+;\mathcal{E}_n)$ iff, as $N \to \infty$, both*

$$\begin{cases} c_N \to 0, \\ \dfrac{d_N}{c_N} \to 0. \end{cases} \tag{5.3.2}$$

Proof: The sufficiency will follow from the standard Lemma 5.3.2 below and the fact that (5.3.2) implies that

$$P_{\xi,\eta}^{N,n} = \delta_{\xi,\eta} + c_N Q_{\xi,\eta} + \circ(c_N),$$

where the error term is small, uniformly with respect to ξ, $\eta \in \mathcal{E}_n$. It follows from exchangeability that for any $f : \{0,1,\ldots,N\} \to \mathbb{R}_+$,

$$\begin{aligned}(N-1)\mathbb{E}[\nu_2 f(\nu_1)] &= \sum_{j=2}^N \mathbb{E}[\nu_j f(\nu_1)] \\ &= \mathbb{E}[(N-\nu_1)f(\nu_1)] \\ &\leq N\mathbb{E}[f(\nu_1)],\end{aligned}$$

hence

$$\mathbb{E}[\nu_2 f(\nu_1)] \leq \frac{N}{N-1}\mathbb{E}[f(\nu_1)]. \tag{5.3.3}$$

From the Markov inequality and (5.3.2), with the notations $(\nu)_2 = \nu(\nu-1)$, $(\nu)_3 = \nu(\nu-1)(\nu-2)$, if $\varepsilon N \geq 2$,

$$\begin{aligned}\mathbb{P}(\nu_1 > \varepsilon N) &\leq \frac{\mathbb{E}[(\nu_1)_3]}{(\varepsilon N)_3}\\ &= \frac{\circ(N\mathbb{E}[(\nu_1)_2])}{\varepsilon^3 N^3},\end{aligned}$$

consequently

$$\mathbb{P}(\nu_1 > \varepsilon N) \leq \varepsilon^{-3} \circ (c_N/N). \tag{5.3.4}$$

Next

$$\begin{aligned}\mathbb{E}[(\nu_1)_2(\nu_2)_2] &\leq \varepsilon N \mathbb{E}[(\nu_1)_2\nu_2; \nu_2 \leq \varepsilon N] + N^2\mathbb{E}[(\nu_1)_2; \nu_2 > \varepsilon N]\\ &\leq \varepsilon N \mathbb{E}[(\nu_1)_2\nu_2] + N^3\mathbb{E}[\nu_1; \nu_2 > \varepsilon N]\\ &\leq \varepsilon N \frac{N}{N-1}\mathbb{E}[(\nu_1)_2] + N^3\frac{N}{N-1}\mathbb{P}(\nu_2 > \varepsilon N),\end{aligned}$$

where we have used (5.3.3) twice in the last inequality. Combining this with (5.3.4), we conclude that for all $\varepsilon > 0$,

$$\begin{aligned}\limsup_{N\to\infty} \frac{\mathbb{E}[(\nu_1)_2(\nu_2)_2]}{N\mathbb{E}[(\nu_1)_2]} &\leq \varepsilon + \limsup_{N\to\infty} \frac{\mathbb{P}(\nu_1 > \varepsilon N)}{c_N/N}\\ &= \varepsilon.\end{aligned}$$

Let $I_1, \ldots, I_n$ denote the parents of n ordered randomly chosen individuals of a given generation. We have the following identities

$$\begin{aligned}\mathbb{P}(I_1 = I_2) &= c_N\\ \mathbb{P}(I_1 = I_2 = I_3) &= d_N\\ \mathbb{P}(I_1 = I_2 \neq I_3 = I_4) &= \frac{\sum_{1\leq i<j\leq N} \mathbb{E}\left[\binom{\nu_i}{2}\binom{\nu_j}{2}\right]}{\binom{N}{4}}\\ &= 3\frac{\mathbb{E}[(\nu_1)_2(\nu_2)_2]}{(N-2)(N-3)}.\end{aligned}$$

Hence we deduce from the last estimate that

$$\lim_{N\to\infty} \frac{\mathbb{P}(I_1 = I_2 \neq I_3 = I_4)}{\mathbb{P}(I_1 = I_2)} = 0, \tag{5.3.5}$$

while (5.3.2) tells us that

$$\lim_{N\to\infty} \frac{\mathbb{P}(I_1 = I_2 = I_3)}{\mathbb{P}(I_1 = I_2)} = 0. \tag{5.3.6}$$

We now conclude, using (5.3.5) and (5.3.6). Let $\xi = (C_{11}, C_{12}, C_2, \dots, C_a)$ and $\eta = (C_1, C_2, \dots, C_a)$, where $C_1 = C_{11} \cup C_{12}$. We have

$$\begin{aligned}\mathbb{P}(I_1 = I_2) &- \mathbb{P}(\{I_1 = I_2\} \cap \{\exists 3 \le m \le a+1;\ I_m = I_1\}) \\ &- \mathbb{P}(\{I_1 = I_2\} \cap \{\exists 3 \le \ell < m \le a+1;\ I_\ell = I_m \ne I_1\}) \\ &\le P^{N,n}_{\xi,\eta} \le \mathbb{P}(I_1 = I_2).\end{aligned}$$

From (5.3.6),

$$\begin{aligned}\mathbb{P}(\{I_1 = I_2\} \cap \{\exists 3 \le m \le a+1;\ I_m = I_1\}) &\le (a-1)\mathbb{P}(I_1 = I_2 = I_3) \\ &= \circ(\mathbb{P}(I_1 = I_2)),\end{aligned}$$

and from (5.3.5),

$$\begin{aligned}\mathbb{P}(\{I_1 = I_2\} \cap \{\exists 3 \le \ell < m \le a+1;\ L_\ell = I_m \ne I_1\}) &\le \binom{a-1}{2} \mathbb{P}(I_1 = I_2 \ne I_3 = I_4) \\ &= \circ(\mathbb{P}(I_1 = I_2))\end{aligned}$$

We have proved that for such a pair (ξ, η), $P^{N,n}_{\xi,\eta} = c_N + \circ(c_N)$. If η' is obtained from ξ by merging more than two classes, then there must be at least either a triple merger or two double mergers, hence from (5.3.6), (5.3.5), $P^{N,n}_{\xi,\eta'} = \circ(c_N)$. Finally, since $|\mathcal{E}_n| < \infty$ and $\sum_{\eta \in \mathcal{E}_n} P^{N,n}_{\xi,\eta} = 1$,

$$\begin{aligned}P^{N,n}_{\xi,\xi} &= 1 - \binom{|\xi|}{2} c_N + \circ(c_N) \\ &= 1 - Q_{\xi,\xi} c_N + \circ(c_N).\end{aligned}$$

□

Lemma 5.3.2. *Let E be a finite set and $\{X_t,\ t \ge 0\}$ a continuous time E–valued Markov chain, with generator $Q = (Q_{x,y})_{x,y \in E}$. Let for each $N \in \mathbb{N}$ X^N be a discrete time Markov chain with transition matrix satisfying*

$$P^N(x,y) = \delta_{x,y} + c_N Q_{x,y} + \circ(c_N),\ x, y \in E,$$

where $c_N \to 0$, as $N \to \infty$. Then whenever $X^N_0 \Rightarrow X_0$,

$$\{X^N_{[t/c_N]},\ t \ge 0\} \Rightarrow \{X_t,\ t \ge 0\} \quad \text{in } D(\mathbb{R}_+; E).$$

Let $\{\mathcal{R}_t^n;\ t \geq 0\}$ start from the trivial partition of $(1, \ldots, n)$. For $2 \leq k \leq n$, let T_k denote the length of the time interval during which there are k branches alive. From the Markov property of the coalescent, and the form of the generator, we deduce that

$$\begin{aligned} &T_n, T_{n-1}, \ldots, T_2 \text{ are independent,} \\ &T_k \simeq \mathcal{E}\text{xp}\left(\binom{k}{2}\right),\ 2 \leq k \leq n, \end{aligned}$$

and consequently the expected time till the Most Recent Common Ancestor in the sample is

$$\begin{aligned} \sum_{k=2}^{n} \frac{2}{k(k-1)} &= 2\sum_{k=2}^{n}\left(\frac{1}{k-1} - \frac{1}{k}\right) \\ &= 2\left(1 - \frac{1}{n}\right). \end{aligned}$$

For $n' > n$, denote by d_n the restriction to $\mathcal{E}_n$ of an element of $\mathcal{E}_{n'}$. Kingman's n–coalescents have the consistency property that

$$d_n\left(\{\mathcal{R}_t^{n'},\ t \geq 0\}\right) \simeq \{\mathcal{R}_t^n,\ t \geq 0\}.$$

This, together with the fact that $\sum_{k\geq 2} T_k < \infty$ a. s., since the series of the expectations converges, allows us to define Kingman's coalescent $\{\mathcal{R}_t,\ t \geq 0\}$ as the limit $\lim_{n\to\infty}\{\mathcal{R}_t^n,\ t \geq 0\}$. It is readily seen that Kingman's coalescent *comes down from infinity*, in the sense that, while $\mathcal{R}_0$ is the trivial partition of $\mathbb{N}^*$, hence $|\mathcal{R}_0| = \infty$, $|\mathcal{R}_t| < \infty$, $\forall t > 0$.

5.3.1 The height and the length of Kingman's coalescent

The *height* of Kingman's n–coalescent is the r. v.

$$H_n = \sum_{k=2}^{n} T_k,$$

where the T_k are as above. This prescribes the law of H_n, which does not obey any simple formula. Note that

$$\mathbb{E}(H_n) = 2\left(1 - \frac{1}{n}\right), \quad \text{Var}(H_n) = \sum_{k=2}^{n} \frac{4}{k^2(k-1)^2}.$$

$\mathbb{E}(H_n) \to 2$ as $n \to \infty$, and $\sup_n \mathrm{Var}(H_n) < \infty$.

The *length* of Kingman's n–coalescent (i. e. the sum of the lengths of the branches of this tree) is the r. v.

$$L_n = \sum_{k=2}^{n} kT_k = \sum_{k=2}^{n} U_k,$$

where the U_k are independent, U_k is an $\mathrm{Exp}((k-1)/2)$ r. v. The distribution function of L_n is given by

Proposition 5.3.3. *For all $x \ge 0$,*

$$\mathbb{P}(L_n \le x) = (1 - e^{-x/2})^{n-1}.$$

This Proposition follows from the fact that the law of L_n is that of the sup over $n-1$ $\mathrm{Exp}(1/2)$ r. v.'s, which is a consequence of the

Proposition 5.3.4. *Let $V_1, V_2, \ldots, V_n$ be i. i. d. Exp(λ) r. v.'s, and $V_{(1)} < V_{(2)} < \cdots < V_{(n)}$ denote the same random sequence, but arranged in increasing order. Then $V_{(1)}, V_{(2)} - V_{(1)}, \ldots, V_{(n)} - V_{(n-1)}$ are independent exponential r. v.'s with respective parameters $n\lambda, (n-1)\lambda, \ldots, \lambda$.*

Proof: For any Borel measurable function $f : \mathbb{R}_+^n \to \mathbb{R}_+$,

$$\begin{aligned}
&\mathbb{E}f(V_{(1)}, V_{(2)} - V_{(1)}, \ldots, V_{(n)} - V_{(n-1)}) \\
&\quad = n!\mathbb{E}\left[f(V_1, V_2 - V_1, \ldots, V_n - V_{n-1}); V_1 < V_2 < \cdots < V_n\right] \\
&\quad = n! \int_{0<x_1<x_2<\cdots<x_n} f(x_1, x_2 - x_1, \ldots, x_n - x_{n-1}) \lambda^n e^{-\lambda \sum_{k=1}^{n} x_k} dx_1 dx_2 \cdots dx_n \\
&\quad = \prod_{k=1}^{n} (k\lambda) \int_0^\infty \cdots \int_0^\infty f(y_1, y_2, \ldots, y_n) \prod_{k=1}^{n} e^{-k\lambda y_{n+1-k}} dy_1 dy_2 \cdots dy_n.
\end{aligned}$$

The result follows. □

5.4 The speed at which Kingman's coalescent comes down from infinity

Consider Kingman's coalescent $\{\mathcal{R}_t,\ t \ge 0\}$ starting from the trivial partition of $\mathbb{N}^*$. Let $R_t = |\mathcal{R}_t|$, $t \ge 0$. Let $\{T_n,\ n \ge 2\}$ be a sequence of independent

r. v.'s, the law of T_n being the exponential law with parameter $\binom{n}{2}$. We let

$$S_n = \sum_{k=n+1}^{\infty} T_k, \quad n \geq 1.$$

Now the process $\{R_t,\ t \geq 0\}$ can be represented as follows.

$$R_t = \sum_{n=1}^{\infty} n \mathbf{1}_{\{S_n \leq t < S_{n-1}\}}.$$

We know that $R_t \to \infty$, as $t \to 0$. We state two results, which give a precise information, as to the speed at which R_t diverges, as $t \to 0$. We first state a strong law of large numbers

Theorem 5.4.1. *As $t \to 0$,*

$$\frac{tR_t}{2} \to 1 \quad a.\ s.$$

We next have a central limit theorem

Theorem 5.4.2. *As $t \to 0$,*

$$\sqrt{\frac{6}{t}} \left(\frac{tR_t}{2} - 1 \right) \Rightarrow N(0,1).$$

Remark 5.4.3. *As we will see in the proof, the behaviour of R_t as $t \to 0$ is intimately connected to the behaviour of S_n, as $n \to \infty$. But while in the classical asymptotic results of probability theory we add more and more random variable as $n \to \infty$, here as n increases, S_n is the sum of less and less random variables (but always an infinite number of those).*

5.4.1 Proof of the strong law of large numbers

We first need to compute some moments of S_n.

Lemma 5.4.4. *We have*

$$\mathbb{E}(S_n) = \frac{2}{n} \tag{5.4.1}$$

$$Var(S_n) = \sum_{k=n}^{\infty} \frac{4}{k^2(k+1)^2} \tag{5.4.2}$$

$$\mathbb{E}\left(|S_n - \mathbb{E}S_n|^4\right) \leq \frac{c}{n^6}, \tag{5.4.3}$$

where c is a universal constant. Moreover

$$n^3\, Var(S_n) \to \frac{4}{3}, \quad as\ n \to \infty. \tag{5.4.4}$$

PROOF: (5.4.1) follows readily from

$$\begin{aligned} \mathbb{E}(S_n) &= \sum_{k=n+1}^{\infty} \frac{2}{k(k-1)} \\ &= \sum_{k=n}^{\infty} \left(\frac{2}{k} - \frac{2}{k+1} \right) \end{aligned}$$

Similarly

$$\begin{aligned} \mathrm{Var}(S_n) &= \sum_{k=n+1}^{\infty} \mathrm{Var}(T_k) \\ &= \sum_{k=n}^{\infty} \frac{4}{k^2(k+1)^2} \end{aligned}$$

This proves (5.4.2). Now (5.4.4) follows from

$$\begin{aligned} \frac{4}{3n^3} = \int_n^{\infty} \frac{4}{x^4} dx \leq \sum_n^{\infty} \frac{4}{(k+1)^4} &\leq \mathrm{Var}(S_n) \\ &\leq \sum_n^{\infty} \frac{4}{k^4} \leq \int_{n-1}^{\infty} \frac{4}{x^4} dx = \frac{4}{3(n-1)^3}. \end{aligned}$$

We finally prove (5.4.3). Note that $\mathbb{E}(|T_k - \mathbb{E}T_k|^4) = 2^4/k^4(k-1)^4$. Moreover

$$\begin{aligned} \mathbb{E}\left(|S_n - \mathbb{E}S_n|^4\right) &= \mathbb{E} \sum_{k=n+1}^{\infty} |T_k - \mathbb{E}T_k|^4 + 6\mathbb{E} \sum_{n<k<\ell} |T_k - \mathbb{E}T_k|^2 |T_\ell - \mathbb{E}T_\ell|^2 \\ &= \sum_{k=n}^{\infty} \frac{2^4}{k^4(k+1)^4} + 4 \times 4! \sum_{n \leq k < \ell} \frac{1}{k^2(k+1)^2 \ell^2 (\ell+1)^2} \\ &\leq \frac{2^4}{7(n-1)^7} + \frac{4 \times 4!}{3^2(n-1)^6}. \end{aligned}$$

□

Theorem 5.4.1 will follow from

Proposition 5.4.5. *As* $n \to \infty$,

$$\frac{S_n}{\mathbb{E}S_n} \to 1 \quad a.\ s.$$

Proof: The result follows from Borel–Cantelli's lemma and the next estimate, where we make use of (5.4.3) and (5.4.1)

$$\mathbb{E}\left(\left|\frac{S_n - \mathbb{E}S_n}{\mathbb{E}S_n}\right|^4\right) \le c\frac{n^4}{n^6} = cn^{-2}.$$

□

Proof of Theorem 5.4.1 All we need to show is that for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left\{\limsup_{t\to 0}\left|\frac{tR_t}{2} - 1\right| > \varepsilon\right\}\right) = 0.$$

