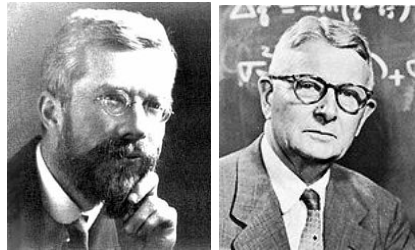# Chapter 5

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) $X_n \to 0$ or ∞, $a.s.$

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 $\{0, \ldots, N\}$ and transition probabilities:

$$P(X_{n+1} = j | X_n = i) = \left(\begin{array}{c} N \\ j \end{array} \right) \left(\frac{i}{N} \right)^j \left(1 - \frac{i}{N} \right)^{N-j}, \quad j = 0, \ldots, N.$$

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 $E_K = \{e_1, \ldots, e_K\}$ is given by a Markov chain X_n with state space $\backslash(E_K)$ (counting measures) and

(5.2) $P(X_{n+1} = (\beta_1, \ldots \beta_K) | X_n = (\alpha_1, \ldots, \alpha_K))$

$$= \frac{N!}{\beta_1! \beta_2! \ldots \beta_K!} \left(\frac{\alpha_1}{N} \right)^{\beta_1} \ldots \left(\frac{\alpha_K}{N} \right)^{\beta_K}$$

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

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

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

Then

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

and

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

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, $Y_1^n, \ldots, Y_N^n$. Assume each individual has a Poisson number of offspring with mean m, $(Z_1, \ldots, Z_N)$ and the offspring is of the same type as the parent. Then

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

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

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

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

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

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 $(Y_1^{n+1}, \ldots, Y_N^{n+1})$ 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 $j = 1, \ldots, K$, $\delta_j \in \mathcal{P}(E_K)$ are absorbing states and $X_n(j)$ is a martingale. Therefore $X_n \to X_\infty$, *a.s.* Since $\mathrm{Var}(X_{n+1}) = N X_n(1 - X_n)$, 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 it 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 $e_1, \ldots, e_K$ at a particular gene we can focus on the set of types given by $E_K^{2\circ}$ where $\frac{K(K+1)}{2}$ 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) $P_{ij} = (2 - \delta_{ij}) p_i p_j.$

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,

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

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

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

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

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

$$(5.12) \quad P(T_{2,1} = n) = \frac{1}{2N}(1 - \frac{1}{2N})^{n-1}, \qquad n = 1, 2 \ldots.$$

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 $N \to \infty$ 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 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}.

$$(5.14)$$

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

Stage III: Resampling. Finally random sampling reduces the population to N adults with proportions P_{ij}^{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 $\mathcal{P}^{\mathcal{N}}(E_K)$. If X_n^N is a Markov chain we defined the transition function

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

where the function π is obtained from (5.14), (5.15), (5.16). XXX

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 nth generation. Then as above $\{X_n^N\}_{n\in\mathbb{N}}$ is a Markov chain.

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

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

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

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

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

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

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,

$$E_{X_n^N}(\frac{X_{n+1}^N}{N}) = \frac{X_n^N}{N}$$

$$E_{X_n^N}[\left(\frac{X_{n+1}^N}{N} - \frac{X_n^N}{N}\right)^2 | \frac{X_n^N}{N}] = \frac{1}{N}\left(\frac{X_n^N}{N}\left(1 - \frac{X_n^N}{N}\right)\right).$$

We can then verify that

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

with

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

$$= \frac{1}{N}E\sum_{k=\lfloor Nt_1\rfloor}^{\lfloor Nt_2\rfloor}p_N(\frac{k}{N})(1 - p_N(\frac{k}{N}))$$

17.4   Topologies on path spaces

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

$$(17.7) \quad \int f d\mu_n \underset{n \to \infty}{\Longrightarrow} \int f d\mu \quad \forall \ f \in C_b(E)$$

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

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

Then $(C_E([0, \infty)), \widetilde{d})$ is also a Polish space. To prove weak convergence in $\mathcal{P}((C_E([0, \infty)), \widetilde{d}))$ 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 $\widetilde{d}$ is a Polish space where

