

# Chapter 5

## Wright-Fisher Processes

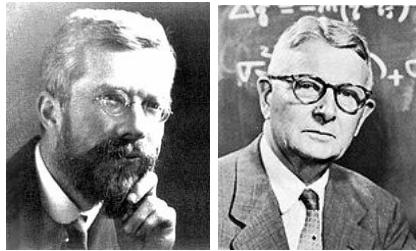


Figure 5.1: Fisher and Wright

### 5.1 Introductory remarks

The BGW processes and birth and death processes we have studied in the previous chapters have the property that

$$(5.1) \quad X_n \rightarrow 0 \text{ or } \infty, \quad a.s.$$

A more realistic model is one in which the population grows at low population densities and tends to a steady state near some constant value. The Wright-Fisher model that we consider in this chapter (and the corresponding Moran continuous time model) assume that the total population remains at a constant level  $N$  and focusses on the changes in the relative proportions of the different types. Fluctuations of the total population, provided that they do not become too small, result in time-varying resampling rates in the Wright-Fisher model but do not change the main qualitative features of the conclusions.

The branching model and the Wright-Fisher idealized models are complementary. The branching process model provides an important approximation in two cases:

- If the total population density becomes small then the critical and near critical branching process provides an useful approximation to compute extinction probabilities.
- If a new type emerges which has a competitive advantage, then the supercritical branching model provides a good approximation to the growth of this type as long as its contribution to the total population is small.

Models which incorporate multiple types, supercritical growth at low densities and have non-trivial steady states will be discussed in a later chapter. The advantage of the idealized models we discuss here is the possibility of explicit solutions.

## 5.2 Wright-Fisher Markov Chain Model

The classical neutral Wright-Fisher (1931) model is a discrete time model of a population with constant size  $N$  and types  $E = \{1, 2\}$ . Let  $X_n$  be the number of type 1 individuals at time  $n$ . Then  $X_n$  is a Markov chain with state space  $\{0, \dots, N\}$  and transition probabilities:

$$P(X_{n+1} = j | X_n = i) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}, \quad j = 0, \dots, N.$$

In other words at generation  $n + 1$  this involves binomial sampling with probability  $p = \frac{X_n}{N}$ , that is, the current empirical probability of type 1. Looking backwards from the viewpoint of generation  $n + 1$  this can be interpreted as having each of the  $N$  individuals of the  $(n + 1)$ st generation “pick their parents at random” from the population at time  $n$ .

Similarly, the neutral  $K$ -allele Wright Fisher model with types  $E_K = \{e_1, \dots, e_K\}$  is given by a Markov chain  $X_n$  with state space  $\setminus(E_K)$  (counting measures) and

$$(5.2) \quad P(X_{n+1} = (\beta_1, \dots, \beta_K) | X_n = (\alpha_1, \dots, \alpha_K)) \\ = \frac{N!}{\beta_1! \beta_2! \dots \beta_K!} \left(\frac{\alpha_1}{N}\right)^{\beta_1} \dots \left(\frac{\alpha_K}{N}\right)^{\beta_K}$$

In this case the binomial sampling is simply replaced by multinomial sampling.

Consider the multinomial distribution with parameters  $(N, p_1, \dots, p_K)$ . Then the moment generating function is given by

$$(5.3) \quad M(\theta_1, \dots, \theta_K) = E(\exp(\sum_{i=1}^K \theta_i X_i)) = \left(\sum_{i=1}^K p_i e^{\theta_i}\right)^N$$

Then

$$(5.4) \quad E(X_i) = Np_i, \quad \text{Var}(X_i) = Np_i(1 - p_i),$$

and

$$(5.5) \quad \text{Cov}(X_i, X_j) = -N p_i p_j, \quad i \neq j.$$

**Remark 5.1** We can relax the assumptions of the Wright-Fisher model in two ways. First, if we relax the assumption of the total population constant, equal to  $N$ , we obtain a Fisher-Wright model with variable resampling rate (e.g. Donnelly and Kurtz [167] and Kaj and Krone [359]).

To introduce the second way to relax the assumptions note that we can obtain the Wright-Fisher model as follows. Consider a population of  $N$  individuals in generation  $n$  with possible types in  $E_K$ ,  $Y_1^n, \dots, Y_N^n$ . Assume each individual has a Poisson number of offspring with mean  $m$ ,  $(Z_1, \dots, Z_N)$  and the offspring is of the same type as the parent. Then

$$\text{conditioned on } \sum_{i=1}^N Z_i = N,$$

the resulting population  $(Y_1^{(n+1)}, \dots, Y_N^{(n+1)})$  is multinomial  $(N; \frac{1}{N}, \dots, \frac{1}{N})$ , that is, we have a multitype (Poisson) branching process conditioned to have constant total population  $N$ . If we then define

$$(5.6) \quad p_{n+1}(i) = \frac{1}{N} \sum_{j=1}^N 1(Y_j^{(n+1)} = i), \quad i = 1, \dots, K,$$

then  $(p_{n+1}(1), \dots, p_{n+1}(K))$  is multinomial  $(N; p_n(1), \dots, p_n(K))$  where

$$(5.7) \quad p_n(i) = \frac{1}{N} \sum_{j=1}^N 1(Y_j^n = i), \quad i = 1, \dots, K.$$

We can generalize this by assuming that the offspring distribution of the individuals is given by a common distribution on  $\mathbb{N}_0$ . Then again conditioned the total population to have constant size  $N$  the vector  $(Y_1^{n+1}, \dots, Y_N^{n+1})$  is exchangeable but not necessarily multinomial. This exchangeability assumption is the basis of the Cannings Model (see e.g. Ewens [243]).

A basic phenomenon of neutral Wright-Fisher without mutation is *fixation*, that is, the elimination of all but one type at a finite random time. To see this note that for each  $j = 1, \dots, K$ ,  $\delta_j \in \mathcal{P}(E_K)$  are absorbing states and  $X_n(j)$  is a martingale. Therefore  $X_n \rightarrow X_\infty$ , a.s. Since  $\text{Var}(X_{n+1}) = NX_n(1 - X_n)$ , this means that  $X_\infty = 0$  or  $1$ , a.s. and  $X_n$  must be 0 or 1 after a finite number of generations (since only the values  $\frac{k}{N}$  are possible).

### 5.2.1 Types in population genetics

The notion of type in population biology is based on the *genotype*. The genotype of an individual is specified by the *genome* and this codes *genetic information* that passes, possibly modified, from parent to offspring (parents in sexual reproduction). The genome consists of a set of *chromosomes* (23 in humans). A chromosome is a single molecule of DNA that contains many genes, regulatory elements and other nucleotide sequences. A given position on a chromosome is called a *locus* (*loci*) and may be occupied by one or more *genes*. Genes code for the production of a protein. The different variations of the gene at a particular locus are called *alleles*. The ordered list of loci for a particular genome is called a *genetic map*. The *phenotype* of an organism describes its structure and behaviour, that is, how it interacts with its environment. The relationship between genotype and phenotype is not necessarily 1-1. The field of *epigenetics* studies this relationship and in particular the mechanisms during cellular development that produce different outcomes from the same genetic information.

*Diploid individuals* have two homologous copies of each chromosome, usually one from the mother and one from the father in the case of sexual reproduction. Homologous chromosomes contain the same genes at the same loci but possibly different alleles at those genes.

### 5.2.2 Finite population resampling in a diploid population

For a diploid population with  $K$ -alleles  $e_1, \dots, e_K$  at a particular gene we can focus on the set of types given by  $E_K^{2\circ}$  where  $\binom{K(K+1)}{2}$  is the set of unordered pairs  $(e_i, e_j)$ . The genotype  $(e_i, e_j)$  is said to be homozygous (at the locus in question) if  $e_i = e_j$ , otherwise heterozygous.

Consider a finite population of  $N$  individuals. Let

$$P_{ij} = \text{proportion of type } (e_i, e_j)$$

Then,  $p_i$ , the proportion of allele  $e_i$  is

$$p_i = P_{ii} + \frac{1}{2} \sum_{j \neq i} P_{ij}.$$

The probability  $\{P_{ij}\}$  on  $E_K^{2\circ}$  is said to be a Hardy-Weinberg equilibrium if

$$(5.8) \quad P_{ij} = (2 - \delta_{ij})p_i p_j.$$

This is what is obtained if one picks independently the parent types  $e_i$  and  $e_j$  from a population having proportions  $\{p_i\}$  ( in the case of sexual reproduction this corresponds to “random mating”).

Consider a diploid Wright-Fisher model of with  $N$  individuals therefore  $2N$  genes with random mating. This means that an individual at generation  $(n + 1)$  has two genes randomly chosen from the  $2N$  genes in generation  $n$ .