But

$$\left\{\limsup_{t\to 0}\left|\frac{tR_t}{2} - 1\right| > \varepsilon\right\} \subset \limsup_{n\to\infty} A_n,$$

where

$$A_n = \left\{\sup_{S_n \le t < S_{n-1}}\left|\frac{tn}{2} - 1\right| > \varepsilon\right\}$$

Now

$$\begin{aligned} A_n &\subset \left\{\left|\frac{nS_n}{2} - 1\right| > \varepsilon\right\} \cup \left\{\left|\frac{nS_{n-1}}{2} - 1\right| > \varepsilon\right\} \\ &\subset \left\{\left|\frac{nS_n}{2} - 1\right| > \varepsilon\right\} \cup \left\{\left|\frac{(n-1)S_{n-1}}{2} - 1\right| > \varepsilon/2\right\}, \end{aligned}$$

as soon as $(\varepsilon + 1)/n \le \varepsilon/2$. But it follows from Proposition 5.4.5 that

$$\mathbb{P}(\limsup_n A_n) = 0.$$

□

5.4.2 Proof of the central limit theorem

Define for each $n \geq 1$ the r. v.

$$Z_n = \sqrt{3n}\frac{S_n - \mathbb{E}S_n}{\mathbb{E}S_n}.$$

Let us admit for a moment the

Proposition 5.4.6. *As $n \to \infty$,*

$$Z_n \Rightarrow N(0,1).$$

PROOF OF THEOREM 5.4.2 Define, for all $t > 0$,

$$\tau(t) = \inf\{0 < s \leq t;\ R_s = R_t\}.$$

Proposition 5.4.6 tells us that, as $t \to 0$,

$$\sqrt{3R_t}\left(\frac{\tau(t)R_t}{2} - 1\right) \Rightarrow N(0,1).$$

Combining with Theorem 5.4.1, we deduce that

$$\sqrt{\frac{6}{t}}\left(\frac{\tau(t)R_t}{2} - 1\right) \Rightarrow N(0,1).$$

It remains to show that

$$\frac{t - \tau(t)}{\sqrt{t}}R_t \to 0 \quad \text{a. s. as } t \to 0.$$

From Theorem 5.4.1, this is equivalent to

$$\frac{t - \tau(t)}{t^{3/2}} \to 0 \quad \text{a. s. as } t \to 0.$$

But

$$\limsup_{t \to 0} \frac{t - \tau(t)}{t^{3/2}} \leq \limsup_{n \to \infty} \frac{T_n}{S_n^{3/2}},$$

and from Proposition 5.4.5, the right hand side goes to zero if and only if $n^{3/2}T_n \to 0$ as $n \to \infty$. We have that $\mathbb{E}(|n^{3/2}T_n|^4) \leq cn^{-2}$, hence from

Bienaymé–Tchebychef and Borel–Cantelli, $n^{3/2}T_n \to 0$ a. s. as $n \to \infty$, and the theorem is proved. □

We finally give the

PROOF OF PROPOSITION 5.4.6 Let φ_n denote the characteristic function of the r. v. Z_n. If we let $c_n = \sqrt{3n}$, $a_n = \sqrt{3n^3}$, we have $Z_n = -c_n + a_n S_n/2$, hence

$$\begin{aligned}
\varphi_n(t) &= e^{-itc_n} \prod_{k=n+1}^{\infty} \mathbb{E}\left[e^{-ita_n T_k/2}\right] \\
&= e^{-itc_n} \prod_{k=n+1}^{\infty} \left(1 - \frac{ita_n}{k(k-1)}\right)^{-1} \\
&= e^{-itc_n} \exp\left\{ \sum_{k=n+1}^{\infty} \log\left(1 + i\frac{ta_n}{k(k-1)} - t^2 \frac{a_n^2}{k^2(k-1)^2} + 0(a_n^3 k^{-6})\right)\right\} \\
&= e^{-itc_n} \exp\left\{ \sum_{k=n+1}^{\infty} \left(i\frac{ta_n}{k(k-1)} - \frac{t^2}{2}\frac{a_n^2}{k^2(k-1)^2} + 0(a_n^3 k^{-6})\right)\right\} \\
&= e^{-itc_n} e^{ita_n/n} \exp\left(-\frac{t^2}{2} \sum_{k=n+1}^{\infty} \frac{3n^3}{k^2(k-1)^2} + 0(n^{-1/2})\right) \\
&\to \exp\left(-t^2/2\right),
\end{aligned}$$

where we have used again the argument leading to (5.4.4). The result follows.

5.5 Duality between Kingman's coalescent and Wright–Fisher's diffusion

We associate to Kingman's coalescent again the process $\{R_t,\ t \geq 0\}$ defined by $R_t = |\mathcal{R}_t|$. $\{R_t,\ t \geq 0\}$ is a pure death process on $\mathbb{N}^*$, with transition from n to $n-1$ happening at rate $\binom{n}{2}$. Consider moreover $\{X_t,\ t \geq 0\}$ a Wright–Fisher diffusion, i. e. the solution of the SDE

$$dX_t = \sqrt{X_t(1-X_t)}dB_t,\ t \geq 0; \quad X_0 = x,$$

where $0 < x < 1$.

Proposition 5.5.1. *The following duality relation holds*

$$\mathbb{E}[X_t^n | X_0 = x] = \mathbb{E}[x^{R_t} | R_0 = n], \; t \geq 0. \tag{5.5.1}$$

PROOF: We fix $n \geq 1$. Define

$$u(t,x) = \mathbb{E}[x^{R_t} | R_0 = n].$$

Since $\{R_t; \; t \geq 0\}$ is a Markov process with gnerator Q defined by

$$Qf(n) = \frac{n(n-1)}{2}[f(n-1) - f(n)],$$

for any $f : \mathbb{N} \to \mathbb{R}_+$,

$$\mathcal{N}_t^f = f(R_t) - f(R_0) - \int_0^t \binom{R_s}{2} [f(R_s - 1) - f(R_s)] ds$$

is a martingale. Let us explicit the above identity for the particular choice $f(n) = x^n$:

$$\begin{aligned} x^{R_t} &= x^n + \int_0^t \frac{R_s(R_s-1)}{2}[x^{R_s-1} - x^{R_s}]ds + \mathcal{N}_t \\ &= x^n + \frac{x(1-x)}{2} \int_0^t R_s(R_s-1)x^{R_s-2} ds + \mathcal{N}_t. \end{aligned}$$

Writing that $\mathbb{E}[\mathcal{N}_t | R_0 = n] = 0$, we deduce that for each $n \in \mathbb{N}$,

$$u(t,x) = u(0,x) + \frac{x(1-x)}{2} \int_0^t \frac{\partial^2 u}{\partial x^2}(s,x) ds.$$

This means that u solves the following linear parabolic PDE

$$\begin{cases} \dfrac{\partial u}{\partial t}(t,x) = \dfrac{x(1-x)}{2} \dfrac{\partial^2 u}{\partial x^2}(t,x) \; t \geq 0, 0 < x < 1; \\ u(0,x) = x^n, \; u(t,0) = 0, \; u(t,1) = 1 \end{cases}$$

It is easily checked that $x \to u(t,x)$ is smooth. We then may apply Itô's calculus to develop $u(t-s, X_s)$, which yields, since u solves the above PDE,

$$u(0, X_t) = u(t,x) + M_t,$$

where M_t is a zero–mean martingale. Taking the expectation in the last identity yields $u(t,x) = \mathbb{E}_x[X_t^n]$. □

We deduce from the above a simple proof of the uniqueness in law of the solution of the Wright–Fisher SDE.

Corollary 5.5.2. *The law of the solution* $\{X_t,\ t \geq 0\}$ *of Wright–Fisher SDE is unique.*

Proof: Since the solution is a homogeneous Markov process, it suffices to show that the transition probabilities are uniquely determined. But for all $t > 0$, $x \in [0, 1]$, the conditional law of X_t, given that $X_0 = x$ is determined by its moments, since X_t is a bounded r. v. The result then follows from Proposition 5.5.1.

Remark 5.5.3. *As* t *gets large, both terms of the identity* (5.5.1) *tend to* x*. The left hand side because it behaves for* t *large as* $\mathbb{P}(X_t = 1) \to x$*, and the right hand side since* $\mathbb{P}(R_t = 1) \to 1$*, as* $t \to \infty$*.*

5.6 The Ewens sampling formula

5.6.1 Coalescence with mutations : the infinite many alleles model

Suppose now that mutations arise on each branch of the coalescence tree, according to a Poisson process with parameter $\theta/2$, see Figure 5.1. Assume that each mutation gives birth to a new type, different for all the others. For instance we may assume that the different types are i. i. d. r. v.'s following the uniform law on $[0, 1]$. We want to record the different types in a sample drawn at present time, we can as well "kill" the lineages which hit a mutation while going backward in time, which changes Figure 5.1 into Figure 5.2, which we can as well change into Figure 5.3. The killed coalescent can be produced by the following procedure : *Any pair of active classes is merged at rate 1, any active class is killed at rate* $\theta/2$. When a class is killed, all its elements are assigned the same (different from all other classes) type. Finish when there are no classes left.

5.6.2 Hoppe's urn

Assume that there are k active classes in the killed coalescent described above. Then the probability that the next (backward in time) event is a

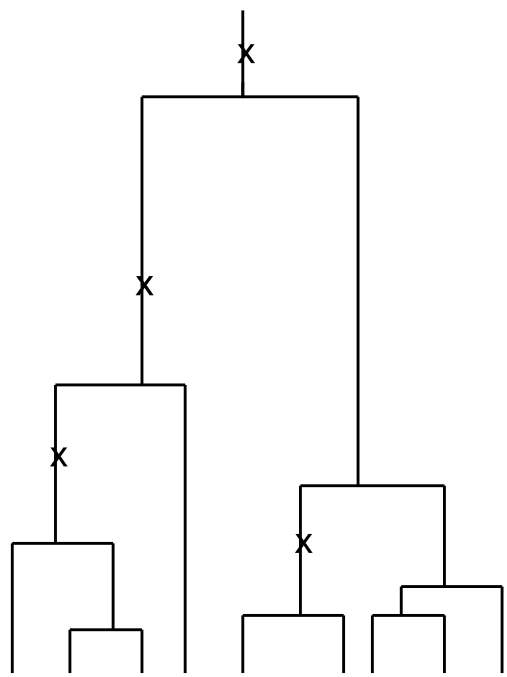

Figure 5.1: The coalescent with mutations

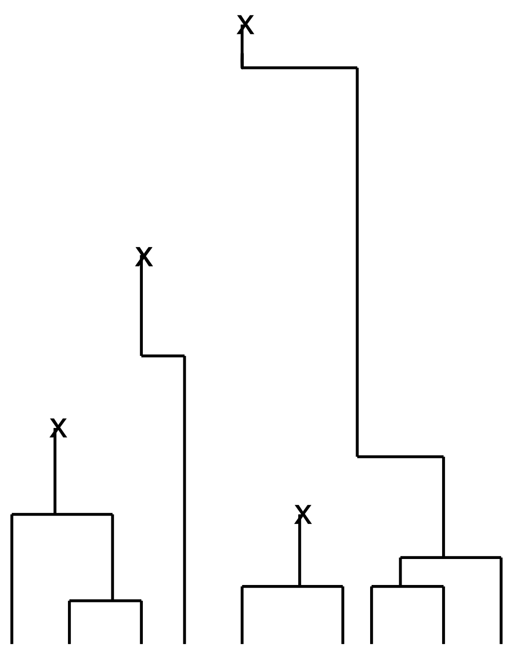

Figure 5.2: The lineages are killed above the mutations

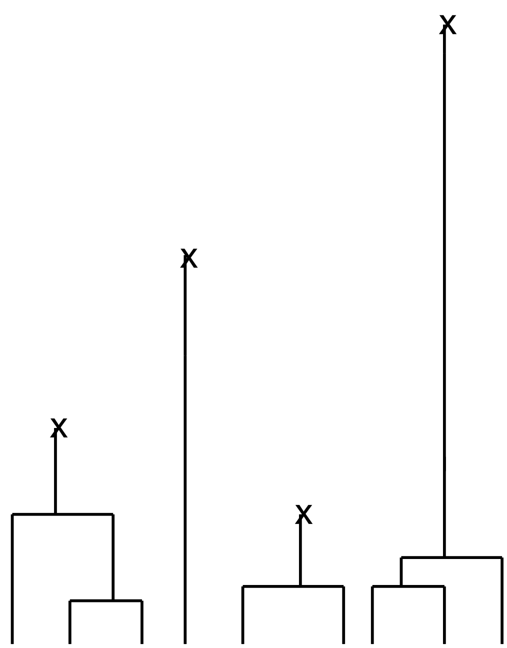

Figure 5.3: Equivalent to Figure 5.2

coalescence is

$$\frac{\binom{k}{2}}{\binom{k}{2} + k\frac{\theta}{2}} = \frac{k-1}{k-1+\theta},$$

and the probability that that event is a mutation (i. e. a killing) is

$$\frac{k\frac{\theta}{2}}{\binom{k}{2} + k\frac{\theta}{2}} = \frac{\theta}{k-1+\theta}.$$

Moreover, given the type of event, all possible coalescence (resp. mutations) are equally likely. The history of a sample of size n is described by n events $e_n, e_{n-1}, \dots, e_1 \in \{\texttt{coal}, \texttt{mut}\}$. Not that the event e_k happens just before (forward in time) k lineages are active, and each of those events corresponds backward in time to the reduction by one of the number of active lineages. The probability to observe a particular sequence is thus

$$\frac{\prod_{k=1}^{n} \left(\theta \mathbf{1}_{\{e_k=\texttt{mut}\}} + (k-1)\mathbf{1}_{\{e_k=\texttt{coal}\}}\right)}{\prod_{k=1}^{n}(k-1+\theta)}. \tag{5.6.1}$$

Hoppe [12] noted that one can generate this sequence *forward in time* using the following urn model.

Hoppe's urn model. We start with an urn containing one unique black ball of mass θ. At each step, a ball is drawn from the urn, with probability proportional to its mass. If the drawn ball is black return it to the urn, together with a ball of mass 1, of a new, not previously used, colour; if the drawn ball is coloured, return it together with another ball of mass 1 of the same colour.

At the k–th step, there are k balls, more precisely $k-1$ coloured balls, plus the black (so called *mutation*) ball. The probability to pick the black ball is thus $\theta/(k-1+\theta)$ while the probablity to pick a coloured ball is $(k-1)/(k-1+\theta)$. If we define

$$e_k = \begin{cases} \texttt{mut}, & \text{if in the } k\text{–step the black ball is drawn,} \\ \texttt{coal}, & \text{otherwise.} \end{cases}$$

Clearly the probability to observe a particular sequence $(e_1, \dots, e_n)$ is given by (5.6.1). Moreover, given that $e_k = \texttt{coal}$, each of the $k-1$ present coloured balls is equally likely to be picked.

Consequently, the distribution of the family sizes generated by the n coloured balls in Hoppe's urn after n steps is the same as the one induced by the n–coalescent in the infinitely–many–alleles mutation model.

Define K_n to be the number of different types oberved in a sample of size n, or equivalently the number of different colours in Hoppe's urn after n steps. Then

$$K_n = X_1 + \cdots + X_n,$$

where

$$X_k = \mathbf{1}_{A_k}, \quad A_k = \{\text{the black ball is drawn at the } k\text{–th step}\},$$

consequently the events $A_1, \ldots, A_n$ are independent, with $\mathbb{P}(A_k) = \theta/(\theta + k - 1)$, $1 \le k \le n$. Consequently

$$\begin{aligned} \mathbb{E}K_n &= \sum_{i=1}^{n} \frac{\theta}{\theta + i - 1} \simeq \theta \log(n), \\ \mathrm{Var}(K_n) &= \sum_{i=1}^{n} \frac{\theta}{\theta + i - 1} \cdot \frac{i-1}{\theta + i - 1} \simeq \theta \log(n), \\ \frac{K_n - \mathbb{E}K_n}{\sqrt{\mathrm{Var}(K_n)}} &\Rightarrow N(0,1), \quad \text{as } n \to \infty. \end{aligned}$$

5.6.3 Ewens' sampling formula

Theorem 5.6.1. *Let $b_1, \ldots, b_n \in \mathbb{N}$ be such that $\sum_{j=1}^{n} j b_j = n$. The probability to observe b_j different types, each with j representatives, ($j = 1, \ldots, n$) in a sample of size n is given by (here $k = \sum_{j=1}^{n} b_j$)*

$$\frac{n!}{1^{b_1} 2^{b_2} \cdots n^{b_n}} \cdot \frac{1}{b_1! b_2! \cdots b_N!} \cdot \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)}. \tag{5.6.2}$$

Remark 5.6.2. *An alternative way to write Ewens' formula is*

$$C(n,\theta) \times \prod_{j=1}^{n} e^{-\theta/j} \frac{(\theta/j)^{b_j}}{b_j!}, \tag{5.6.3}$$

with

$$C(n,\theta) = \frac{n! \exp[\theta \sum_{j=1}^{n} 1/j]}{\theta(\theta+1)\cdots(\theta+n-1)}.$$

Hence the distribution of the type spectrum $(B_1, \dots, B_n)$ in a sample of size n is the product of the measures Poi(θ/j), $j = 1, \dots, n$, conditionned on $\sum_{j=1}^n jB_j = n$. For the justification of that statement, see Remark 5.6.3 below.

