Diffusion limits in population models

F. Redig*

June 8, 2006

Abstract

Diffusion limits of models such as the Wright Fisher model, the Moran model and its modifications are introduced. The approach is to show convergence of the generators. Stationary distributions, absorbtion probabilities and the concept duality are discussed.

1 Diffusion limit: random walk example

Consider simple random walk on the integers $\mathbb{Z}$

$$X_n = \sum_{i=1}^{n} \epsilon_i$$

where X_i are i.i.d. with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$. This is a Markov chain on $\mathbb{Z}$ with transition probability operator

$$Pf(x) = \mathbb{E}\left(f(X_1)|X_0 = x\right) = \frac{1}{2}\left(f(x+1) + f(x-1)\right)$$

for $f : \mathbb{Z} \to \mathbb{R}$. From the central limit theorem, we immediately conclude that

$$\frac{1}{n} X_{[n^2 t]} \to N(0,t)$$

where $N(0,t)$ denotes a normally distributed random variable with mean 0 and variance t. We can "lift up" this convergence to convergence of the whole process $\{\frac{1}{n} X_{[n^2 t]}, t \geq 0\}$ to Brownian motion.

A convenient associated continuous time random walk pure jump process is defined by letting the jumps of the random walk X_n be performed on the event times of an independent mean one Poisson process, i.e.,

$$X_t = \sum_{i=1}^{N_t} \epsilon_i$$

*Mathematisch Instituut Universiteit Leiden, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands, redig@math.leidenuniv.nl

This process has generator

$$Lf(x) = \lim_{t\downarrow 0} \frac{\mathbb{E}_x f(X_t) - f(x)}{t} = (P - I)f(x) = \frac{1}{2}\left(f(x+1) + f(x-1) - 2f(x)\right)$$

where $\mathbb{E}_x$ denotes expectation in the random walk starting from $X_0 = x$. If we define

$$S_t f(x) = \mathbb{E}_x f(X_t)$$

then we call S_t the semigroup of the process $\{X_t : t \geq 0\}$ and we have the relation

$$S_t = e^{tL} = \sum_{n=0}^{\infty} \frac{t^n}{n!} e^{-t} P^n = \sum_{n=0}^{\infty} \mathbb{P}(N_t = n) P^n$$

and

$$L = \lim_{t\downarrow 0} \frac{S_t - I}{t}$$

Consider now the rescaled process

$$X^n(t) := \frac{1}{n} X_{n^2 t}$$

this is a pure jump process on the real line $\mathbb{R}$ making jumps of size $1/n$ with generator

$$L_n f(x) = \frac{n^2}{2}\left(f(x + \frac{1}{n}) + f(x - \frac{1}{n}) - 2f(x)\right)$$

In the limit $n \to \infty$ this converges to

$$\lim_{n\to\infty} L_n f(x) = \frac{1}{2} f''(x) := \mathcal{L}f(x)$$

We do not specify here details about domains, conditions on f, etc. $\mathcal{L}$ is the generator of Brownian motion $\{W_t : t \geq 0\}$, i.e.,

$$\begin{aligned}
\mathcal{L}f(x) &= \frac{1}{2} f''(x) = \lim_{t\downarrow 0} \frac{\mathbb{E}_x f(W_t) - f(x)}{t} \\
&= \lim_{t\downarrow 0} \frac{1}{t} \int \frac{e^{-(y-x)^2/2t}}{\sqrt{2\pi t}} (f(y) - f(x)) dy = \lim_{t\downarrow 0} \frac{1}{t} \frac{S_t f(x) - f(x)}{t}
\end{aligned}$$

This convergence of the generators of the rescaled processes to the generator of Brownian motion implies weak convergence of the associated processes, i.e.,

$$\{X_n(t) : t \geq 0\} \to \{W_t : t \geq 0\}$$

the meaning of the arrow here is "weak convergence in path space". In words this means the following: the process $\{X_n(t) : t \geq 0\}$ starting from $x \in \mathbb{R}$ gives a unique distribution $\mathbb{P}^n_x$ on trajectories (right continuous with left limits) and the Brownian motion gives a measure $\mathbb{P}_x^{bm}$ on continuous trajectories; the convergence then means that for all continuous functions on trajectories

$$\int f(\omega)\mathbb{P}_x^n(d\omega) \to \int f(\omega)\mathbb{P}_x^{bm}(d\omega)$$

In particular this implies that all finite dimensional distributions of $\{X_n(t) : t \geq 0\}$ converge to the corresponding finite dimensional distributions of the Brownian motion. For more details about weak convergence in path space, and in particular the good topology on trajectories we refer to the excellent monograph [2].

In our context of population models, we start from a discrete model, rescale appropriately, show convergence of the generators of the rescaled processes to the generator of a diffusion process. This (if worked out in full detail) implies weak convergence in path space of the associated processes. For this we use the Trotter-Kato theorem, see e.g. [7], chapter 1, but we do not spell out these details here (i.e., showing uniform convergence, specifying domains, etc.).