In order to introduce the notions of identity by descent and genealogy we assume that in generation 0 each of the  $2N$  genes correspond to different alleles. Now consider generation  $n$ . What is the probability,  $F_n$ , that an individual is homozygous, that is, two genes selected at random are of the same type (homozygous)? This will occur only if they are both descendants of the same gene in generation 0.

First note that in generation 1, this means that an individual is homozygous only if the same allele must be selected twice and this has probability  $\frac{1}{2N}$ . In generation  $n + 1$  this happens if the same gene is selected twice or if different genes are selected from generation  $n$  but they are identical alleles. Therefore,

$$(5.9) \quad F_1 = \frac{1}{2N}, \quad F_n = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F_{n-1}.$$

Let  $H_n := 1 - F_n$  (heterozygous). Then

$$(5.10) \quad H_1 = 1 - \frac{1}{2N}, \quad H_n = \left(1 - \frac{1}{2N}\right)H_{n-1}, \quad H_n = \left(1 - \frac{1}{2N}\right)^n$$

Two randomly selected genes are said to be *identical by descent* if they are the same allele. This will happen if they have a *common ancestor*. Therefore if  $T_{2,1}$  denotes the time in generations back to the common ancestor we have

$$(5.11) \quad P(T_{2,1} > n) = H_n = \left(1 - \frac{1}{2N}\right)^n, \quad n = 0, 1, 2, \dots,$$

$$(5.12) \quad P(T_{2,1} = n) = \frac{1}{2N} \left(1 - \frac{1}{2N}\right)^{n-1}, \quad n = 1, 2, \dots$$

Similarly, for  $k$  randomly selected genes they are identical by descent if they all have a common ancestor. We can consider the time  $T_{k,1}$  in generations back to the most recent common ancestor of  $k$  individuals randomly sampled from the population. We will return to discuss the distribution of  $T_{k,1}$  in the limit as  $N \rightarrow \infty$  in Chapter 9.

### 5.2.3 Diploid population with mutation and selection

In the previous section we considered only the mechanism of resampling (genetic drift). In addition to genetic drift the basic genetic mechanisms include mutation, selection and recombination. In this subsection we consider the Wright-Fisher model incorporating mutation and selection.

For a diploid population of size  $N$  with mutation, selection and resampling the *reproduction cycle* can be modelled as follows (cf [225], Chap. 10). We assume that in generation 0 individuals have genotypic proportions  $\{P_{ij}\}$  and therefore the proportion of type  $i$  (in the population of  $2N$  genes) is

$$p_i = P_{ii} + \frac{1}{2} \sum_{j \neq i} P_{ij}.$$

Stage I:

In the first stage diploid cells undergo meiotic division producing haploid gametes (single chromosomes), that is, meiosis reduces the number of sets of chromosomes from two to one. The resulting *gametes* are haploid cells; that is, they contain one half a complete set of chromosomes. When two gametes fuse (in animals typically involving a sperm and an egg), they form a *zygote* that has two complete sets of chromosomes and therefore is diploid. The zygote receives one set of chromosomes from each of the two gametes through the *fusion* of the two gametes. By the assumption of random mating, then in generation 1 this produces zygotes in *Hardy-Weinberg proportions*  $(2 - \delta_{ij})p_i p_j$ .

Stage II: Selection and Mutation.

*Selection.* The resulting zygotes can have different viabilities for survival. The viability of  $(e_i, e_j)$  has viability  $V_{ij}$ . The proportions of surviving zygotes are proportional to the product of the viabilities and the Hardy-Weinberg proportions, that is,

$$(5.13) \quad P_{k,\ell}^{\text{sel}} = \frac{V_{k\ell} \cdot (2 - \delta_{k\ell})p_k p_\ell}{\sum_{k' \leq \ell} (2 - \delta_{k'\ell'}) V_{k'\ell'} p_{k'} p_{\ell'}}$$

*Mutation.* We assume that each of the 2 gametes forming zygote can (independently) mutate with probability  $p_m$  and that if a gamete of type  $e_i$  mutates then it produces a gamete of type  $e_j$  with probability  $m_{ij}$ .

$$\begin{aligned} (5.14) \quad P_{ij}^{\text{sel,mut}} &= (1 - \frac{1}{2}\delta_{ij}) \sum_{k \leq \ell} (m_{ki} m_{\ell j} + m_{kj} m_{\ell i}) P_{k\ell}^{\text{sel}} \\ &= (1 - \frac{1}{2}\delta_{ij}) \sum_{k \leq \ell} (m_{ki} m_{\ell j} + m_{kj} m_{\ell i}) \frac{V_{k\ell} \cdot (2 - \delta_{k\ell})p_k p_\ell}{\sum_{k' \leq \ell} (2 - \delta_{k'\ell'}) V_{k'\ell'} p_{k'} p_{\ell'}} \end{aligned}$$

Stage III: Resampling. Finally random sampling reduces the population to  $N$  adults with proportions  $P_{ij}^{\text{next}}$  where

$$(5.15) \quad (P_{ij}^{\text{next}})_{i \leq j} \sim \frac{1}{N} \text{multinomial}(N, (P_{ij}^{\text{sel,mut}})_{i \leq j}).$$

We then obtain a population of  $2N$  gametes with proportions

$$(5.16) \quad p_i^{\text{next}} = P_{ii}^{\text{next}} + \frac{1}{2} \sum_{j \neq i} P_{ij}^{\text{next}}.$$

Therefore we have defined the process  $\{X_n^N\}_{n \in \mathbb{N}}$  with state space  $\mathcal{P}^N(E_K)$ . If  $X_n^N$  is a Markov chain we defined the transition function

$$P(X_{n+1}^N = (p_1^{\text{next}}, \dots, p_K^{\text{next}}) | X_n^N = (p_1, \dots, p_K)) = \pi_{p_1, \dots, p_K}(p_1^{\text{next}}, \dots, p_K^{\text{next}})$$

where the function  $\pi$  is obtained from (5.14), (5.15), (5.16). See Remark 5.7.

## 5.3 Diffusion Approximation of Wright-Fisher

### 5.3.1 Neutral 2-allele Wright-Fisher model

As a warm-up to the use of diffusion approximations we consider the case of 2 alleles  $A_1, A_2$ , ( $k = 2$ ). Let  $X_n^N$  denote the number of individuals of type  $A_1$  at the  $n$ th generation. Then as above  $\{X_n^N\}_{n \in \mathbb{N}}$  is a Markov chain.

**Theorem 5.2** (*Neutral case without mutation*) *Assume that  $N^{-1}X_0^N \rightarrow p_0$  as  $N \rightarrow \infty$ . Then*

$$\{p_N(t) : t \geq 0\} \equiv \{N^{-1}X_{[Nt]}^N, t \geq 0\} \implies \{p(t) : t \geq 0\}$$

where  $\{p(t) : t \geq 0\}$  is a Markov diffusion process with state space  $[0, 1]$  and with generator

$$(5.17) \quad Gf(p) = \frac{1}{2}p(1-p)\frac{d^2}{dp^2}f(p)$$

if  $f \in C^2([0, 1])$ . This is equivalent to the pathwise unique solution of the SDE

$$\begin{aligned} dp(t) &= \sqrt{p(t)(1-p(t))}dB(t) \\ p(0) &= p_0. \end{aligned}$$

**Proof.** Note that in this case  $X_{n+1}^N$  is Binomial( $N, p_n$ ) where  $p_n = \frac{X_n^N}{N}$ . Then from the Binomial formula,

$$\begin{aligned} E_{X_n^N}\left(\frac{X_{n+1}^N}{N}\right) &= \frac{X_n^N}{N} \\ E_{X_n^N}\left[\left(\frac{X_{n+1}^N}{N} - \frac{X_n^N}{N}\right)^2 \mid \frac{X_n^N}{N}\right] &= \frac{1}{N} \left(\frac{X_n^N}{N} \left(1 - \frac{X_n^N}{N}\right)\right). \end{aligned}$$

We can then verify that

$$(5.18) \quad \{p_N(t) := N^{-1}X_{[Nt]}^N : t \geq 0\} \text{ is a martingale}$$

with

$$\begin{aligned} (5.19) \quad E(p_N(t_2) - p_N(t_1))^2 &= E \sum_{k=\lfloor Nt_1 \rfloor}^{\lfloor Nt_2 \rfloor} (p_N(\frac{k+1}{N}) - p_N(\frac{k}{N}))^2 \\ &= \frac{1}{N} E \sum_{k=\lfloor Nt_1 \rfloor}^{\lfloor Nt_2 \rfloor} p_N(\frac{k}{N})(1 - p_N(\frac{k}{N})) \end{aligned}$$

and then that

$$(5.20) \quad M_N(t) = p_N^2(t) - \frac{1}{N} \sum_{k=\lfloor Nt_1 \rfloor}^{\lfloor Nt_2 \rfloor} p_N\left(\frac{k}{N}\right)\left(1-p_N\left(\frac{k}{N}\right)\right)$$

is a martingale.