Proof: We consider an n–coalescent, in which the sampled individuals are artificially labelled $1, 2, \dots, n$. Mutations happen at rate $\theta/2$ on each lineage. Once a lineage has hit a mutation, it is killed, and we do not follow it anymore backward in type. We consider what we call a *protocol*, which is a sequence $e_1, e_2, \dots, e_n$ of n succesive elementary events, which each will be either of type

$$\texttt{mut}(i), \text{ i. e. lineage } i \text{ hits a mutation event, or}$$
$$\texttt{coal}(i \to j), \text{ i. e. lineage } i \text{ coalesces into lineage } j \ (\neq i)$$

for some i (and possibly some j) $\in \{1, \dots, n\}$. Here we keep track of whom coalesces into whom. Both elementary events $\texttt{mut}(i)$ and $\texttt{coal}(i \to j)$ make i inactive, hence that lineage cannot appear in another elementary event further in the protocol, which clearly must satisfy that consistency condition.

The coalescence rate is $1/2$ per ordered pair of alive lineages, and the mutation rate is $\theta/2$ per alive lineage. Before the m–th elementary event, there are $n-m+1$ active lineages so the probability of observing a particular e_m equals

$$\frac{1/2}{(n-m)(n-m+1)/2 + \theta(n-m+1)/2} = \frac{1}{(n-m+1)(n-m+\theta)}, \quad \text{if } e_m \text{ is a coalescence,}$$

$$\frac{\theta/2}{(n-m)(n-m+1)/2 + \theta(n-m+1)/2} = \frac{\theta}{(n-m+1)(n-m+\theta)}, \quad \text{if } e_m \text{ is a mutation.}$$

Hence the probability of observing a given consistent protocol which contains $k(\leq n)$ mutation events, which means that it describes a possible history of a sample which contains k different types, is

$$\frac{\theta^k}{\prod_{m=1}^n (n-m+1)(n-m+\theta)} = \frac{\theta^k}{n!\theta(\theta+1)\cdots(\theta+n-1)}. \tag{5.6.4}$$

For a given type spectrum $b_1, \dots, b_n$ with $\sum_{j=1}^n b_j = k$, and of course $\sum_{j=1}^n jb_j = n$, we need to compute how many consitent protocols yield exactly that type spectrum.

Assume that there are k artificially labelled types. Consider the corresponding family sizes $n_\ell \in \mathbb{N}$ for $\ell = 1, \ldots, k$ such that $\sharp\{\ell,\ n_\ell = j\} = b_j$, $j = 1, 2, \ldots, n$. We now use a more precise description of our protocol, in the sense that we write $\mathtt{mut}_\ell(i)$ for the elementary event which consists in the fact hat the lineage i hits a mutation which produces the new type ℓ. This enriched protocol must satisfy the new consistency condition that each of the types from the set $\{1, \ldots, k\}$ appears exactly in one mutation event.

One way to describe an enriched protocol which produces a sample with n_ℓ representatives of type ℓ, $\ell = 1, \ldots, k$ is as follows :

1. Fix the order in which the lineages become inactive. There are $n!$ possibilities.

2. Assign a type to each lineage. This amounts to putting n distinguishable balls (the lineages) into k distinguishable boxes (the types), such that n_ℓ balls land in box ℓ, $\ell = 1, \ldots, k$. There are $\frac{n!}{n_1!n_2!\cdots n_k!}$ possibilities.

3. Prescribe the coalescence program within each type. Consider type ℓ. Assume that in step 2 we have decided that the lineages $i_1, i_2, \ldots, i_{n_\ell}$ are of that type, and that step 1 prescribes that i_1 is lost first, i_2 is lost second, etc. We still are free to choose into which of the $n_\ell - 1$ other lineages the lineage i_1 coalesces, into which of the $n_\ell - 2$ other lineages the lineage i_2 coalesces, etc, leading to $(n_\ell - 1)!$ possible choices, this for each type, so that alltogether the number of possible choices at this step is
$$(n_1 - 1)! \times (n_2 - 1)! \times \cdots \times (n_k - 1)!.$$

Combining the number of possible choices at each step yields

$$n! \times \frac{n!}{n_1!n_2!\cdots n_k!} \times (n_1-1)! \times (n_2-1)! \times \cdots \times (n_k-1)! = \frac{(n!)^2}{\prod_{\ell=1}^k n_\ell} = \frac{(n!)^2}{\prod_{j=1}^n j^{b_j}}$$

different enriched protocols with labelled types, which produce n_ℓ representatives of type ℓ $(\ell = 1, \ldots, k)$ in the sample.

From an enriched protocol, we deduce a standard one by ignoring the labels. Now $b_1! \cdot b_2! \cdots b_n!$ labelled protocols correspond to the same unlabelled one. Consequently, we find that there are

$$\frac{(n!)^2}{\prod_{j=1}^n (b_j! j^{b_j})} \tag{5.6.5}$$

different protocols. Multiplying (5.6.4) with (5.6.5) yields (5.6.2). □

Remark 5.6.3. *We want to justify the statement in Remark 5.6.2. All we need to show is that is $B_1, \dots, B_n$ are independent, each B_j is Poisson with parameter θ/j, then*

$$\mathbb{P}\left(\sum_{j=1}^{n} jB_j = n\right) = \frac{\theta(\theta+1)\cdots(\theta+n-1)}{n!\exp[\theta\sum_{j=1}^{n} 1/j]}.$$

But the left hand side of the above equals

$$\sum_{k_1,\dots,k_n;\sum jk_j=n} e^{-\theta/j}\,(\theta/j)^{k_j}/k_j! = \exp[-\theta\sum_{j=1}^{n} 1/j]\sum_k \alpha(n,k)\theta^k,$$

where

$$\alpha(n,k) = \sum_{k_1,\dots,k_n;\sum k_j=k,\ \sum jk_j=n}\left(\prod_{j=1}^{n} j^{k_j}k_j!\right)^{-1}.$$

It remains to show that

$$\theta(\theta+1)\cdots(\theta+n-1) = n!\sum_{k=1}^{n}\alpha(n,k)\theta^k.$$

Let $s(n,k) = n!\alpha(n,k)$. Splitting the last term in the above left hand side into θ plus $n-1$, we deduce that

$$s(n,k) = s(n-1,k-1) + (n-1)s(n-1,k).$$

This shows that $s(n,k)$ can be interpreted as the number of permutations of $\{1,\dots,n\}$ which contain exactly k cycles. Now that number is given by

$$\begin{aligned} s(n,k) &= \sum_{k_1,\dots,k_n,\sum k_j=k,\sum jk_j=n}\frac{n!}{\prod_{j=1}^{n}(jk_j)!}\times\prod_{j=1}^{n}\left(\frac{(jk_j)!}{(j!)^{k_j}}\frac{1}{k_j!}\left[(j-1)!\right]^{k_j}\right) \\ &= n!\sum_{k_1,\dots,k_n,\sum k_j=k,\sum jk_j=n}\prod_{j=1}^{n}\frac{1}{j^{k_j}k_j!}. \end{aligned}$$

Indeed in the above formula,

$$\frac{n!}{\prod_{j=1}^{n}(jk_j)!}$$

is the number of possibilities of choosing the elements for the cycles of size j*,* j *varying from* 1 *to* n*,*

$$\frac{(jk_j)!}{(j!)^{k_j}}\frac{1}{k_j!}$$

is the number of ways in which one can distribute the jk_j *elements in the* k_j *cycles of size* j*, and*

$$[(j-1)!]^{k_j}$$

is the number of different possible orderings of the elements in the k_j *cycles of size* j*.*

5.7 Bibliographical comments

This chapter is mainly inspired by the presentation of the same material in [6].

Chapter 6

The look–down approach to Wright–Fisher and Fleming–Viot models

6.1 Introduction

Here we shall first define an alternative to the Wright–Fisher model, namely the continuous–time Moran model. We shall then present the look–down construction due to Donnelly and Kurtz [9] (see also [10]), and show that this particular version of the Moran model converges a. s., as the population size N tends to infinity, towards the Wright–Fisher diffusion.

6.2 The Moran model

Consider a population of fixed size N, which evolves in continuous time according to the following rule. For each ordered pair (i, j) with $1 \leq i \neq j \leq N$, at rate $1/2N$ individual i gives birth to an individual who replaces individual j, independently of the other ordered pairs. This can be graphically represented as follows. For each ordered pair (i, j) we draw arrows from i to j at rate $1/2N$. If we denote by $\mathcal{P}$ the set of ordered pairs of elements of the set $\{1, \dots, N\}$, μ the counting measure on $\mathcal{P}$, and λ the Lebesgue measure on $\mathbb{R}_+$, the arrows constitute a Poisson process on $\mathcal{P} \times \mathbb{R}_+$ with intensity measure $(2N)^{-1}\mu \times \lambda$.

Consider the Harris diagram for the Moran model in Figure 6.1. Time flows down. If we follow the diagram backward from the bottom to the top, and coalesce any pair of individuals whenever they find a common ancestor, we see that starting from the trivial partition

$$\{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\}\},$$

after the first arrow has been reached we get

$$\{\{1\},\{2\},\{3,4\},\{5\},\{6\},\{7\},\{8\},\{9\}\},$$

next

$$\{\{1\},\{2\},\{3,4\},\{5\},\{6,9\},\{7\},\{8\}\},$$

next

$$\{1,3,4\},\{2\},\{5\},\{6,9\},\{7\},\{8\}\},$$

the fourth arrow does not modify the partition, next

$$\{\{1,3,4,5\},\{2\},\{6,9\},\{7\},\{8\}\},$$

next

$$\{\{1,3,4,5\},\{2\},\{6,7,9\},\{8\}\},$$

the next arrow has no effect, then

$$\{\{1,3,4,5,8\},\{2\},\{6,7,9\}\}$$

and the last arrow (the first on from the top) has no effect.

It is not hard to see that the coalescent which is imbedded in the Moran model looked at backward in time is exactly Kingman's coalescent – here more precisely Kingman's N–coalescent.

Suppose now that as in the preceding chapter the population includes two types of individuals, type a and type A. Each offspring is of the same type as his parent, we do not consider mutations so far. Denote

$$Y_t^N = \text{ number of type } A \text{ individuals at time } t.$$

Provided we specify the initial number of type A indviduals, the above model completely specifies the law of $\{Y_t^N,\ t \geq 0\}$. We now introduce the *proportion* of type A individuals in rescaled time, namey

$$X_t^N = N^{-1} Y_{Nt}^N, \quad t \geq 0.$$

Note that in this new time scale, the above Poisson process has the intensity measure $\mu \times \lambda$. We have, similarly as in Theorem 2.1.1,

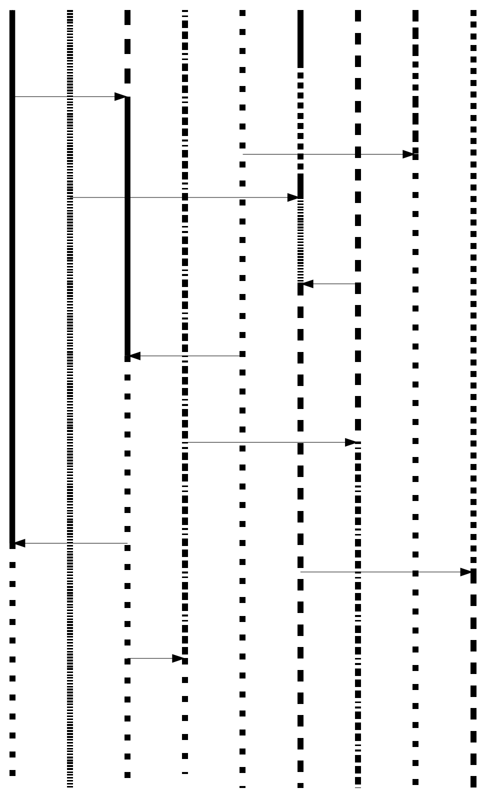

Figure 6.1: The Moran model

Theorem 6.2.1. *Suppose that* $X_0^N \Rightarrow X_0$, *as* $N \to \infty$. *Then* $X^N \Rightarrow X$ *in* $D(\mathbb{R}_+; [0,1])$, *where* $\{X_t,\ t \geq 0\}$ *solves the SDE*

$$dX_t = \sqrt{X_t(1-X_t)}dB_t,\ t \geq 0.$$

PROOF: As for Theorem 2.1.1, the proof goes through two steps.
PROOF OF TIGHTNESS One needs to show that $\mathbb{E}[|X_t^N - X_s^N|^4] \leq c(t-s)^2$.
IDENTIFICATION OF THE LIMIT Note that the process $\{Z_t^N := Y_{Nt}^N,\ t \geq 0\}$ is a jump Markov process with values in the finite set $\{0, 1, 2, \dots, N\}$, which, when in state k, jumps to

1. $k-1$ at rate $k(N-k)/2$,
2. $k+1$ at rate $k(N-k)/2$.

In other words if Q^N denotes the infinitesimal generator of this process,

$$Q^N f(Z_t^N) = Z_t^N(N - Z_t^N)\left[\frac{f(Z_t^N+1) + f(Z_t^N-1)}{2} - f(Z_t^N)\right].$$

In other words,

$$\begin{aligned}\mathbb{E}\left[f(X_{t+\Delta t}^N) - f(X_t^N)|X_t^N = x\right] &= N^2x(1-x)\left[\frac{f\left(x+\frac{1}{N}\right) + f\left(x-\frac{1}{N}\right)}{2} - f(x)\right]\Delta t + o(\Delta t)\\ &= \frac{x(1-x)}{2}f''(x)\Delta t + o(\Delta t),\end{aligned}$$

since from two applications of the order two Taylor expansion,

$$\frac{f\left(x+\frac{1}{N}\right) + f\left(x-\frac{1}{N}\right)}{2} - f(x) = \frac{1}{2N^2}f''(x) + o(N^{-2}).$$

□

6.3 The look–down construction

The construction which we are going to present here is often called in the literature the *modified* look–down construction.

Let us again consider first the case where the size N of the population is finite and fixed. We redraw the Harris diagram of Moran's model, forbidding

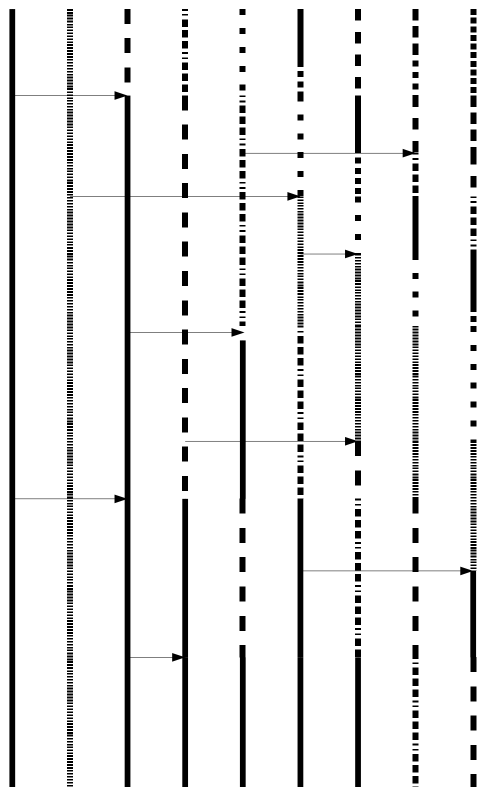

Figure 6.2: The look–down construction

half of the arrows. We consider only arrows from left to right. Considering immediately the rescaled time, for each $1 \leq i < j \leq N$, we put arrows from i to j at rate 1 (twice the above 1/2). At such an arrow, the individual at level i puts a child at level j. Individuals previously at levels $j, \ldots, N-1$ are shifted one level up; individual at site N dies.

Note that in this construction the level one individual is immortal, and the genealogy is not exchangeable.

However the partition at time t induced by the ancestors at time 0 is exchangeable, since going back each pair coalesces at rate 1.

Consider now the case where there are two types of individuals, type a, represented by black, and type A, represented by red. We want to choose the types of the N individuals at time 0 in an exchangeable way, with the constraint that the proportion of type red individuals is given. One possibility is to draw whithout replacement N balls from an urn where we have put k red balls and $N-k$ black balls. At each draw, each of the balls which remain in the urn has the same probability of being chosen.

It follows from the above considerations that at each time $t > 0$, the types of the N individuals are exchangeable.

6.4 a. s. convergence to the Wright–Fisher diffusion

Our goal now is to take the limit in the above quantities as $N \to \infty$. The look–down construction can be defined directly with an infinite population. The description is the same as above, except that we start with an infinite number of lines, and no individual dies any more.

