

HW5 SOLUTIONS

PIERRE YVES GAUDREAU LAMARRE

1. Existence of Limits Along Subsequences

(a). Let $(M_t^{(i)})_{t \geq 0}$ (for $1 \leq i \leq \binom{N}{2}$) be the counting processes associated with the reproduction times of every possible $\binom{N}{2}$ pairs of individuals. More precisely, for every $0 \leq s < t$, the random variable $M_t^{(i)} - M_s^{(i)}$ represents the number of times that the pair of individuals labelled " i " has reproduced in the time interval $(s, t]$. By assumption, $(M_t^{(i)})_{t \geq 0}$ are i.i.d. Poisson processes with parameter 1.

By definition of the Moran model, the random times $\tau_1 < \tau_2 < \dots$ correspond to the arrival times in the process

$$(1.1) \quad M_t := \sum_{i=1}^{\binom{N}{2}} M_t^{(i)}, \quad t \geq 0,$$

that is, the combination of reproduction times of every pair of individuals. The result then follows directly from the classical *superposition principle* for Poisson processes, whose general statement we recall here:

Theorem 1.1 (Superposition Principle). *Let $\tilde{N}^{(1)}, \dots, \tilde{N}^{(k)}$ be independent Poisson processes with respective parameters $\lambda_1, \dots, \lambda_k > 0$. Then, $\tilde{N}^{(1)} + \dots + \tilde{N}^{(k)}$ is a Poisson process with parameter $\lambda_1 + \dots + \lambda_k$.*

(b). Let $g := (g_1, g_2)$ represent the genes of the two individuals that are about to reproduce, and let $h := (h_1, h_2)$ be the genes of their children. We know that

$$g \in \{(A, A), (A, B), (B, A), (B, B)\},$$

and since both siblings must inherit the same gene,

$$h \in \{(A, A), (B, B)\}.$$

By assumption,

$$(1.2) \quad \mathbf{P}[h = (A, A) | g = (g_1, g_2)] = \begin{cases} 1 & \text{if } (g_1, g_2) = (A, A), \\ 1/2 & \text{if } (g_1, g_2) = (A, B), \\ 1/2 & \text{if } (g_1, g_2) = (B, A), \\ 0 & \text{if } (g_1, g_2) = (B, B), \end{cases}$$

and

$$(1.3) \quad \mathbf{P}[h = (B, B) | g = (g_1, g_2)] = \begin{cases} 0 & \text{if } (g_1, g_2) = (A, A), \\ 1/2 & \text{if } (g_1, g_2) = (A, B), \\ 1/2 & \text{if } (g_1, g_2) = (B, A), \\ 1 & \text{if } (g_1, g_2) = (B, B). \end{cases}$$

Going from (A, A) to (A, A) or from (B, B) to (B, B) leaves the proportion of A genes unchanged, going from $(A, B), (B, A)$ to (A, A) increases it by $1/N$, and going from $(A, B), (B, A)$ to (B, B) decreases it by $1/N$. Since these are the only possible transitions, we conclude that the increments of $X^{(N)}$ take values in $\{-1/N, 0, 1/N\}$.

We now compute the transition probabilities. Let $(Z_n)_{n \in \mathbb{N} \cup \{0\}}$ be the process representing the proportion of A genes after each reproduction. That is,

$$(1.4) \quad Z_n = X_{\tau_n}^{(N)}, \quad n \geq 0$$

(where we assume by convention that $\tau_0 = 0$). By (1.2) and (1.3), we know that

$$\begin{aligned} \mathbf{P}[Z_{n+1} - Z_n = -1/N] &= \frac{1}{2} \mathbf{P}[\text{the parents selected at } \tau_n \text{ have different genes}], \\ \mathbf{P}[Z_{n+1} - Z_n = 0] &= \mathbf{P}[\text{the parents selected at } \tau_n \text{ have the same genes}], \\ \mathbf{P}[Z_{n+1} - Z_n = 1/N] &= \frac{1}{2} \mathbf{P}[\text{the parents selected at } \tau_n \text{ have different genes}]. \end{aligned}$$

