# Asymptotic Theory of the Planar Stepping Stone Model

Ted Cox
BIRS, August 2004

## Contents

We focus on

- spatially explicit stepping stone model on the torus in $\mathbb{Z}^2$
- coalescing random walks on the torus in $\mathbb{Z}^2$
- effect of spatial structure on some measures of kinship

Based on work with Rick Durrett and Iljana Zähle:

*The stepping stone model: new formulas expose old myths (with R. Durrett).* **Ann. Appl. Probab.**, 12 (2002), 1348-1377.

*The stepping stone model II: Genealogies and the infinite sites model (with R. Durrett and I. Zähle).* To appear, **Ann. Appl. Probab.**

# 1 Discrete Time Stepping Stone Model

Model for the evolution of gene frequencies in a population.

Colonies of individuals subject to migration and mutation.

- **Time** discrete, $n = 0, 1, \ldots$
- **Colonies** located at sites $x \in \mathcal{S}$
- **Colony size** $N$ (diploid), $2N$ (haploid case)
- **Types** mutations produce new types
- **Mutation probability** $\mu > 0$
- **Migration probability** $\nu \in (0, 1]$
- **Migration kernel** $q(x, y)$, $x, y \in \mathcal{S}$, $q(x, x) = 0$

## Formation of generation $n+1$ at colony $C_x$

Each individual, independently:

- prob. $\mu$: assumes a new type
- prob. $1-\mu$: assumes type of individual chosen at random from colony $C_y$ with probability

$$p(x,y) = (1-\nu)I(x,y) + \nu q(x,y)$$

**Note**

- Wright-Fisher type model
- keep kernel $q(x,y)$ fixed, allow other parameters to vary

## Duality

We can trace **lineages** of individuals backward in time, following paths of **coalescing random walks**.

## Two lineages

$W_n^1, W_n^2$ : independent walks on $\mathcal{S}$, kernel $p(x, y)$.

$U_n^1, U_n^2$ : independent, uniformly distributed on $\{1, 2, \ldots, 2N\}$.

**Note.** In $k$ steps of $W_n^1$, about $\nu k$ are steps to different colonies.

## Hitting Times

Starting with walks at 0 and x,

$T_{0x} =$ time required for two lineages to **reside in the same colony**

$= \inf\{n : W_n^1 = W_n^2\}$

$t_{0x} =$ time required for two lineages to **coalesce**

$= \inf\{n : (W_n^1, U_n^1) = (W_n^2, U_n^2)\}$

Given the distribution of $t_{0x}$, we can compute

$$\begin{aligned} h &= \text{probability of } \textit{identity by descent} \text{ for two} \\ &\quad \text{(different) individuals picked at random} \\ &\quad \text{from entire population} \\ &= E(1-\mu)^{2t_0} \end{aligned}$$

$$\begin{aligned} \phi(x) &= \text{probability of } \textit{identity by descent} \text{ for two} \\ &\quad \text{(different) individuals picked at random} \\ &\quad \text{from colonies } C_0 \text{ and } C_x \\ &= E(1-\mu)^{2t_{0x}} \end{aligned}$$

$$\begin{aligned} F_{ST} &= \text{Wright's statistic} \\ &= \frac{\phi(0)-h}{1-h} \qquad \text{(following Nei (1975)} \end{aligned}$$

# 2 The Single Colony Case

One colony, migration parameter $\nu = 0$

$t_0$ is **geometric**,

$$P(t_0 = k) = (\frac{1}{2N})(1 - \frac{1}{2N})^{k-1}, \qquad k = 1, 2, \ldots$$

For large $N$, $\quad \dfrac{t_0}{2N} \overset{\mathrm{d}}{\approx} \mathcal{E}(1)$

**Calculation of $h$**

$$\begin{aligned} h &= E(1-2\mu)^{2t_0} \\ &= E((1-2\mu)^{2N})^{t_0/2N} \\ &\approx E(e^{-4N\mu\mathcal{E}(1)}) \\ &= \frac{1}{1+4N\mu} \end{aligned}$$

## $s > 2$ lineages

Consider $s$ distinct lineages (coalescing rw's).