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

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

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

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*

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

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 τ_n bounded by $T < \infty$ and $\delta_n \downarrow 0$ as $n \to \infty$*

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

  or

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

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

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 $L : (D(L) \times \mathbb{R} \times [0, \infty) \times \Omega) \to \mathbb{R}$ with the following properties:

1. for every $(x, t, \omega) \in \mathbb{R} \times [0, \infty) \times \Omega$ the mapping $\phi \to L(\phi, x, t, \omega)$ is a linear functional on $D(L)$ and $L(\phi, \cdot, t, \omega) \in D(L)$,

2. for every $\phi \in D(L)$, $(x, t, \omega) \to L(\phi, x, t, \omega)$ is $\mathcal{B}(\mathbb{R}) \times \mathcal{P}$-measurable, where $\mathcal{P}$ is the predictable σ-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^ϕ defined by

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

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

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

The functions

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

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

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, α^m, β^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 $\Omega^m\}_{m \in \mathbb{N}}$ such that for every $m \in \mathbb{N}, x \in \mathbb{R}, \omega \in \Omega^m$,*

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

and for every $T > 0$,

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

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

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

4. *if we set*

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

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

and

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

Corollary 17.13

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

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

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

and

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

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

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

then the local characteristics are given by

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

The Joffe-Métivier criterion implies that if

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

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

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

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 $t \to \int f(x)X_t(\omega, dx)$ 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 $0 \le t \le T$, there are bounded predictable processes $\{v_i(\cdot) : i = 1, 2\}$ such that for each n

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

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

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

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

Proof. (a) See e.g. Dawson, [139] Section 3.6.
(b) Given stopping times τ_n and $\delta_n \downarrow 0$ as $n \to \infty$.

$$E_n\left[(x(\tau_n + \delta_n) - x(\tau_n))^2\right]$$
$$= \{E_n[x^2(\tau_n + \delta_n) - x^2(\tau_n)] - 2E_n[x(\tau_n)(x(\tau_n + \delta_n) - x_n(\tau_n))]\}$$
$$\le E_n[\int_{\tau_n}^{\tau_n + \delta_n} |v_{2,n}(s)|ds + 2K \int_{\tau_n}^{\tau_n + \delta_n} |v_{1,n}(s)|ds]$$
$$\le \delta_n \sup_n E_n(\sup_s(|v_{2,n}(s)| + |v_{1,n}(s)|))$$
$$\to 0 \ \text{ as } \delta_n \to 0.$$

The result then follows by Aldous' condition.
(c)

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

Hence for any $\varepsilon > 0$

$$\limsup_{\delta_n \to 0} \; E_n \left[(x(\tau_n + \delta_n) - x(\tau_n))^2 \right]$$

$$\leq \lim_{n \to \infty} \delta_n \sup_n E_n (\sup_s (|v_{f_\varepsilon, n}(s)| + |v_{1,n}(s)|)) + 2\varepsilon$$

$$= 2\varepsilon.$$

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*

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

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 $\delta_n \to 0$ as $n \to \infty$, then for any limit point P_∞, $P_\infty(C([0,T], \mathbb{R})) = 1$.*

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) $d_{\mathrm{H}}(K_1, K_2) = \inf\{\varepsilon \geq 0 : K_1 \subset V_\varepsilon(K_2), \; K_2 \subset V_\varepsilon(K_1)\}$

where $V_\varepsilon(K)$ denotes the ε-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 $f : X \to M$ and $g : Y \to M$ and where d_{Haus} denotes Hausdorff distance between subsets in M. d_{GH} is a pseudometric 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, μ) where (X, r) is a metric space such that $(\mathrm{supp}(\mu), \mathrm{r})$ 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 $n \leq \infty$

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

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, μ) (representative of equivalence class) is defined by

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

Since this depends only on the equivalence class it defined the mapping $\kappa \to \nu^\kappa$ 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 χ in $\mathbb{M}$ if and only if $\Phi(\chi_n)$ converges to $\Phi(\chi)$ in $\mathbb{R}$ for all polynomial in Π.

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 $\mathbb{M}$

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