Let  $P_{p_N}^N \in \mathcal{P}(D_{[0,1]}([0, \infty)))$  denote the probability law of  $\{p_N(t)\}_{t \geq 0}$  with  $p_N(0) = p_N$ . From this we can prove that the sequence  $\{P_{p_N(0)}^N\}_{N \in \mathbb{N}}$  is tight on  $\mathcal{P}(D_{[0,\infty)}([0, 1]))$ . To verify this as in the previous chapter we use Aldous criterion  $P_{p_N(0)}^N(p_N(\tau_N + \delta_N) - p_N(\tau_N) > \varepsilon) \rightarrow 0$  as  $N \rightarrow \infty$  for any stopping times  $\tau_N \leq T$  and  $\delta_N \downarrow 0$ . This follows easily from the strong Markov property, (5.19) and Chebyshev's inequality. Since the processes  $p_N(\cdot)$  are bounded it then follows that for any limit point  $P_{p_0}$  of  $P_{p_N(0)}^N$  we have

(5.21)

$\{p(t)\}_{t \geq 0}$  is a bounded martingale with  $p(0) = p_0$  and with increasing process  

$$\langle p \rangle_t = \int_0^t p(s)(1-p(s))ds.$$

Since the largest jump of  $p_N(\cdot)$  goes to 0 as  $N \rightarrow \infty$  the limiting process is continuous (see Theorem 17.14 in the Appendix). Also, by the Burkholder-Davis-Gundy inequality we have

$$(5.22) \quad E((p(t_2) - p(t_1))^4) \leq \text{const} \cdot (t_2 - t_1)^2,$$

so that  $p(t)$  satisfies Kolmogorov's criterion for a.s. continuous.

We can then prove that there is a unique solution to this *martingale problem*, that is, for each  $p$  there exists a unique probability measure on  $C_{[0,\infty)}([0, \infty))$  satisfying (5.21) and therefore this defines a Markov diffusion process with generator (5.17).

The uniqueness can be proved by determining all joint moments of the form

$$(5.23) \quad E_p((p(t_1))^{k_1} \dots (p(t_\ell))^{k_\ell}), \quad 0 \leq t_1 < t_2 < \dots < t_\ell, \quad k_i \in \mathbb{N}$$

by solving a closed system of differential equations. It can also be proved using duality and this will be done in detail below (Chapter 7) in a more general case.)

■ We now give an illustrative application of the diffusion approximation, namely the calculation of expected fixation times.

**Corollary 5.3** (*Expected fixation time.*) *Let  $\tau := \inf\{t : p(t) \in \{0, 1\}\}$  denote the fixation time of the diffusion process. Then*

$$E_p[\tau] = f(p) = -[p \log p + (1-p) \log(1-p)].$$

**Proof.** Let  $f \in C^2([0, 1])$ ,  $f(0) = f(1) = 0$ . Let  $g(p) := \int_0^\infty T_s f(p) ds$ , and note that as  $f \uparrow 1_{(0,1)}$  this converges to the expected time spent in  $(0, 1)$ . Since  $p(t) \rightarrow \{0, 1\}$  as  $t \rightarrow \infty$ , a.s., we can show that

$$G \left( \int_0^t T_s f(p) ds \right) = \int_0^t GT_s f(p) ds = T_t f(p) - f(p) \rightarrow 0 - f(p) \text{ as } t \rightarrow \infty,$$

that is,

$$Gg(p) = -f(p)$$

where  $G$  is given by (5.17).

Applying this to a sequence of  $C^2$  functions increasing to  $1_{(0,1)}$  we get

$$\begin{aligned} E_p(\tau) &= \int_0^\infty P_p(\tau > t) dt \\ &= \int_0^\infty T_t 1_{(0,1)}(p) dt \\ &= g(p) \end{aligned}$$

We then obtain  $g(p)$  by solving the differential equation  $Gg(p) = -1$  with boundary conditions  $g(0) = g(1) = 0$  to obtain

$$E_p[\tau] = g(p) = -[p \log p + (1-p) \log(1-p)].$$

■

Let  $\tau_N$  denote the fixation time for  $N^{-1}X_{[Nt]}$ . We want to show that

$$(5.24) \quad E_{\frac{X_0^N}{N}}[\tau_N] \rightarrow E_{p_0}[\tau] \text{ if } \frac{X_0^N}{N} \rightarrow p \text{ and } N \rightarrow \infty.$$

However note that  $\tau$  is not a continuous function on  $D([0, \infty), [0, 1])$ . The weak convergence can be proved for  $\tau^\varepsilon = \inf\{t : p(t) \notin (\varepsilon, 1-\varepsilon)\}$  (because there is no “slowing down” here). To complete the proof it can be verified that for  $\delta > 0$

$$(5.25) \quad \lim_{\varepsilon \rightarrow 0} \limsup_{N \rightarrow \infty} P(|\tau_N^\varepsilon - \tau_N| > \delta) = 0$$

(see Ethier and Kurtz, [225], Chapt. 10, Theorem 2.4).

### 2-allele Wright-Fisher with mutation

For each  $N$  consider a Fisher-Wright population of size  $M_N$  and with mutation rates  $m_{12} = \frac{u}{N} A_1 \rightarrow A_2$  and  $m_{21} = \frac{v}{N} A_2 \rightarrow A_1$ .

In this case  $X_{n+1}^N$  is Binomial( $M_N, p_n$ ) with

$$(5.26) \quad p_n = \left(1 - \frac{u}{N}\right) \frac{X_n^N}{M_N} + \frac{v}{N} \left(1 - \frac{X_n^N}{M_N}\right).$$

We now consider

$$(5.27) \quad p_N(t) = \frac{1}{M_N} X_{\lfloor Nt \rfloor}.$$

If we assume that

$$(5.28) \quad \gamma = \lim_{N \rightarrow \infty} \frac{N}{M_N},$$

then the diffusion approximation is given by the diffusion process  $p_t$  with generator

$$(5.29) \quad Gf(p) = \frac{\gamma}{2} p(1-p) \frac{\partial^2}{\partial p^2} + [-up + v(1-p)] \frac{\partial}{\partial p}.$$

In this case the domain of the generator involves boundary conditions at 0 and 1 (see [225], Chap. 8, Theorem 1.1) but we will not need this.

**Remark 5.4** Note that the diffusion coefficient is proportional to the inverse population size. Below for more complex models we frequently think of the diffusion coefficient in terms of inverse effective population size.

### Error estimates

Consider a haploid Wright-Fisher population of size  $M$  with mutation rates  $m_{12} = u$ ,  $m_{21} = v$ .

Let  $p_t^{(M,u,v)}$  denote the diffusion process with generator (5.29) with  $\gamma = \frac{1}{M}$ . Then if  $\alpha, \beta \geq 0$ , the law of

$$(5.30) \quad \{Z_t^{(\alpha,\beta)}\}_{t \geq 0} := p_t^{(M, \frac{\alpha}{M}, \frac{\beta}{M})}$$

is independent of  $M$  and is a Wright-Fisher diffusion with generator

$$(5.31) \quad Gf(p) = \frac{\gamma}{2} p(1-p) \frac{\partial^2}{\partial p^2} + [-\alpha p + \beta(1-p)] \frac{\partial}{\partial p}.$$

The assumption of mutation rates of order  $O(\frac{1}{N})$  corresponds to the case in which both mutation and genetic drift are of the same order and appear in the limit as population sizes goes to  $\infty$ . Other only one of the two mechanisms appears in the limit as  $N \rightarrow \infty$ .

On the other hand one can consider the diffusion process as an approximation to the finite population model. Ethier and Norman ([228]) obtained an estimate of the error due to the diffusion approximation in the calculation of the expected value of a smooth function of the nth generation allelic frequency.

To formulate their result consider the Wright-Fisher Markov chain model  $\{X_n^{(M,u,v)}\}$  with population size  $M$  and one-step mutation probabilities  $m_{12} = u$ ,  $m_{21} = v$  and  $p_t^{(M,u,v)}$  the Wright-Fisher diffusion with generator (5.29) with  $\gamma = \frac{1}{M}$ .