$\zeta_n^s =$ number of remaining lineages at time $n$

For large $N$, can show

$$\zeta^s_{[2Nt]}) \overset{\mathrm{d}}{\approx} D^s_t, \qquad t > 0$$

where $D^s_t, t \geq 0$ on $\{1, 2, \ldots, s\}$ is the pure death process, with

$$k \to 1 \text{ at rate } \binom{k}{2}$$

Can also keep track (on time scale $2N$) of which of the $s$ random walks have coalesced , leading to **Kingman's coalescent**.

Consistent with previous $s = 2$ case?

$$\zeta_{[2Nt]} = 2 \quad \Longleftrightarrow \quad t_0 > [2Nt]$$

and

$$P(D^2_t = 2) = P(\mathcal{E}(1) > t) = e^{-t}$$

# 3 $\mathcal{S} =$ the torus $\Lambda(L)$ in $\mathbb{Z}^d$

$\Lambda(L) = (-L/2, L]^2 \cap \mathbb{Z}^2$ (wrap around)

Dynamics: each individual, independently:

- prob. $\mu$: assumes a new type
- prob. $1-\mu$: assumes type of individual chosen at random from colony $C_y$ with prob.

$$p^L(x,y) = (1-\nu)I(x,y) + \nu q^L(x,y)$$

where

$$q^L(x,y) = \sum_{z \in \mathbb{Z}^2} q(x, y + Lz)$$

Assumptions: $q(x,y) = q(y,x) = q(0, y-x)$, finite range, covariance matrix $\sigma^2 I$

Other models: island model, circular, ...

Seek distributions of

$$t^L_{0x} \quad \text{and} \quad \zeta^s_n$$

as $L \to \infty$, and possibly $N \to \infty$, $\nu \to 0$

# 4 Migration rate vs. system size

We consider the following "regimes"

Migration rate vs. system size

$$N\nu \gg \log L$$

$$N\nu \approx \log L$$

$$N\nu \ll \log L$$

Why? Let $\tilde{t}^L_{00}$ be coalescent time for two walks starting in same colony. Then

$$t^L_{0x} \stackrel{\mathrm{d}}{=} T^L_{0x} + \tilde{t}^L_{00}$$

and we will see that

$$\tilde{t}^L_{00} \approx 2NL^2$$

and if $|x| \approx L$ then

$$T^L_{0x} \approx \frac{L^2 \log L}{\nu}$$

Previous work: $2N = 1$, $\nu = 1$, nearest-neighbor kernel

# Migration rate vs. system size

$$T_{0x}^L \ll \tilde{t}_{00}^L \qquad \frac{N\nu}{\log L} \to \infty$$

$$T_{0x}^L \asymp \tilde{t}_{00}^L \qquad \frac{4\pi\sigma^2 N\nu}{\log L} \to \alpha \in (0,\infty)$$

$$T_{0x}^L \gg \tilde{t}_{00}^L \qquad \frac{N\nu}{\log L} \to 0$$

## Behavior of $t_{00}^L$:    high migration rate

**Theorem (C., Durrett)** *Assume* $\dfrac{N\nu}{\log L} \to \infty$ *as* $L \to \infty$. *Then for any* $t > 0$,

$$\sup_{x \in \Lambda_L} \left| P\left( \frac{t_{0,x}^L}{2NL^2} > t \right) - e^{-t} \right| \to 0.$$

That is, uniformly in $x \in \Lambda(L)$, for large $L$,

$$\frac{t_{0,x}^L}{2NL^2} \overset{\mathrm{d}}{\approx} \mathcal{E}(1)$$

We use

**Theorem (Strobeck (1987))**

$$E(t_{00}^L) = 2NL^2 (= \textit{population size})$$

## Spatial scales

$x_1, x_2, \ldots, x_s \in \Lambda(L)$

For $i \neq j$,

$$|x_i - x_j| \approx L^{\beta}, \qquad 0 \leq \beta \leq 1$$

$$\beta = 1 \iff \text{uniformly spread out on torus}$$

## Behavior of $T_{0x}^L$

**Theorem (C., Durrett)** *Suppose $x = x_L$ satisfies*

$$\frac{\log^+ |x|}{\log L} \to \beta \in [0,1] \qquad \textit{as } L \to \infty.$$

*For all $t > 0$, uniformly for $\nu \in (0,1]$,*

$$P\left( \frac{T_{0x}^L}{\left(\frac{L^2 \log L}{2\pi\sigma^2\nu}\right)} > t \right) \to \beta e^{-t}$$

