Diffusion processes and coalescent trees

Bob Griffiths

University of Oxford

The frequency of a mutation

The relative frequency $\{X(t), t \geq 0\}$ of genes of type a in a population of two types a and A is modelled by a diffusion process with generator

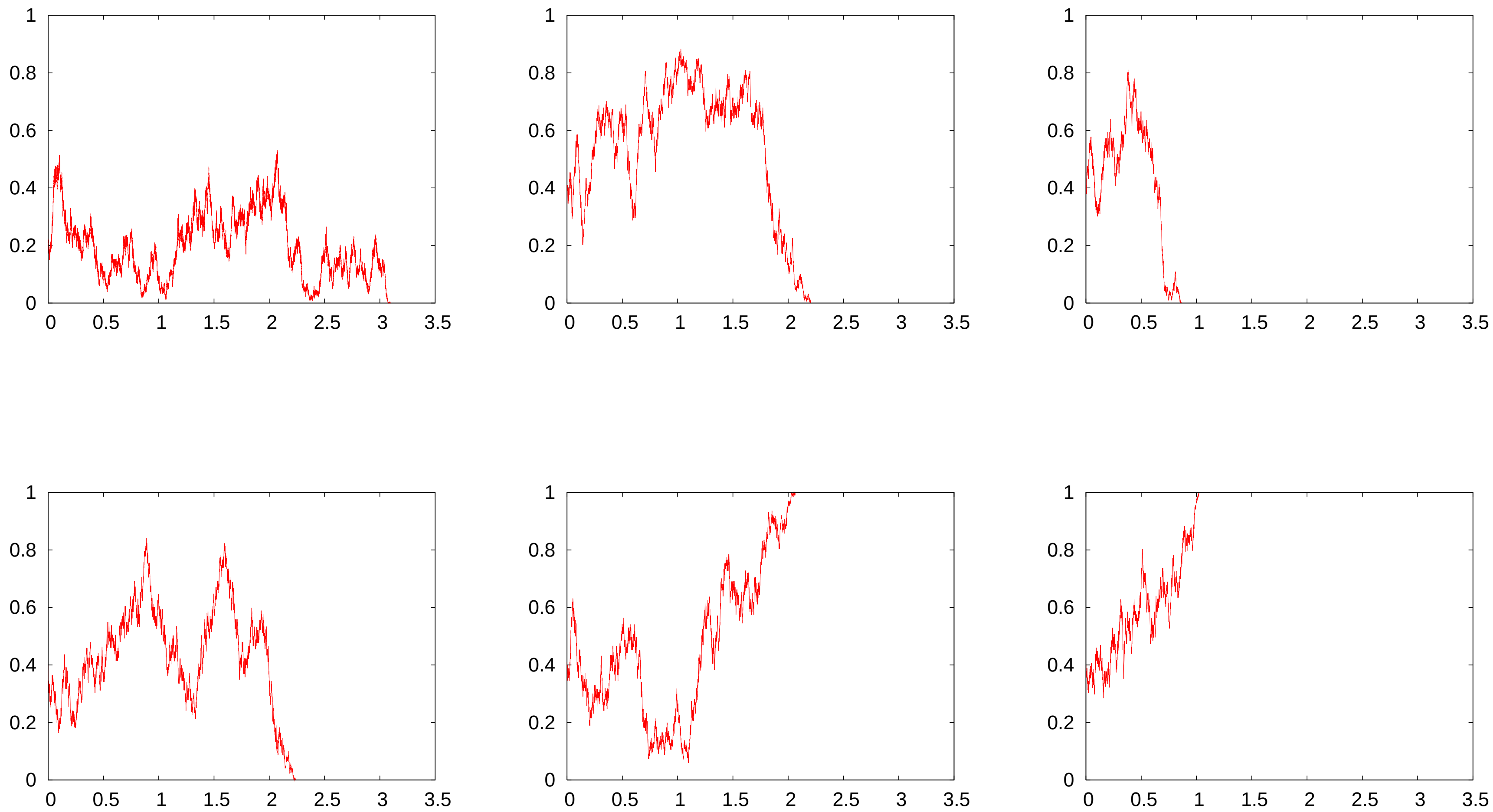
$$L = \frac{1}{2}x(1-x)\frac{\partial^2}{\partial x^2}$$

Simplest model with no mutation or selection.

Denote $\Delta X(t) = X(t + \Delta t) - X(t)$

$$\begin{aligned} E(\Delta X(t) \mid X(t) = x) &= 0\Delta t + o(\Delta t) \\ \mathrm{Var}(\Delta X(t) \mid X(t) = x) &= x(1-x)\Delta t + o(\Delta t) \end{aligned}$$

Kimura (1955), Bochner (1954) in a different context.



$X(t)$, Neutral model, no mutation

Wright-Fisher model N genes, Types a, A. Fixed population size of N genes. Offspring in a generation choose their parents at random with replacement from the prior generation and inherit their type.

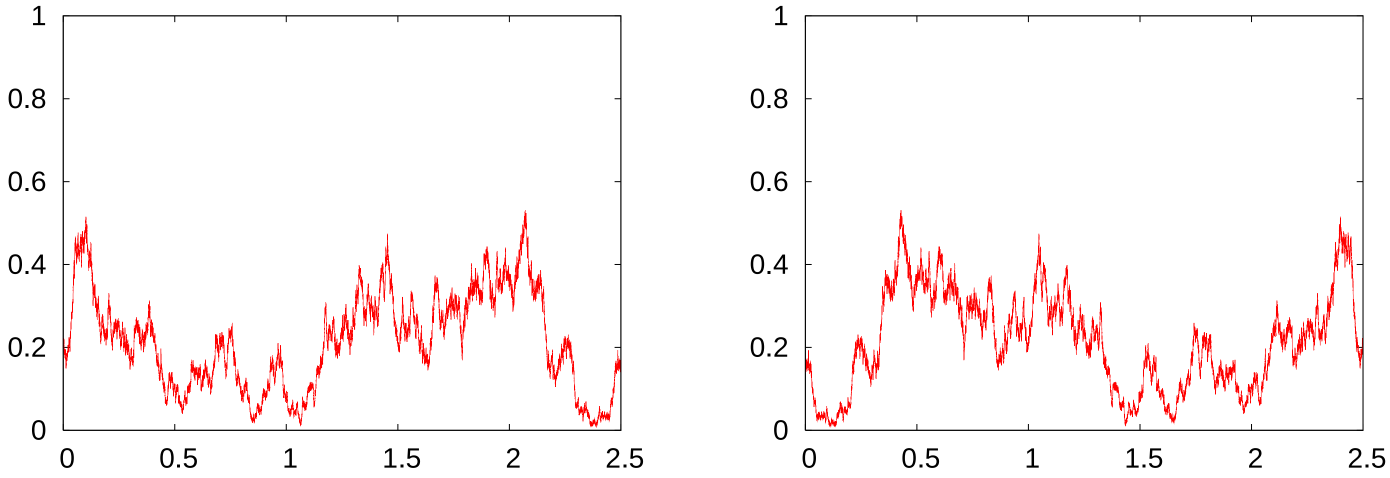
$X^N(\tau), \tau = 0, 1, \ldots$ is the relative frequency of a genes in generation τ.