At time τ_n , there are

$$N^2 Z_n (1 - Z_n)$$

ways of choosing two parents with different genes. Since there are $\binom{N}{2} = \frac{N(N-1)}{2}$ ways of choosing two parents, we therefore conclude that

$$(1.5) \quad \mathbf{P}[Z_{n+1} - Z_n = -1/N | Z_n] = \frac{Z_n(1 - Z_n)}{(1 - 1/N)},$$

$$(1.6) \quad \mathbf{P}[Z_{n+1} - Z_n = 0 | Z_n] = 1 - \frac{2Z_n(1 - Z_n)}{(1 - 1/N)},$$

$$(1.7) \quad \mathbf{P}[Z_{n+1} - Z_n = 1/N | Z_n] = \frac{Z_n(1 - Z_n)}{(1 - 1/N)}.$$

(c). Since $X^{(N)}$ takes values in $[0, 1]$, it is integrable, and we can make it adapted to its own filtration. Thus, it only remains to prove the martingale property. Let us define M and Z as in (1.1) and (1.4) respectively, so that

$$(1.8) \quad X_t^{(N)} = Z_{M_t}, \quad t \geq 0.$$

By the total probability rule, for any $0 \leq s \leq t$,

$$\begin{aligned} &\mathbf{E} \left[X_t^{(N)} - X_s^{(N)} | X_s^{(N)} \right] \\ &= \sum_{m,n=0}^{\infty} \mathbf{E} \left[X_t^{(N)} - X_s^{(N)} | X_s^{(N)}, M_s = m, M_t = m + n \right] \mathbf{P} \left[M_s = m, M_t = m + n | X_s^{(N)} \right] \\ &= \sum_{m,n=0}^{\infty} \mathbf{E} \left[Z_{m+n} - Z_m | Z_m \right] \mathbf{P} \left[M_s = m, M_t = m + n | X_s^{(N)} \right], \end{aligned}$$

where the last equality follows from a combination of (1.8) and the fact that the reproduction times and the selection of genes are independent. By (1.5)–(1.7), we know that $(Z_n)_{n \in \mathbb{N} \cup \{0\}}$ is a discrete-time martingale. Thus $\mathbf{E} [Z_{m+n} - Z_m | Z_m] = 0$ for every $m, n \geq 0$, concluding the proof of the martingale property for $X^{(N)}$.

(d).

Step 1: Tightness. Let $(Y_t^{(N)})_{t \geq 0}$ be the linear interpolation process of $X^{(N)}$'s increments. We recall that $Y^{(N)}$ takes values in $C[0, T]$, and that

$$(1.9) \quad \|Y^{(N)} - X^{(N)}\|_{\infty} := \sup_{t \in [0, T]} |Y_t^{(N)} - X_t^{(N)}| \leq 1/N.$$

We want to prove that every subsequence of $Y^{(N)}$ has a further subsequence that converges in distribution in the space $C[0, T]$. According to Prokhorov's theorem, it suffices

to prove that the sequence $(Y^{(N)})_{N \in \mathbb{N}}$ is tight, that is, for every $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset (C[0, T], \|\cdot\|_\infty)$ and $N_\varepsilon \in \mathbb{N}$ such that

$$(1.10) \quad \inf_{N \geq N_\varepsilon} \mathbf{P} \left[Y^{(N)} \in K_\varepsilon \right] \geq 1 - \varepsilon.$$

Step 2: From $X^{(N)}$ to $Y^{(N)}$. At this point, we must prove that the condition stated in the problem statement, namely,

$$(1.11) \quad \forall C, \delta > 0, \exists \eta \in (0, T] : \liminf_{N \rightarrow \infty} \mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |X_t^{(N)} - X_s^{(N)}| \leq C \right] \geq 1 - \delta,$$

implies that $Y^{(N)}$ is tight in the sense of (1.10). As a first step toward showing this, we prove that (1.11) implies that

$$(1.12) \quad \forall C, \delta > 0, \exists \eta \in (0, T] : \liminf_{N \rightarrow \infty} \mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |Y_t^{(N)} - Y_s^{(N)}| \leq C \right] \geq 1 - \delta.$$