Note that the possibility of doing the same construction for $N = \infty$ is related to the fact that in any finite interval of time, if we restrict ourselves to the first N individuals, the evolution is determined by finitely many arrows. This would not be the case with the standard Moran model, which could not be described in the case $N = \infty$. Indeed in the Moran model with infinitely many individuals, there would be infinitely many arrows towards any individual i, in any time interval of positive length. This is a great advantage of the look–down construction.

Consider now the case of two types of individuals. Suppose that the initial colours of the various individuals at time $t = 0$ are i. i. d., each red with probability x, black with probability $1 - x$. Define

$$\eta_t(k) = \begin{cases} 1, & \text{if the } k\text{–th individual is red at time } t; \\ 0, & \text{if the } k\text{–th individual is black at time } t. \end{cases}$$

$\{\eta_0(k),\ k \geq 1\}$ are i. i. d. Bernoulli random variables, while at each $t > 0$, $\{\eta_t(k),\ k \geq 1\}$ is an exchangeable sequence of $\{0,1\}$–valued random variables. A celebrated theorem due to de Finetti (see Corollary 7.5.6 below) says that an exchangeable sequence of $\{0,1\}$–valued r. v. is a mixture of i. i. d. Bernoulli. Consequently the following limit exists a. s.

$$X_t = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \eta_i(t) = \lim_{N\to\infty} X_t^N. \tag{6.4.1}$$

Before stating the main theorem of this section, let us establish three auxiliary results which we shall need in its proof.

Proposition 6.4.1. *Let* $\{\xi_1, \xi_2, \dots\}$ *be a countable exchangeable sequence of* $\{0,1\}$*–valued r. v.'s and* $\mathcal{T}$ *denote its tail* σ*–field. Let* $\mathcal{H}$ *be some additional* σ*–algebra. If conditionally upon* $\mathcal{T} \vee \mathcal{H}$*, the r. v.'s are exchangeable, then conditionally upon* $\mathcal{T} \vee \mathcal{H}$ *they are i. i. d.*

Proof: Let $f : \{0,1\}^n \to \mathbb{R}$ be an arbitrary mapping. It follows from the assumption that

$$\begin{aligned}\mathbb{E}\left(f(\xi_1, \dots, \xi_n)|\mathcal{T} \vee \mathcal{H}\right) &= \mathbb{E}\left(N^{-1}\sum_{k=1}^{N} f(\xi_{(k-1)n+1}, \dots, \xi_{kn})\Big|\mathcal{T} \vee \mathcal{H}\right) \\ &= \mathbb{E}\left(f(\xi_1, \dots, \xi_n)|\mathcal{T}\right),\end{aligned}$$

where the second equality follows from the fact that the quantity inside the previous conditional expectation converges a. s. to $\mathbb{E}\left(f(\xi_1, \dots, \xi_n)|\mathcal{T}\right)$ as $N \to \infty$, as a consequence of exchangeability and de Finetti's theorem (see Corollary 7.5.6 below). □

Lemma 6.4.2. *Let* $\{X_n,\ n \geq 1\}$ *and* X *be real–valued random variables such that* $X_n \to X$ *a. s. as* $n \to \infty$*, and* $A, B \in \mathcal{F}$*. If*

$$\mathbb{P}(A \cap C) = \mathbb{P}(B \cap C), \quad \forall C \in \sigma(X_n),\ \forall n \geq 1,$$

then

$$\mathbb{P}(A \cap C) = \mathbb{P}(B \cap C), \quad \forall C \in \sigma(X).$$

Proof: The assumption implies that for all $f \in C_b(\mathbb{R})$, all $n \geq 1$,

$$\mathbb{E}[f(X_n); A] = \mathbb{E}[f(X_n); B],$$

from which we deduce by bounded convergence that

$$\mathbb{E}[f(X); A] = \mathbb{E}[f(X); B].$$

The result follows. □

Let S_n denote the group of permutations of $\{1, 2, \dots, n\}$. If $\pi \in S_n$, $a \in \{0,1\}^n$, we shall write $\pi^*(a) = (a_{\pi(1)}, \dots, a_{\pi(n)})$. Recall that a partition $\mathcal{P}$ of $\{1, \dots, n\}$ induces an equivalence relation, whose equivalence classes are the blocks of the partition. Hence we shall write $i \simeq_{\mathcal{P}} j$ whenever i and j are in the same block of $\mathcal{P}$. Finally we write $\#\mathcal{P}$ for the number of blocks of the partition $\mathcal{P}$.

Proposition 6.4.3. *For all $n \geq 1$, $0 < r < s$, $a \in \{0,1\}^n$, p such that $0 \leq np \leq n$ is an integer, $\pi \in S_n$,*

$$\mathbb{P}(\{\eta_s^n = a\} \cap \{X_r^n = p\}) = \mathbb{P}(\{\eta_s^n = \pi^*(a)\} \cap \{X_r^n = p\}).$$

PROOF: Denote by $\mathcal{P}_a$ the set of partitions $\mathcal{P}$ of $\{1, 2, \dots, n\}$ which are such that $i \simeq_{\mathcal{P}} j \Rightarrow a_i = a_j$.

The arrows between time r and time s in the look–down construction pointing to levels between 2 and n prescribe in particular which individuals at time s have the same ancestor back at time r. This corresponds to a partition $\{1, 2, \dots, n\}$ which is the result of the coalescent process backward from time s to time r. $\{\text{coal}_s^r = \mathcal{P}\}$ is the event that that partition is $\mathcal{P}$. Suppose that $\#\mathcal{P} = k$. The look–down construction prescribes that the block containing 1 carries the type of the individual siting on level 1 at time r, the block containing the smallest level not in that first block carries the type of the individual sitting on level 2 at time r, ... Thus the event

$$\{\eta_s^n = a\} \cap \{\text{coal}_s^r = \mathcal{P}\},$$

which is non empty iff $\mathcal{P} \in \mathcal{P}_a$, determines the values of $\eta_r^n(1), \dots, \eta_r^n(k)$ if $k = \#\mathcal{P}$. There is a finite (possibly zero) number of possible values b for η_r^n which respect both the above condition and the restriction $n^{-1}\sum_{i=1}^n b_i = p$. We denote by $\mathcal{A}_{r,s}(a, \mathcal{P}, p) \subset \{0,1\}^n$ the set of those b's. Note that this set is empty if the restriction $n^{-1}\sum_{i=1}^n b_i = p$ contradicts the conditions $\eta_s^n = a$ and $\text{coal}_s^r = \mathcal{P}$.

We then have

$$\{\eta_s^n = a\} \cap \{X_r^n = p\} = \bigcup_{\mathcal{P} \in \mathcal{P}_a} \bigcup_{b \in \mathcal{A}(a,\mathcal{P},p)} \{\text{coal}_s^r = \mathcal{P}\} \cap \{\eta_r^n = b\},$$

and from the independence of coal_s^r and η_r^n

$$\mathbb{P}(\eta_s^n = a, X_r^n = p) = \sum_{\mathcal{P} \in \mathcal{P}_a} \sum_{b \in \mathcal{A}_{r,s}(a,\mathcal{P},p)} \mathbb{P}(\text{coal}_s^r = \mathcal{P})\mathbb{P}(\eta_r^n = b).$$

Similarly, if $\pi^*(\mathcal{P})$ is defined by $i \simeq_{\mathcal{P}} j \Leftrightarrow \pi(i) \simeq_{\pi^*(\mathcal{P})} \pi(j)$,

$$\begin{aligned}\mathbb{P}(\eta_s^n = \pi^*(a), X_r^n = p) &= \sum_{\mathcal{P} \in \mathcal{P}_{\pi^*(a)}} \sum_{b \in \mathcal{A}_{r,s}(\pi^*(a),\mathcal{P},p)} \mathbb{P}(\text{coal}_s^r = \mathcal{P})\mathbb{P}(\eta_r^n = b) \\ &= \sum_{\mathcal{P} \in \mathcal{P}_a} \sum_{b \in \mathcal{A}_{r,s}(\pi^*(a),\pi^*(\mathcal{P}),p)} \mathbb{P}(\text{coal}_s^r = \pi^*(\mathcal{P}))\mathbb{P}(\eta_r^n = b),\end{aligned}$$

We now describe a one–to–one correspondence ρ_π between $\mathcal{A}_{r,s}(a,\mathcal{P},p)$ and $\mathcal{A}_{r,s}(\pi^*(a),\pi^*(\mathcal{P}),p)$. Suppose that $\#\mathcal{P}=k$. Then $\#\pi^*(\mathcal{P})=k$ as well. Let $b\in\mathcal{A}_{r,s}(a,\mathcal{P},p)$. The values of b_j, $1\le j\le k$ are specified by the pair $(a,\mathcal{P})$. We define $b'=\rho_\pi(b)$ as follows. $b'_1,\dots,b'_k$ are specified by $(\pi^*(a),\pi^*(\mathcal{P}))$. Note that the definitions of $\pi^*(a)$ and $\pi^*(\mathcal{P})$ imply that

$$(b'_1,\dots,b'_k)=(b_{\pi'(1)},\dots,b_{\pi'(k)}),\quad \text{for some } \pi'\in S_k.$$

We complete the definition of b' by the conditions

$$b'_j=b_j,\quad k<j\le n.$$

Clearly there exists $\pi''\in S_n$ such that $b'=\pi''^*(b)$.

Consequently

$$\begin{aligned}\sum_{b\in\mathcal{A}_{r,s}(\pi^*(a),\pi^*(\mathcal{P}),p)} &\mathbb{P}(\text{coal}_s^r=\pi^*(\mathcal{P}))\mathbb{P}(\eta_r^n=b)\\ &=\sum_{b\in\mathcal{A}_{r,s}(a,\mathcal{P},p)}\mathbb{P}(\text{coal}_s^r=\pi^*(\mathcal{P}))\mathbb{P}(\eta_r^n=\rho_\pi(b))\\ &=\sum_{b\in\mathcal{A}_{r,s}(a,\mathcal{P},p)}\mathbb{P}(\text{coal}_s^r=\mathcal{P})\mathbb{P}(\eta_r^n=b),\end{aligned}$$

where the last identity follows from the fact that both η_r^n and coal_s^r are exchangeable. The result follows from the three identities proved above. □

We can now prove

Theorem 6.4.4. *The* $[0,1]$*–valued process* $\{X_t,\ t\ge 0\}$ *defined by* (6.4.1) *possesses a continuous modification which is a weak sense solution of the Wright–Fisher SDE, i. e. there exists a standard Brownian motion* $\{B_t,\ t\ge 0\}$ *such that*

$$dX_t=\sqrt{X_t(1-X_t)}dB_t,\ t\ge 0.$$

Proof: Need to add : $\exists$ a continuous modification.

Step 1 We first need to show that $\{X_t,\ t\ge 0\}$ is a Markov process. We know that conditionally upon $X_s=x$, the $\eta_s(k)$ are i. i. d. Bernoulli with parameter x. Now for any $t>s$, X_t depends only upon the $\eta_s(k)$ and the arrows which are drawn between time s and time t, which are independent from $\{X_r,\ 0\le r\le s\}$. So all we need to show is that conditionally upon $\sigma(X_r,\ 0\le r\le s)$, the $\eta_s(k)$ are i. i. d. In view of Proposition 6.4.1, it suffices to prove that conditionally upon $\sigma(X_r,\ 0 \le r \le s)$, the $\eta_s(k)$ are exchangeable. This will follow from the fact that the same is true conditionally upon $\sigma(X_{r_1}, \dots, X_{r_k}, X_s)$ for all $k \ge 1$, $0 \le r_1 < r_2 < \cdots < r_k < s$.

We first prove this in the case $k = 1$. Write $\eta_s^n = (\eta_s(1), \dots, \eta_s(n))$. All we have to show is that for all $n \ge 1$, $a \in \{0,1\}^n$, $\pi \in S_n$, if $\pi^*(a) = (a_{\pi(1)}, \dots, a_{\pi(n)})$, $A_r \in \sigma(X_r)$ and $A_s \in \sigma(X_s)$,

$$\mathbb{P}\left(\{\eta_s^n = a\} \cap A_r \cap A_s\right) = \mathbb{P}\left(\{\eta_s^n = \pi^*(a)\} \cap A_r \cap A_s\right). \tag{6.4.2}$$

In view of Lemma 6.4.2, a sufficient condition for (6.4.2) is that for all $m \ge n$, $p, q > 0$ such that $0 \le mp, mq \le m$ are integers,

$$\mathbb{P}\left(\{\eta_s^n = a\} \cap \{X_r^m = p, X_s^m = q\}\right) = \mathbb{P}\left(\{\eta_s^n = \pi^*(a)\} \cap \{X_r^m = p, X_s^m = q\}\right),$$

and clearly it suffices to prove that result for $n = m$, which is done in Proposition 6.4.3.

A similar proof shows that the $\eta_s(k)$ are conditionally exchangeable given $\sigma(X_{r_1}, \dots, X_{r_k}, X_s)$. The Markov property of the process $\{X_t,\ t \ge 0\}$ is established.

Step 2 It remains to show that the process $\{X_t,\ t \ge 0\}$ has the right transition probability, which will follow (see Proposition 5.5.1) from the fact that for all $n \ge 1$, $x \in [0,1]$,

$$\mathbb{E}_x[X_t^n] = \mathbb{E}_n[x^{R_t}].$$

For all $n \ge 1$,

$$X_t^n = \mathbb{P}(\eta_t(1) = \cdots = \eta_t(n) = 1 | X_t),$$

consequently

$$\begin{aligned}
\mathbb{E}_x[X_t^n] &= \mathbb{E}_x\left[\mathbb{P}(\eta_t(1) = \cdots = \eta_t(n) = 1 | X_t)\right] \\
&= \mathbb{P}_x(\eta_t(1) = \cdots = \eta_t(n) = 1) \\
&= \mathbb{P}_x(\text{the ancestors at time 0 of } 1, \dots, n \text{ are red}) \\
&= \mathbb{E}_n\left[x^{R_t}\right],
\end{aligned}$$

where $\{R_t,\ t \ge 0\}$ is a pure death continuous–time process, which jumps from k to $k-1$ at rate $k(k-1)/2$. □

6.5 The MRCA

In this section, we consider that all individuals are different. Let first ignore mutations. Consider an individual, e. g. the one who was at position k at time 0. As soon as an arrow points to a vertical line indexed by some $1 \le \ell \le k$, this type is pushed to the right. Moreover, the mean number of pushes to the right per time unit is $k(k-1)/2$ for an individual at position k. This number is quadratic in the position. It is easy to deduce that each individual other than the one at position 1 at time 0 is pushed to $+\infty$ in finite time, i. e. out of the population (this is an explosion in terms of ODE). At the time when the individual who occupied initially the position 2 is pushed out, all the individuals are sons of the one who occupied initially position 1. The expectation of the time until the individual initially at position 2 is pushed out equals

$$\sum_{k=2}^{\infty} \frac{2}{k(k-1)} = 2\sum_{k=2}^{\infty} \left(\frac{1}{k-1} - \frac{1}{k} \right) = 2.$$

Note that in the two–types case, the probability that type A eventually gets fixed equals the probability that the individual 1 at time 0 has type A, which is x, as should be. We also get that the mean time to fixation, given that the type A gets fixed, equals

$$2(1-x)\sum_{k=1}^{\infty} \frac{x^{k-1}}{k} = 2\frac{1-x}{x} \log\left(\frac{1}{1-x} \right),$$

since the type of the individuals are i. i. d., all having type A with probability x. Compare with the results of section 4.6.

Let us go back to the situation where all individuals have distinct types. At time S_1, which is exponential with parameter 1, the first individual gives birth to a son who occupies site 2. That son has pushed away all other individuals who were originally in the population by time T_1, which is exponential with expectation 2. In fact $T_1 - S_1$ is exponential with parameter 1, and independent of S_1. At time S_2, which is such that the pair (S_1, S_2) has the same law as (S_1, T_1), the individual at site 1 gives birth to a second child, who pushes the first child and its progeny to the right, and eventually suppresses that subpopulation by time T_2, etc. Not counting the individual at site 1 in the population, we have that at successive times $\{T_k,\ k \ge 1\}$ the

Most Recent Common Ancestor of the whole population changes for a more recently born individual.

Consider the process which describes the age of the MRCA. The trajectory of this process is clearly made of segments of straight lines with slope one, interrupted at the fixation times of the successive sons of the individual located at site 1 by negative jumps, see figure 6.3.

The expectation of the heights of the minima (this is the age of the MRCA at the time of fixation) is 1, while the expectation of the heights of the maxima is 2 (this is the age of a MRCA, just before it gets replaced by a new one). Now if $t > 0$ is arbitrary, what is the age of the current MRCA, and the waiting time until a new one takes over ?