$$\begin{aligned} E(\Delta X^N(\tau) \mid X^N(\tau) = x) &= 0 \\ \text{Var}(\Delta X^N(\tau) \mid X^N(\tau) = x) &= x(1-x) \cdot \frac{1}{N} \end{aligned}$$

Higher moments of $\Delta X^N(\tau)$ are of order smaller than $\frac{1}{N}$.

Measure time in units of N generations, so $\delta t = \frac{1}{N}$, then

$$\{X^N([Nt]), t \geq 0\} \Longrightarrow \{X(t), t \geq 0\}$$



Reversibility of $X(t)$

Transition density in the neutral model

$$f(p,x;t) = x^{-1}(1-x)^{-1}\sum_{i=2}^{\infty} e^{-\frac{1}{2}i(i-1)t}P_i(p)P_i(x)$$

where $\{P_i(\cdot)\}$ are scaled Gegenbauer polynomials, orthogonal on

$$m(x) = x^{-1}(1-x)^{-1}$$

Reversibility before hitting 0 or 1

$$m(x)f(x,p;t) = m(p)f(p,x;t)$$

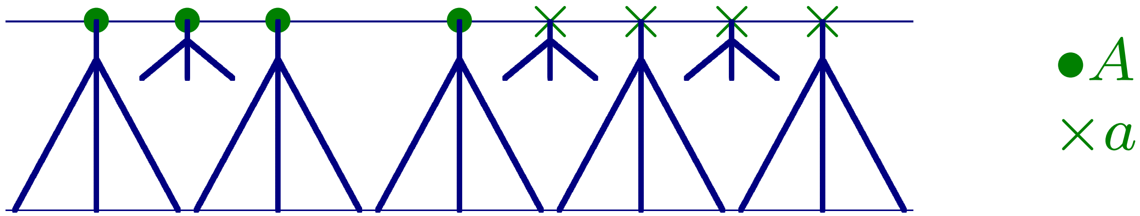
Genealogical form of the transition density

$$f(p,x;t) = \sum_{\ell=1}^{\infty} q_\ell(t) \sum_{k=1}^{\ell} \binom{\ell}{k} p^k (1-p)^{\ell-k} \times \frac{\Gamma(\ell)}{\Gamma(k)\Gamma(\ell-k)} x^{k-1}(1-x)^{\ell-k-1}$$

where $\{q_\ell(t)\}$ are transition probabilities of a death process (coalescent process) $\{A(t),\ t \geq 0\}$ with death rates $\mu_k = \binom{k}{2}$ and $A(0) = \infty$.

G (1980), Ethier and G (1993).

Time t back, $A(t) = \ell$ ancestors



Current population

The relative frequency of family sizes in the ℓ lines is Dirichlet(1,1,. . .,1).

Line types are chosen from a, A independently with probability $p, 1-p$.
If k of the l lines are of type a, then the density of a is

$$\frac{\Gamma(\ell)}{\Gamma(k)\Gamma(\ell-k)} x^{k-1}(1-x)^{\ell-k-1}$$

Two-allele models with mutation

Wright-Fisher model with genes of type a, A
Mutation $a \longrightarrow A$ with probability u per gene per generation
Mutation $A \longrightarrow a$ with probability v per gene per generation

Diffusion process $\{X(t), t \geq 0\}$ for the relative frequency of a genes at time t.

Generator

$$L = \frac{1}{2}x(1-x)\frac{\partial^2}{\partial x^2} + \frac{1}{2}(-\alpha x + \beta(1-x))\frac{\partial}{\partial x}$$

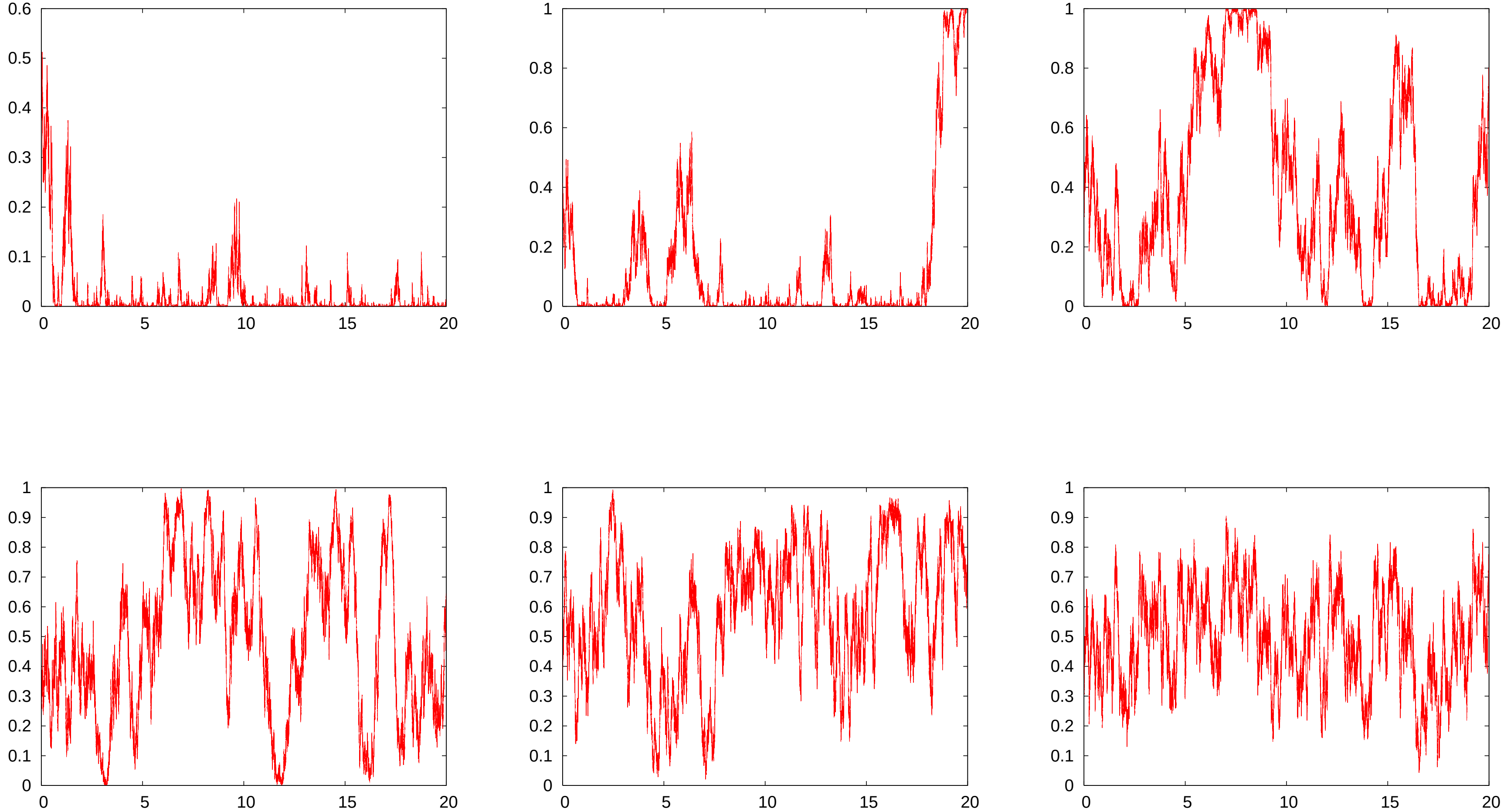
where $\alpha = 2Nu$, $\beta = 2Nv$.

