

# ORFE 527: Stochastic Calculus

## Homework 5

Zachary Hervieux-Moore

Saturday 8<sup>th</sup> April, 2017

The purpose of this homework is to complete the proof that the gene frequency in the *Moran model* converges to the solution of

$$dX_t = \sqrt{X_t(1 - X_t)}dB_t, \quad X_0 = x_0$$

as the number of individuals  $N$  becomes large.

Recall that in the Moran model with  $N$  individuals

- each individual carries one of the two genes  $A$  or  $B$ ;
- each pair of individuals is assigned an independent exponential random variable of parameter 1;
- at the (random) time given by the smallest of these random variables, the corresponding pair of individuals produces 2 children who both carry one of their parents' genes (chosen at random) and the parents die;
- the procedure is repeated indefinitely.

We write  $X_t^{(N)}$  for the frequency of gene  $A$  (i.e. the proportion of individuals carrying it) at time  $t$  in the Moran model with  $N$  individuals. We assume throughout that  $X_0^{(N)}$  is non-random and that the limit  $x_0 := \lim_{N \rightarrow \infty} X_0^{(N)}$  exists.

**Exercise 1:** (Existence of limits along subsequences) The goal of this exercise is to verify that, on every fixed time interval  $[0, T]$ , every subsequence of processes  $X^{(N)}$ ,  $N \in \mathcal{N}$  has a further subsequence converging in distribution.

- a) Show that the sequence of (random) time  $\tau_1 \leq \tau_2 \leq \dots$  at which new children are born forms a Poisson process with parameter  $\binom{N}{2}$ .
- b) Show that at the times when new children are born the gene frequency  $X^{(N)}$  either goes up by  $\frac{1}{N}$ , or goes down by  $\frac{1}{N}$ , or stays the same. Compute the probabilities of the three possibilities.
- c) Use parts a) and b) to verify that  $X^{(N)}$  is a martingale.
- d) Use a combination of Prokhorov's Theorem with Arzela-Ascoli Theorem to check that it is enough to verify

$$\forall C > 0, \delta > 0 \exists \eta \in (0, T] : \\ \liminf_{N \rightarrow \infty} \mathbb{P} \left( \sup_{0 \leq s, t \leq T, |t-s| \leq \eta} |X_t^{(N)} - X_s^{(N)}| \leq C \right) \geq 1 - \delta$$

to obtain the goal of the exercise (i.e. the existence of limits along subsequences of  $X^{(N)}$ ,  $N \in \mathbb{N}$  on  $[0, T]$ ).

*Hint:* note that even though the paths of  $X^{(N)}$  are not continuous, one can find for each path a continuous path which deviates from  $X^{(N)}$  by at most  $\frac{1}{N}$ . Then, check that Prokhorov's Theorem applies to the laws of the approximate continuous paths.

- e) Prove the above statement by fixing  $C > 0$ ,  $\delta > 0$ , viewing the interval  $[0, T]$  as the union of intervals  $[0, \eta]$ ,  $[\eta, 2\eta]$ ,  $\dots$ , using the union bound, and finally Doob's inequality for submartingales.

**Answer:**

- a) Let  $Y_{ij}$  denote the random variable that individual  $i$  and  $j$  meet,

$$\begin{aligned}
\mathbb{P}(\tau_i \leq t) &= \mathbb{P}(\min_{i,j} Y_{ij} \leq t) \\
&= 1 - \mathbb{P}(\min_{i,j} Y_{ij} \geq t) \\
&= 1 - \mathbb{P}(\text{all } Y_{ij} \geq t) \\
&= 1 - \prod_{i,j \in N} \mathbb{P}(Y_{ij} \geq t) \quad \text{by independence} \\
&= 1 - \prod_{i,j \in N} e^{-t} \\
&= 1 - e^{-\sum_{i,j \in N} t} \\
&= 1 - e^{-\binom{N}{2}t}
\end{aligned}$$

Where the last line is the CDF for an exponential with parameter  $\binom{N}{2}$ .