Let $C, \delta > 0$ be fixed. Then, we know there is some $\eta \in (0, T]$ such that

$$\liminf_{N \rightarrow \infty} \mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |X_t^{(N)} - X_s^{(N)}| \leq C/2 \right] \geq 1 - \delta.$$

According to (1.9), for any $s, t \in [0, \infty)$ and $N \in \mathbb{N}$,

$$\begin{aligned} |Y_t^{(N)} - Y_s^{(N)}| &\leq |X_t^{(N)} - X_s^{(N)}| + |Y_t^{(N)} - X_t^{(N)}| + |Y_s^{(N)} - X_s^{(N)}| \\ &\leq |X_t^{(N)} - X_s^{(N)}| + 2/N \end{aligned}$$

In particular, given that $2/N \leq C/2$ for large enough N ,

$$\begin{aligned} &\liminf_{N \rightarrow \infty} \mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |Y_t^{(N)} - Y_s^{(N)}| \leq C \right] \\ &\geq \liminf_{N \rightarrow \infty} \mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |X_t^{(N)} - X_s^{(N)}| \leq C - 2/N \right] \\ &\geq \liminf_{N \rightarrow \infty} \mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |X_t^{(N)} - X_s^{(N)}| \leq C/2 \right] \geq 1 - \delta, \end{aligned}$$

as desired.

Step 3: Condition on $Y^{(N)}$ Implies Tightness. We conclude the proof with the claim that (1.12) implies tightness. It is easy to see that (1.12) implies that

$$(1.13) \quad \forall C, \delta > 0, \exists \eta \in (0, T], N_0 \in \mathbb{N} : \inf_{N \geq N_0} \mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |Y_t^{(N)} - Y_s^{(N)}| \leq C \right] \geq 1 - \delta.$$

That (1.13) is equivalent to tightness by the Arzelá-Ascoli theorem is a classical result in probability; the standard reference is [2, Theorem 7.3]; for a reference that is available online on the Princeton University Library website¹, see [1, Theorem 5.3].

(e). We now prove (1.11). For the entirety of this proof, we assume that we have fixed some arbitrary $C > 0$ and $0 < \delta < 1$.

¹<https://catalog.princeton.edu/catalog/9028275>.

Step 1: Union Bound and Suprema on Smaller Intervals. Let $0 < \eta < T$ be fixed, and let $0 \leq s \leq t \leq T$ be such that $|t - s| \leq \eta$. Then, there is a unique integer $0 \leq k \leq \lfloor T/\eta \rfloor$ such that $k\eta \leq s \leq (k+1)\eta$ and $k\eta \leq t \leq (k+2)\eta$. On the one hand, if $k\eta \leq t \leq (k+1)\eta$, then

$$|X_t^{(N)} - X_s^{(N)}| \leq 2 \max \{ |X_t^{(N)} - X_{k\eta}^{(N)}|, |X_s^{(N)} - X_{k\eta}^{(N)}| \}.$$

On the other hand, if $(k+1)\eta \leq t \leq (k+2)\eta$, then

$$|X_t^{(N)} - X_s^{(N)}| \leq 3 \max \{ |X_t^{(N)} - X_{(k+1)\eta}^{(N)}|, |X_{(k+1)\eta}^{(N)} - X_{k\eta}^{(N)}|, |X_s^{(N)} - X_{k\eta}^{(N)}| \}.$$