The stationary distribution of the diffusion process is Beta

$$\phi(x;\alpha,\beta) = B(\alpha,\beta)^{-1}x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1$$

satisfying the forward equation

$$\frac{1}{2}\frac{\partial^2}{\partial x^2}\Big\{x(1-x)\phi(x;\alpha,\beta)\Big\} - \frac{\partial}{\partial x}\Big\{\mu(x)\phi(x;\alpha,\beta)\Big\} = 0$$



$X(t)$, Neutral model, with mutation, $\alpha = \beta = 0.01, 0.1, 0.5, 1.0, 2.0, 5.0$

Wright-Fisher diffusion process with d types

$X(t) = (X_1(t)\ldots, X_d(t))$ are gene frequencies of d types, labelled $1, 2, \ldots, d$ at time $t \geq 0$.
$|X(t)| = X_1(t) + \cdots + X_d(t) = 1$.
Mutations $i \to j$ occur at rate $\frac{1}{2}\epsilon_j$, $i, j = 1, \ldots d$.

Backwards generator

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} x_i(\delta_{ij} - x_j)\frac{\partial^2}{\partial x_i \partial x_j} + \frac{1}{2}\sum_{i=1}^{d}(\epsilon_i - |\epsilon| x_i)\frac{\partial}{\partial x_i}$$

Stationary distribution is Dirichlet

$$\mathcal{D}(x, \epsilon) = \frac{\Gamma(|\epsilon|)}{\Gamma(\epsilon_1)\cdots\Gamma(\epsilon_d)} x_1^{\epsilon_1 - 1} \cdots x_d^{\epsilon_d - 1}$$

for $x_1, \ldots x_d > 0$ and $|x| = \sum_1^d x_i = 1$

Transition density ($x \rightarrow y$ in time t) is

$$f(x,y;t) = \mathcal{D}(y,\epsilon)\Big\{1 + \sum_{\{n;n\in\mathbb{Z}_+^{d-1},|n|>0\}} \rho_{|n|}(t)Q_n^\circ(x)Q_n^\circ(y)\Big\}$$

Eigenfunctions $Q_n^\circ(x)$ are orthonormal polynomials on the Dirichelt distribution.

Eigenvalues are $\rho_{|n|}(t) = \exp\{-\frac{1}{2}|n|(|n| + |\epsilon| - 1)t\}$ repeated $\binom{|n|+d-2}{|n|}$ times corresponding to eigenvectors $Q_n^\circ(x)$, G (1979).

Karlin and McGregor (1975) - Moran model version whose stationary distribution is the Dirichlet-Multinomial. Polynomials are Hahn multivariable analogues explicitly constructed.

Backward generator

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} x_i(\delta_{ij} - x_j)\frac{\partial^2}{\partial x_i \partial x_j} + \frac{1}{2}\sum_{i=1}^{d}(\epsilon_i - |\epsilon| x_i)\frac{\partial}{\partial x_i}$$

Forward generator

$$\widetilde{\mathcal{L}} = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} x_i(\delta_{ij} - x_j) - \frac{1}{2}\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(\epsilon_i - |\epsilon| x_i)$$

Eigenfunction equations

$$\begin{aligned} \mathcal{L} Q_n^\circ(x) &= -\frac{1}{2}|n|(|n| + |\epsilon| - 1) Q_n^\circ(x) \\ \widetilde{\mathcal{L}} \mathcal{D}(x,\epsilon) Q_n^\circ(x) &= -\frac{1}{2}|n|(|n| + |\epsilon| - 1)\mathcal{D}(x,\epsilon) Q_n^\circ(x) \end{aligned}$$

Mutant families in an infinite-leaf coalescent tree

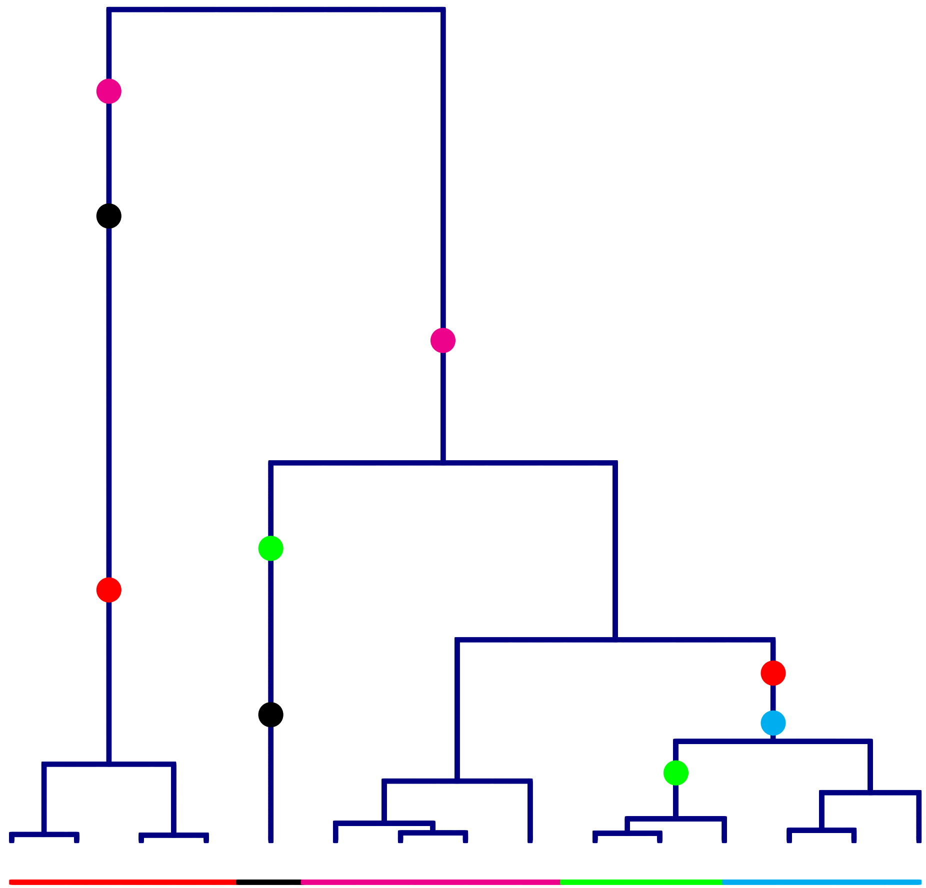
Mixture distribution arising from the coalescent