**Theorem 5.5** (*Ethier and Norman [228]*) *Assume that  $f \in C^6([0, 1])$ . Then for  $n \in \mathbb{N}_0$ ,*

$$(5.32) \quad \begin{aligned} & |E_x(f(X_n^{(M,u,v)}) - E_x(f(p_n^{(M,u,v)}))| \\ & \leq \frac{\max(u, v)}{2} \cdot \|f^{(1)}\| + \frac{1}{M} \left( \frac{1}{8} \|f^{(2)}\| + \frac{1}{216\sqrt{3}} \|f^{(3)}\| \right) \\ & \quad + \frac{9 \max(u^2, v^2)}{2} \left( \sum_{j=1}^6 \|f^{(j)}\| \right) + \frac{7}{16M^2} \sum_{j=2}^6 \|f^{(j)}\| \end{aligned}$$

where  $\|f^{(j)}\|$  is the sup of the  $j$ th derivative of  $f$ .

We do not include a proof but sketch the main idea. Let

$$(5.33) \quad (S_n f)(x) := E_x[f(X_n^{(M,u,v)})],$$

$$(5.34) \quad (T_t f)(x) := E_x[f(p_t^{(M,u,v)})].$$

If  $g \in C_b^6([0, \infty))$ , then we have the Taylor expansions

$$(5.35) \quad (T_1 g)(x) = g(x) + (Gg)(x) + \frac{G^2 g(x)}{2} + \omega_2 \frac{\|G^3 g\|}{6}, \quad |\omega_2| \leq 1$$

and

$$(5.36)$$

$$(S_1 g)(x) = g(x) + \sum_{j=1}^5 E_x[(X_1^{(M,u,v)} - x)^j] \frac{g^{(j)}(x)}{j!} + \omega_1 E_x[(X_1^{(M,u,v)} - x)^6] \frac{\|g^{(6)}\|}{6!}, \quad |\omega_1| \leq 1.$$

We then obtain

$$(5.37) \quad \|S_1 g - T_1 g\|_M \leq \sum_{j=1}^6 \gamma_j \|g^{(j)}\|$$

for some constants  $\gamma_j$ .

The proof is then completed using the inequality

$$(5.38) \quad \|S_n f - T_n f\|_M \leq \sum_{k=0}^{n-1} \|(S_1 - T_1) T_k\|_M$$

where  $\|\cdot\|_M$  is the sup norm on  $\{\frac{j}{M} : j = 1, \dots, M\}$ .

### 5.3.2 K-allele Wright-Fisher Diffusion

Now consider the  $K$ -allele Wright-Fisher Markov chain  $\{X_k^{2N}\}_{k \in \mathbb{N}}$  with  $2N$  gametes present in each generation and assume that the mutation rates and fitnesses satisfy

$$(5.39) \quad m_{ij} = \frac{q_{ij}}{2N}, \quad i \neq j, \quad m_{ii} = 1 - \frac{m}{N}, \quad m = \sum_j q_{ij}$$

$$(5.40) \quad V_{ij} = 1 + \frac{\sigma_{ij}}{2N} + O\left(\frac{1}{N^2}\right).$$

We now consider the Markov process with state space

$$(5.41) \quad \Delta_{K-1} := \{(p_1, \dots, p_K) : p_i \geq 0, \sum_{i=1}^K p_i = 1\}.$$

defined by

$$(5.42) \quad \{p^{2N}(t) : t \geq 0\} \equiv \left\{ \frac{1}{2N} X_{[2Nt]}^{2N}, t \geq 0 \right\}.$$

**Theorem 5.6** Assume that  $2N^{-1} X_0^{2N} \rightarrow p$  as  $N \rightarrow \infty$  in  $\Delta_{K-1}$ .

Then the laws of the càdlàg processes  $\{p_N(t) := \frac{1}{2N} X_t^{2N}\}_{t \geq 0}$  are tight and for any limit point and function  $f(p) = f(p_1, \dots, p_{K-1}) \in C^2(\Delta_{K-1})$ ,

$$(5.43) \quad M_f(t) := f(p(t)) - \int_0^t G^K f(p(s)) ds \quad \text{is a martingale}$$

where

$$(5.44) \quad \begin{aligned} & G^K f(p) \\ &= \frac{1}{2} \sum_{i,j=1}^{K-1} p_i (\delta_{ij} - p_j) \frac{\partial^2 f(p)}{\partial p_i \partial p_j} \\ &+ \sum_{i=1}^{K-1} \left[ \left( m \left( \sum_{j=1, j \neq i}^K q_{ji} p_j - p_i \right) + p_i \left( \sum_{j=1}^K \sigma_{ij} p_j - \sum_{k,\ell} \sigma_{k\ell} p_k p_\ell \right) \right] \frac{\partial f(p)}{\partial p_i}. \end{aligned}$$

The martingale problem (5.43) has a unique solution which determines a Markov diffusion process  $\{p(t) : t \geq 0\}$  called the K-allele Wright-Fisher diffusion.

**Proof.** Following the pattern of the 2-allele neutral case the proof involves three steps which we now sketch.

Step 1. The tightness of the probability laws  $P^N$  of  $\{p^{2N}(\cdot)\}$  on  $D_{\Delta_{K-1}}([0, \infty))$  can be proved using Aldous criterion.

Step 2. Proof that for any limit point of  $P^N$  and  $i = 1, \dots, K$

$$(5.45) \quad M_i(t) := p_i(t) - p_i(0) - \int_0^t \left[ m \left( \sum_{j=1}^K q_{ji} p_j(s) - p_i(s) \right) + p_i(s) \left( \sum_{j=1}^K \sigma_{ij} p_j(s) - \sum_{k,\ell} \sigma_{k\ell} p_k(s) p_\ell(s) \right) \right] ds$$

is a martingale with quadratic covariation process

$$(5.46) \quad \langle M_i, M_j \rangle_t = \frac{1}{2} \int_0^t p_j(s) (\delta_{ij} - p_i(s)) ds$$

To verify this, let  $\mathcal{F}_{\frac{k}{2N}} = \sigma\{p_i^{2N}(\frac{\ell}{2N}) : \ell \leq k, i = 1, \dots, K\}$ . Then we have for  $k \in \mathbb{N}$

$$(5.47) \quad \begin{aligned} & E[p_i^{2N}(\frac{k+1}{2N}) - p_i^{2N}(\frac{k}{2N}) \mid \mathcal{F}_{\frac{k}{2N}}] \\ &= \frac{1}{2N} \left[ m \left( \sum_{j=1, j \neq i}^K \frac{q_{ji}}{m} p_j^{2N}(\frac{k}{2N}) - p_i^{2N}(\frac{k}{2N}) \right) \right. \\ & \quad \left. + \left( \sum_{j=1}^K \sigma_{ij} p_j^{2N}(\frac{k}{2N}) - \sum_{k,\ell=1}^K \sigma_{k\ell} p_k^{2N}(\frac{k}{2N}) p_\ell^{2N}(\frac{k}{2N}) \right) \right] \\ &+ o(\frac{1}{2N}) \end{aligned}$$

$$(5.48) \quad Cov(p_i^{2N}(\frac{k+1}{2N}), p_j^{2N}(\frac{k+1}{2N}) \mid \mathcal{F}_{\frac{k}{2N}}) = \frac{p_i^{2N}}{2N}(\frac{k}{2N})(\delta_{ij} - p_j^{2N}(\frac{k}{2N})) + o(\frac{1}{N})$$

**Remark 5.7** The Markov property for  $X_n^N$  follows if in the resampling step the  $\{P_{ij}^{\text{sel,mut}}\}$  are in Hardy-Weinberg proportions which implies that the  $\{p_i^{\text{next}}\}$  are

$$(5.49) \quad \text{multinomial}(2N, (p_1^{\text{sel,mut}}, \dots, p_K^{\text{sel,mut}})).$$

This is true without selection or with multiplicative selection  $V_{ij} = V_i V_j$  (which leads to haploid selection in the diffusion limit) but not in general. In the diffusion limit this can sometimes be dealt with by the  $O(\frac{1}{N^2})$  term in (5.40). In general the diffusion limit result remains true but the argument is more subtle. The idea is that the selection-mutation changes the allele frequencies more slowly than the mechanism of Stages I and III which rapidly bring the frequencies to Hardy-Weinberg equilibrium - see [225], Chap. 10, section 3.