- b) Suppose that both parents have gene  $A$ . This occurs with probability  $X_t^{(N)^2}$ . Then both offspring have gene  $A$  with probability 1. Likewise, if both parents have gene  $B$  the children have gene  $B$  with probability 1. This occurs with probability  $(1 - X_t^{(N)})^2$ . In both these cases, the gene frequency does not change. Now suppose one parent has gene  $A$  and the other has gene  $B$ . This occurs with probability  $2X_t^{(N)}(1 - X_t^{(N)})$ . The children have both gene  $A$  with probability  $1/2$ , which increases the gene frequency. The children both get gene  $B$  with probability  $1/2$ , which decreases the gene frequency. Putting this all together

$$\begin{aligned}
\mathbb{P}(\text{gene frequency stays the same}) &= X_t^{(N)^2} + (1 - X_t^{(N)})^2 \\
\mathbb{P}(\text{gene frequency increases}) &= X_t^{(N)}(1 - X_t^{(N)}) \\
\mathbb{P}(\text{gene frequency decreases}) &= X_t^{(N)}(1 - X_t^{(N)})
\end{aligned}$$

Now, the increase or decrease is  $\frac{1}{N}$  because, by design, when the gene frequency changes, it can only be 1 individual switching its gene. That is parents  $AB$  produce children  $AA$  or  $BB$ . Thus only 1 in  $N$  individuals is changing genes. That is, with parents  $AB$ , the frequency increases by  $\frac{1}{N}$  if  $AA$  is produced or decreases by  $\frac{1}{N}$  if  $BB$  is produced.

c) Since  $X_t^{(N)} \in [0, 1]$  then  $\mathbb{E}[|X_t^{(N)}|] < \infty$ . Now we show the martingale property. Let  $N_t$  denote the number of hitting times  $\tau_i$  between  $s$  and  $t$ . We suppose that  $\tau_i \leq s \leq \tau_{i+1} \leq \dots \leq \tau_{i+N_t} \leq t \leq \tau_{i+N_t+1}$ . We have

$$X_t^{(N)} = X_s^{(N)} + \sum_{i=1}^{N_t} (X_{\tau_{i+1}} - X_{\tau_i})$$

We define a new process  $Z_i = X_{\tau_{i+1}} - X_{\tau_i}$  which we note are i.i.d. Thus,

$$\mathbb{E}[X_t^{(N)} - X_s^{(N)} | \mathcal{F}_s] = \mathbb{E}\left[\sum_{i=1}^{N_t} Z_i | \mathcal{F}_s\right]$$

We can now apply Wald's equality because  $N_t$  is a nonnegative random variable and the  $Z_i$ 's are i.i.d.

$$\begin{aligned} \mathbb{E}[X_t^{(N)} - X_s^{(N)} | \mathcal{F}_s] &= \mathbb{E}[N_t | \mathcal{F}_s] \mathbb{E}[Z_i | \mathcal{F}_s] \\ &= \mathbb{E}[N_t] \mathbb{E}[Z_i | \mathcal{F}_s] \end{aligned}$$

Since  $N_t$  is independent of  $\mathcal{F}_s$ . We also note that  $\mathbb{E}[N_t] < \infty$  as the  $\tau_i$  are Poisson. Now, we use the tower property on  $\mathcal{F}_{\tau_i} \subseteq \mathcal{F}_s$

$$\mathbb{E}[Z_i | \mathcal{F}_s] = \mathbb{E}[\mathbb{E}[Z_i | \mathcal{F}_{\tau_i}] | \mathcal{F}_s]$$

Now, using part b), we have

$$\begin{aligned} \mathbb{E}[Z_i | \mathcal{F}_{\tau_i}] &= \mathbb{E}[X_{\tau_{i+1}} - X_{\tau_i} | \mathcal{F}_{\tau_i}] \\ &= \mathbb{E}[X_{\tau_{i+1}} | \mathcal{F}_{\tau_i}] - X_{\tau_i} \\ &= X_{\tau_i} \mathbb{P}(\text{no change in gene frequency}) \\ &\quad + (X_{\tau_i} + \frac{1}{N}) \mathbb{P}(\text{increase in gene frequency}) \\ &\quad + (X_{\tau_i} - \frac{1}{N}) \mathbb{P}(\text{decrease in gene frequency}) - X_{\tau_i} \\ &= X_{\tau_i} (X_{\tau_i}^{(N)^2} + (1 - X_{\tau_i}^{(N)})^2) \\ &\quad + (X_{\tau_i} + \frac{1}{N}) (X_{\tau_i}^{(N)} (1 - X_{\tau_i}^{(N)})) \\ &\quad + (X_{\tau_i} - \frac{1}{N}) (X_{\tau_i}^{(N)} (1 - X_{\tau_i}^{(N)})) - X_{\tau_i} \\ &= X_{\tau_i} - X_{\tau_i} \\ &= 0 \end{aligned}$$

Therefore

$$\mathbb{E}[X_t^{(N)} - X_s^{(N)} | \mathcal{F}_s] = \mathbb{E}[N_t] \mathbb{E}[\mathbb{E}[0 | \mathcal{F}_{\tau_i}] | \mathcal{F}_s] = 0$$

And so we conclude that  $X^{(N)}$  is a martingale.