2 Limiting diffusions: general strategy

We start from a model with N diploid individuals. For the models we deal with here, this is the same as a population of $2N$ individuals of type A_1 or A_2 (two alleles). The state space variable is the number of A_1 alleles, and the model we start with is a Markov chain on $\{0, \ldots, 2N\}$, with transition probability operator

$$P_N f(i) = \sum_{j=1}^{2N} p_N(i,j) f(j)$$

the associated continuous time pure jump process has generator

$$L_N f(i) = (P_N - I) f(i) = \sum_{j=1}^{2N} p_N(i,j)(f(j) - f(i))$$

One then considers the rescaled process

$$X_N(t) = \frac{1}{2N} X_{2Nt}$$

which is a continuous-time Markov (pure jump) process on $\{0, \frac{1}{2N}, \ldots, 1\} \subseteq [0,1]$ with generator

$$\mathcal{L}_N f(\frac{i}{2N}) = 2N \sum_{j=1}^{2N} p_N(i,j) \left(f(\frac{j}{2N}) - f(\frac{i}{2N}) \right)$$

Expanding this expression up to second order, we obtain

$$\mathcal{L}_N f(\frac{i}{2N}) = \sum_j p_N(i,j)(j-i) f'(\frac{i}{2N}) + \frac{1}{2} \sum_j p_N(i,j) \frac{(j-i)^2}{2N} f''(\frac{i}{2N}) + \mathcal{R}_N$$

If we choose the transition probabilities $p_N(i,j)$ appropriately, then the rest term $\mathcal{R}_N$ will vanish in the limit $N \to \infty$. In that case in the limit, we obtain a second order differential operator which we can associate to a diffusion process, where the first order part corresponds to the drift of the diffusion ("deterministic component") and the second order part corresponds to the noise ("random component").

If we choose $i = i_N$ such that $i_N/(2N) \to x \in [0,1]$ as $N \to \infty$, and if we choose our transition probabilities such that

$$\lim_{N\to\infty} \sum_j p_N(i_N, j)(j-i) = M(x) \tag{2.1}$$

and

$$\lim_{N\to\infty} \sum_j p_N(i,j)\frac{(j-i)^2}{2N} = V(x) \tag{2.2}$$

with $M, V : [0,1] \to \mathbb{R}$, then the limiting generator becomes

$$\lim_{N\to\infty} \mathcal{L}_N f(\frac{i_N}{2N}) = M(x)f'(x) + \frac{1}{2}V(x)f''(x) \tag{2.3}$$

As a consequence we have that the rescaled processes

$$\{\frac{1}{2N}X_{2Nt} : t \geq 0\}$$

starting with $X_0 = i_N$ converge as $N \to \infty$ (in the sense previously explained) to the diffusion process

$$\{x_t : t \geq 0\}$$

with generator

$$\mathcal{L}f(x) = M(x)f'(x) + \frac{1}{2}V(x)f''(x)$$

starting from $x = \lim_{N\to\infty} i_N/(2N)$ This in turn implies that the same limit theorem for the *discrete time* rescaled process, i.e.,

$$\{\frac{1}{2N}X_{[2Nt]} : t \geq 0\}$$

converges to the same limit (by a standard so-called random time change argument which we do not spell out here).

3 Example 1: Wright Fisher diffusion

In the Wright Fisher model, the next generation is obtained by choosing at random a parent from the previous generation and adopting its (allelic) type. This means that the transition probabilities are given by

$$p_N(i,j) = \binom{2N}{j}\left(\frac{i}{2N}\right)^j\left(1-\frac{i}{2N}\right)^{2N-j} \tag{3.1}$$

In the corresponding Markov chain, X_{n+1} has then distribution $Bin(2N, i/(2N))$. This implies

$$\mathbb{E}(X_{n+1}|X_n) = X_n$$

i.e., $\{X_n, n \in \mathbb{N}\}$ is a *martingale*. Therefore, in the limiting diffusion no drift term will be present. More precisely if we choose the initial conditions $i = i_N$ such that $i_N/(2N) \to x \in [0,1]$, then, returning to (2.1), (2.2), we obtain

$$\sum_j p_N(i_N, j)(j - i) = 0$$

for all N, and hence $M(x) = 0$. Next,

$$V(x) = \lim_{N\to\infty} \sum_j p_N(i,j)\frac{(j-i)^2}{2N} = \lim_{N\to\infty} \frac{i_N}{2N}\left(1 - \frac{i_N}{2N}\right) = x(1-x)$$

Here, the first equality uses that the variance of a $Bin(m,p)$ is equal to $mp(1-p)$. Finally, the rest term $\mathcal{R}_N$ in (2.1) is dominated by

$$\mathcal{R}_N \le 2N\|f'''\|_\infty \sum_j \frac{(j-i)^3}{2N^3}$$

If $X_0 = i \simeq 2Nx$, then X_1 is $Bin(2N, x) \simeq N(2Nx, 2Nx(1-x))$, therefore

$$\mathbb{E}_{i_N}(X_1 - X_0)^3 \simeq \mathbb{E}_{i_N}(X_1 - Nx)^3 \le CN^{3/2}$$