Then for each  $N$  and  $i$

$$(5.50) \quad M_i^{2N}(t) := p_i^{2N}(t) - p_i^{2N}(0) - \int_0^t \left[ m \left( \sum_{j=1, j \neq i}^K q_{ji} p_j^{2N}(s) - p_i^{2N}(s) \right) + p_i^{2N}(s) \left( \sum_{j=1}^K \sigma_{ij} p_j^{2N}(s) - \sum_{k,\ell} \sigma_{k\ell} p_k^{2N}(s) p_\ell^{2N}(s) \right) \right] ds^N + o\left(\frac{1}{N}\right)$$

is a martingale and for  $i, j = 1, \dots, K$

$$(5.51) \quad E[(M_i^{2N}(t_2) - M_i^{2N}(t_1))(M_j^{2N}(t_2) - M_j^{2N}(t_1))] \\ = \frac{1}{2N} E \sum_{k=\lfloor 2Nt_1 \rfloor}^{\lfloor 2Nt_2 \rfloor} p_i^{2N}\left(\frac{k}{2N}\right) \left( \delta_{ij} - p_j^{2N}\left(\frac{k}{2N}\right) \right) + o\left(\frac{1}{N}\right).$$

Step 3. Proof that there exists a unique probability measure on  $C_{\Delta_{K-1}}([0, \infty))$  such that (5.45) and (5.46) are satisfied.

Uniqueness can be proved in the neutral case,  $\sigma \equiv 0$ , by showing that moments are obtained as unique solutions of a closed system of differential equations.

■

**Remark 5.8** *The uniqueness when  $\sigma$  is not zero follows from the dual representation developed in the next chapter.*

## 5.4 Stationary measures

A special case of a theorem in Section 8.3 implies that if the matrix  $(q_{ij})$  is irreducible, then the Wright-Fisher diffusion is ergodic with unique stationary distribution.

### 5.4.1 The Invariant Measure for the neutral K-alleles WF Diffusion

Consider the neutral  $K$ -type Wright-Fisher diffusion with type-independent mutation (Kingman's "house-of-cards" mutation model) with generator

$$G^K f(p) = \frac{1}{2} \sum_{i,j=1}^{K-1} p_i (\delta_{ij} - p_j) \frac{\partial^2 f(p)}{\partial p_i \partial p_j} + \frac{\theta}{2} \sum_{i=1}^{K-1} (\nu_i - p_i) \frac{\partial f(p)}{\partial p_i}.$$

where the *type-independent mutation* kernel is given by  $\nu \in \Delta_{k-1}$ .

**Theorem 5.9** (*Wright [615], Griffiths [280]*) *The Dirichlet distribution  $D(p_1, \dots, p_n)$  on  $\Delta_{K-1}$  with density*

$$\Pi_K(dp) = \frac{\Gamma(\theta_1 + \dots + \theta_K)}{\Gamma(\theta_1) \dots \Gamma(\theta_K)} p_1^{\theta_1-1} \dots p_K^{\theta_K-1} dp_1 \dots dp_{K-1}$$

$$\theta_j = \theta\nu_j, \nu \in \mathcal{P}(1, \dots, K)$$

is a reversible stationary measure for the neutral  $K$ -alleles WF diffusion with  $\gamma = 1$ .

In the case  $K = 2$  this is the Beta distribution

$$(5.52) \quad \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)} x_1^{\theta_1-1} (1-x_1)^{\theta_2-1} dx_1.$$

**Proof.** (cf. [232]) Reversibility and stationarity means that when  $\Pi_K$  is the initial distribution, then  $\{p(t) : 0 \leq t \leq t_0\}$  has the same distribution as  $\{p(t_0 - t) : 0 \leq t \leq t_0\}$ . In terms of the strongly continuous semigroup  $\{T(t)\}$  on  $C(\Delta_{K-1})$  generated by  $G$  a necessary and sufficient condition (see Fukushima and Stroock (1986) [259]) for reversibility with respect to  $\Pi_K$  is that

$$\int g T(t) f d\Pi_K = \int f T(t) g d\Pi_K \quad \forall f, g \in C(\Delta_{K-1}), t \geq 0$$

or equivalently that

$$\int g G f d\Pi_K = \int f G g d\Pi_K \quad \forall f, g \in D(G)$$

or for  $f, g$  in a core for  $G$  (see Appendix I).

Since the space of polynomials in  $p_1, \dots, p_K$  is a core for  $G$  it suffices by linearity to show that

$$\int g G f d\Pi = \int f G g d\Pi \quad \forall f = f_\alpha, g = f_\beta$$

where  $f_\alpha = p_1^{\alpha_1} \dots p_K^{\alpha_K}$ . Let  $|\alpha| = \sum \alpha_i$ .

Then

$$\begin{aligned} & \int f_\beta G f_\alpha d\Pi_K \\ &= \frac{1}{2} \int \left[ \sum_{i=1}^K \alpha_i (\alpha_i + \theta_i - 1) f_{\alpha+\beta-e^i} - |\alpha|(|\alpha| + \sum_{i=1}^K \theta_i - 1) f_{\alpha+\beta} \right] d\Pi_K \\ &= \frac{1}{2} \left\{ \sum_{i=1}^K \frac{\alpha_i (\alpha_i + \theta_i - 1)}{\alpha_i + \beta_i + \theta_i - 1} - \frac{|\alpha|(|\alpha| + \sum \theta_i - 1)}{|\alpha| + |\beta| + \sum \theta_i - 1} \right\} \\ &\cdot \frac{\Gamma(\alpha_1 + \beta_1 + \theta_1) \dots \Gamma(\alpha_K + \beta_K + \theta_K)}{\Gamma(|\alpha| + |\beta| + \sum \theta_i - 1)} \frac{\Gamma(\sum \theta_i)}{\Gamma(\theta_1) \dots \Gamma(\theta_K)}. \end{aligned}$$

To show that this is symmetric in  $\alpha, \beta$ , let  $h(\alpha, \beta)$  denote the expression within  $\{ \dots \}$  above. Then

$$\begin{aligned} h(\alpha, \beta) - h(\beta, \alpha) &= \sum \frac{\alpha_i^2 - \beta_i^2 + (\alpha_i - \beta_i)(\theta_i - 1)}{\alpha_i + \beta_i + \theta_i - 1} - \frac{|\alpha|^2 - |\beta|^2 + (|\alpha| - |\beta|)(\sum \theta_i - 1)}{|\alpha| + |\beta| + \sum \theta_i - 1} \\ &= \sum (\alpha_i - \beta_i) - (|\alpha| - |\beta|) \\ &= 0 \end{aligned}$$

■

**Corollary 5.10** *Consider the mixed moments:*

$$m_{k_1, \dots, k_K} = \int \dots \int_{\Delta_{K-1}} p_1^{k_1} \dots p_K^{k_K} \Pi_K(dp)$$

*Then*

$$m_{k_1, \dots, k_K} = \frac{\Gamma(\theta_1) \dots \Gamma(\theta_K)}{\Gamma(\theta_1 + \dots + \theta_K)} \frac{\Gamma(\theta_1 + \dots + \theta_K + k_1 + \dots + k_K)}{\Gamma(\theta_1 + k_1) \dots \Gamma(\theta_K + k_K)}.$$

#### Stationary measure with selection

If selection (as in (5.44)) is added then the stationary distribution is given by the “Gibbs-like” distribution

$$(5.53) \quad \Pi_\sigma(dp) = C \exp \left( \sum_{i,j=1}^K \sigma_{ij} p_i p_j \right) \Pi_K(dp_1 \dots dp_{K-1})$$

and this is reversible. (This is a special case of a result that will be proved in a later section.)

#### 5.4.2 Convergence of stationary measures of $\{p^N\}_{N \in \mathbb{N}}$

It is of interest to consider the convergence of the stationary measures of the Wright-Fisher Markov chains to (5.53). A standard argument applied to the Wright-Fisher model is as follows.