$$f(x,y;t) = \sum_{|l|=0}^{\infty} q_{|l|}^{|\epsilon|}(t) \sum_{\{l:|l| \text{ fixed}\}} \binom{|l|}{l} \prod_{1}^{d} x_i^{l_i} \mathcal{D}(y, \epsilon + l)$$

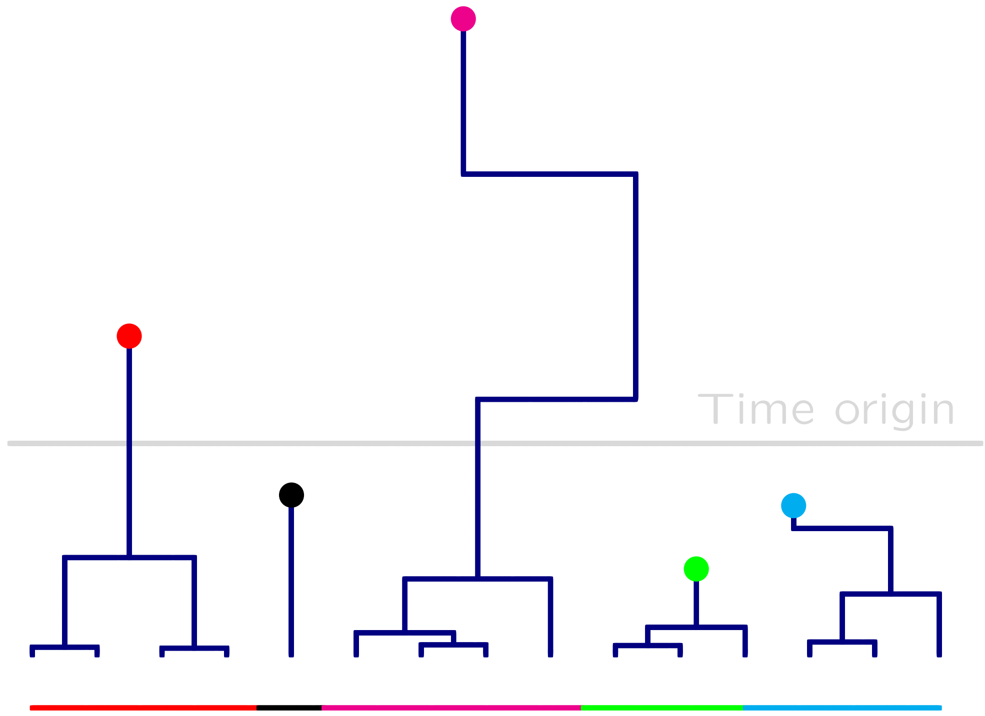
$q_{|l|}^{|\epsilon|}(t)$ is the distribution of $L^{|\epsilon|}(t)$, the number of non-mutant founder lineages at time t back. $L^{|\epsilon|}(t)$ is a death process back in time, starting from infinity, where lineages are reduced by coalescence or mutation from $k \to k-1$ at rate $k(k-1)/2 + k|\epsilon|/2$. Families are either from founder lineages or new mutations, giving the Dirichlet mixture.

$f(x,y;t)$ and the eigenfunction form are identical.

Infinite-leaf coalescent tree with mutations, type j with probability $\epsilon_j/|\epsilon|$



Forest of non-mutant ancestral lineages, to defining mutations



Ancestral lineages are lost back in time by coalescence
or mutation at rates $\binom{i}{2}$ and $\frac{i\theta}{2}$ while i non-mutant lineages.

Dirichlet family sizes

$$\sum_{\{l:|l| \text{ fixed}\}} \binom{|l|}{l} \prod_{1}^{d} x_i^{l_i} \mathcal{D}(y, \epsilon + l)$$

$|l|$ non-mutant founder lineages are divided into $l = (l_1, \dots, l_d)$ numbers of types $1, \dots, d$ with probability $\binom{|l|}{l} \prod_1^d x_i^{l_i}$.
Let $U = (U_1, \dots, U_l)$ be their relative family sizes in the leaves of the tree, and $V = (V_1, \dots, V_d)$ be the frequencies of families derived from new mutations on the tree edges in $(0, t)$.
$U \oplus V = (U_1, \dots, U_l, V_1, \dots, V_d)$ is $\mathcal{D}(u \oplus v, (1, \dots, 1) \oplus \epsilon)$.
$\mathcal{D}(y, \epsilon + l)$ is obtained by adding Dirichlet parameters corresponding to types $1, \dots, d$.