6.6 Mutations, the infinitely many sites model

Suppose now that all individuals are subject to mutations, which arise according to a Poisson process with rate $\theta/2$. We assume that each new mutation hit a new (never hit before) site on the genome. What is the number of mutations which fixate at the time when a new MRCA takes over ? If we call M that random number, we have that

$$\mathbb{P}(M = m) = \left(\frac{\theta}{2+\theta}\right)^m \frac{2}{2+\theta}, \quad m = 0, 1, 2, \ldots$$

Let us go back to the finite size population model (size $= N$), in the case of the Moran model. We ask the question : how many mutations do we expect that are carried by i individuals ? Mutations appear somewhere at rate $N\theta/2$, which means that singleton families are born at rate $N\theta/2$. The size of each family changes according to the Moran dynamics, i. e. it jumps

$$\begin{aligned} &\text{from } i \text{ to } i+1 \text{ at rate } i(N-i)/2, \\ &\text{from } i \text{ to } i-1 \text{ at rate } i(N-i)/2. \end{aligned}$$

The size of the family carrying a given mutation is a $\{1, 2, 3, \ldots, N-1\}$–valued continuous time birth and death process. N is an absorbing state. When the family size reaches that state, the mutation is fixed in the population. We no longer consider it as a mutation. The process describing the family sizes of all the mutations is a continuous time jump Markov process $\{X_t,\ t \geq 0\}$ with values in the state space

$$E = \{\emptyset\} \cup \cup_{n=1}^{\infty} E_n, \quad \text{where } E_n \subset \{1, 2, \ldots, N-1\}^n$$

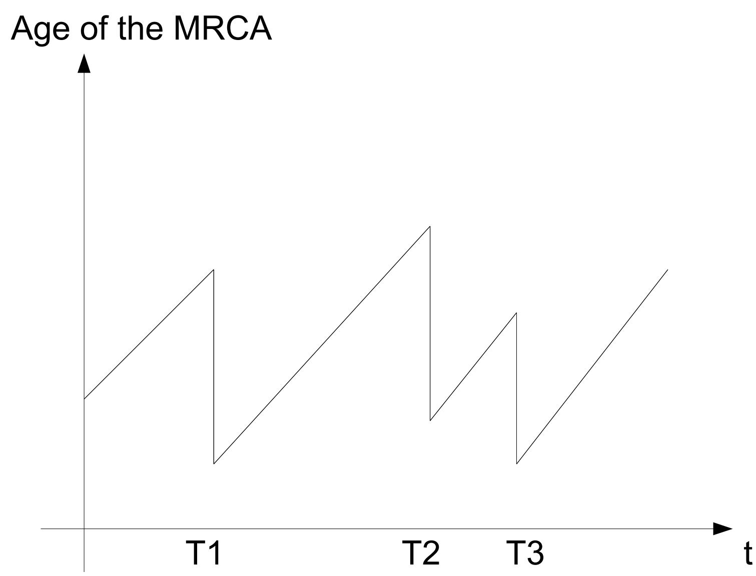

Figure 6.3: The age of the MRCA

is the set of increasing sequences of length n, made of numbers between 1 and $N-1$. The infinitesimal generator of the process $\{X_t\}$ is not easy to describe. The reason is that, while the size of each family changes only by jumps of ± 1, as noted above, the number of families whose size changes at a given jump time is arbitrary (it depends upon the number of mutations carried by the two individuals concerned by the arrow of the Moran model), so that at a jump time of t least one family size, the jumps of the concerned coordinates of X_t can be anything as $0, 1, 2, 3, \ldots$.

For $1 \le i \le N-1$, let $\Phi_i(x) := \#\{k,\ x_k = i\}$. Then $K_i(t) = \Phi_i(X_t)$ denote the number of mutations which are carried by exactly i individuals in the population at time t. For each $1 \le i \le N-1$, an increase (or a decrease) of $K_i(t)$is necessarily compensated by a decrease (or an increase) either of $K_{i-1}(t)$or of $K_{i+1}(t)$ (or of both), such that at a jump time, i. e. a time t where $\sum_j |\Delta K_j(t)| \neq 0$, there exists $1 \le i \le N-1$ such that

$$\Delta K_{i-1}(t) + \Delta K_i(t) + \Delta K_{i+1}(t) = 0.$$

Note that $\Delta K_0(t)$ and $\Delta K_N(t)$ should be interpreted in a very peculiar way. Indeed, 0 and N are sinks where families get lost. Of course, 0 is also a source, in the sense that families of size 1 get birn at rate $\theta/2$.

Write $f_i(t) := \mathbb{E}[K_i(t)]$. From the above considerations follow the fact that

$$\begin{cases} \dfrac{d}{dt} f_1(t) = \dfrac{\theta N}{2} - (N-1) f_1(t) + (N-2) f_2(t) \\ \dfrac{d}{dt} f_i(t) = \dfrac{(i+1)(N-i-1)}{2} f_{i+1}(t) + \dfrac{(i-1)(N-i+1)}{2} f_{i-1}(t) \\ \qquad\qquad - i(N-i) f_i(t), \quad 1 < i < N-1, \\ \dfrac{d}{dt} f_{N-1}(t) = (N-2) f_{N-2}(t) - (N-1) f_{N-1}(t). \end{cases}$$

At equilibrium the time derivatives vanish. If we write $\xi_i = i f_i$, $1 \le i \le N-1$, we deduce from the last of the above identities that $\xi_{N-1} = \xi_{N-2}$, from the middle one that $\xi_{i+1} - \xi_i = \xi_i - \xi_{i-1}$, $1 < i < N-1$, hence the first identity implies that $\xi_1 = \xi_2 = \cdots = \xi_{N-1} = \theta$. In other words

Proposition 6.6.1. *In the Moran model with mutations appearing each at a new site at rate $\theta/2$, at equilibrium the mean number of families of size i is $f_i = \theta/i$, $1 \le i \le N-1$.*

Let us now reconsider this result from the point of view of the coalescent picture. A mutation which is shared by i individuals is a mutation which, in the N–coalescent, appears in a segment which supports exactly i leaves. Since mutations happen at rate $\theta/2$ on each branch of the coalescent tree, we have (see figure 6.4)

Corollary 6.6.2. *The expected total length of edges in the N coalescent supporting exactly i leaves is $2/i$, $1 \leq i \leq N-1$.*

It is remarkable that this quantity does not depend upon N. As N increases, there are more such branches, but they tend to be shorter.

Consider now the case $N = \infty$. Sample n individuals in this population, and ask how many mutations should we expect which are shared by exactly i individuals in the sample, where $1 \leq i \leq n$. Let us emphasize that in the case $i = n$, we consider only those mutations which have not been fixed in the total population.

Proposition 6.6.3. *The expected number of mutations which are shared by exactly i individuals in the sample (and which in particular are not fixed in the total infinite population), equals θ/i, $1 \leq i \leq n$.*

First proof of Proposition 6.6.3 If $i \leq n-1$, the result is clearly that given by Proposition 6.6.1 (just let the population size N in this Proposition be the present sample size n).We only have to establish the result for $i = n$. In that case, we look for the expected number of mutations which appear in the coalescent tree above the MRCA of the n sample, but below the MRCA of the infinite population. The expected height of the n coalescent is $2(1-1/n)$, while the expected height of the infinite coalescent is 2. Hence the expected length of segment supporting the n leaves of the sample, and not supporting the full infinite coalescent is $2/n$. Since mutations arise at rate $\theta/2$, the expected number of mutations which are shared by the whole sample, but not fixed in the total population, is θ/n. □

Exercise 6.6.4. Second proof of Proposition 6.6.3 *Prove the result of corollary 6.6.2 directly, by computing $\mathbb{E}[L_{n,i}]$, where $L_{n,i}$ denotes the total length of all those edges which support exactly i leaves, in a tree with n leaves, by recurrence on both n and i. Deduce the result of Proposition 6.6.3.*

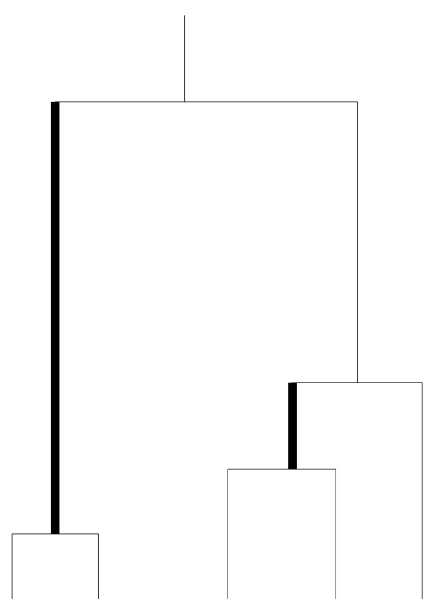

Figure 6.4: The edges supporting exactly 2 leaves

6.7 Expected number of families of a given size in the infinite population model

The aim of this section is to show that when N is infinite, the density of the expected number of families of relative size x, $0 < x < 1$, is $f(x) = \theta/x$. We will then deduce Proposition 6.6.3 as a consequence of this result.

Suppose first that N is very large, and that a mutation occurs at time $t = 0$. What is the fate of that mutation at a given time $t > 0$? The probability that the relative size of that family belongs to $A \in \mathcal{B}[0,1]$ is

$$\mathbb{P}_{1/N}(X_t^N \in A),$$

where $\{X_t^N,\ t \geq 0\}$ stands for the continuous–time solution of the Moran model, which is well approximated by the Wright–Fisher diffusion. Note that by time homogeneity of whatever model we use, $\forall A \in \mathcal{B}([0,1])$,

$$\mathbb{P}_{-t,1/N}(X_0^N \in A) = \mathbb{P}_{1/N}(X_t^N \in A).$$

If mutations have appeared for ever at the constant rate $\theta/2$ on each lineage, then the expected number of families of relative size in $A \in \mathcal{B}[0,1]$ at time $t = 0$ is given by

$$\frac{\theta N}{2}\int_{-\infty}^{0} \mathbb{P}_{t,1/N}(X_0^N \in A)dt = \frac{\theta N}{2}\int_{0}^{\infty} \mathbb{P}_{1/N}(X_t^N \in A)dt$$

The main result of this section is

Theorem 6.7.1. *Let $\{X_t^N,\ t \geq 0\}$ be the solution of Moran model. Then $\forall A \in \mathcal{B}(0,1)$, as $N \to \infty$,*

$$\frac{\theta N}{2}\int_{0}^{\infty} \mathbb{P}_{1/N}(X_t^N \in A)dt \to \int_A f(x)dx,$$

where

$$f(x) = \frac{\theta}{x}, \quad x \in [0,1].$$

Proof: In order to make our method of proof more transparent, we first do the

PROOF OF THE RESULT WITH X^N REPLACED BY ITS LIMIT X Let us compute the limit of

$$N\int_0^\infty \mathbb{P}_{1/N}(X_t \in A)dt = N\mathbb{E}_{1/N}\int_0^{\tau_{0,1}} \mathbf{1}_A(X_t)dt, \qquad (6.7.1)$$

where $\{X_t,\ t \geq 0\}$ denotes the solution of the Wright–Fisher SDE and $\tau_{0,1} := \inf\{t > 0, X_t \in \{0,1\}\}$. From the occupation times formula (see Proposition 4.2.5), if

$$\varphi(x) = [x(1-x)]^{-1}\mathbf{1}_A(x),$$

we have

$$\begin{aligned}\int_0^t \mathbf{1}_A(X_s)ds &= \int_0^t \varphi(X_s)d\langle X, X\rangle_s \\ &= \int_A L_t^y \frac{dy}{y(1-y)}.\end{aligned}$$

In other words,

$$\mathbb{E}_x\int_0^{\tau_{0,1}} \mathbf{1}_A(X_t)dtdt = \int_A \mathbb{E}\left(L^y_{\tau_{0,1}}\right)\frac{dy}{y(1-y)}. \qquad (6.7.2)$$

It remains to compute the expectation of the local time. From the very definition of the local time of the diffusion X,

$$|X_t - y| = |x-y| + \int_0^t \text{sign}(X_s - y)dX_s + L_t^y.$$

Consequently

$$\mathbb{E}_x(|X_{t\wedge\tau_{0,1}} - y|) = |x-y] + \mathbb{E}_x\left(L^y_{t\wedge\tau_{0,1}}\right).$$

We can take the limit as $t \to \infty$ by bounded convergence for the left hand side and monotone convergence for the right hand side, yielding

$$\mathbb{E}_x(|X_{\tau_{0,1}} - y|) = |x-y] + \mathbb{E}_x\left(L^y_{\tau_{0,1}}\right).$$

From the results in section 4.6,

$$\mathbb{P}_x(X_{\tau_{0,1}} = 0) = 1-x, \quad \mathbb{P}_x(X_{\tau_{0,1}} = 1) = x.$$

The last computations yield, in the case $0 < x < y < 1$,

$$\mathbb{E}_x\left(L^y_{\tau_{0,1}}\right) = 2x(1-y). \tag{6.7.3}$$

The result follows from (6.7.1), (6.7.2) and (6.7.3).

Proof of the Theorem We shall mimic the above proof, the arguments being simpler when we replace the diffusion X_t by the continuous time jump Markov process X^N_t with values in $\{0, 1/N, 2/N, \dots, (N-1)/N, 1\}$. Consider the evolution of the process $|X^N_t - y|$, where $y = k/N$, $1 \le k \le N-1$. First note that the rate of the jumps is $X^N_t(1 - X^N_t)$. If $X^N_t \neq y$ just before the jump of X^N_t, then the jump of $|X^N_t - y|$ is $1/N$ with probability $1/2$, and $-1/N$ with probability $1/2$. On the other hand, if $X^N_t = y$ just before the jumps of X^N_t, then the jump of $|X^N_t - y|$ is $1/N$. Consequently, if $\tau_{0,1} = \inf\{t,\ X^N_t \in \{0,1\}\}$

$$\begin{aligned}\mathbb{E}_x(|X^N_{t\wedge\tau_{0,1}} - y|) &= |x-y| + N\mathbb{E}_x\int_0^{t\wedge\tau_{0,1}} X^N_s(1-X^N_s)\mathbf{1}_{\{X^N_s=y\}}ds,\\ &= |x-y| + Ny(1-y)\mathbb{E}_x\int_0^{t\wedge\tau_{0,1}}\mathbf{1}_{\{X^N_s=y\}}ds,\end{aligned}$$

and letting $t \to \infty$, we deduce by bounded and monotone convergence

$$\mathbb{E}_x(|X^N_{\tau_{0,1}} - y|) = |x-y] + Ny(1-y)\mathbb{E}_x\int_0^{\tau_{0,1}}\mathbf{1}_{\{X^N_s=y\}}ds.$$

Since the jumps of NX^N_t are 1 with probability $1/2$ and -1 with probability $1/2$, the embedded chain is the symmetric random walk, and we know from Exercise 2.13 page 51 in [21] that if $x = \ell/N$, $\mathbb{P}_x(X^N_{\tau_{0,1}} = 0) = 1-x$, $\mathbb{P}_x(X^N_{\tau_{0,1}} = 1) = x$. Finally if $0 < x < y < 1$, we have that

$$\mathbb{E}_x\int_0^{\tau_{0,1}}\mathbf{1}_{\{X^N_s=y\}}ds = \frac{2}{N}\frac{x}{y}.$$

In particular we have for $1 \le k \le N-1$

$$\frac{\theta}{2}N\int_0^\infty \mathbb{P}_{1/N}\left(X^N_t = \frac{k}{N}\right)dt = \frac{\theta}{k}.$$

Finally if A is a subinterval of $(0,1)$ and f stands for the function in the statement,

$$\begin{aligned}\frac{\theta N}{2}\int_0^\infty \mathbb{P}_{1/N}(X_t^N\in A)dt &= \sum_{k;\frac{k}{N}\in A}\frac{\theta N}{2}\int_0^\infty \mathbb{P}_{1/N}\left(X_t^N=\frac{k}{N}\right)dt\\ &= \sum_{k;\frac{k}{N}\in A}\frac{1}{N}f\left(\frac{k}{N}\right)\\ &\to \int_A f(y)dy,\end{aligned}$$

as $N\to\infty$. □

Remark 6.7.2. *The above result should be understood as follows. We assume that there has been since time $-\infty$ a Poissonian flow of mutations affecting each individual at rate θ. This means a countable number of mutations. At most a countable number of those are still present in the population at time 0 (i. e. they have neither been fixed nor gone extinct). For $i\geq 1$, denote by ξ_i the fraction of the total population which carries the mutation number i. Let $g:(0,1)\to\mathbb{R}_+$ be a Borel measurable function. In the limit as the size N of the population goes to ∞,*

$$\mathbb{E}\sum_{i=1}^\infty g(\xi_i)=\theta\int_0^1 g(y)\frac{dy}{y}.$$

Remark 6.7.3. *Theorem 6.7.1 could be proved by a more analytic argument. Note that*

$$\int_0^\infty \mathbb{P}_x(X_t\in A)dt=\int_A G(x,y)dy,$$

where $G(x,y)$ denotes the Green kernel of Markov process $\{X_t\}$. G is the fundamental solution of the differential equation

$$\mathcal{L}u+f=0,\quad u(0)=u(1)=0.$$

In other words, G solves

$$\frac{1}{2}x(1-x)G''_{xx}(x,y)+\delta_{x-y}=0,\quad G(0,y)=G(1,y)=0.$$

The solution reads

$$G(x,y) = \begin{cases} 2\frac{x}{y}, & \text{if } x < y; \\ 2\frac{1-x}{1-y}, & \text{if } x \geq y. \end{cases}$$

The same approach can be used to prove that

$$\mathbb{E}\int_0^\infty \mathbf{1}_{\{X_s^N = y\}} ds = \frac{2x}{Ny}.$$

Indeed, if we call $G^N(x,y)$ *the left hand side, we have, with* Q^N *the infinitesimal generator induced by the Moran model,*

$$Q^N G_N(x,y) + \mathbf{1}_{\{x=y\}} = 0,\ x,y \in \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, 1\},\ G(0,y) = G(1,y) = 0.$$

It is easily seen that the solution of this difference equation is

$$G(x,y) = \begin{cases} \frac{2x}{Ny} & , \text{ for } x \leq y; \\ \frac{2(1-x)}{N(1-y)} & , \text{ for } x > y. \end{cases}$$

Exercise 6.7.4. *Show that in the more general case of diffusion of the form*

$$dX_t = h(X_t)dt + \sqrt{X_t} dB_t$$

which admits 0 *and* 1 *as exit points,*

$$G(x,y) = 2m(y)\frac{S(x)(S(1) - S(y))}{S(1)},$$

where S *is a scale function of the above diffusion, defined for an arbitrary* $z \in [0,1]$ *as*

$$S(x) \int_z^x \exp\left\{-\int_z^y 2[h(v)/g(v)]dv\right\} dy,$$

note that $S(X_t)$ *is a martingale, and* $m(x)$ *is the density of the speed measure of the same difusion, defined as*

$$m(x) = [g(x)s(x)]^{-1}, \quad \text{where } s(x) = S'(x).$$

Show finally that as $N \to \infty$,

$$N\int_0^\infty \mathbb{P}_{1/N}(X_t \in A)dt \to \int_A f(y)dy,$$

where $f(y) = 2m(y)\mathbb{P}_y(X_{\tau_{0,1}} = 0)$.