**Theorem 5.11** Convergence of Stationary Measures. *Assume that the diffusion limit,  $p(t)$ , has a unique invariant measure,  $\nu$  and that  $\nu_N$  is an invariant measure for  $p^N(t)$ . Then*

$$(5.54) \quad \nu_N \implies \nu \text{ as } N \rightarrow \infty.$$

**Proof.** Denote by  $\{T_t\}_{t \geq 0}$  the semigroup of the Wright-Fisher diffusion. Since the state space is compact, the space of probability measure is compact. and therefore the sequence  $\nu_N$  is tight  $M_1(\Delta_{K-1})$ . Given a limit point  $\tilde{\nu}$  and a subsequence  $\nu_{N'}$  that converges weakly to  $\tilde{\nu} \in M_1(\Delta_{K-1})$  it follows that for  $f \in C(\Delta_{K-1})$ ,

$$\begin{aligned} \int T(t) f d\tilde{\nu} &= \lim_{N' \rightarrow \infty} \int T(t) f d\nu_{N'} \text{ (by } \nu_{N'} \Rightarrow \nu) \\ &= \lim_{N' \rightarrow \infty} \int T_{N'}(2N't) f d\nu_{N'} \text{ (by } p_N \Rightarrow p) \\ &= \lim_{N' \rightarrow \infty} \int f d\nu_{N'} \text{ (by inv. of } \nu_{N'}) \\ &= \int f d\tilde{\nu} \text{ (by } \nu_{N'} \Rightarrow \tilde{\nu}). \end{aligned}$$

Therefore  $\tilde{\nu}$  is invariant for  $\{T(t)\}$  and hence  $\tilde{\nu} = \nu$  by assumption of the uniqueness of the invariant measure for  $p(t)$ . That is, any limit points of  $\{\nu_N\}$  coincides with  $\nu$  and therefore  $\nu_N \Rightarrow \nu$ . ■

### Properties of the Dirichlet Distribution

1. Consistency under merging of types.

Under  $D(\theta_1, \dots, \theta_n)$ , the distribution of  $(X_1, \dots, X_k, 1 - \sum_{i=1}^k X_i)$  is

$$D(\theta_1, \dots, \theta_k, \theta_{k+1} + \dots + \theta_n)$$

and the distribution of  $\frac{X_k}{1 - \sum_{i=1}^{k-1} X_i} = \frac{X_k}{\sum_{i=k}^n X_i}$  is  $Beta(\theta_k, \sum_{i=k+1}^n \theta_i)$ .

2. Bayes posterior under random sampling

Consider the n-dimensional Dirichlet distribution,  $\mathcal{D}(\alpha)$  with parameters  $(\alpha_1, \dots, \alpha_n)$ . Assume that some phenomena is described by a random probability vector  $p = (p_1, \dots, p_n)$ . Let  $\mathcal{D}(\alpha)$  be the “prior distribution of the vector  $p$ . Now let us assume that we take a sample and observe that  $N_i$  of the outcome are  $i$ . Now compute the posterior distribution of  $p$  given the observations  $N = (N_1, \dots, N_n)$  as follows: Using properties of the Dirichlet distribution we can show that it is

$$\begin{aligned} P(p \in dx | N) &= \frac{1}{Z} \frac{x_1^{\alpha_1} \dots x_n^{\alpha_n} x_1^{N_1} \dots x_n^{N_n}}{\int x_1^{\alpha_1} \dots x_n^{\alpha_n} x_1^{N_1} \dots x_n^{N_n} dx_1 \dots dx_n} \\ &= \frac{1}{Z'} x_1^{(\alpha_1 + N_1)} \dots x_n^{(\alpha_n + N_n)}. \end{aligned}$$

That is,

$$(5.55) \quad P(p \in \cdot | N) \text{ is } \mathcal{D}(\alpha_1 + N_1, \dots, \alpha_n + N_n).$$

# Chapter 18

## Appendix III: Markov Processes

### 18.1 Operator semigroups

See Ethier-Kurtz, [222] Chap.1.

Consider a strongly continuous semigroup  $\{T_t\}$  with generator  $G$  and domain  $D(G)$ . A subset  $D_0 \subset D(G)$  is a *core* if the closure of  $G|_{D_0}$  equals  $G$ . If  $D_0$  is dense and  $T_t : D_0 \rightarrow D_0$  for all  $t$ , then it is a core.

**Theorem 18.1** (*Kurtz semigroup convergence Theorem [222], Chap. 1, Theorem 6.5*) Let  $L, L_n$  be Banach spaces and  $\pi_n : L \rightarrow L_n$  is a bounded linear mapping and  $\sup_n \|\pi_n\| < \infty$ . We say  $f_n \in L_n \rightarrow f \in L$  if  $\lim_{n \rightarrow \infty} \|f_n - \pi_n f\| = 0$ .

For  $n \in \mathbb{N}$  let  $T_n$  be a contraction on a Banach space  $L_n$ , let  $\varepsilon_n > 0$ ,  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$ . Let  $\{T(t)\}$  be a strongly continuous contraction semigroup on  $L$  with generator  $A$  and let  $D$  be a core for  $A$ . Then the following are equivalent:

(a) For each  $f \in L$ ,  $T_n^{\lfloor t/\varepsilon_n \rfloor} \pi_n f \rightarrow T(t)f$ , for all  $t \geq 0$ , uniformly on bounded intervals.

(b) For each  $f \in D$  there exists  $f_n \in L_n$  such that  $f_n \rightarrow f$  and  $A_n f_n \rightarrow Af$ .

**Theorem 18.2** [222] Chap. 4, Theorem 2.5.

Let  $E$  be locally compact and separable. For  $n = 1, 2, \dots$  let  $\{T_n(t)\}$  be a Feller semigroup on  $C_0(E)$  and suppose that  $X_n$  is a Markov process with semigroup  $\{T_n(t)\}$  and sample paths in  $D_E([0, \infty))$ . Suppose that  $\{T(t)\}$  is a Feller semigroup on  $C_0(E)$  such that for each  $f \in C_0(E)$

$$(18.1) \quad \lim_{n \rightarrow \infty} T_n(t)f = T(t)f, \quad t \geq 0.$$

If  $\{X_n(0)\}$  has limiting distribution  $\nu \in \mathcal{P}(E)$ , then there is a Markov process  $X$  corresponding to  $\{T(t)\}$  with initial distribution  $\nu$  and sample paths in  $D_E([0, \infty))$  with initial distribution  $\nu$  and sample paths in  $D_E([0, \infty))$  and  $X_n \Rightarrow X$ .

## 17.4 Topologies on path spaces

**Definition 17.9** Let  $\mu_i, \mu \in \mathcal{M}_f$ . Then  $(\mu_n)_{n \in \mathbb{N}}$  converges weakly to  $\mu$  as  $n \rightarrow \infty$ , denoted  $\mu_n \Rightarrow \mu$  iff and only if

$$(17.7) \quad \int f d\mu_n \xrightarrow[n \rightarrow \infty]{} \int f d\mu \quad \forall f \in C_b(E)$$

Given a Polish space  $(E, d)$  we consider the space  $C_E([0, \infty))$  with the metric

$$(17.8) \quad \tilde{d}(f, g) = \sum_{n=1}^{\infty} 2^{-n} \sup_{0 \leq t \leq n} |f(t) - g(t)|.$$

Then  $(C_E([0, \infty)), \tilde{d})$  is also a Polish space. To prove weak convergence in  $\mathcal{P}((C_E([0, \infty)), \tilde{d}))$  it suffices to prove tightness and the convergence of the finite dimensional distributions.

Similarly the space  $D_E([0, \infty)$  of càdlàg functions from  $[0, \infty)$  to  $E$  with the Skorohod metric  $\tilde{d}$  is a Polish space where

$$(17.9) \quad \tilde{d}(f, g) = \inf_{\lambda \in \Lambda} \left( \gamma(\lambda) + \int_0^\infty e^{-u} \left( 1 \wedge \sup_t d(f(t \wedge u), g(t \wedge u)) \right) \right)$$

where  $\Lambda$  is the set of continuous, strictly increasing functions on  $[0, \infty)$  and for  $\lambda \in \Lambda$ ,

$$(17.10) \quad \gamma(\lambda) = 1 + \left( \sup_t |t - \lambda(t)| \vee \sup_{s \neq t} \left| \frac{\log(\lambda(s) - \lambda(t))}{s - t} \right| \right).$$

**Theorem 17.10 (Ethier-Kurtz) (Ch. 3, Theorem 10.2)** Let  $X_n$  and  $X$  be processes with sample paths in  $D_E([0, \infty)$  and  $X_n \Rightarrow X$ . Then  $X$  is a.s. continuous if and only if  $J(X_n) \Rightarrow 0$  where

$$(17.11) \quad J(x) = \int_0^\infty e^{-u} \left[ \sup_{0 \leq t \leq u} d(X(t), x(t-)) \right].$$

### 17.4.1 Sufficient conditions for tightness

**Theorem 17.11 (Aldous (1978))** Let  $\{P_n\}$  be a sequence of probability measures on  $D([0, \infty), \mathbb{R})$  such that

- for each fixed  $t$ ,  $P_n \circ X_t^{-1}$  is tight in  $\mathbb{R}$ ,
- given stopping times  $\tau_n$  bounded by  $T < \infty$  and  $\delta_n \downarrow 0$  as  $n \rightarrow \infty$

$$(17.12) \quad \lim_{n \rightarrow \infty} P_n(|X_{\tau_n + \delta_n} - X_{\tau_n}| > \varepsilon) = 0,$$

or

- $\forall \eta > 0 \exists \delta, n_0$  such that

$$(17.13) \quad \sup_{n \geq n_0} \sup_{\theta \in [0, \delta]} P_n(|X_{\tau_n + \theta} - X_{\tau_n}| > \varepsilon) \leq \eta.$$

Then  $\{P_n\}$  are tight.