Eigenfunction expansion

$$f(x, y; t) = \mathcal{D}(y, \epsilon)\Big\{1 + \sum_{1}^{\infty} \rho_{|n|}(t) Q_{|n|}(x, y)\Big\}$$

n-Kernel polynomials

$$Q_{|n|}(x, y) = \sum_{\{n; |n| \text{ fixed }\}} Q_n^\circ(x) Q_n^\circ(y)$$

invariant under which orthogonal polynomial set is used.

$$Q_{|n|}(x, y) = (|\epsilon| + 2|n| - 1) \sum_{m=0}^{n} (-1)^{|n|-m} \frac{(|\epsilon| + m)_{(|n|-1)}}{m!(|n| - m)!} \xi_m$$

where $\xi_m = \sum_{|l|=m} \binom{m}{l} \frac{|\epsilon|_{(m)}}{\prod_1^d \epsilon_{i(l_i)}} \prod_{1}^{d} (x_i y_i)^{l_i}$

Two forms of the density

$$f(x, y; t) = \mathcal{D}(y, \epsilon)\Big\{1 + \sum_{1}^{\infty} \rho_{|n|}(t) Q_{|n|}(x, y)\Big\}$$

$\rho_{|n|}(t)$ are eigenvalues.

$$f(x, y; t) = \sum_{|l|=0}^{\infty} q_{|l|}^{|\epsilon|}(t) \sum_{\{l:|l| \text{ fixed}\}} \binom{|l|}{l} \prod_{1}^{d} x_i^{l_i} \mathcal{D}(y, \epsilon + l)$$

$q_{|l|}^{|\epsilon|}(t)$ are death process transition probabilities.
Fleming-Viot Dirichlet process version in Ethier & G (1993).

Poisson Dirichlet random measure

$$\mu = \sum_{i=1}^{\infty} x_i \delta_{\xi_i}$$

where $\{x_i\}$ is PD(θ) and independent of $\{\xi_j\}$ which are *i.i.d.* $\nu_0 \in \mathcal{P}(S)$, with S a compact metric space.

Stationary distribution of the random measure

$$\Pi_{\theta,\nu_0}(\cdot) = P(\mu \in \cdot)$$

Fleming-Viot process with type space S, and mutation operator

$$(Af)(x) = \frac{\theta}{2} \int_S (f(\xi) - f(x))\nu_0(d\xi)$$

The Fleming-Viot process with type space S and mutation operator A has transition function $P(t,\mu,d\nu)$ for given $\mu \in \mathcal{P}(S)$

$$\begin{aligned} P(t,\mu,.) &= q_0^\theta(t)\Pi_{\theta,\nu_0}(\cdot) \\ &\quad + \sum_{n=1}^{\infty} q_n^\theta(t) \int_{S^n} \mu^n(dy_1 \times \cdots \times dy_n) \\ &\qquad\qquad \Pi_{n+\theta,(n+\theta)^{-1}\{n\eta_n(y_1,\ldots,y_n)+\theta\nu_0\}}(\cdot) \end{aligned}$$

where $\eta_n(y_1,\ldots,y_n)$ as the empirical measure of points $y_1,\ldots,y_n \in S$,

$$\eta_n(y_1,\ldots,y_n) = n^{-1}(\delta_{y_1} + \cdots + \delta_{y_n})$$

Characterization of Markov processes with Beta stationary distributions and polynomial eigenfunctions

Characterization of distributions with polynomial eigenfunctions

$$B(\alpha,\beta)^{-1}y^{\alpha-1}(1-y)^{\beta-1} \times \Big\{1+\sum_{n=1}^{\infty}\omega_n h_n R_n^{(\alpha,\beta)}(x)R_n^{(\alpha,\beta)}(y)\Big\},\ 0<x,y<1$$

is a probability distribution for a sequence of constants $\{\omega_n\}$ if and only if $\{\omega_n\}$ is a positive definite sequence.

$\{R_n^{(\alpha,\beta)}(x)\}$ are Jacobi polynomials on the Beta distribution with $R_n^{(\alpha,\beta)}(1)=1$. Scaling by $\sqrt{h_n}$ makes them orthonormal.

Bochner (1954) A bounded sequence $\{c_n\}$ is positive definite if

$$\sum a_n h_n R_n^{(\alpha,\beta)}(x) \geq 0, \ \sum |a_n| h_n < \infty$$

implies that

$$\sum a_n c_n h_n R_n^{(\alpha,\beta)}(x) \geq 0$$

Gasper (1972) Let $\alpha < \beta$. If either $1/2 \leq \alpha$ or $\alpha + \beta \geq 2$, then a sequence ρ_n is positive definite if and only if

$$\rho_n = \mathbb{E}[R_n^{(\alpha,\beta)}(Z)]$$

for some random variable Z in $[0, 1]$.

Bochner (1954): A family of bounded sequences $\{c_n(t)\}$, $0 \le t < \infty$ is called a homogeneous stochastic process if

(i) $\{c_n(t)\}$ is positive definite for each t
(ii) $c_n(t)$ is continuous in t
(iii) $c_n(0) = c_0(t) = 1$
(iv) $c_n(t+s) = c_n(t)c_n(s)$

Gasper (1972): Under the conditions on α and β, it is N & S that $c_n(t) = e^{-d_n t}$ with

$$d_n = \sigma n(n+\alpha+\beta-1) + \int_0^{1-} \frac{1 - R_n^{(\alpha,\beta)}(z)}{1-z}\, d\nu(z)$$

where $\sigma \ge 0$ and ν is a finite measure on $[0,1)$.
Sequences $\{c_n(t)\}$ characterize eigenvalues of transition probability functions with polynomials as eigenfunctions.

Subordinated Diffusion Process Let $\{X(t)\}$ be the two allele Wright-Fisher diffusion process with eigenvalues $e^{-\frac{1}{2}n(n+\theta-1)t}$, where $\theta = \alpha + \beta$

Let $\{Z(t)\}$ be a Lévy process with Laplace transform

$$\mathbb{E}[\exp(-\phi Z(t))] = \exp\Big\{ - t\int_0^\infty \frac{1-e^{-y\phi}}{y} G(dy)\Big\}$$

where $\int_1^\infty \frac{1}{y}G(dy) < \infty$.

Then $\{X(Z(t))\}$ has polynomial eigenvectors and eigenvalues

$$c_n(t) = \exp\Big\{ - t\int_0^\infty \frac{1-e^{-yn(n+\theta-1)/2}}{y} G(dy)\Big\}$$

Markov Process $\{\widetilde{X}(t)\}$ with transition density

$$B(\alpha,\beta)^{-1}y^{\alpha-1}(1-y)^{\beta-1}$$
$$\times\left\{1+\sum_{n=1}^{\infty}e^{-nt}h_nR_n^{(\alpha,\beta)}(x)R_n^{(\alpha,\beta)}(y)\right\},\ 0<x,y<1$$

Poisson kernel in orthogonal polynomial theory

$$1+\sum_{n=1}^{\infty}r^nh_nR_n^{(\alpha,\beta)}(x)R_n^{(\alpha,\beta)}(y),\ 0<x,y<1,|r|<1$$

$\widetilde{X}(t)=X(Z(t))$, where $\{Z(t)\}$ is a Lévy process with Laplace transform

$$\exp\left\{-t\left[\sqrt{2\lambda+(\theta-1)^2/4}-\sqrt{(\theta-1)^2/4}\right]\right\},\ \ \theta\geq 1$$