We can now discuss again the question which was adressed in Proposition 6.6.3. How many families of size i do we expect to find in a sample of size n, where $1 \le i \le n$ (we consider only mutations which are not fixed in the population) ?

Third proof of Proposition 6.6.3 Remind that there are at most countably many mutations which are present in our model. If one of those is present in the population with the proportion $0 < y < 1$, then the probability that that particular mutation is carried by i individuals in a sample of size n is

$$\binom{n}{i} y^i (1-y)^{n-i}.$$

The quantity which we are looking at is then the integral of the above against $f(y)dy$, i. e.

$$\begin{aligned}\int_0^1 \binom{n}{i} y^i (1-y)^{n-i} f(y)dy &= \theta \frac{n}{i} \int_0^1 \binom{n-1}{i-1} y^{i-1}(1-y)^{n-i} dy \\ &= \frac{\theta}{i},\end{aligned}$$

where the computation of the integral on the right hand side is easily done by integration by parts. □

In fact the above result is also true for $i = 0$. In an infinite population at equilibrium, the expected number of mutation which is carried by no individual from a sample of size $n \ge 1$ equals $+\infty$. We also have

Proposition 6.7.5. *The expectation of the number of families of relative size $y \in [a, b]$, where $0 \le a < b \le 1$, to which a randomly selected individual belongs equals $\theta(b - a)$.*

Proof: The probability that a randomly selected individual belongs to a family of relative size y is precisely y. Summing over all families of relative size between a and b and taking the expectation is the same as integrating from a to b with respect to $f(y)dy$. □

6.8 The case of two population which have been separated for a while

Consider now the case of a infinite population which splits into two infinite populations, say population 1 and population 2, which between time $-t$ and time 0 have no contact (so that there is no exchange of genetic material between them).

Consider one individual, let us call him A, in population 2, and a sample of size m in population 1, both randomly chosen at time 0. We ask the question : what is the expectation of the number of mutations carried A which are shared by exactly i individuals in the sample from population 1, $i = 0, 1, \ldots, m$. We consider only mutations which are present in population 1, in particular which have not gone extinct in that population (this restriction is relevant in the case $i = 0$) and have also not been fixed (this restriction is relevant in the case $i = m$). The answer is given by

Proposition 6.8.1. *The expected number of mutations which are shared between A and exactly i individuals from a sample of size m from population 1, is*

$$\frac{\theta}{m+1}e^{-t}, \quad \forall 0 \le i \le m.$$

PROOF: Consider the mutations which the ancestor at time $-t$ of individual A was carrying. From Proposition 6.7.5, the density of the expectation of the number of families to which this ancestor was belonging at time $-t$ is constant over the interval $[0,1]$ of relative family sizes, equal to θ. According to Lemma 6.8.2 below, the density of those family sizes is constant equal to $e^{-t}\theta$ at time 0. Now we sample m individuals at random. The probability that exactly $0 \le i \le m$ among them share a mutation which is carried by a family of relative size y is

$$\binom{m}{i} y^i (1-y)^{m-i}.$$

Hence the answer to our question is given by the integral of that quantity against the uniform density $e^{-t}\theta$ on the interval $[0,1]$, i. e.

$$e^{-t}\theta \binom{m}{i} \int_0^1 y^i (1-y)^{m-i} dy = \frac{\theta}{m+1} e^{-t}.$$

□

Lemma 6.8.2. *Let $\{X_t,\ t \ge 0\}$ denote the Wright–Fisher diffusion started at time $t = 0$ with the uniform probability on the interval $[0,1]$, i. e. the Lebesgue measure λ on $[0,1]$. Then for each $t > 0$ the measure*

$$\mu_t := e^{-t}\lambda + \frac{1-e^{-t}}{2}(\delta_0 + \delta_1)$$

is the law of X_t*.*

Proof: It is sufficient to show that for each $n \geq 0$,

$$\mathbb{E}_\lambda[X_t^n] = \int_{[0,1]} x^n \mu_t(dx).$$

Clearly

$$\int_{[0,1]} x^n \mu_t(dx) = \frac{e^{-t}}{n+1} + \frac{1-e^{-t}}{2} = \frac{1}{2}\left(1 - \frac{n-1}{n+1}e^{-t}\right).$$

But from the duality formula between the Wright–Fisher diffusion and Kingman's coalescent (see Proposition 5.5.1)

$$\mathbb{E}_x[X_t^n] = \mathbb{E}_n[x^{R_t}].$$

Hence

$$\begin{aligned}\mathbb{E}_\lambda[X_t^n] &= \int_0^1 \mathbb{E}_n[x^{R_t}]dx \\ &= \mathbb{E}_n\left[\frac{1}{R_t+1}\right].\end{aligned}$$

The result then follows from the next Exercise. □

Exercise 6.8.3. *1. Prove that* $2\mathbb{E}_2\left[\frac{1}{R_t+1}\right] = 1 - e^{-t}/3$*.*

2. Compute $2\mathbb{E}_n\left[\frac{1}{R_t+1}\right]$ *in terms of* $2\mathbb{E}_{n-1}\left[\frac{1}{R_t+1}\right]$*, by conditioning upon* T_n*, the time spend by the process* R_t *in the state* n*.*

3. Deduce by recurrence upon n *that*

$$2\mathbb{E}_n\left[\frac{1}{R_t+1}\right] = 1 - \frac{n-1}{n+1}e^{-t}.$$

We could also show that the collection $\{\mu_t,\ t \geq 0\}$ given in the statement solves the Fokker–Planck equation associated to X_t, i. e.

Exercise 6.8.4. *Prove that for all* $f \in C^2(\mathbb{R})$,

$$\frac{d}{dt} < \mu_t, f > = < \mu_t, Lf >, \quad \forall t \geq 0,$$

where

$$Lf(x) = \frac{1}{2}x(1-x)f''(x).$$

Note that deducing the Lemma 6.8.2 from Exercise 6.8.4 would imply proving a uniqueness result for the Fokker–Planck equation, which can be done but is not very easy.

6.9 Selection

In this section we want to discuss how selection can be introduced into the look–down construction. Suppose that the individuals in the infinite population are of two different types, say *red* or *black*, and that the *red* individuals have a selective advantage which we quantify by a positive parameter α. We introduce the following additional feature in the look–down construction : the *black* individuals die at rate α. One way of doing so is to put crosses on Harris' diagram on each level at rate α, independently of the arrows. Whenever a *red* individual hits a cross, nothing happens. Whenever a *black* individual hits a cross, he dies and the empty level is filled by shifting all individuals on its right one level down.

Suppose that the initial colours are i. i. d., *red* with probability x, *black* with probability $1-x$, and denote again by X_t the proportion of *red* individuals in the population at time t. We want to show the analog of Theorem 6.4.4

Theorem 6.9.1. *The* $[0,1]$*–valued process* $\{X_t,\ t \geq 0\}$ *solves the Wright–Fisher SDE with selection*

$$dX_t = \alpha X_t(1-X_t)dt + \sqrt{X_t(1-X_t)}dB_t,\ t \geq 0;\ X_0 = x.$$

PROOF:

□

Now conditioning upon fixation of the advantageous allele amounts to colouring the particle 1 *red* at time $t = 0$.

6.10 The Fleming–Viot measure–valued process

Consider again the look–down construction of the Moran model, where now each individual at time $t = 0$ is assigned, independently of all other individuals, a type which is the value of a $\mathcal{U}[0,1]$ random variable. In other words, the $\{\eta_0(k),\ k \geq 1\}$ are i. i. d. $\mathcal{U}[0,1]$'s. Arrows are drawn from k to ℓ at rate 1, for all $1 \leq k < \ell$. The evolution of the $\eta_t(k)$ is described as in the preceding section. For ech $N \geq 1$, let $\{\rho_t^N,\ t \geq 0\}$ denote the measure–valued process defined as

$$\rho_t^N([a,b]) = N^{-1} \sum_{k=1}^{N} \mathbf{1}_{\{\eta_k(t) \in [a,b]\}}, \quad 0 \leq a < b \leq 1.$$

It is easily seen that $\rho_0^N \Rightarrow \mathcal{U}[0,1]$, as $N \to \infty$.

Consider now the distribution function of the random mesures ρ_t^N. For each $0 \leq z \leq 1$, the result of the previous section shows that for all $t \geq 0$, $\rho_t^N[0,z] \to X_t(z)$, where $\{X_t(z),\ t \geq 0\}$ is a Wright–Fisher diffusion starting from $X_0(z) = z$. From this a. s. convergence and obvious inequalities concerning the $\rho_t^N[0,z]$, we deduce that $z \to \rho_t[0,z]$ is a. s. increasing for all $t \geq 0$. Moreover we know that for each $t \geq 0$, $X_t(z) \to 0$ as $z \to 0$, and $X_t(z) \to 1$ as $z \to 1$, a. s. Clearly ρ_t is the distribution function of a random probability on $[0,1]$, which we still denote ρ_t. That measure valued process is the a. s. weak limit of the approximating ρ_t^N. We have the following characterization of that process.

Proposition 6.10.1. *For any bounded Borel measurable function $g : [0,1] \to \mathbb{R}$, $\{\int_{[0,1]} g(x)\rho_t(dx),\ t \geq 0\}$ is a martingale whose quadratic variation $V_t(g)$ is given as*

$$V_t(g) = \frac{1}{2} \int_0^t ds \int \int_{[0,1]^2} [g(x) - g(y)]^2 \rho_s(dx)\rho_s(dy). \tag{6.10.1}$$

Proof: Consider the case where $g = \mathbf{1}_{[0,z]}$. In that case, $\int_{[0,1]} g(x)\rho_t(dx) = X_t(z)$, which is a Wright–Fisher diffusion, hence a martingale with the quadratic variation

$$\int_0^t ds \left[\int_{[0,z]} \rho_s(dx) - \left(\int_{[0,z]} \rho_s(dx) \right)^2 \right],$$

which is easily identified with the right hand side of 6.10.1 for our particular choice of g. The formula is easily extended to g's which are step functions, and to arbitrary bounded Borel measurable g's by approximation. □

We easily deduce from the previous result

Proposition 6.10.2. *For each Borel subset $A \subset [0,1]$, $\{\rho_t(A),\ t \geq 0\}$ is a martingale whose quadratic variation is given by*

$$\langle \rho(A) \rangle_t = \int_0^t \left[\rho_s(A) - \rho_s(A)^2\right] ds.$$

If $A, B \subset [0,1]$ and $A \cap B = \emptyset$, then

$$\langle \rho(A), \rho(B) \rangle_t = -\int_0^t \rho_s(A)\rho_s(B) ds.$$

We now prove the following result, which is reminiscent of some of the properties of the Feller branching diffusion, see chapter 1 above.

Theorem 6.10.3. *Assume that $\rho_0 = \mathcal{U}[0,1]$. Then a. s. for all $t > 0$, the support of the probability measure ρ_t is a finite set.*

PROOF: This result is a consequence of the duality between Kingman's coalescent and the Fleming–Viot process, and the coming down from infinity property of Kingman's coalescent.

Consider the m first individuals in the look–down construction at time $t > 0$. Going back in time, each pair of those individuals coalesce at rate 1. Thus the types of the individuals $\{1, 2, \dots, m\}$ at time t coincide with the types of the individuals $\{\xi_t(1), \xi_t(2), \dots, \xi_t(m)\}$ where for $1 \leq k \leq m$, $\xi_t(k)$ is the number between 1 and m of the individual to which following backward the look–down graphical representation leads to. The important fact is that $\xi_t(k) = \xi_t(\ell)$ if and only if the lineages k and ℓ belong to the same bloc of Kingman's m–coalescent partition at time t. If we write $\mathbb{P}_{\mathcal{U}}$ for the law of the Fleming–Viot process started with $\rho_0 = \mathcal{U}[0,1]$, and $\mathbb{P}_m$ for the law of the m–Kigman coalescent, we have

$$\begin{aligned}
&\mathbb{E}_{\mathcal{U}} \int_{[0,1]^m} f(y_1, y_2, \dots, y_m)\rho_t(dy_1)\rho_t(dy_2)\cdots\rho_t(dy_m) \\
&\quad = \mathbb{E}_m \int_{[0,1]^m} f(y_{\xi_t(1)}, y_{\xi_t(2)}, \dots, y_{\xi_t(m)})\rho_0(dy_1)\rho_0(dy_2)\cdots\rho_0(dy_m).
\end{aligned}$$

Note that the values of the indices $\xi_t(k)$ are biased towards 1 by the look–down construction, but this does not change the value of the above right hand side. The important fact is that the number of different values taken by that sequence of indices remains bounded a. s., as $m \to \infty$.

6.11 Bibliographical comments

Section 6.2 to 6.9 were inspired by a set of beautiful lectures given by Anton Wakolbinger in La Londe les Maures in septembre 2008. I am indebted to Vlada Limic for suggestions which led to the proof of the Markov property in Theorem 6.4.4. Section 6.10 started from a lecture given by Vlada Limic.

Chapter 7

Appendix

7.1 A Ray–Knight theorem

Consider a function $g \in C^1(\mathbb{R};\mathbb{R}_+)$ with compact support equal to $[0,a]$. We then have

Lemma 7.1.1. *The two–point boundary value problem*

$$\begin{cases} F''(t) = g(t)F(t), \quad 0 \le t \le a; \\ F(0) = 1, \quad F'(a) = 0; \end{cases} \tag{7.1.1}$$

has a unique decreasing and positive solution $\{F_g(t),\ 0 \le t \le a\}$.

PROOF: For any $z \in \mathbb{R}$, denote by $\{F_z(t),\ 0 \le t \le a\}$ the solution of the ODE with initial condition

$$\begin{cases} F''(t) = g(t)F(t), \quad 0 \le t \le a; \\ F(0) = 1, \quad F'(0) = z. \end{cases}$$

The mapping $z \to F'_z(a)$ is continuous, it is not hard to show that it is strictly increasing, $F'_0(a) > 0$ and $F'_z(a) < 0$ for $z < 0$, large enough in absolute value. Hence there is a unique $z_0 < 0$ such that $F'_{z_0}(a) = 0$. It is not hard to show that $F_{z_0}(a) > 0$. It is enough for our purpose to show that $F_{z_0}(a) \ge 0$. If that would not be the case, there would exist $0 < s < a$ such that $F_{z_0}(s) = 0$, $F'_{z_0}(s) \le 0$, $F_{z_0}(t) < 0$ for $s < t \le a$. Then from the equation $F''_{z_0}(t) < 0$ for $s < t \le a$, and consequently $F'_{z_0}(t) < 0$ for $s < t \le a$, which contradicts the condition $F'_{z_0}(a) = 0$. It is easily seen that $\{F_{z_0}(t),\ 0 \le t \le a\}$ has all the required properties, and that it is the unique solution of the two–point boundary value problem (7.1.1). □

Definition 7.1.2. *For each $x \geq 0$, $\delta \in \mathbb{R}$, we denote by QF_x^δ the law of the solution of the SDE*

$$Z_t^\delta = x + \delta \int_0^t Z_s ds + 2 \int_0^t \sqrt{Z_s^\delta} dB_s. \tag{7.1.2}$$

Consider now the process

$$X_s = B_s + \frac{1}{2} L_s^0, \; s \geq 0,$$

where $\{B_s, \; s \geq 0\}$ is a one–dimensional standard Brownian motion, and $\{L_s^t, \; s \geq 0, \; t \geq 0\}$ denotes the local time of X accumulated at level t by time s, hence X is "reflected Brownian motion".