d) Noting the hint, we define our continuous approximation as

$$\tilde{X}_t^{(N)} = \tilde{X}_0^{(N)} + \sum_{i=1}^{N_t} (X_{\tau_i}^{(N)} - X_{\tau_{i-1}}^{(N)}) + \frac{t - \tau_{N_t}}{\tau_{N_t+1} - \tau_{N_t}} (X_{\tau_{N_t+1}} - X_{\tau_{N_t}})$$

Which is indeed continuous and deviates from  $X^{(N)}$  by at most  $\frac{1}{N}$ . Now, define the set

$$K_n^{C+\frac{1}{N}} = \{\tilde{X}^{(N)} : \sup_{[0,T]} |\tilde{X}| \leq K, \sup_{0 \leq s, t \leq T, |t-s| \leq \eta} |\tilde{X}_t^{(N)} - \tilde{X}_s^{(N)}| \leq C + \frac{1}{N}\}$$

Then, by Arzela-Ascoli,  $K_n$  is pre-compact. We then have by assumption

$$\mathbb{P}(\tilde{X}^{(N)} \in K_n^{C+\frac{1}{N}}) \geq \mathbb{P}(X^{(N)} \in K_n^C) \geq 1 - \delta$$

Thus, by Prokhorov's theorem (using the closure of  $K_n$ ) we have that there exists a convergent subsequence  $N_k$  with

$$\tilde{X}^{(N_k)} \xrightarrow{d} \tilde{X}^\infty$$

Now we show that  $X^{N_k}$  converges to the same thing. Pick  $f \in C([0, T])$ . As  $[0, T]$  is compact, then  $f$  is uniformly continuous. Then for all  $\epsilon > 0$ , there exists  $\eta > 0$  such that  $|x - y| < \eta$  implies  $|f(x) - f(y)| \leq \epsilon/2$ . Now, for  $k$  large enough we have

$$\begin{aligned} \mathbb{E}[f(\tilde{X}^{(N_k)})] - \epsilon/2 &\leq \mathbb{E}[f(X^{(N_k)})] \leq \mathbb{E}[f(\tilde{X}^{(N_k)})] + \epsilon/2 \\ \text{and } \mathbb{E}[f(\tilde{X}^\infty)] - \epsilon/2 &\leq \mathbb{E}[f(\tilde{X}^{(N_k)})] \leq \mathbb{E}[f(\tilde{X}^\infty)] + \epsilon/2 \end{aligned}$$

Putting these inequalities together yields

$$\mathbb{E}[f(\tilde{X}^\infty)] - \epsilon \leq \mathbb{E}[f(X^{(N_k)})] \leq \mathbb{E}[f(\tilde{X}^\infty)] + \epsilon$$

And so, for all  $f \in C([0, T])$  we have

$$\begin{aligned}\mathbb{E}[f(X^{(N_k)})] &\rightarrow \mathbb{E}[f(\tilde{X}^\infty)] \\ \implies X^{(N_k)} &\xrightarrow{d} \tilde{X}^\infty\end{aligned}$$

and we conclude that along subsequences of  $X^{(N)}$  we have the existence of limits, namely  $\tilde{X}^\infty$ .

e) Fix  $C > 0$  and  $\delta > 0$ , define  $K_t = \max\{k : \eta k \leq T\}$ , then

$$\begin{aligned}&\mathbb{P}\left(\sup_{0 \leq s, t \leq T, |t-s| \leq \eta} |X_t^{(N)} - X_s^{(N)}| \leq C\right) \\ &= 1 - \mathbb{P}\left(\sup_{0 \leq s, t \leq T, |t-s| \leq \eta} |X_t^{(N)} - X_s^{(N)}| \geq C\right) \\ &\geq 1 - \mathbb{P}\left(\bigcup_{k=0}^{K_t} \sup_{0 \leq s, t \leq T, |t-s| \leq \eta} |X_t^{(N)} - X_{k\eta}| + |X_{k\eta} - X_s^{(N)}| \geq C\right)\end{aligned}$$