$Z(t)$ is a tilted positive stable process with index $\frac{1}{2}$ and density

$$\frac{t}{2\pi z^3} e^{-\frac{1}{2}\frac{t}{\sqrt{z}}} \cdot e^{-\frac{1}{8}(\theta-1)^2 z + \frac{1}{2}|(\theta-1)|t}, \; z > 0$$

$Z(t)$ is a Lévy process with Laplace transform

$$\exp\Big\{ - \frac{t}{\sqrt{2\pi}} \int_0^\infty \frac{e^{-x(\theta-1)^2/8}}{x^{3/2}} \big(1 - e^{-x\lambda}\big)\, dx \Big\}$$

The eigenvalues of $\widetilde{X}(t)$ are, for $\theta \geq 1$,

$$\mathbb{E}\Big[\exp\Big\{ -\frac{1}{2} n(n+\theta-1) Z(t) \Big\} \Big] = \exp\{-nt\}$$

$\widetilde{X}(t)$ has a generator $\widetilde{L}$ satisfying

$$-\widetilde{L} = \sqrt{2(-L) + (\theta - 1)^2/4} - \sqrt{(\theta - 1)^2/4}$$

A series expansion with positive coefficients of L is

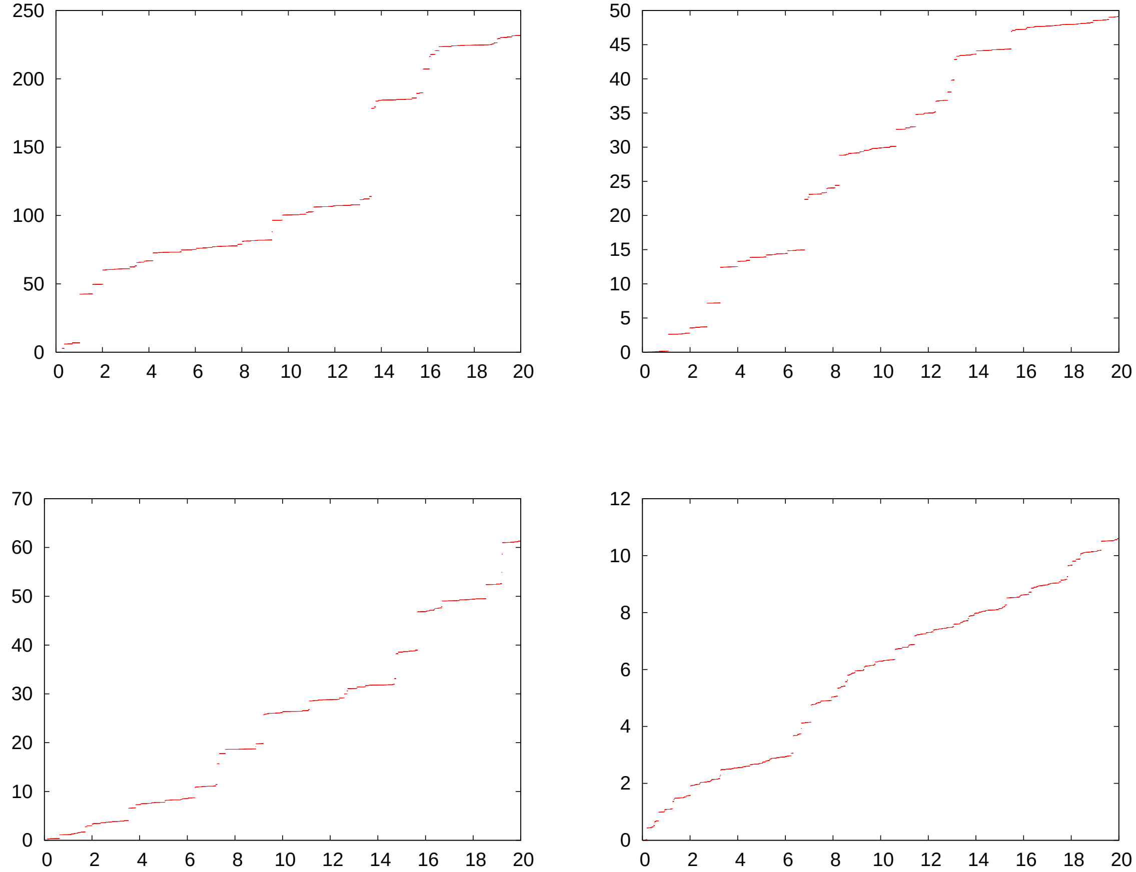
$$\widetilde{L} = \sqrt{c} \sum_{k=1}^{\infty} \frac{2^k}{k!} \left(\tfrac{1}{2}\right)_{(k-1)} \left(\frac{L}{c}\right)^k$$

where $c = (\theta - 1)^2/4$.

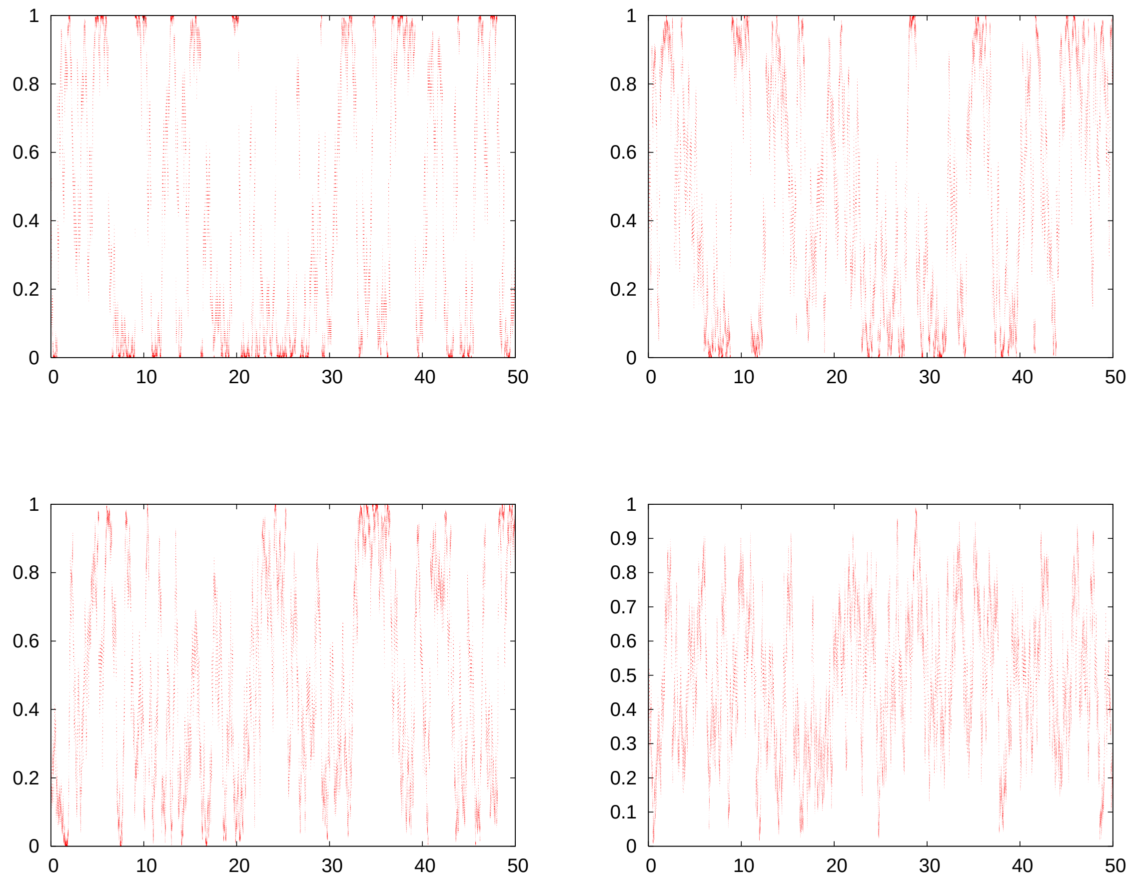
$$L = \frac{1}{2}\Big(- \widetilde{L}^2 + (\theta - 1)\widetilde{L}\Big)$$