That is, if $|x| \approx L^\beta$ and $\tau_L = \dfrac{L^2 \log L}{2\pi\sigma^2\nu}$ then

$$\frac{t_{0,x}^L}{\tau_L} \overset{\mathrm{d}}{\approx} \begin{cases} 0 & \text{prob. } 1-\beta \\ \mathcal{E}(1) & \text{prob. } \beta \end{cases}$$

**Low migration rate:** $N\nu = O(\log L)$

**Theorem (C., Durrett)** *Assume* $\dfrac{4\pi\sigma^2 N\nu}{\log L} \to \alpha \in [0,\infty)$ *and* $x = x_L$ *satisfies* $\dfrac{\log^+ |x|}{\log L} \to \beta \in [0,1]$ *as* $L \to \infty$. *Then for any* $t > 0$*,*

$$P\left(\frac{t^L_{0,x}}{\tau_L} > (1+\alpha)t\right) \to \left(\beta + (1-\beta)\frac{\alpha}{1+\alpha}\right) e^{-t}.$$

That is, if $\dfrac{4\pi\sigma^2 N\nu}{\log L} \approx \alpha$ and $|x| \approx L^\beta$ then

$$\frac{t^L_{0,x}}{(1+\alpha)\tau_L} \overset{\mathrm{d}}{\approx} \begin{cases} 0 & \text{prob. } \dfrac{1-\beta}{1+\alpha} \\[2em] \mathcal{E}(1) & \text{prob. } \dfrac{\beta+\alpha}{1+\alpha} \end{cases}$$

$$\tau_L = \frac{L^2 \log L}{2\pi\sigma^2\nu}$$

**Example.** Approximation of $\phi(x)$ using

$$\frac{t_{0,x}^L}{(1+\alpha)\tau_L} \overset{\mathrm{d}}{\approx} \begin{cases} 0 & \text{prob. } \dfrac{1-\beta}{1+\alpha} \\ \\ \mathcal{E}(1) & \text{prob. } \dfrac{\beta+\alpha}{1+\alpha} \end{cases}$$

where $\dfrac{4\pi\sigma^2 N\nu}{\log L} \approx \alpha$ and $|x| \approx L^\beta$

$$\begin{aligned} \phi(x) &= E((1-\mu)^{2t_{0x}^L}) \\ &= E(((1-\mu)^{2(1+\alpha)\tau_L})^{t_{0x}^L/(1+\alpha)\tau_L}) \\ &\approx \frac{1-\beta}{1+\alpha} + \frac{\beta+\alpha}{1+\alpha} E(((1-\mu)^{2(1+\alpha)\tau_L})^{\mathcal{E}(1)}) \\ &= \frac{1-\beta}{1+\alpha} + \frac{\beta+\alpha}{1+\alpha} \frac{1}{2(1+\alpha)\mu\tau_L} \end{aligned}$$

**Some special cases.**

1. $N\nu \ll \log L$ and $|x| \approx L$ ($\alpha = 0$ and $\beta = 1$)

$$h = \phi(x) \approx \frac{1}{1 + 2\mu\tau_L} = \frac{1}{1 + 4N_e\mu}$$

where

$$N_e = \frac{\tau_L}{2} = \frac{L^2 \log L}{4\pi\sigma^2\nu}$$

For $N = 20$, $L = 50$, $\nu = 0.1$ ($N\nu = 2$), $q(x, y)$ uniform on $[-2, 2]^2 \setminus (0, 0)$, $\sigma^2 = 50/24$,
$N_e = 3736 \ll NL^2 = 50,000$.

Compare with **the island model:** there are $k$ colonies,

$$p(x, y) = \begin{cases} 1 - \nu & x = y \\ \nu/(k-1) & x \neq y \end{cases}$$

$N_e = Nk \left(1 + \dfrac{(k-1)^2}{4N\nu k^2}\right) >$ actual population size

2. Correlation with distance $N\nu \ll \log L$ ($\alpha = 0$) and $|x| \approx L^\beta$ or $\beta \approx \log|x|/\log L$,

$$\begin{aligned}\phi(x) &= E(1-\mu)^{2t_{0,x}^L} \\ &\approx (1-\beta) + \beta\frac{1}{2\mu\tau_L}\end{aligned}$$

$$\begin{aligned}\phi(0) - \phi(x) &\approx 1 - \left((1-\beta) + \frac{\beta}{2\mu\tau_L}\right) \\ &= \beta(1 - \frac{\beta}{2\mu\tau_L}) \\ &\approx \beta 2\mu\tau_L \\ &\approx \log|x|\frac{L^2}{\pi\sigma^2\nu}\end{aligned}$$