Therefore,

$$\mathcal{R}_N \le 2N\|f'''\|_\infty \sum_j \frac{(j-i)^3}{(2N)^3} \le C'N^{-1/2}$$

Hence, the limiting diffusion has generator

$$\mathcal{L}f(x) = \frac{1}{2}x(1-x)f''(x) \tag{3.2}$$

A related (to Wright Fisher model) continuous time process is the so-called Moran model. In this model individual types are updated sequentially (on the event times of independent Poisson processes). More precisely, the Moran model is a birth and death process on $\{0, \ldots, 2N\}$ with generator

$$Lf(i) = (2N)\frac{i}{2N}\left(1 - \frac{i}{2N}\right)(f(i+1) + f(i-1) - 2f(i)) \tag{3.3}$$

For this model, the limiting diffusion is the same Wright Fisher diffusion. The advantage of this model is that it is easy to introduce selection into it (see later).

4 Properties and applications of Wright Fisher diffusion

4.1 Duality

If we apply the WF generator (3.2) to the function $\Psi(n,x) := x^n$, then we obtain

$$\mathcal{L}\Psi(n,x) = \frac{n(n-1)}{2}(\Psi(n-1,x) - \Psi(n,x)) := \hat{\mathcal{L}}\Psi(\cdot,x)(n) \tag{4.1}$$

where now $\hat{\mathcal{L}}$ is the generator of a process on the natural numbers, i.e.,

$$\hat{\mathcal{L}}f(n) = \frac{n(n-1)}{2}(f(n-1) - f(n)) \tag{4.2}$$

This is a much simpler (compared to the WF diffusion) process called "Kingman's coalescent". It starts at a natural number n and goes down one unit at rate $n(n-1)/2$. The relation (4.1) can be turned into a corresponding relation for the processes:

$$\mathbb{E}_x(\Psi(n,x_t)) = \hat{\mathbb{E}}_n(\Psi(n_t,x)) \tag{4.3}$$

where $\mathbb{E}_x$ denotes expectation in the WF diffusion starting from $x_0 = x$ and $\hat{\mathbb{E}}_n$ denotes expectation in Kingman's coalescent (the process with generator (4.2) starting from $n_0 = n$. The advantage of this relation (called "duality") lies in the fact that the process $\{n_t : t \geq 0\}$ is much simpler that the WF-diffusion and gives full information about all the moments of the WF diffusion. Since polynomials are dense (remember that the state space of the WF diffusion is the interval $[0,1]$), knowing all the moments gives full information about the process.

As an application, remember that the Wright Fisher Markov chain has two traps $0, 2N$, and eventually fixates to one of these traps. By the (bounded) martingale property we have

$$X_n \to X_\infty$$

with

$$2N\mathbb{P}_i(X_\infty = 2N) = \mathbb{E}_i(X_\infty) = \mathbb{E}_i(X_0) = i$$

therefore the probability that one fixates at the state $2N$ equals $i/2N$.

In the corresponding diffusion, the fixating states are zero and one, and a measure for not being fixated yet is the heterozygosity:

$$h(x) = x(1-x) = \Psi(1,x) - \Psi(2,x)$$

Using duality, we have

$$\mathbb{E}_x(h(x_t)) = \hat{\mathbb{E}}_1\Psi(n_t,x) - \hat{\mathbb{E}}_2\Psi(n_t,x)$$

In the dual process 1 is a trap (the rate to jump down is zero for $n = 10$, i.e., $\mathbb{E}_1\Psi(n_t,x) = \Psi(1,x) = x$, and from the state $n = 2$, the dual process jumps down at rate $2(2-1)/2 = 1$, hence starting from 2, the state at time t is 2 with probability $1-e^{-t}$ and 1 with probability e^{-t}. This gives

$$\hat{\mathbb{E}}_2 \Psi(n_t, x) = x - x(1-x)e^{-t}$$

and hence,

$$\mathbb{E}_x(h(x_t)) = x(1-x)e^{-t}$$

the WF thus fixates exponentially fast.

Remark 4.4. *In the discrete time WF Markov chain (the process with transition probabilities* (3.1)*) we also have duality in the form*

$$\mathbb{E}_i H_N(X_t, k) = \hat{\mathbb{E}}_k H_N(i, \xi_t)$$

with duality function

$$H_N(i,k) = \frac{\binom{i}{k}}{\binom{N}{k}}$$

and ξ_t the so-called ancestral process, *i.e., the number of ancestors t generations backwards in time. Therefore, convergence to WF diffusion after rescaling corresponds to convergence to the Kingman's coalescent of the ancestral process.*

4.2 Time to fixation

Consider the WF Markov chain and call τ_N the random time at which a fixating state is reached. We then have the simple relation

$$\mathbb{E}_i(\tau_N) = \sum_{j=1}^{2N} p_N(i,j)\mathbb{E}_i(\tau_N | X_1 = j) = 1 + \sum_j p_N(i,j)\mathbb{E}_j(\tau_N)$$