Eigenfunctions

$$\widetilde{L} R_n^{(\alpha,\beta)}(x) = -n R_n^{(\alpha,\beta)}(x)$$



Tilted Stable subordinator, index half ($\theta = 1.0, 1.5, 2.0, 5.0$)



Subordinated Wright-Fisher diffusion ($\theta = 1.0, 1.5, 2.0, 5.0$)

If $\theta < 1$ then for $n \geq 1$,

$$\mathbb{E}\Big[\exp\Big\{-\frac{1}{2}n(n+\theta-1)Z(t)\Big\}\Big] = \exp\{-nt\} \times \exp\{t(1-\theta)\}$$

Let $\widetilde{f}(x,y;t)$ be the transition density of $\widetilde{X}(t)$, then the transition density with eigenvalues $\exp\{-nt\}$, $n \geq 0$ is

$$e^{-t(1-\theta)}\widetilde{f}(x,y;t) + (1-e^{-t(1-\theta)})B(\alpha,\beta)^{-1}y^{\alpha-1}(1-y)^{\beta-1}$$

Subordinated process $X(\widehat{Z}(t))$ where $\widehat{Z}(t)$ is a similar process to $Z(t)$ but has an extra state ∞. $Z(t)$ is killed by a jump to ∞ at a rate $(1-\theta)$.

Subordinated genealogical form of the transition density

$$\sum_{\ell=1}^{\infty} q_\ell^\theta(t) \sum_{k=1}^{\ell} \binom{\ell}{k} p^k (1-p)^{\ell-k} B(\alpha+k, \beta+\ell-k)^{-1} x^{\alpha+k-1}(1-x)^{\beta+\ell-k-1}$$

where $\{q_\ell^\theta(t)\}$ are transition probabilities of a death process $\{A^\theta(t),\, t \geq 0\}$ with death rates $\mu_k = \binom{k}{2} + \frac{k\theta}{2}$ and $A^\theta(0) = \infty$.

Subordinated process $\widetilde{X}(t) = X(Z(t))$ has a similar form for the transition density, with $q_\ell^\theta(t)$ replaced by $\mathbb{E}(q_l^\theta(Z(t))$, transition functions of the subordinated death process $A^\theta(Z(t))$.

Subordinated death process $A^{\theta}(Z(t))$ example

$Z(t)$ is a tilted positive stable process with index $\frac{1}{2}$.
$\widetilde{A}^{\theta}(t) = A^{\theta}(Z(t))$ is a Markov process with $\widetilde{A}^{\theta}(0) = \infty$ and a *pgf* of

$$G_{\widetilde{A}^{\theta}(t)}(s) = \Big(\frac{1-4pqs}{1-4pq}\Big)^{-\frac{1}{2}} \times \Big(\frac{1-\sqrt{1-4pqs}}{2ps}\Big)^{\theta-1},\ \theta \geq 0,$$

where $p = e^{-t}/(1+e^{-t})$ and $q = 1/(1+e^{-t})$.

Random walk on $\mathbb{Z}$ with transitions $j \longrightarrow j+1$ with probability p and $j \longrightarrow j-1$ with probability $q = 1-p$ and $q \geq p$.

Let the number of steps to hit $-\theta$, starting from 0 be $2\xi + \theta$.

Then ξ has a *pgf* of

$$H(s) = \Big(\frac{1 - \sqrt{1-4pqs}}{2ps} \Big)^{\theta}$$

$\widetilde{A}(t)+1$ has the same distribution as the size-biassed distribution of ξ, with *pgf*

$$G_{\widetilde{A}(t)} = \frac{sH'(s)}{H'(1)}$$

The probability distribution of $\widetilde{A}^{\theta}(t)$ is

$$\binom{2j+\theta-1}{j}\Big(\frac{z}{1+z}\Big)^{j}\Big(\frac{1}{1+z}\Big)^{j+\theta}(1-z)$$

for $j = 0, 1, \ldots$, where $z = e^{-t}$.

The probability distribution of $A^{\theta}(t)$ is

$$\sum_{k=j}^{\infty} \rho_k^{\theta}(t)(-1)^{k-j}\frac{(2k+\theta-1)(j+\theta)_{(k-1)}}{j!(k-j)!}$$

for $j = 0, 1, \ldots$, where $\rho_k^{\theta}(t) = e^{-k(k+\theta-1)t/2}$.

General characterization of eigenvalues

$$R_n^{(\alpha,\beta)}(x) = {}_2F_1(-n, n+\theta-1; \alpha; 1-y)$$

Eigenvalues

$$\int_0^{1-} \frac{1 - R_n^{(\alpha,\beta)}(y)}{1-y}\, dG(y)$$
$$= -c \sum_{k=1}^{n} \frac{(-n)_{(k)}(n+\theta-1)_{(k)}}{\alpha_{(k)}} \frac{\mu_{k-1}}{k!}$$
$$= -c \sum_{k=1}^{n} \frac{\prod_{j=0}^{k-1} \big(-n(n+\theta-1) + j(j+\theta-1)\big)}{\alpha_k} \frac{\mu_{k-1}}{k!}$$

where $\int_0^{1-}(1-y)^k\, dG(y) = c\mu_k$.