### 17.4.2 The Joffe-Métivier criteria for tightness of D-semimartingales

We recall the Joffe Métivier criterion ([352]) for tightness of locally square integrable processes.

A càdlàg adapted process  $X$ , defined on  $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$  with values in  $\mathbb{R}$  is called a *D-semimartingale* if there exists a càdlàg function  $A(t)$ , a linear subspace  $D(L) \subset C(\mathbb{R})$  and a mapping  $L : (D(L) \times \mathbb{R} \times [0, \infty) \times \Omega) \rightarrow \mathbb{R}$  with the following properties:

1. for every  $(x, t, \omega) \in \mathbb{R} \times [0, \infty) \times \Omega$  the mapping  $\phi \rightarrow L(\phi, x, t, \omega)$  is a linear functional on  $D(L)$  and  $L(\phi, \cdot, t, \omega) \in D(L)$ ,
2. for every  $\phi \in D(L)$ ,  $(x, t, \omega) \rightarrow L(\phi, x, t, \omega)$  is  $\mathcal{B}(\mathbb{R}) \times \mathcal{P}$ -measurable, where  $\mathcal{P}$  is the predictable  $\sigma$ -algebra on  $[0, \infty) \times \Omega$ , ( $\mathcal{P}$  is generated by sets of the form  $(s, t] \times F$  where  $F \in \mathcal{F}_s$  and  $s, t$  are arbitrary)
3. for every  $\phi \in D(L)$  the process  $M^\phi$  defined by

$$(17.14) \quad M^\phi(t, \omega) := \phi(X_t(\omega) - \phi(X_0(\omega)) - \int_0^t L(\phi, X_{s-}(\omega), s, \omega) dA_s,$$

is a locally square integrable martingale on  $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ ,

4. the functions  $\psi(x) := x$  and  $\psi^2$  belong to  $D(L)$ .

The functions

$$(17.15) \quad \beta(x, t, \omega) := L(\psi, x, t, \omega)$$

$$(17.16) \quad \alpha(x, t, \omega) := L((\psi)^2, x, t, \omega) - 2x\beta(x, t, \omega)$$

are called the *local characteristics of the first and second order*.

**Theorem 17.12** *Let  $X^m = (\Omega^m, \mathcal{F}^m, \mathcal{F}_t^M, P^m)$  be a sequence of D-semimartingales with common  $D(L)$  and associated operators  $L^m$ , functions  $A^m, \alpha^m, \beta^m$ . Then the sequence  $\{X^m : m \in \mathbb{N}\}$  is tight in  $D_{\mathbb{R}}([0, \infty)$  provided the following conditions are satisfied:*

1.  $\sup_m E|X_0^m|^2 < \infty$ ,

2. there is a  $K > 0$  and a sequence of positive adapted processes  $\{C_t^m : t \geq 0\}$  on  $\Omega^m$  such that for every  $m \in \mathbb{N}, x \in \mathbb{R}, \omega \in \Omega^m$ ,

$$(17.17) \quad |\beta_m(x, t, \omega)|^2 + \alpha_m(x, t, \omega) \leq K(C_t^m(\omega) + x^2)$$

and for every  $T > 0$ ,

$$(17.18) \quad \sup_m \sup_{t \in [0, T]} E|C_t^m| < \infty, \text{ and } \lim_{k \rightarrow \infty} \sup_m P^m(\sup_{t \in [0, T]} C_t^m \geq k) = 0,$$

3. there exists a positive function  $\gamma$  on  $[0, \infty)$  and a decreasing sequence of numbers  $(\delta_m)$  such that  $\lim_{t \rightarrow 0} \gamma(t) = 0$ ,  $\lim_{m \rightarrow \infty} \delta_m = 0$  and for all  $0 < s < t$  and all  $m$ ,

$$(17.19) \quad (A^m(t) - A^m(s)) \leq \gamma(t - s) + \delta_m.$$

4. if we set

$$(17.20) \quad M_t^m := X_t^m - X_0^m - \int_0^t \beta_m(X_{s-}^m, s, \cdot) dA_s^m,$$

then for each  $T > 0$  there is a constant  $K_T$  and  $m_0$  such that for all  $m \geq m_0$ , then

$$(17.21) \quad E(\sup_{t \in [0, T]} |X_t^m|^2) \leq K_T(1 + E|X_0^m|^2),$$

and

$$(17.22) \quad E(\sup_{t \in [0, T]} |M_t^m|^2) \leq K_T(1 + E|X_0^m|^2),$$

### Corollary 17.13

Assume that for  $T > 0$  there is a constant  $K_T$  such that

$$(17.23) \quad \sup_m \sup_{t \leq T, x \in \mathbb{R}} (|\alpha_m(t, x)| + |\beta_m(t, x)|) \leq K_T, \text{ a.s.}$$

$$(17.24) \quad \sum_m (A^m(t) - A^m(s)) \leq K_T(t - s) \text{ if } 0 \leq s \leq t \leq T,$$

and

$$(17.25) \quad \sup_m E|X_0^m|^2 < \infty,$$

and  $M_t^m$  is a square integrable martingale with  $\sup_m E(|M_T^m|^2) \leq K_T$ . The the  $\{X^m : m \in \mathbb{N}\}$  are tight in  $D_{\mathbb{R}}([0, \infty)$ .

### Criteria for continuous processes

Now consider the special case of probability measures on  $C([0, \infty), \mathbb{R}^d)$ . This criterion is concerned with a collection  $(X^{(n)}(t))_{t \geq 0}$  of semimartingales with values in  $\mathbb{R}^d$  with continuous paths. First observe that by forming

$$(17.26) \quad (< X^{(n)}(t), \lambda >)_{t \geq 0}, \quad \lambda \in \mathbb{R}^d$$

we obtain  $\mathbb{R}$ -valued semi-martingales. If for every  $\lambda \in \mathbb{R}^d$  the laws of these projections are tight on  $C([0, \infty), \mathbb{R})$  then this is true for  $\{\mathcal{L}[X^{(n)}(t)]_{t \geq 0}, n \in \mathbb{N}\}$ . The tightness criterion for  $\mathbb{R}$ -valued semimartingales is in terms of the so-called local characteristics of the semimartingales.

For Itô processes the local characteristics can be calculated directly from the coefficients. For example, if we have a sequence of semimartingales  $X^n$  that are also a Markov processes with generators:

$$(17.27) \quad L^{(n)} f = \left( \sum_{i=1}^d a_i^n(x) \frac{\partial}{\partial x_i} + \sum_{i=1}^d \sum_{j=1}^d b_{i,j}^n(x) \frac{\partial^2}{\partial x_i \partial x_j} \right) f$$

then the local characteristics are given by

$$(17.28) \quad a^n = (a_i^n)_{i=1, \dots, d}, \quad b^n = (b_{i,j}^n)_{i,j=1, \dots, d}.$$

The Joffe-Métivier criterion implies that if

$$(17.29) \quad \sup_n \sup_{0 \leq t \leq T} E[|a^n(X^{(n)}(t))|^2 + |b^n(X^{(n)}(t))|^2] < \infty,$$

$$(17.30) \quad \lim_{k \rightarrow \infty} \sup_n P(\sup_{0 \leq t \leq T} (|a^n(X^{(n)})(t)| + |b^n(X^{(n)})(t)|) \geq k) = 0$$

then  $\{\mathcal{L}[X^{(n)}(t)]_{t \geq 0}, n \in \mathbb{N}\}$  are tight in  $C([0, \infty), \mathbb{R})$ . See [352] for details.