(These two estimates follow from the triangle inequality.) Thus, it follows from a union bound that for every $N \in \mathbb{N}$,

$$\begin{aligned} (1.14) \quad & \mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |X_t^{(N)} - X_s^{(N)}| \geq 3C \right] \\ & \leq \mathbf{P} \left[\max_{k \in \mathbb{N}, k \leq T/\eta} \sup_{k\eta \leq t \leq (k+1)\eta} |X_t^{(N)} - X_{k\eta}^{(N)}| \geq C \right] \\ & \leq \sum_{k=0}^{\lfloor T/\eta \rfloor} \mathbf{P} \left[\sup_{0 \leq t \leq \eta} |X_{k\eta+t}^{(N)} - X_{k\eta}^{(N)}| \geq C \right]. \end{aligned}$$

Step 2: Martingale Inequality. Since $X^{(N)}$ is a martingale, for every $c > 0$, the process

$$X_{c+t}^{(N)} - X_c^{(N)}, \quad t \geq 0$$

is a martingale. Thus, by Jensen's inequality, for every $p \geq 1$, the process

$$|X_{c+t}^{(N)} - X_c^{(N)}|^p, \quad t \geq 0$$

is a submartingale (the process is trivially integrable since $X^{(N)}$ takes values in $[0, 1]$). Therefore, by Doob's submartingale inequality, we can estimate the tail probabilities of the suprema appearing in (1.14) by

$$(1.15) \quad \mathbf{P} \left[\sup_{0 \leq t \leq \eta} |X_{k\eta+t}^{(N)} - X_{k\eta}^{(N)}| \geq C \right] \leq \frac{\mathbf{E} \left[|X_{(k+1)\eta}^{(N)} - X_{k\eta}^{(N)}|^p \right]}{C^p}.$$

Step 3: Fluctuations of $X^{(N)}$ by a Random Walk Coupling. Let S_0, S_1, S_2, \dots be a simple symmetric random walk on \mathbb{Z} , that is, $S_0 = 0$, and

$$S_n := \sum_{k=1}^n I_k, \quad n \geq 1,$$

where I_1, I_2, \dots are i.i.d. uniform random variables on $\{-1, 1\}$.

Recall the definition of the sequence Z_0, Z_1, Z_2, \dots in (1.4), which represents the increments in the proportion of A genes at successive reproduction times. Moreover, for every $n \in \mathbb{N}$, let

$$R_n := \sum_{k=1}^n \mathbf{1}_{\{\text{the parents selected at } \tau_k \text{ have different genes}\}}, \quad n \in \mathbb{N}$$

count the number of times that the two parents have different genes in the first n reproductions. By definition of the Moran model, it is clear that for every $m \geq 0$, we can couple Z and S in such a way that

$$Z_{m+n} - Z_m = \frac{S_{R_{m+n}-R_m}}{N} = \sum_{k=1}^{R_{m+n}-R_m} \frac{I_k}{N},$$

where the I_k are assumed to be independent of R . That is, we assume that the randomness that determines which of the two parents gets to pass on their gene (when the parents don't have the same gene) is driven by the random walk S .

Implementing this in the case $p = 4$ in (1.15), if we denote the reproduction-times Poisson process M as in (1.1), then an application of the total probability rule yields

$$\mathbf{E} \left[\left| X_{(k+1)\eta}^{(N)} - X_{k\eta}^{(N)} \right|^4 \right] = \frac{1}{N^4} \sum_{k=0}^{\infty} \mathbf{E}[S_{R_k}^4] \mathbf{P}[M_{(k+1)\eta} - M_{k\eta} = k].$$

The fourth moment of the simple symmetric random walk is easily computed as

$$\mathbf{E}[S_n^4] = n(3n - 2), \quad n \in \mathbb{N}.$$

Since these moments are increasing in n and $R_k \leq k$ for all k , we get that there exists some constant $\tilde{C} > 0$ independent of all other parameters such that

$$(1.16) \quad \mathbf{E} \left[\left| X_{(k+1)\eta}^{(N)} - X_{k\eta}^{(N)} \right|^4 \right] \leq \frac{1}{N^4} \sum_{k=0}^{\infty} k(3k - 2) \mathbf{P}[M_{(k+1)\eta} - M_{k\eta} = k] \\ = \frac{\eta \binom{N}{2} + 3\eta^2 \binom{N}{2}^2}{N^4} \leq \tilde{C} \left(\frac{\eta}{N^2} + \eta^2 \right).$$