Therefore if we call $f(i) = \mathbb{E}_i(\tau_N)$ then we have

$$(P_N - I)f = -1$$

or

$$2N(P_N - I)f = -2N$$

Therefore, if $f(i) = \psi(i/2N)$ and we put $x = i/2N$, we obtain from the convergence of the generators (after rescaling)

$$2N(P_N - I)f(i) \simeq \frac{1}{2}x(1-x)\psi''(x) \tag{4.5}$$

The solution of

$$\frac{1}{2}x(1-x)\psi''(x) = -2N$$

is

$$\psi(x) = -4N(x\log x + (1-x)\log(1-x))$$

Therefore (not specifying the details about the symbol $\simeq$ in the approximation (4.5)) we obtain

$$\mathbb{E}_i(\tau_N) \simeq 4NH(x)$$

where

$$H(x) = -(x \log x + (1-x)\log(1-x))$$

is the "entropy" of the initial state. In [3] an upper bound of the form

$$\mathbb{E}_i(\tau_N) \leq CH\left(\frac{i}{2N}\right)$$

for all i, N (with C not depending on N or i) is derived for Markov chains satisfying a so-called mean condition and weak variation condition (including Wright Fisher and more general models).

In the WF diffusion, if we call τ the time to fixation (=hitting time of zero or one), then we have the *equality*

$$\mathbb{E}_x(\tau) = 2H(x)$$

The function H is special because if we apply the WF generator to it we find

$$\mathcal{L}H(x) = \frac{1}{2}x(1-x)H''(x) = -1/2$$

Therefore

$$H(x_{t\wedge\tau}) - H(x_0) + \frac{1}{2}(t\wedge\tau) := M_t$$

is a mean zero martingale, which gives, after taking the expectation and the limit $t \to \infty$, using $H(x_\tau) = 0$,

$$\mathbb{E}_x(\tau) = 2\mathbb{E}_x H(x_0) = 2H(x)$$

5 General diffusion limits

For the WF diffusion, the state eventually fixates, in particular there is no "equilibrium distribution". As we saw in section 2, depending on the choice of the transition probabilities $p_N(i,j)$ several diffusion limits can come out. In general the generator of the limiting diffusion process has the form

$$\mathcal{L}f(x) = M(x)f'(x) + \frac{1}{2}V(x)f''(x) \tag{5.1}$$

The associated diffusion process $\{x_t : t \geq 0\}$ satisfies the stochastic differential equation

$$dx_t = M(x_t)dt + \sqrt{V(x_t)}dW_t \tag{5.2}$$

We will not go into details here about the precise meaning and solvability of this equation. We just add a few lines here about the relation between the stochastic calculus formulation and the definition via the generator. These lines can be skipped without any consequences for further reading. First we recall Ito's formula which tells that

$$f(x_t) - f(x_0) = \int_0^t f'(x_s)dx_s + \frac{1}{2}\int_0^t f''(x_s)d < x, x >_s \tag{5.3}$$

Here $< x, x >_s$ is the so-called quadratic variation process, defined as the unique increasing process such that

$$x_t^2 - x_0^2 - < x, x >_t$$

is a martingale. To become more concrete, suppose $V(x) = 1$, $M(x) = 0$, then x_t is simply brownian motion, and one then knows (or verifies by hand) that $x_t^2 - x_0^2 - t$ is a martingale, so in that case $< x, x >_t = t$. If the drift M is not zero, then it still does not affect the quadratic variation process, i.e., if $V(x) = 1$, $M(x)$ arbitrary, then still $< x, x >_t = t$. Finally, one has the relation

$$\left\langle \int_0^t \psi(x_s)dx_s, \int_0^t \varphi(x_s)dx_s \right\rangle = \int_0^t \psi(x_s)\varphi(x_s)d < x, x >_s$$

where in the left hand side for two different processes $xandy$, $< x, y >$ can be defined via polarization

$$< x, y >_t = \frac{1}{4}\left(< (x + y), (x + y) >_t - < (x - y), (x - y) >_t\right)$$

Therefore, in the general case $< x, x >_t = \int_0^t V(x_s)ds$, and hence we find, starting from (5.3) that

$$\begin{aligned} f(x_t) - f(x_0) &= \int_0^t f'(x_s)dx_s + \frac{1}{2}\int_0^t f''(x_s)d < x, x >_s \\ &= \int_0^t f'(x_s)dx_s + \frac{1}{2}\int_0^t f''(x_s)V(x_s)ds \\ &= \int_0^t f'(x_s)M(x_s)ds + \int_0^t f'(x_s)\sqrt{V(x_s)}dW_t + \frac{1}{2}\int_0^t f''(x_s)V(x_s)ds \end{aligned}$$

The stochastic integral involving dW_t is a mean zero martingale: this is a general property of the Ito integral which follows straight from its definition.