**Theorem 17.14** (Ethier-Kurtz [225] Chapt. 3, Theorem 10.2) *Let*

$$(17.31) \quad J(x) = \int_0^\infty e^{-u} [J(x, u) \wedge 1] du, \quad J(x, u) = \sup_{0 \leq t \leq u} d(x(t), x(t-)).$$

*Assume that a sequence of processes  $X_n \Rightarrow X$  in  $D_E([0, \infty))$ . Then  $X$  is a.s. continuous if and only if  $J(X_n) \Rightarrow 0$ .*

#### 17.4.3 Tightness of measure-valued processes

**Lemma 17.15** (Tightness Lemma).

(a) *Let  $E$  be a compact metric space and  $\{P_n\}$  a sequence of probability measures on  $D([0, \infty), M_1(E))$ . Then  $\{P_n\}$  is compact if and only if there exists a linear separating set  $D \subset C(E)$  such that  $t \mapsto \int f(x) X_t(\omega, dx)$  is relatively compact in  $D([0, \infty), [-\|f\|, \|f\|])$  for each  $f \in D$ .*

(b) Assume that  $\{P_n\}$  is a family of probability measures on  $D([0, \infty), [-K, K])$  such that for  $0 \leq t \leq T$ , there are bounded predictable processes  $\{v_i(\cdot) : i = 1, 2\}$  such that for each  $n$

$$M_{i,n}(t) := x(\omega, t)^i - \int_0^t v_{i,n}(\omega, s)ds, \quad i = 1, 2$$

are  $P_n$ -square integrable martingales with

$$\sup_n E_n(\sup_s (|v_{2,n}(s)| + |v_{1,n}(s)|)) < \infty.$$

Then the family  $\{P_n\}$  is tight.

(c) In (b) we can replace the  $i = 2$  condition with: for any  $\varepsilon > 0$  there exists  $f$  and  $v_{f,n}$  such that

$$\sup_{[-K, K]} |f_\varepsilon(x) - x^2| < \varepsilon$$

and

$$M_{f,n}(t) := f_\varepsilon(x(\omega, t)) - \int_0^t v_{f,n}(\omega, s)ds$$

$$\sup_n E_n(\sup_s |v_{f,n}(s)|) < \infty.$$

**Proof.** (a) See e.g. Dawson, [139] Section 3.6.

(b) Given stopping times  $\tau_n$  and  $\delta_n \downarrow 0$  as  $n \rightarrow \infty$ .

$$\begin{aligned} & E_n [(x(\tau_n + \delta_n) - x(\tau_n))^2] \\ &= \{E_n[x^2(\tau_n + \delta_n) - x^2(\tau_n)] - 2E_n[x(\tau_n)(x(\tau_n + \delta_n) - x_n(\tau_n))]\} \\ &\leq E_n \left[ \int_{\tau_n}^{\tau_n + \delta_n} |v_{2,n}(s)| ds + 2K \int_{\tau_n}^{\tau_n + \delta_n} |v_{1,n}(s)| ds \right] \\ &\leq \delta_n \sup_n E_n(\sup_s (|v_{2,n}(s)| + |v_{1,n}(s)|)) \\ &\rightarrow 0 \quad \text{as } \delta_n \rightarrow 0. \end{aligned}$$

The result then follows by Aldous' condition.

(c)

$$\begin{aligned} & E_n [(x(\tau_n + \delta_n) - x(\tau_n))^2] \\ &= \{E_n[x^2(\tau_n + \delta_n) - x^2(\tau_n)] - 2E_n[x(\tau_n)(x(\tau_n + \delta_n) - x_n(\tau_n))]\} \\ &\leq E_n(f(x(\tau_n + \delta_n)) - f(x(\tau_n))) + 2K \int_{\tau_n}^{\tau_n + \delta_n} |v_{1,n}(s)| ds + 2\varepsilon \\ &\leq E_n \left[ \int_{\tau_n}^{\tau_n + \delta_n} |v_{f,n}(s)| ds + 2K \int_{\tau_n}^{\tau_n + \delta_n} |v_{1,n}(s)| ds \right] + 2\varepsilon \\ &\leq \delta_n \sup_n E_n(\sup_s (|v_{f,n}(s)| + |v_{1,n}(s)|)) + 2\varepsilon \end{aligned}$$

Hence for any  $\varepsilon > 0$

$$\begin{aligned} & \lim_{\delta_n \rightarrow 0} \sup_n E_n [(x(\tau_n + \delta_n) - x(\tau_n))^2] \\ & \leq \lim_{n \rightarrow \infty} \delta_n \sup_n E_n (\sup_s (|v_{f_\varepsilon, n}(s)| + |v_{1, n}(s)|)) + 2\varepsilon \\ & = 2\varepsilon. \end{aligned}$$

and the result again follows from Aldous criterion. ■

**Remark 17.16** These results can be also used to prove tightness in the case of non-compact  $E$ . However in this case an additional step is required, namely to show that for  $\varepsilon > 0$  and  $T > 0$  there exists a compact subset  $K_{T, \varepsilon} \subset E$  such that

$$P_n[D([0, T], K_{T, \varepsilon})] > 1 - \varepsilon \quad \forall n.$$

**Remark 17.17** Note that if  $P_n$  is a tight sequence of probability measures on  $D([0, T], \mathbb{R})$  such that  $P_n(\sup_{0 \leq s \leq T} |x(s) - x(s-)| \leq \delta_n) = 1$  and  $\delta_n \rightarrow 0$  as  $n \rightarrow \infty$ , then for any limit point  $P_\infty$ ,  $P_\infty(C([0, T], \mathbb{R})) = 1$ .

## 17.5 The Gromov-Hausdorff metric on the space of compact metric spaces

Let  $E$  be a metric space and  $B_1, B_2$  two subsets. Then the Hausdorff distance is defined by

$$(17.32) \quad d_H(K_1, K_2) = \inf\{\varepsilon \geq 0 : K_1 \subset V_\varepsilon(K_2), K_2 \subset V_\varepsilon(K_1)\}$$

where  $V_\varepsilon(K)$  denotes the  $\varepsilon$ -neighbourhood of  $K$ . This defines a pseudometric,  $d_H(B_1, B_2) = 0$  iff they have the same closures.

If  $X$  and  $Y$  are two compact metric spaces. The *Gromov-Hausdorff metric*  $d_{GH}(X, Y)$  is defined to be the infimum of all numbers  $d_H(f(X), g(Y))$  for all metric spaces  $M$  and all isometric embeddings  $f : X \rightarrow M$  and  $g : Y \rightarrow M$  and where  $d_{\text{Haus}}$  denotes Hausdorff distance between subsets in  $M$ .  $d_{GH}$  is a pseudo-metric with  $d_{GH}(K_1, K_2) = 0$  iff they are isometric.

Now let  $(\mathbb{K}, d_{GH})$  denote the class of compact metric spaces (modulo isometry) with the Gromov-Hausdorff metric. Then  $(\mathbb{K}, d_{GH})$  is complete.

See Gromov [289] and Evans [235] for detailed expositions on this topic.

### 17.5.1 Metric measure spaces

The notion of *metric measure space* was developed by Gromov [289] (called mm spaces there). It is given by a triple  $(X, r, \mu)$  where  $(X, r)$  is a metric space such that  $(\text{supp}(\mu), r)$  is complete and separable and  $\mu \in \mathcal{P}(X)$  is a probability measure

on  $(X, r)$ . Let  $\mathbb{M}$  be the space of equivalence classes of metric measure spaces (whose elements are not themselves metric spaces - see remark (2.2(ii)) in [287]) with equivalence in the sense of measure-preserving isometries. The distance matrix map is defined for  $n \leq \infty$

$$(17.33) \quad X^n \rightarrow \mathbb{R}_+^{\binom{n}{2}}, \quad ((x_i)_{i=1,\dots,n}) \rightarrow (r(x_i, x_j))_{1 \leq i < j \leq n}$$

and we denote by  $R(X, r)$  the map that sends a sequence of points to its infinite distance matrix.

Then the *distance matrix distribution of*  $(X, r, \mu)$  (representative of equivalence class) is defined by

$$(17.34) \quad \nu^{(X, r, \mu)} := R^{(X, r)} - \text{pushforward of } \mu^{\otimes \mathbb{N}} \in \mathcal{P}(\mathbb{R}_+^{\binom{\mathbb{N}}{2}}).$$

Since this depends only on the equivalence class it defined the mapping  $\kappa \rightarrow \nu^\kappa$  for  $\kappa \in \mathbb{M}$ . Gromov [289] (Section 3 $\frac{1}{2}$ .5) proved that a metric measure space is characterized by its distance matrix distribution.

Greven, Pfaffelhuber and Winter (2008) [287] introduced the *Gromov-weak topology*. In this topology a sequence  $\{\chi_n\}$  converges Gromov-weakly to  $\chi$  in  $\mathbb{M}$  if and only if  $\Phi(\chi_n)$  converges to  $\Phi(\chi)$  in  $\mathbb{R}$  for all polynomial in  $\Pi$ .

In [287], Theorem 1, they proved that  $\mathbb{M}$  equipped with the Gromov-weak topology is Polish.

An important subclass is the set of ultrametric measure spaces given by the closed subset of  $\mathbb{M}$

$$(17.35) \quad \mathbb{U} := \{ u \in \mathbb{M} : u \text{ is ultra-metric}\}.$$