By union bound

$$\begin{aligned}&\geq 1 - \sum_{k=0}^{K_t} \mathbb{P}\left(\sup_{0 \leq s, t \leq T, |t-s| \leq \eta} |X_t^{(N)} - X_{k\eta}| + |X_{k\eta} - X_s^{(N)}| \geq C\right) \\ &\geq 1 - \sum_{k=0}^{K_t} \mathbb{P}\left(\sup_{k\eta \leq t \leq (k+1)\eta} |X_t^{(N)} - X_{k\eta}| + \sup_{(k-1)\eta \leq t \leq k\eta} |X_{k\eta} - X_s^{(N)}| \geq C\right) \\ &\geq 1 - \sum_{k=0}^{K_t} \mathbb{P}\left(\sup_{k\eta \leq t \leq (k+1)\eta} |X_t^{(N)} - X_{k\eta}| \geq C/2\right) \\ &\quad + \mathbb{P}\left(\sup_{(k-1)\eta \leq t \leq k\eta} |X_{k\eta} - X_s^{(N)}| \geq C/2\right) \\ &\geq 1 - 2 \sum_{k=0}^{K_t} \mathbb{P}\left(\sup_{k\eta \leq t \leq (k+1)\eta} |X_t^{(N)} - X_{k\eta}| \geq C/2\right) \\ &\geq 1 - 2 \sum_{k=0}^{K_t} \mathbb{P}\left(\sup_{0 \leq t \leq \eta} |X_t^{(N)} - X_0| \geq C/2\right)\end{aligned}$$

Where the last line results because, we have  $\mathbb{P}(\sup_{k\eta \leq t \leq (k+1)\eta} |X_t^{(N)} - X_{k\eta}| \geq C/2) = \mathbb{P}(\sup_{0 \leq t \leq \eta} |X_t^{(N)} - X_0| \geq C/2)$  by stationarity. Now, we apply Doob's inequality for submartingales (as  $X^{(N)}$  is a martingale)

$$\begin{aligned} &\geq 1 - 4 \sum_{k=0}^{K_t} \frac{\mathbb{E}[|X_{k\eta} - X_0|]}{C} \\ &\geq 1 - 4 \left\lceil \frac{T}{\eta} \right\rceil \frac{1}{NC} \quad \text{for } \eta \text{ small enough} \end{aligned}$$

Where the last line is justified because we can shrink  $\eta$  down to the point where it is just after the first meeting. Thus, the difference is at most  $\frac{1}{N}$ . Now, taking  $N \rightarrow \infty$  achieves the result that

$$\liminf_{N \rightarrow \infty} \mathbb{P} \left( \sup_{0 \leq s, t \leq T, |t-s| \leq \eta} |X_t^{(N)} - X_s^{(N)}| \leq C \right) \geq 1 - \delta$$

**Exercise 2:** (Uniqueness for the limiting SDE) We have checked in class that any limit  $X$  of a subsequence of  $X^{(N)}$ ,  $N \in \mathbb{N}$  on  $[0, T]$  solves the martingale problem associated with the Moran model SDE.

- a) Prove that any such  $X$  is a weak solution of the SDE
- b) Prove that the weak solution of the SDE is unique, e.g. by showing that the strong solution of the SDE is unique and using the Yamada-Watanabe Theorem.
- c) Conclude that the (full) sequence  $X^{(N)}$ ,  $N \in \mathbb{N}$  converges to the unique weak solution of the weak solution of the SDE.

**Answer:**

- a) By Stroock-Varadhan theorem, since  $X$  solves the martingale problem for the SDE, then  $X$  is a weak solution of the SDE as well
- b) By Yamada-Watanabe, strong uniqueness implies weak uniqueness. Thus, we only have to check if this has a unique strong solution. We bound  $\sigma(\cdot)$

$$\begin{aligned}
 |\sigma(t, X) - \sigma(t, \tilde{X})| &= |\sqrt{X(1-X)} - \sqrt{\tilde{X}(1-\tilde{X})}| \\
 &\leq \sqrt{|X - X^2 - \tilde{X} + \tilde{X}^2|} \\
 &\leq \sqrt{|X - \tilde{X}|} \quad \text{since } X, \tilde{X} \in [0, 1]
 \end{aligned}$$

Thus picking  $h(u) = \sqrt{u}$ , we have that  $h(0) = 0$ ,  $h$  increasing, and  $\int_0^\epsilon \frac{1}{h(u)^2} du = \infty$  and so by the other Yamada-Watanabe theorem, we have strong uniqueness and so weak uniqueness as well.

- c) Since all the subsequences converge to the same limiting point by uniqueness of the previous part, then the original sequence must converge to the same limit point. That is, we can just pick

$$C = \max\{C_k : |X^{(N_k)} - X^\infty| < \epsilon \text{ for all } N_k \geq C_k\}$$

for all subsequence  $(N_k)$  with  $(N) = \bigcup_{N_k} (N_k)$ . That is the union of all the converging subsequences contains any  $N$  since  $X^{(N)}$  is pre-compact. Then we have

$$|X^{(N)} - X^\infty| < \epsilon \quad \forall N \geq C$$



So we have that the full sequence converges and so, by uniqueness in the previous part, the full sequence  $X^{(N)}$  converges to the unique weak solution.