3. Wright's statistic $F_{ST}$

$$\frac{4\pi\sigma^2 N\nu}{\log L} \to \alpha > 0$$

$$F_{ST} = \frac{\phi(0) - h}{1 - h}$$

Approximate $h$ ($\beta = 1$) and $\phi(0)$ ($\beta = 0$) as before, get

$$F_{ST} \approx \frac{1}{1+\alpha} \approx \frac{1}{8\pi\sigma^2 N\mu} \log L^2$$

Crow and Aoki (1984), numerical studies, $F_{ST}$ proportional to the log of the number of colonies.

## 5 Times $t \ll \tau_L = \dfrac{L^2 \log L}{2\pi\sigma^2\nu}$

If $\dfrac{4\pi\sigma^2 N\nu}{\log L} \approx \alpha$ and $|x| \approx L^\beta$ then

$$\frac{t^L_{0,x}}{(1+\alpha)\tau_L} \overset{\mathrm{d}}{\approx} \begin{cases} 0 & \text{prob. } \dfrac{1-\beta}{1+\alpha} \\ \\ \mathcal{E}(1) & \text{prob. } \dfrac{\beta+\alpha}{1+\alpha} \end{cases}$$

Change to continuous time version of model (Moran type).

For $0 < \delta \leq 1$ and $c > 0$ let

$$\Gamma(L, c, \delta) = (L^\delta / \log L, c\delta L^\delta \log L)$$

**Theorem (C. Durrett, Zähle)** *Assume* $2N\nu\pi\sigma^2 / \log L \to \alpha \in [0, \infty)$ *as* $L \to \infty$*. For any fixed* $\beta_0 > 0$*, uniformly in* $\beta_0 \leq \beta \leq \gamma \leq 1$ *and* $|x| \in \Gamma(L, c, \beta)$*,*

$$P\left(t_{0x}^L > \frac{L^{2\gamma}}{2\nu}\right) \to \frac{\beta + \alpha}{\gamma + \alpha}$$

*as* $L \to \infty$*.*

# 6 $s > 2$ lineages

$h_L = (1+\alpha)\tau_L = (1+\alpha)L^2 \log L/(2\pi\sigma^2\nu)$

$$\mathcal{G}(L,n,1) = \Big\{ A = \{x_1, \dots x_s\} : \\ \forall i, x_i \in \Lambda(L), \, \forall i \neq j, |x_i - x_j| \geq \frac{L}{\log L} \Big\}.$$

**Theorem (C. Durrett, Zähle)** *Assume* $2N\nu\pi\sigma^2/\log L \to \alpha \in [0,\infty)$ *as* $L \to \infty$*. Uniformly in* $t \geq 0$*,* $A \in \mathcal{G}(L,s,1)$*,*

$$\Big| P(|\zeta^A_{h_L t}| = k) - P(D^s_t = k) \Big| \to 0.$$

$$\mathcal{G}(L,n,c,\delta) = \Big\{ A = \{x_1, \dots x_n\} : \\ \forall i, x_i \in \Lambda(L),\, \forall i \neq j, |x_i - x_j| \in \Gamma(L,c,\delta) \Big\}.$$

**Theorem (C. Durrett, Zähle)** *Assume* $2N\nu\pi\sigma^2/\log L \to \alpha \in [0,\infty)$ *as* $L \to \infty$. *Uniformly in* $\beta_0 \le \beta \le \gamma \le 1$ *and* $A \in \mathcal{G}(L,s,c,\beta)$,

$$P\left(|\zeta^A_{\frac{L^{2\gamma}}{2\nu}}| = k\right) - P\left(D^s_{\log\frac{\gamma+\alpha}{\beta+\alpha}} = k\right) \to 0$$

*Uniformly for* $t \ge 0$ *and* $A \in \mathcal{G}(L,s,c,\beta)$,

$$\Pr\left(|\zeta^A_{\frac{L^2}{2\nu}+h_L t}| = k\right) - P\left(D^s_{\log\frac{1+\alpha}{\beta+\alpha}+t} = k\right) \to 0.$$