Therefore

$$f(x_t) - f(x_0) - \int_0^t f'(x_s)M(x_s)ds - \frac{1}{2}\int_0^t f''(x_s)V(x_s)ds = \mathcal{M}_t$$

with $\mathcal{M}_t$ a mean zero martingale. This is the more probabilistic way of saying that x_t is a Markov process with generator $\mathcal{L}$, namely that for $f : [0, 1] \to \mathbb{R}$ (in the domain of the generator),

$$f(x_t) - f(x_0) - \int_0^t \mathcal{L}f(x_s)ds = f(x_t) - f(x_0) - \int_0^t f'(x_s)M(x_s)ds - \frac{1}{2}\int_0^t f''(x_s)V(x_s)ds$$

is a mean zero martingale. One then formulates this by saying that the martingale problem posed by $\mathcal{L}$ has x_t as its (unique) solution. In cases where individuals have more than two types, or more generally have a type which is a variable taking values in a general space, then the corresponding diffusion is on the space of probability measures on the type space (measure valued diffusion). In this more general (and difficult) context, it is preferable to work with the martingale problem formulation rather than with generators.

5.1 Invariant measures

If x_t is the diffusion process with generator (5.1) then we have that its associated semigroup

$$S_t f(x) = \text{“}e^{tL}\text{”} f(x) = \mathbb{E}_x(f(x_t))$$

satisfies

$$\frac{d}{dt} S_t f = L S_t f = S_t L f \tag{5.4}$$

Here the exponential e^{tL} has been put in quotes because it is not defined via its series expansion (because L is not a bounded operator). However, it is useful to remember that this relation formally holds (after proper reinterpretation of the exponential via e.g. Hille -Yosida theorem), and that computations based on it usually “end well”.

If we start the diffusion x_t from an initial distribution μ of x_0 (i.e., μ is a probability distribution on $[0,1]$), then the distribution μ_t of x_t satisfies by definition

$$\int S_t f d\mu = \int f d\mu_t$$

which gives, using (5.4)

$$\frac{d}{dt} \int f d\mu_t = \int L f d\mu_t$$

If μ has a density ψ, i.e., $d\mu(x) = \psi(x)dx$, then the density ψ_t at time t satisfies

$$\frac{d}{dt} \int f \psi_t dx = \int L f \psi_t dx = \int f L^* \psi_t dx$$

where $*$ denotes adjoint in L^2. therefore we obtain the so-called Kolmogorov forward equation for the evolution of a probability density:

$$\frac{d}{dt} = L^* \psi_t \tag{5.5}$$

It is now easy to read of L^* from L (once more, not worrying about domain questions)

$$L^* = \left(M \frac{d}{dx} + V \frac{d^2}{dx^2} \right)^*$$

now use $(AB)^* = B^* A^*$ and $(d/dx)^* = -(d/dx)$ (=partial integration), to find

$$L^* f = \left(-\frac{d}{dx}(Mf) + \frac{d^2}{dx^2}(Vf) \right)$$

Therefore, the equation for the evolution of a probability density becomes

$$\frac{d}{dt}\psi_t(x) = \left(-\frac{d}{dx}(M(x)\psi_t(x)) + \frac{d^2}{dx^2}(V(x)\psi_t(x))\right) \tag{5.6}$$

In particular, for the stationary distribution we have $\psi_t = \psi$ and we obtain the equation

$$\left(-\frac{d}{dx}(M(x)\psi(x)) + \frac{d^2}{dx^2}(V(x)\psi(x))\right) = 0 \tag{5.7}$$

which gives

$$\psi(x) = \frac{C}{V(x)} e^{\int_0^x \frac{2M}{V}(y)dy} \tag{5.8}$$

where the constant C is determined by the normalization

$$C = \frac{1}{\int_0^1 \frac{1}{V(x)} e^{\int_0^x \frac{2M}{V}(y)dy} dx}$$

Remark 5.9. 1. *A priori it is not clear that there are no other solutions than* (5.8) *of the equation for the stationary distribution. They have to be non-negative and normalizable to one (in order to be candidate probability density). In the cases we will consider here there will in fact always be* a unique stationary distribution*, so we do not have to worry about other solutions. In cases where one has duality, it is easy to show that there is a unique stationary distribution (the corresponding dual birth and death chain fixates).*

2. *In the definition of the generator we never specified boundary conditions. This can of course be important, and in the computation of the adjoint we implicitely assumed that the domain of the generator are functions such that* $f(0) = f(1) = 0$.