Remark 7.1.3. *Brownian motion reflected on $\mathbb{R}_+$ is by definition the unique process X such that*

$$X_s = B_s + K_s, \quad K_0 = 0,$$
$$X_s \geq 0, \; K \textit{ is continuous and increasing}, \; \int_0^s X_r dK_r = 0.$$

It is well known and easy to verify that $K_s = -\inf_{0 \leq r \leq s} B_r$. But this is not the expression for K_s which we are seeking here.

It follows from Tanaka's formula that

$$X_s = X_s^+ = \int_0^s \mathbf{1}_{\{X_r > 0\}} dB_r + \frac{1}{2} L_s^0 = B_s + \frac{1}{2} L_s^0,$$

if L^0 denotes the local time at level 0 of the process X. Comparing the two above formulas, we deduce that $K_s = \frac{1}{2} L_s^0$. This justifies the above definition of reflected Brownian motion.

For $x > 0$, let

$$\tau_x = \inf\{s > 0; \; L_s^0 > x\}.$$

The following result is due to Ray and Knight

Theorem 7.1.4. *The law of the process $\{L_{\tau_x}^t, \; t \geq 0\}$ is Q_x^0.*

PROOF: Let g and F_g be as in Lemma 7.1.1. Define $f \in C(\mathbb{R}^2)$ as

$$f(y,x) = F_g(x)\exp[-F_g'(0)y/2].$$

It follows from Itô's formula that

$$\begin{aligned} f(L_s^0, X_s) = 1 &+ \int_0^s \left[f_y'(L_r^0, X_r) + \frac{1}{2} f_x'(L_r^0, X_r) \right] dL_r^0 \\ &+ \int_0^s f_x'(L_r^0, X_r) dB_r \\ &+ \frac{1}{2} \int_0^s f_{x^2}''(L_r^0, X_r) dr. \end{aligned}$$

Clearly in the dL_r^0 integral, X_r can be replaced by 0. Then from the choice of the function f, the dL_r^0 integral vanishes. Moreover

$$f_{x^2}''(L_r^0, X_r) = g(X_r) f(L_r^0, X_r).$$

Then

$$\exp\left(-\frac{1}{2}\int_0^s g(X_r)dr\right) f(L_s^0, X_s) = 1 + \int_0^s \exp\left(-\frac{1}{2}\int_0^r g(X_{r'})dr'\right) f_x'(L_r^0, X_r) dB_r.$$

It is easily seen that this last stochastic integral is a bounded martingale, hence by optional stopping and the occupation times formula

$$\begin{aligned} \mathbb{E}\left[\exp\left(-\frac{1}{2}\int_0^{\tau_x} g(X_s)ds\right)\right] &= \mathbb{E}\left[\exp\left(-\frac{1}{2}\int_0^{\infty} g(t) L_{\tau_x}^t dt\right)\right] \\ &= \exp\left(\frac{x}{2} F_g'(0)\right). \end{aligned}$$

The result follows from the next Lemma. □

Lemma 7.1.5. *For any g and F_g as in Lemma 7.1.1, if $\{Z_t,\ t \geq 0\}$ denotes the solution of (7.1.2) with $\delta = 0$,*

$$\mathbb{E}\left[\exp\left(-\frac{1}{2}\int_0^{\infty} Z_t g(t) dt\right)\right] = \exp\left(F_g'(0)\frac{x}{2}\right).$$

Proof: Let $H_g(t) = F_g'(t)/F_g(t)$. Note that $H_g'(t) = g(t) - H_g^2(t)$. Hence

$$H_g(t)Z_t = H_g(0)x + \int_0^t H_g(s)dZ_s + \int_0^t g(s)Z_s ds - \int_0^t H_g^2(s)Z_s ds,$$

and the following is a local martingale

$$\begin{aligned} M_t &= \exp\left(\frac{1}{2}\int_0^t H_g(s)dZ_s - \frac{1}{2}\int_0^t Z_s H_g^2(s)ds\right) \\ &= \exp\left(\frac{1}{2}\left\{H_g(t)Z_t - H_g(0)x - \int_0^t g(s)Z_s ds\right\}\right) \end{aligned}$$

which is bounded on $[0, a]$. The result follows by writing

$$\mathbb{E}(M_a) = \mathbb{E}(M_0).$$

□

Remark 7.1.6. *We could as well have defined X as Brownian motion, instead of reflected Brownian motion. The result would have been the same, and the proof esntially identical, except that we should have started with developing the quantity $f(L_s^0, (X_s)^+)$. It is not surprising that the result is the same for the local time of Brownian motion and the local time of Brownian motion reflected in $\mathbb{R}_+$, see Lemma 7.1.8 below.*

Consider now for each $\delta \in \mathbb{R}$, $a > 0$, the process

$$X_s^{\delta,a} = \frac{\delta}{2}s + B_s + \frac{1}{2}L_s^0(\delta, a) - \frac{1}{2}L_s^a(\delta, a), \quad s \geq 0, \tag{7.1.3}$$

where $L(\delta, a)$ denotes the local time of $X^{\delta,a}$, which is Brownian motion with drift $\delta/2$, reflected in the interval $[0, a]$. We next define

$$\tau_x(\delta, a) = \inf\{s > 0,\ L_s^0(\delta, a) > x\}.$$

In the case $\delta < 0$, the reflection at 0 is necessary, if we want to make sure that the process $L^0(\delta, a)$ reaches the value x. Similarly, the reflection at a is necessary in the case $\delta > 0$. We first need to show that the reflection at a does not distort our object of study. This follows from

Proposition 7.1.7. *For all $0 < a < b$, the two processes $\{L_{\tau_x}^t(\delta, a),\ 0 \leq t \leq a\}$ and $\{L_{\tau_x}^t(\delta, b),\ 0 \leq t \leq a\}$ have the same law.*

Proposition 7.1.7 is an immediate consequence of the next Lemma. Before stating it, we need to introduce some notations. For any $a > 0$, $\varphi \in C(\mathbb{R}_+; \mathbb{R}_+)$, define

$$\begin{aligned} A^a_s(\varphi) &= \int_0^s \mathbf{1}_{\{\varphi(r)\le a\}} dr, \ \ s \ge 0, \\ C^a_s(\varphi) &= \inf\{r \ge 0, \ A^a_r(\varphi) > s\}, \ \ s \ge 0, \\ \pi_a(\varphi) &= \varphi \circ C^a(\varphi). \end{aligned}$$

Lemma 7.1.8. *For any $0 < a < b$, $\pi_a(X^{\delta,b})$ has the same law as $X^{\delta,a}$.*

Proof: In this proof, we shall use the notations

$$A(s) = A^a_s(X^{\delta,b}), \quad C(s) = C^a_s(X^{\delta,b}).$$

Note that

$$\begin{aligned} \pi_a(X^{\delta,b})_s &= X^{\delta,b}_{C(s)} \\ &= \frac{\delta}{2}C(s) + B_{C(s)} + \frac{1}{2}L^0_{C(s)} - \frac{1}{2}L^b_{C(s)} \\ &= \int_0^{C(s)} \mathbf{1}_{\{X^{\delta,b}_r \le a\}}(dB_r + \frac{\delta}{2}dr) + \frac{1}{2}L^0_{C(s)} \\ &\quad + \int_0^{C(s)} \mathbf{1}_{\{X^{\delta,b}_r > a\}}(dB_r + \frac{\delta}{2}dr) - \frac{1}{2}L^b_{C(s)}. \end{aligned}$$

Tanaka's formula tells us that

$$(X^{\delta,b}_r - a)^+ = \int_0^r \mathbf{1}_{\{X^{\delta,b}_{r'} > a\}} dX^{\delta,b}_{r'} + \frac{1}{2}L^a_r,$$

which choosing $r = C(s)$ gives

$$\int_0^{C(s)} \mathbf{1}_{\{X^{\delta,b}_r > a\}} dX^{\delta,b}_r + \frac{1}{2}L^a_{C(s)} = 0,$$

from which we deduce that

$$\int_0^{C(s)} \mathbf{1}_{\{X^{\delta,b}_r > a\}}(dB_r + \frac{\delta}{2}dr) - \frac{1}{2}L^b_{C(s)} + \frac{1}{2}L^a_{C(s)} = 0.$$

Combining this and the previous identity gives

$$X^{\delta,b}_{C(s)} = \int_0^{C(s)} \mathbf{1}_{\{X^{\delta,b}_r \leq a\}}(dB_r + \frac{\delta}{2}dr) + \frac{1}{2}L^0_{C(s)} - \frac{1}{2}L^a_{C(s)}.$$

Define $\tilde{X}^{\delta,b}_s = X^{\delta,b}_{C(s)}$. We have

$$\tilde{X}^{\delta,b}_s = B'_s + \frac{\delta}{2}A \circ C(s) + \frac{1}{2}L^0_{C(s)} - \frac{1}{2}L^a_{C(s)}.$$

Note that $A \circ C(s) = s$ and B' is a standard Brownian motion. Indeed B' is a continuous local martingale (this follows from Doob's optional stopping theorem), whose quadratic variation process equals

$$\langle B' \rangle_s = \int_0^{C(s)} \mathbf{1}_{\{X^{\delta,b}_r \leq a]\}} dr = A \circ C(s) = s.$$

Let $K_s := \frac{1}{2}L^0_{C(s)} - \frac{1}{2}L^a_{C(s)}$. We note that we have

$$\begin{aligned} \tilde{X}^{b,\delta}_s &= B'_s + \frac{\delta}{2}s + K_s, \ X_s \in [0,a], \ \forall s \geq 0, \text{ a. s.}, \\ K_s &= \int_0^s \mathbf{1}_{\{\tilde{X}_r = 0\}} d|K|_r - \int_0^s \mathbf{1}_{\{\tilde{X}_r = a\}} d|K|_r \\ \int_0^s \mathbf{1}_{\{\tilde{X}_r \notin \{0,a\}\}} d|K|_r &= 0. \end{aligned}$$

This means that $\tilde{X}$ is Brownian motion plus drift equal to $\delta/2$, reflected in the interval $[0,a]$, in other words, if $\tilde{L}(b,\delta)$ denotes the local time of $\tilde{X}^{b,\delta}$,

$$\tilde{X}^{b,\delta}_s = B'_s + \frac{\delta}{2}s + \frac{1}{2}\tilde{L}^0_s(b,\delta) - \frac{1}{2}\tilde{L}^a_s(b,\delta).$$

The result follows. □

It follows from Proposition 7.1.7 that there exists a continuous process $\{L^t_{\tau_x}(\delta), \ t \geq 0\}$ such that for each $a > 0$, $\{L^t_{\tau_x}(\delta), \ 0 \leq t \leq a\}$ has the same law as $\{L^t_{\tau_x}(\delta, a), \ 0 \leq t \leq a\}$. We now show

Theorem 7.1.9. *The law of* $\{L^t_{\tau_x}(\delta), \ t \geq 0\}$ *is* QF^δ_x.

PROOF: We first show that for some $a>0$, the law of $\{L^t_{\tau_x}(\delta),\ 0\le t\le a\}$ is the restriction of QF^δ_x to $C([0,a])$. For the rest of this proof, we choose $a>0$ arbitrary if $\delta<0$, and $0<a<2/\delta$ if $\delta>0$.

Since we shall be using Girsanov's theorem, in the proof we shall index the expectation by δ (writing $\mathbb{E}^\delta$), and drop the index δ from the processes themselves.

Let again g be a continuous function from $\mathbb{R}_+$ into $\mathbb{R}_+$, with support included in $[0,a]$, for some arbitrary $a>0$. It follows from Girsanov's theorem that

$$\mathbb{E}^\delta\left[\exp\left(-\int_0^{\tau_x\wedge n} g(X^a_s)ds\right)\right]=\mathbb{E}^0\left[\exp\left(\frac{\delta}{2}B_{\tau_x\wedge n}-\frac{\delta^2}{8}\tau_x\wedge n-\int_0^{\tau_x\wedge n} g(X^a_s)ds\right)\right].$$

Under $\mathbb{P}^0$

$$-\frac{x}{2}\le-\frac{1}{2}L^0_{\tau_x\wedge n}\le B_{\tau_x\wedge n}\le a+\frac{1}{2}L^a_{\tau_x\wedge n}\le a+\frac{1}{2}L^a_{\tau_x},$$

so we can take the limit as $n\to\infty$ in the above identity with the help of Lebesgue's dominated convergence theorem, provided either $\delta<0$, or

$$\mathbb{E}^0\left[\exp\left(\frac{\delta}{4}L^a_{\tau_x}\right)\right]<\infty. \tag{7.1.4}$$

But we know exactly the law of $L^a_{\tau_x}$ under $\mathbb{P}^0$. Indeed, from the results in section 1.6, $L^a_{\tau_x}$ is the sum $Y_1+\cdots+Y_Z$, where $Z,Y_1,Y_2,\dots$ are mutually independent, Z is Poisson with parameter $x/2a$, and the Y_k are exponential with parameter $1/2a$. Hence (7.1.4) holds iff $\delta<2/a$, which is exactly our standing assumption.

$$\begin{aligned}
\mathbb{E}^\delta\left[\exp\left(-\int_0^{\tau_x} g(X^a_s)ds\right)\right]&=\mathbb{E}^0\left[\exp\left(\frac{\delta}{2}B_{\tau_x}-\frac{\delta^2}{8}\tau_x-\int_0^{\tau_x} g(X^a_s)ds\right)\right]\\
&=\mathbb{E}^0\left[\exp\left(\frac{\delta}{4}\left(L^a_{\tau_x}-x\right)-\frac{\delta^2}{8}\tau_x-\int_0^{\tau_x} g(X^a_s)ds\right)\right]\\
&=\mathbb{E}^0\left[\exp\left(\frac{\delta}{4}\left(L^a_{\tau_x}-x\right)-\int_0^a\left(g(t)+\frac{\delta^2}{8}\right)L^t_{\tau_x}dt\right)\right]\\
&=\mathbb{E}^0\left[\exp\left(\frac{\delta}{2}\left(Z^x_a-x\right)-\int_0^a\left(g(t)+\frac{\delta^2}{8}\right)Z^x_t dt\right)\right]\\
&=\mathbb{E}^0\left[\exp\left(\frac{\delta}{2}\int_0^a\sqrt{Z^x_t}dB_t-\int_0^a\left(g(t)+\frac{\delta^2}{8}\right)Z^x_t dt\right)\right]\\
&=\mathbb{E}^\delta\left[\exp\left(-\int_0^a g(t)Z^x_t dt\right)\right],
\end{aligned}$$

where we have used (7.1.3) for the second equality and the fact that from Theorem 7.1.4 under $\mathbb{P}^0$ the law of $\{L^t_{\tau_x}(a),\ 0 \le t \le a\}$ is Q^0_x restricted to $C([0,a])$ at the fourth equality, and again Girsanov's theorem at the last equality. The reason why we are in a position to apply Girsanov's theorem is that the same argument which leads to (7.1.4) shows that for some $c > 0$ (e. g. $c = \delta/4$, but this is unimportant)

$$\mathbb{E}^0\left[\exp\left(cZ^x_t\right)\right] = \mathbb{E}^0\left[\exp\left(cL^t_{\tau_x}\right)\right] < \infty, \quad 0 \le t \le a.$$

Now an argument similar to the one leading to Proposition 7.1.7 shows that the law of the portion of the trajectory of $X^{\delta,b}$ which lives between the levels a and b (where $a < b$) is the same as the law of $\frac{\delta}{2}s + B_s$ reflected in $[a,b]$. This says in particular that $\{L^t_{\tau_x}(\delta,b),\ a \le t \le b\}$ is conditionally independent of $\{L^t_{\tau_x}(\delta,b),\ 0 \le t \le a\}$, given $L^a_{\tau_x}(\delta,b)$, and establishes the Markov property of the process $\{L^t_{\tau_x}(\delta),\ t \ge 0\}$.