Step 4: Putting Everything Together. By combining (1.14), (1.15), and (1.16), we obtain the upper bound

$$\mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |X_t^{(N)} - X_s^{(N)}| \geq 3C \right] \leq (\lfloor T/\eta \rfloor + 1) \frac{\tilde{C}}{C^4} \left(\frac{\eta}{N^2} + \eta^2 \right)$$

for every $N \in \mathbb{N}$ and $\eta < T$. Thus, if we choose η small enough so that

$$\eta < T \quad \text{and} \quad (\lfloor T/\eta \rfloor + 1) \frac{\eta^2 \tilde{C}}{C^4} \leq \frac{\delta}{2},$$

then there exists some $N_0 \in \mathbb{N}$ (depending on T , C , and δ) large enough so that

$$\sup_{N \geq N_0} \mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |X_t^{(N)} - X_s^{(N)}| \geq 3C \right] \leq \delta,$$

namely, we want N_0 large enough so that

$$(\lfloor T/\eta \rfloor + 1) \frac{\eta^2 \tilde{C}}{C^4 N^2} \leq \frac{\delta}{2} \quad \text{for every } N \geq N_0.$$

This clearly implies that

$$\limsup_{N \rightarrow \infty} \mathbf{P} \left[\sup_{|s-t| \leq \eta, 0 \leq s, t \leq T} |X_t^{(N)} - X_s^{(N)}| \geq 3C \right] \leq \delta,$$

and since C and δ were arbitrary, the latter is equivalent to (1.11); hence the proof is complete.

2. Uniqueness for the Limiting SDE

(a). It was shown in class that the limit point X of any convergent subsequence of $Y^{(N)}$ solves the martingale problem associated with the SDE

$$(2.1) \quad dX_t = \sqrt{X_t(1 - X_t)} dB_t, \quad X_0 = x_0, \quad t \in [0, T].$$

The existence of a unique solution then readily follows from the general form of the so-called Stroock-Varadhan martingale problem, as stated, for instance, in [3, Chapter 5, Proposition 4.6].

(b). Define the function $\sigma(x) := \sqrt{x(1-x)}$ for $x \in [0, 1]$. It can be noted that for every $x, y \in [0, 1]$, one has

$$|\sigma(x) - \sigma(y)| \leq \sqrt{|x(1-x) - y(1-y)|} = \sqrt{|(x-y)(1-x-y)|}.$$

Given that $|1-x-y| \in [0, 1]$, this yields

$$|\sigma(x) - \sigma(y)| \leq \sqrt{|x-y|}.$$

Let $h(x) := \sqrt{|x|}$ for $x \geq 0$. This function h is strictly increasing, $h(0) = 0$, and we note that for every $\varepsilon > 0$,

$$\int_0^\varepsilon h(x)^{-2} dx = \int_0^\varepsilon \frac{1}{x} dx = \infty.$$

Thus, by the Yamada-Watanabe criterion (i.e., [3, Chapter 5, Theorem 2.13]), strong uniqueness holds for (2.1), which then implies weak uniqueness.

(c). Problem 1 implies that every subsequence of the continuous version $Y^{(N)}$ has a further subsequence that converges in distribution to some limiting process X . Problem 2 (a) implies that every such limit point X is a weak solution of the SDE (2.1). By Problem 2 (b), (2.1) has a unique weak solution, which implies in particular that every limit point X must have the same distribution. In short, we have proved that every every subsequence of $Y^{(N)}$ has a further subsequence that converges in distribution to the one and same limit, hence the full sequence converges in distribution to that limit.

References

- [1] Patrick Billingsley. *Weak convergence of measures: Applications in probability*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1971. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 5.
- [2] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [3] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.