6 Backward equation and fixation probabilities

We say that the diffusion process $\{x_t : t \geq 0\}$ has a transition probability density if its semigroup can be written in integral kernel form, i.e., if

$$S_t f(x) = \int p_t(x,y) f(y) dy$$

The Brownian motion e.g. has this property with

$$p_t(x,y) = \frac{e^{-(x-y)^2/2t}}{\sqrt{2\pi t}}$$

From the equation

$$\frac{d}{dt} S_t f = L S_t f$$

we then infer that the transition probability density satisfies

$$\frac{\partial p_t(x,y)}{\partial t} = Lp_t(\cdot,y)(x) \tag{6.1}$$

where the operator L acts on the x-variable. E.g., for Brownian motion one verifies directly that

$$\frac{\partial p_t(x,y)}{\partial t} = \frac{\partial^2}{\partial x^2}(p_t(x,y))$$

Equation (6.1) is called the Kolomogorov backward equation for the transition probability density. If one considers $p_t(x,y)$ as a function of t and y (i.e., fixing x), then it satisfies the Kolmogorov forward equation

$$\frac{\partial p_t(x,y)}{\partial t} = L^* p_t(x,\cdot)(y)$$

The limit

$$\lim_{t\to\infty} p_t(x,1) = u(x)$$

is the probability that the diffusion eventually fixates in the state 1. Taking the limit $t \to \infty$ in (6.1), we find that $u(x)$ satisfies the equation

$$Lu(x) = 0 \tag{6.2}$$

with boundary conditions $u(1) = 1$, $u(0) = 0$. Using now the explicit expression (5.1) for the generator yields the equation

$$M(x)u'(x) + \frac{1}{2}V(x)u''(x) = 0 \tag{6.3}$$

with solution (satisfying the boundary conditions)

$$u(x) = \frac{\int_0^x e^{-\int_0^y 2M/V dy}}{\int_0^1 e^{-\int_0^y 2M/V dy}} \tag{6.4}$$

E.g., in the Wright Fisher case, $M = 0, V = x(1-x)$, and we recover

$$u(x) = x$$

i.e., the fixation probability of one allele equals its initial proportion. This can of course also be seen directly from the fact that x_t is a martingale which eventually fixates at 0 or 1, i.e.,

$$u(x) = \mathbb{E}_x(x_\infty) = \mathbb{E}_x(x_0) = x$$

7 Diffusion limits for Wright Fisher with mutation and or selection

To introduce mutation in the WF Markov chain, in the next generation an individual chooses at random a parent from the previous generation, and then changes the type with probability v if allelic type A_1 was chosen and with probability u is allelic type A_2 was chosen. If initially there are i individuals (alleles) of type A_1, then in the next generation the number of A_1 individuals (alleles) is $Bin(2N, p_i)$ distributed with

$$p_i = p_i(u,v) = \frac{i}{2N}(1-v) + (1-\frac{i}{2N})u$$

the transition probabilities then become

$$p_N(i,j) = \binom{2N}{j} p_i^j (1-p_i)^{2N-j}$$

In order to obtain a diffusion limit after rescaling, we have to let u and v depend on N, as we will see now. Choose $i = i_N$ such that $i_N/(2N) \to x$, and return to equation (2.1)

$$M(x) = \lim_{N\to\infty} \sum_j p_N(i_N, j)(j-i) = \lim_{N\to\infty}(2Np_i - i) = \lim_{N\to\infty}(-i_N v_N + (2N - i_N)u_N) \tag{7.1}$$

In order to obtain a finite limit, we choose

$$u = u_N = q/4N, \ v = v_N = r/4N$$

which gives after taking the limit $N \to \infty$

$$M(x) = -\frac{r}{2}x + \frac{q}{2}(1-x) \tag{7.2}$$

For the variance we obtain

$$V(x) = \lim_{N\to\infty} \sum_j p_N(i,j)\frac{(j-i)^2}{2N} = \lim_{N\to\infty} p_i(1-p_i) = x(1-x)$$

The generator of the limiting diffusion is then given by

$$\mathcal{L}f(x) = (\frac{1}{2}q(1-x) - \frac{1}{2}rx)f'(x) + \frac{1}{2}x(1-x)f''(x) \tag{7.3}$$

corresponding to the stochastic differential equation

$$dx_t = (\frac{1}{2}q(1-x_t) - \frac{1}{2}rx_t)dt + \sqrt{x_t(1-x_t)}dW_t \tag{7.4}$$

For the stationary distribution of this diffusion, we use (5.7), (5.8):

$$\psi(x) = \frac{C}{x(1-x)} e^{\int_0^x \frac{q(1-y)-ry}{y(1-y)}dy} = Cx^{q-1}(1-x)^{r-1} \tag{7.5}$$

This is known as the density of the Beta distribution and $C = C(q,r) = 1/B(q,r)$ with B the Beta function.

Crucial distinction is to be made here between the case $q, r < 1$ (low mutation rate), where the old fixating states $0, 1$ are the "maxima" versus $q, r > 1$ (high mutation rate) where the maximum (without quotes) is at

$$x = \frac{q-1}{q+r-2}$$

If the mutation $A_1 \to A_2$ and $A_2 \to A_1$ occur with equal probabilities, then the maximum is always at the "maximum entropy state" $x = 1/2$, for $q = r > 1$.