The generator corresponding to a process with these eigenvalues is

$$\widehat{L} = c\sum_{k=1}^{\infty} \frac{\prod_{j=0}^{k-1}\big(2L + j(j+\theta-1)\big)}{\alpha_{(k)}} \frac{\mu_{k-1}}{k!}$$

This is a positive series in L.

Laguerre diffusion $\{Y(t), t \geq 0\}$

Branching process with immigration

Generator

$$L = x\frac{\partial^2}{\partial x^2} + (-bx + c)\frac{\partial}{\partial x}, \;\; b = 1, c > 0$$

Transition functions, for $x, y > 0$

$$\begin{aligned} f(x, y; t) &= \frac{y^{c-1}}{\Gamma(c)}e^{-y}\sum_{n=0}^{\infty} e^{-nt}L_n^{(c-1)}(x)L_n^{(c-1)}(y)\frac{n!}{c_{(n)}} \\ &= \sum_{k=0}^{\infty} e^{-\mu x}\frac{(\mu x)^k}{k!}\frac{(1-r)^{-(c+k)}}{\Gamma(c+k)}y^{c+k-1}e^{-y/(1-r)} \end{aligned}$$

where $\{L_n^{(c-1)}(x)\}$ are Laguerre polynomials, $r = e^{-t}$ and $\mu = r/(1-r)$.

Transition functions of a Markov process

$$\widetilde{f}(x,y;t) = \frac{y^{c-1}}{\Gamma(c)}e^{-y}\sum_{n=0}^{\infty} e^{-td_n}L_n^{(c-1)}(x)L_n^{(c-1)}(y)\frac{n!}{c_{(n)}}$$

are characterized by their eigenvalues having the representation

$$d_n = \int_0^1 \frac{1-y^n}{1-y}\,dG(y) = an + \int_0^{1-} \frac{1-y^n}{1-y}\,dG(y),\ a \geq 0$$

Sarmanov (1968), G (1969).

A Markov process with transition functions $\widetilde{f}(x, y; t)$ has a characterization as a subordinated process

$$\widetilde{Y}(t) = Y\big(at + Z(t)\big)$$

where $\{Z(t), t \geq 0\}$ is a Lévy process with Laplace transform

$$\exp\Big\{ - \int_0^{1-} \frac{1 - e^{-\lambda(-\log y)}}{1 - y}\, dG(y) \Big\}$$

$$\int_0^{1-} 1 \wedge (-\log y) \frac{1}{1-y}\, dG(y) < \infty \;\equiv\; \int_0^{1-} dG(y) < \infty$$

References

Bochner, S. (1949). Diffusion equation and stochastic processes. *Proc. Nat. Acad. Sc.* 35, 197–233.

Bochner, S. (1954). Positive zonal functions of spheres. *Proc. Nat. Acad. Sc.* 40, 1141–1147.

Donnelly, P., Kurtz, T. G. (1996) A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Appl. Prob.* 24, 698–742.

Donnelly, P.J., Kurtz, T.G. (1999) Particle representations for measure-valued population models. *Ann. Prob.* 24, (1999), 166–205.

Durrett, R. (2002) Probability Models for DNA Sequence Evolution. Springer-Verlag, New York

Etheridge, A.M. (2000) A introduction to superprocesses. University Lecture Series, Vol. 20, American Mathematical Society. Providence, Rhode Island.

Ethier, S.N., Griffiths, R.C. (1987) The infinitely many sites model as a measure valued diffusion. *Ann. Prob.* 15, 515-545.

Ethier, S.N., Griffiths, R. C. (1993) The transition function of a Fleming–Viot process. *Ann. Prob.* 21, 1571–1590.

Ethier, S.N., Kurtz, T.G. (1993) Fleming-Viot processes in population genetics. *SIAM J. Control Optimization* 31, 345–386.

Griffiths, R.C. (1979) A transition density expansion for a multi-allele diffusion model. *Adv. Appl. Prob.* 11, 310–325.

Hoppé, F. (1984) Polya-like urns and the Ewens sampling formula. *J. Math. Biol.* 20, 91–99.

Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoret. Popul. Biol.* 3, 87–112.

Ewens, W.J. (2004) Mathematical Population Genetics. Springer-Verlag, New York.

Griffiths, R.C. (1969) The canonical correlation coefficients of bivariate gamma distributions. *Ann. Math. Statist.* 40, 1401-1408.

Griffiths, R.C. (1980) Lines of descent in the diffusion approximation of neutral Wright-Fisher models. *Theoret. Popul. Biol.* 17, 37–50.

Griffiths, R.C. (2006) Coalescent lineage distributions. *Adv. Appl. Prob.* 38, 405–429.

Griffiths, R. C., Lessard, S. (2005) Ewens' sampling formula and related formulae: Combinatorial proofs, extensions to variable population size and applications to ages of alleles. *Theoret. Popul. Biol.* 68,167–177.

Joyce, P., Tavaré, S. (1987) Cycles, permutations and the structures of the Yule process with immigration, *Stochastic Proc. Appl.* 25, 309–314.

Kimura, M. (1955) Solution of a process of random genetic drift with a continuous model. *Proc. Natl. Acad. Sci.* 41, 144–150.

Kingman, J.F.C. (1982) The coalescent. *Stochastic Proc. Appl.* 13, 235–248.

Kingman, J.F.C. (1993) Poisson Processes. Oxford University Press.

Karlin, S. and McGregor, J. (1975). Linear Growth Models with many types and multidimensional Hahn Polynomials. *Theory and Application of Special Functions.* 261–268.

Koornwinder T. H., Schwartz A. L. (1997). Product formulas and associated hypergroups for orthogonal polynomials on the simplex and on a parabolic biangle. *Constr. Approx.* 13, 537–567.

Sarmanov, I. O. (1968). A generalized symmetric gamma correlation. *Dokl. Akad. Nauk SSSR.* 179, 1276–1278; *Soviet Math.Dokl.* 9, 547–550.

Tavaré, S. (1984) Line-of-descent and genealogical processes, and their application in population genetics models. *Theoret. Popul. Biol.* 26, 119–164.

Tavaré, S. Zeitouni. (2004) Lectures on Probability Statistics Lecture Notes in Mathematics 1837 Springer-Verlag Berlin Heidelberg

Watterson, G.A. (1996) Motoo Kimura's use of diffusion theory in population genetics. *Theoret. Popul. Biol.* 49, 154–188.