The above argument allows us to identify the conditional law of $\{L^{a+t}_{\tau_x}(\delta,2a),$ $0 \le t \le a\}$ given that $L^a_{\tau_x}(\delta,2a) = y$ as the law QF^δ_y, restricted to $C([0,a])$. Combining this with the previous result tells us that the law of $\{L^t_{\tau_x}(\delta),\ 0 \le t \le 2a\}$ is QF^δ_x restricted to $C([0,2a])$. Repeating the argument gives the same statement with $2a$ replaced by ka, for any $k \in \mathbb{N}$. The Theorem follows. □

7.2 Tightness in C

We prove the following particular case of Theorem 12.3 in Billingsley [4] page 95 :

Theorem 7.2.1. *A sufficient condition for a sequence $\{X^n_t, : t \ge 0\}_{n\ge1}$ of continuous $\mathbb{R}^d$–valued martingales to be tight in $C(\mathbb{R}_+;\mathbb{R}^d)$ is that*

1. *The sequence $\{X^n_0\}_{n\ge1}$ is tight.*

2. *There exists $c > 0$ such that for all $0 \le s \le t$,*

$$\mathbb{E}\left(|X^n_t - X^n_s|^4\right) \le c|t-s|^2. \tag{7.2.1}$$

PROOF: It suffices to prove tightness in $C([0,T];\mathbb{R}^d)$ for all $T > 0$. For the sake of simplifying our notations, we shall prove tightness in $C := C([0,1])$. The denote by

$$w(x,\delta) = \sup_{|t-s|\le\delta} |x(t) - x(s)|$$

the modulus of continuity of an arbitrary element $x \in C$. Let us first check that a sufficient condition for tightness in C of the random sequence $\{X^n\}$ is, together with *1*, the fact that for all ε, $\eta > 0$, we can find $\delta > 0$ such that

$$\mathbb{P}(w(X^n,\delta) \ge \varepsilon) \le \eta, \ \forall n \ge 1. \tag{7.2.2}$$

Indeed, in that case $\forall \eta > 0$, $k \ge 1$, there exists $\delta_k > 0$ such that

$$\mathbb{P}\left(w(X^n,\delta_k) > \frac{1}{k}\right) \le \frac{\eta}{2^k}.$$

Hence

$$\mathbb{P}\left(\bigcap_k \left\{w(X^n,\delta_k) \le \frac{1}{k}\right\}\right) \ge 1 - \eta.$$

Consequently, combining with *1*, we deduce that for M large enough, $\mathbb{P}(X^n \in K) \ge 1 - 2\eta$, where

$$K := \left\{x, |x(0)| \le M \text{ and } w(x,\delta_k) \le \frac{1}{k}, \ \forall k \ge 1\right\}$$

is compact from the Arzelà–Ascoli theorem.

For each $n \ge 1$, $0 \le t \le 1$, $\varepsilon > 0$, $\delta > 0$, define the event

$$A^n_{t,\delta,\varepsilon} = \left\{\sup_{t\le s\le(t+\delta)\wedge 1} |X^n_s - X^n_t| \ge \varepsilon\right\}.$$

Since each $\{X^n_t, \ 0 \le t \le 1\}$ is a martingale, it follows from Doob's inequality and (7.2.1) that

$$\mathbb{P}(A^n_{t,\delta,\varepsilon}) \le c'\frac{\delta^2}{\varepsilon^4}.$$

Now the pairs (s,t) which count in the definition of $w(X^n,\delta)$ belong either to the same or to two neighbouring intervals of the form $[i\delta,(i+1)\delta]$. Consequently

$$\begin{aligned}\mathbb{P}(w(X^n,\delta)\geq 3\varepsilon) &\leq \mathbb{P}\left(\cup_{i<\delta^{-1}} A^n_{i\delta,\delta,\varepsilon}\right)\\ &\leq \sum_{i<\delta^{-1}} \mathbb{P}(A^n_{i\delta,\delta,\varepsilon})\\ &\leq c'\frac{\delta}{\varepsilon^4}.\end{aligned}$$

(7.2.2) clearly now follows by an appropriate choice of δ. □

7.3 A martingale representation theorem

We state and prove the result in dimension 1, since we will need it only in that case.

Theorem 7.3.1. *Let $\{M_t,\ t\geq 0\}$ be a one dimensional continuous martingale, defined on a probability space $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbb{P})$, with associated increasing process*

$$\langle M\rangle_t = \int_0^t A_s ds,$$

where $\{A_t,\ t\geq 0\}$ is an $\mathcal{F}_t$–progressively measurable $\mathbb{R}_+$–valued process. Then there exists, possibly on an enlarged probability space $(\Omega',\mathcal{F}',\mathbb{P}')$, a standard Brownian motion such that

$$M_t = \int_0^t \sqrt{A}_s dB_s,\ t\geq 0. \tag{7.3.1}$$

PROOF: Let $(\Omega',\mathcal{F}',\mathbb{P}') = (\Omega\times C(\mathbb{R}_+),\mathcal{F}\otimes\mathcal{C},\mathbb{P}\times\mathcal{W})$, where $\mathcal{C}$ denote the Borel σ field over $C(\mathbb{R}_+)$, and $\mathcal{W}$ the Wiener measure on $(C(\mathbb{R}_+),\mathcal{C})$. Let $\{W_t,\ t\geq 0\}$ denote the canonical process defined on $(C(\mathbb{R}_+),\mathcal{C},\mathcal{W})$. Define the two following $\mathcal{F}_t$–progressively measurable processes

$$\text{For } t\geq 0,\quad a_t = \begin{cases} A_t^{-1/2}, & \text{if } A_t > 0,\\ 0, & \text{if } A_t = 0.\end{cases}$$

$$b_t = \begin{cases} 0, & \text{if } A_t > 0,\\ 1, & \text{if } A_t = 0.\end{cases}$$

Let now

$$B_t = \int_0^t a_s dM_s + \int_0^t b_s dW_s,\ t \geq 0.$$

It is easy to check that $\{B_t,\ t \geq 0\}$ is a continuous martingale. Moreover, since M and W are independent,

$$\begin{aligned} \langle B \rangle_t &= \int_0^t a_s^2 A_s ds + \int_0^t b_s^2 ds \\ &= \int_0^t \mathbf{1}_{\{A_s > 0\}} ds + \int_0^t \mathbf{1}_{\{A_s = 0\}} ds \\ &= t, \end{aligned}$$

hence $\{B_t,\ t \geq 0\}$ is a Brownian motion, and (7.3.1) holds.

7.4 The Lévy–Khinchin formula

Definition 7.4.1. *The law of the real–valued r. v. X is said to be infinitely divisible if for every $n > 1$, there exist n i. i. d. r. v.'s $X_1^n, \dots, X_n^n$ such that*

$$\mathcal{L}(X) = \mathcal{L}(X_1^n + \cdots + X_n^n).$$

The characteristic function of an infinitely divisible r. v. X can be written as

$$\varphi_X(u) = \mathbb{E}\left[\exp(iuX)\right] = \exp(-\Psi(u)),$$

with a unique *characteristic exponent* $\Psi \in C(\mathbb{R}; \mathbb{C})$ satisfying $\Psi(0) = 0$, specified by the celebrated Lévy–Khintchin formula (see e. g. [7])

Theorem 7.4.2. *A function $\Psi : \mathbb{R} \in \mathbb{C}$ is the characteristic exponent of an infinitely divisible distribution on $\mathbb{R}$ iff there are $\alpha \in \mathbb{R}$, $\beta \geq 0$, Λ a measure on $\mathbb{R}\backslash\{0\}$, called the Lévy measure, which satisfies $\int (1 \wedge |x|^2)\Lambda(dx) < \infty$, such that*

$$\Psi(u) = i\alpha u + \beta u^2 + \int_{\mathbb{R}} (1 - e^{iux} + iux\mathbf{1}_{\{|x| \leq 1\}})\Lambda(dx). \tag{7.4.1}$$

In the particular case of a positive valued infinitely divisible r. v., we prefer to explicit the Laplace exponent of X, i. e. for $\lambda \geq 0$

$$\psi(\lambda) = \log\left\{-\exp\left[-\lambda X\right]\right\}$$

which takes the form

$$\psi(\lambda) = d\lambda + \int_0^\infty (1 - e^{-\lambda x})\Lambda(dx), \tag{7.4.2}$$

again with $\lambda \geq 0$. Here $d \geq 0$ is called the drift coefficient, and the Lévy measure Λ is in this case a measure on $(0, +\infty)$ which satisfies

$$\int_{\mathbb{R}_+} (1 \wedge x)\Lambda(dx) < \infty.$$

7.5 de Finetti's theorem

A permutation π of the set $\{1, 2, \ldots\}$ is said to be finite if $|\{i,\ \pi(i) \neq i\}| < \infty$. Let us formulate the

Definition 7.5.1. *The countably infinite sequence $\{X_n,\ n \geq 1\}$ is said to be exchangeable if for all finite permutation π of $\{1, 2, \ldots\}$,*

$$(X_1, X_2, \ldots) \overset{\mathcal{L}}{=} (X_{\pi(1)}, X_{\pi(2)}, \ldots).$$

It is not too hard to show that

Lemma 7.5.2. *Given a countably infinite sequence of r. v.'s $\{X_1, X_2, \ldots\}$, the three following properties are equivalent*

1. *The sequence $\{X_1, X_2, \ldots\}$ is exchangeable.*
2. *For all $n > 1$,*

$$(X_1, \ldots, X_{n-1}, X_n, X_{n+1}, \ldots) \overset{\mathcal{L}}{=} (X_n, \ldots, X_{n-1}, X_1, X_{n+1}, \ldots).$$

3. *For all sequence $\{n_i,\ i \geq 1\}$ of distinct integers,*

$$(X_1, X_2, X_3, \ldots) \overset{\mathcal{L}}{=} (X_{n_1}, X_{n_2}, X_{n_3}, \ldots).$$

Let us recall the well–known "reversed martingale convergence theorem" (see e. g. [5])

Theorem 7.5.3. *Let $\{\mathcal{G}_n,\ n \geq 1\}$ be a decreasing sequence of sub–σ–fields of $\mathcal{F}$, $\mathcal{G} = \cap_n \mathcal{G}_n$. Then for any integrable r. v. Z, $\mathbb{E}(Z|\mathcal{G}_n) \to \mathbb{E}(Z|\mathcal{G})$ a. s.*

We now prove an easy lemma

Lemma 7.5.4. *Let Y be a bounded r. v., and $\mathcal{H} \subset \mathcal{G}$ be two sub–σ–fields of $\mathcal{F}$. Then $\mathbb{E}\left[\mathbb{E}(Y|\mathcal{H})^2\right] = \mathbb{E}\left[\mathbb{E}(Y|\mathcal{G})^2\right]$ (or a fortiori $\mathbb{E}(Y|\mathcal{H}) \stackrel{\mathcal{L}}{=} \mathbb{E}(Y|\mathcal{G})$) implies that $\mathbb{E}(Y|\mathcal{H}) = \mathbb{E}(Y|\mathcal{G})$ a. s.*

Proof: The result follows readily from the identity

$$\mathbb{E}\left[(\mathbb{E}(Y|\mathcal{G}) - \mathbb{E}(Y|\mathcal{H}))^2\right] = \mathbb{E}\left[(\mathbb{E}(Y|\mathcal{G}))^2\right] - \mathbb{E}\left[(\mathbb{E}(Y|\mathcal{H}))^2\right].$$

□

We now state the celebrated de Finetti's theorem. Our proof follows one of the proofs given in [1]. See also [5] for the case of $\{0,1\}$–valued r. v. 's.

Theorem 7.5.5. *An exchangeable (countably infinite) sequence $\{X_n,\ n \geq 1\}$ of r. v.'s is a mixture of i. i. d. sequences, in the sense that conditionally upon $\mathcal{T}$ (the tail σ–field of the sequence $\{X_n\}$, the X_n are i. i. d.*

Proof: For each $n \geq 0$, let $\mathcal{G}_n := \sigma(X_{n+1}, X_{n+2}, \ldots)$, and let $\mathcal{T} := \cap_n \mathcal{G}_n$ the tail σ–field. By exchangeability, for all $n \geq 2$,

$$(X_1, X_2, X_3, \ldots) \stackrel{\mathcal{L}}{=} (X_1, X_{n+1}, X_{n+2}, \ldots).$$

Consequently for any bounded Borel measurable function $\varphi : \mathbb{R} \to \mathbb{R}$, $n \geq 2$,

$$\mathbb{E}(\varphi(X_1)|\mathcal{G}_1) \stackrel{\mathcal{L}}{=} \mathbb{E}(\varphi(X_1)|\mathcal{G}_n).$$

Theorem 7.5.3 implies that

$$\mathbb{E}(\varphi(X_1)|\mathcal{G}_n) \to \mathbb{E}(\varphi(X_1)|\mathcal{T}) \quad \text{a. s., as } n \to \infty.$$

We deduce that

$$\mathbb{E}(\varphi(X_1)|\mathcal{G}_1) \stackrel{\mathcal{L}}{=} \mathbb{E}(\varphi(X_1)|\mathcal{T}).$$

Now Lemma 7.5.4 implies that the equality holds a. s. This implies that

$$X_1 \quad \text{and} \quad \mathcal{G}_1 \quad \text{are conditionally independent given } \mathcal{T}.$$

The same argument applied to $(X_n, X_{n+1}, \ldots)$ says that for all $n \geq 1$,

$$X_n \quad \text{and} \quad \mathcal{G}_n \quad \text{are conditionally independent given } \mathcal{T}.$$

This implies that the whole sequence $\{X_n,\ n \geq 1\}$ is conditionally independent given $\mathcal{T}$. Now exchangeability says that for all $n \geq 1$,

$$(X_1, X_{n+1}, X_{n+2}, \ldots) \overset{\mathcal{L}}{=} (X_n, X_{n+1}, X_{n+2}, \ldots).$$

So for the same φ's as above,

$$\mathbb{E}(\varphi(X_1)|\mathcal{G}_n) = \mathbb{E}(\varphi(X_n)|\mathcal{G}_n) \quad \text{a. s..}$$

Taking the conditional expectation given $\mathcal{T}$ yields

$$\mathbb{E}(\varphi(X_1)|\mathcal{T}) = \mathbb{E}(\varphi(X_n)|\mathcal{T}) \quad \text{a. s..}$$

Hence, conditionally upon $\mathcal{T}$, the X_n are also identically distributed. □

We now deduce the

Corollary 7.5.6. *Let $\{X_n,\ n \geq 1\}$ be an exchangeable (countably infinite) sequence of $\{0, 1\}$–valued r. v.'s. Then, conditionally upon*

$$a.\ s.\ \lim_n n^{-1} \sum_{k=1}^{n} X_k = x,$$

the X_n are i. i. d. Bernoulli with parameter x.

Bibliography

[1] David Aldous, Exchangeability and related topics, in *Ecole d'Ete St Flour 1983* Lecture Notes in Math. **1117**, 1–198, Springer 1985.

[2] Jean Bertoin, Joaquim Fontbona, Servet Martinez, On prolific individuals in a supercritical continuous state branching process, Preprint 2007.

[3] Jean Bertoin, Jean–Francois Le Gall, The Bolthausen–Sznitman coalescent and the genealogy of continuous–state branching processes, *Probab. Theory Relat. Fields* **117**, 249–266, 2000.

[4] Patrick Billingsley, *Convergence of probability measures*, Wiley 1968.

[5] Patrick Billingsley, *Probability and measures*, 3d ed. Wiley 1995.

[6] Matthias Birkner, Stochastic models from population biology, lecture notes for a course at TU Berlin, summer 2005 http://www.wias-berlin.de/people/birkner/smpb-30.6.05.pdf

[7] L. Breiman : *Probability*, Addison–Wesley, 1968. New edition SIAM 1992.

[8] Jean–Franois Delmas, Height process for super–critical continuous state branching process, *Markov Proc. and Rel. Fields*, **14**, pp. 309–326, 2008.

[9] P. Donnelly, T. Kurtz, A countable representation of the Fleming–Viot measure–valued diffusion, *Annals Probab.* **24**, 698–742, 1996.

[10] Alison Etheridge, *An introduction to superprocesses*, Memoirs of the Amer. Math. Soc. 2000.

[11] Stewart Ethier, Thomas Kurtz, *Markov processes*, Wiley 1986.

[12] Fred Hoppe, Polya–like urns and the Ewens sampling formula, *J. Math. Biol.* **20**, 91–94, 1984.

[13] Nobuyuki Ikeda, Shinzo Watanabe, *Stochastic differential equations and diffusion processes*, North–Holland 1989.

[14] Olav Kallenberg, *Foundations of modern probability*, Springer 1997.

[15] Samuel Karlin, Howard Taylor, *A second course in stochastic processes*, Acad. Press, 1981.

[16] Amaury Lambert, *Population dynamics and random genealogies*, Lecture notes for a course in Guanajuato, Mexico (CIMAT), 2006.

[17] Jean–Francois Le Gall, *Spatial branching processes, random snakes and partial differential equations*, Lectures in Mathematics, **ETH** Zürich, Birkhäuser, 1999.

[18] Russell Lyons, Yuval Peres, *Probability on trees and networks*, a book in progress, http://mypage.iu.edu/ rdlyons/prbtree/prbtree.html

[19] Neil O'Connell, Yule process approximation for the skeleton of a branching process, *J. Appl. Probab.* **30**, 725–729, 1993.

[20] Neil O'Connell, The genealogy of branching processes and the age of our most recent common ancestor, *Adv. Appl. Probab.* **27**, 418–442, 1995.

[21] Etienne Pardoux, *Processus de Markov et applications*, Dunod, 2007.

[22] Daniel Revuz, Marc Yor, *Continuous martingales and Brownian motion*, Springer 1991.