7.1 A note on duality

The generator (when mutation is present) is of the form

$$\mathcal{L} = \alpha(\theta - x)\frac{d}{dx} + \frac{1}{2}x(1-x)\frac{d^2}{dx^2}$$

with $\alpha = (q-r)/2$, $\theta = q/(q-r)$. Applying this to $\Psi(n,x) = x^n$ gives

$$\mathcal{L}\Psi(n,x) = \alpha\theta n\Psi(n-1,x) - \alpha n\Psi(n,x) + \frac{1}{2}n(n-1)(\Psi(n-1,x) - \Psi(n,x))$$

If $\theta = 1$ (i.e., r=0), then this is of the form

$$\mathcal{L}\Psi(n,\cdot)(x) = \hat{\mathcal{L}}\Psi(\cdot,x)(n)$$

with

$$\hat{\mathcal{L}}f(n) = (\alpha n + \frac{1}{2}n(n-1))(f(n-1) - f(n))$$

i.e., an extra dead rate linear in n is added to the Kingman's coalescent. To transform from $\theta = 1$ to arbitrary θ can be done by Girsanov's formula which becomes very simple in this case. It gives the Radon Nikodym derivative of the process with $\theta \neq 1$ w.r.t. the process with $\theta = 1$:

$$\frac{d\mathbb{P}^{\theta,[0,t]}}{d\mathbb{P}^{1,[0,t]}} = \exp(x_t(\theta-1) + \frac{1}{2}(\theta-1)^2 x_t(1-x_t))$$

So this means that all expectations in the process $\theta \neq 1$ can be computed from the process with $\theta = 1$ via a simple exponential transformation.

8 Diffusion limits with selection

To introduce selection, it is convenient to work with the Moran model rather than with WF model. In the Moran model every individual (allele) has a Poisson clock; if this clock rings the individual is replaced by another individual (usually interpreted as death and giving birth to a new individual). If the number of A_1 alleles is i, then the new individual coming from a parent A_2 is of type A_1 with probability $i/2N$ and from A_1 to A_2 with probability $(1-\frac{i}{2N})$. The rate to create an extra A_1 allele thus equals $b_i = (2N-i)i/2N$ and the rate to create an extra A_2 allele (is minus an A_1 allele) is $d_i = i(1-\frac{i}{2N})$ (note that both rates are in fact the same). The Moran model is then the birth and death chain $\{X_t : t \geq 0\}$ on $\{0, \ldots, 2N\}$ with generator

$$Lf(i) = d_i(f(i-1) - f(i)) + c_i(f(i+1) - f(i)) \tag{8.1}$$

We once more rescale in order to obtain a diffusion limit:

$$X_N(t) := \frac{1}{2N} X_{2Nt} \tag{8.2}$$

This rescaled process has generator

$$L_N f(x) = 2N\left(d_{2Nx}(f(x-\frac{1}{N}) - f(x)) + c_{2Nx}(f(x+\frac{1}{N}) - f(x))\right) \tag{8.3}$$

with $x = i/2N$. For the Moran model, $d_i = c_i$ and hence the first term in the Taylor expansion of this expansion cancels, whereas the second term will give as before $\frac{1}{2}x(1-x)f''(x)$.

It is now easy to introduce selection in this model. Every individual has its Poisson clock, but now the individuals of type A_1 have a slower Poisson clock. This means that an individual of type A_1 has a smaller rate to change type, and hence A_1 has a larger fitness. In this modified version of the Moran model one chooses the birth and death rates as

$$b_i(s) = b_i = i(2N-i)/(2N)$$

and

$$d_i(s) = (1-s)i(2N-i)/(2N)$$

where $s \in (0,1)$ is the fitness parameter $s = 0$ means both alleles have the same fitness and $s \to 1$ means that A_1 has a much larger fitness. In order to obtain a diffusion limit, one chooses $s = s_N$ Taylor expansion up to second order of (8.3) gives

$$\begin{aligned} L_N f(x) &= 2N\left(d_{2Nx}(f(x-\frac{1}{N}) - f(x)) + c_{2Nx}(f(x+\frac{1}{N}) - f(x))\right) \\ &= x(1-x)2Ns_N f'(x) + \frac{1}{2}x(1-x)(1-s_N)f''(x) + \mathcal{R}_N \end{aligned}$$

Therefore, if we choose $s_N = \sigma/4N$, we obtain as the limiting diffusion generator

$$\mathcal{L}f(x) = \frac{\sigma}{2}x(1-x)f'(x) + \frac{1}{2}x(1-x)f''(x) \tag{8.4}$$

which corresponds to the stochastic differential equation

$$dx_t = \frac{\sigma}{2}x_t(1-x_t)dt + \sqrt{x_t(1-x_t)}dW_t \tag{8.5}$$

If we now combine mutation and selection, we obtain the limiting diffusion

$$dx_t = M(x_t)dt + \sqrt{V(x_t)}dW_t \tag{8.6}$$

with drift

$$M(x) = \frac{q}{2}(1-x) - \frac{r}{2}x + \frac{\sigma}{2}x(1-x) \tag{8.7}$$

and variance

$$V(x) = x(1-x) \tag{8.8}$$

Using formula (5.8) we obtain for the density of stationary distribution

$$\psi(x) = Cx^{q-1}(1-x)^{r-1}e^{\frac{\sigma}{2}x} \tag{8.9}$$

9 A note on reversibility and open questions

9.1 Reversibility

In the diffusion limits which we obtained, we could always identify explicitely the invariant measure. This is because one dimensional diffusions are always reversible, and for a reversible diffusion there is a simple relation between drift, variance and the invariant measure.

By reversibility, we mean the following. If we start the diffusion $\{x_t : t \geq 0\}$ from its stationary measure μ, then we obtain a stationary process which can be extended to negative times (by stationarity). More precisely, we define x_{-t} via

$$\mathbb{E}_\mu(f(x_{-t})g(x_0)) = \mathbb{E}_\mu(f(x_0)g(x_t))$$

or in semigroup notation

$$\int S_{-t}fgd\mu = \int fS_tgd\mu$$

i.e., $S(-t) = S^*(t)$ where $*$ is now in $L^2(\mu)$.

The process is said to be reversible if one of the following three equivalent conditions hold

1. For all f, g
$$\int (S_tf)gd\mu = \int f(S_tg)d\mu$$
2. For all f, g (in the domain of the generator)
$$\int (Lf)gd\mu = \int f(Lg)d\mu$$
3. In distribution we have the equality
$$\{x_t : t \in \mathbb{R}\} = \{x_{-t} : t \in \mathbb{R}\}$$

If the diffusion admits a transition probability densities $p_t(x, y)$, and the invariant measure is absolutely continuous with probability density ψ (i.e., $\mu(dx) = \psi(x)dx$) then reversibility is equivalent with the so-called detailed balance condition

$$\psi(x)p_t(x, y) = p_t(y, x)\psi(y)$$

A diffusion process on $\mathbb{R}^d$ is of the form

$$dX_t = M(X_t)dt + \sqrt{V(X_t)}dW_t$$

where now the drift $M(X_t)$ is a vector, and the "variance" $V(X_t)$ a positive matrix.

If we have the relation

$$M(x) = -\frac{1}{2}V(x).\nabla U + \frac{1}{2}\nabla.V(x) \tag{9.1}$$

for some $U : \mathbb{R}^d \to \mathbb{R}$ such that $\int e^{-U(x)}dx < \infty$, and where

$$\nabla U = \sum_{i=1}^{d} \frac{\partial U}{\partial x_i}\vec{e_i}$$

$$\nabla.V(x) = \sum_{i=1}^{d} \frac{\partial V_{ij}(x)}{\partial x_i}\vec{e_j}$$

then the measure

$$\mu(dx) = e^{-U(x)}dx$$

is the unique reversible measure for the diffusion $\{X_t : t \geq 0\}$. In fact this relation (9.1) is necessary and sufficient for reversibility. In $d = 1$, we can solve for U and recover

$$e^{-U(x)} = Ce^{-\int_0^x \frac{M-\frac{1}{2}V'}{\frac{1}{2}V}} = \frac{C}{V(x)}e^{-\int_0^x \frac{2M}{V}}$$

what we already found in (5.7). In $d > 1$, (9.1) is satisfied if $V(x)^{-1}(M(x) - \frac{1}{2}\nabla.V(x))$ is curl-free. If this is not satisfied, it becomes a non-trivial problem to identify the stationary measure or even prove existence and or uniqueness of it. Multidimensional diffusions occur naturally as diffusion limits of Markov chain models in the multi-allelic case.

9.2 Questions

This is an incomplete list: please add and or remove questions.

1. Condition WF on survival and take the diffusion limit: what are the effective mutation rate giving the same diffusion limit ?

2. How to introduce selection in the WF model (without having to pass via the Moran model). Gives this the same diffusion limit for weak selection

3. How to deal with fluctuating population sizes: what becomes then the good rescaling in order to obtain a diffusion limit.

4. Are there microscopic models for selection which after a proper rescaling lead to the same diffusion limit?

5. What can be said about corrections to the diffusion limit for large but finite N ?

References

[1] Barbour A.D., Ethier, S.N. and Griffiths R.C., A transition function expansion for a diffusion model with selection, Ann. Prob. **10** 123-162 (2000).

[2] Billingsley, P. , *Convergence of probability measures*, Second edition, Wiley, (1999).

[3] Buss, S.R., and Clote, P., Solving the Fisher-Wright and coalescence problems with a discrete Markov chain analysis, Applied Prob. Trust **36**, 1175-1197 (2004).

[4] Donnelly, P. and Rodrigues, E. R., Convergence to stationarity on the Moran model, J. Appl. Prob. , **37**, 705-717 (2000).

[5] Etheridge, A., Stochastic models in genetic evolution, lecture notes available at homepage of the author.

[6] Den Hollander, F, Stochastic models for genetic evolution, lecture notes in preparation.

[7] Liggett, T.M., *Interacting particle systems*, Springer Berlin, (1985)