Numerical study on the population genetic drift problems

Xingye Yue

Soochow University, Suzhou

Joint work with Minxin Chen, Chun Liu, David Waxman and Shixin Xu

26-Dec-2013 Taibei

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Model of genetic drift

Random genetic drift

Random genetic drift occurs when genes of a given type are transmitted to the next generation with random variation in their number.

Model assumptions

- A population always has N adults
- Each adult carries two genes, each of which may be A or B (sexual diploid)

$$X_t = \frac{\text{Number of } A \text{ genes in } t \text{ generation}}{\text{Total number of genes in } t \text{ generation}}$$

- X_t undergoes a random walk on $[\frac{0}{2N}, \frac{1}{2N}, \frac{2}{2N}, \cdots, \frac{2N}{2N}]$ with 'sticky' ends

Wright-Fisher (1930) model of genetic drift

Random genetic drift

- X_t = Proportion of A genes in t generation
- For $t+1$ generation

$$X_{t+1} = \frac{Binom(2N, X_t)}{2N}$$

- leads to

$$f_{t,n} = Prob\left(X_t = \frac{n}{2N}\right)$$

obeying

$$f_{t+1,n} = \sum_{m=0}^{2N} W_{n,m} f_{t,m},$$

– Chapman-Kolmogorov relation

Diffusion approximation

- Approximate time t as continuous
- Approximate frequency X_t as $X(t)$
- $f_{t,n} \to f(x,t)$: the probability density function of $X(t)$
- Then $f(t+1,n) = \sum_{m=0}^{2N} W_{n,m} f_{t,m} \to$ diffusion equation, in a simplest case

$$\frac{\partial}{\partial t} f - \frac{1}{4N} \frac{\partial^2}{\partial x^2} [x(1-x) f(x,t)] = 0$$

– Fokker-Planck equation

'Diffusion' or convection-diffusion

- By time rescale

$$\frac{\partial}{\partial t} f - \frac{\partial^2}{\partial x^2}[x(1-x)f(x,t)] = 0, \ \ x \in (0,1)$$

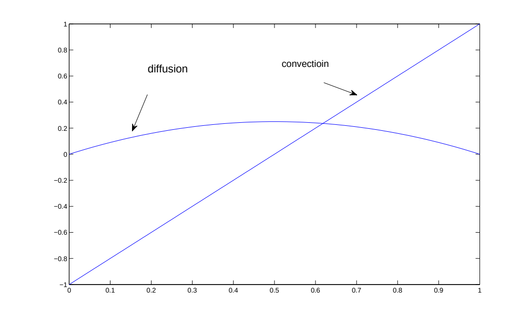
- Actually fall into the framework of convection-diffusion

$$\frac{\partial}{\partial t} f - \frac{\partial}{\partial x}[x(1-x)\frac{\partial}{\partial x} f(x,t)] + \frac{\partial}{\partial x}((2x-1)f) = 0$$

Diffusion coef. : $a = x(1-x)$
Convection coef : $b = (2x-1)$

Degenerate and convection-dominated

Degenerate: $a(0) = a(1) = 0$



Outline

A simple convection-diffusion problem

Convection-diffusion equation

When $\varepsilon \ll |b|$, convection is dominated.

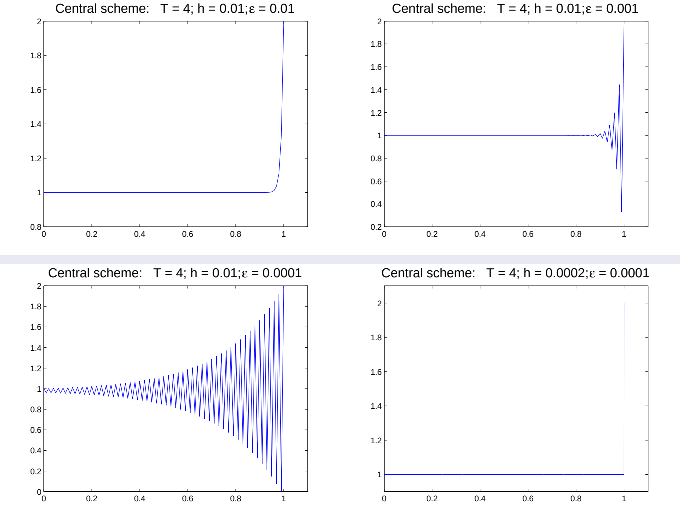
$$
\begin{aligned}
& u_t - \varepsilon u_{xx} + b u_x = 0 \qquad \text{in } (0,1) \times (0,T)\ , \\
& u(x,0) = u_0(x) \qquad x \in (0,1), \\
& u(0,t) = 1; \quad u(1,t) = 2 \qquad t \in (0,T)
\end{aligned}
$$

Central different scheme

Gvien $U_i^{n-1} = u(x_i, t_{n-1})$, $i = 1, 2, \ldots, M$, wants $U_i^n = u(x_i, t_n)$ such that

$$
\begin{aligned}
& \frac{U_i^n - U_i^{n-1}}{\tau} - \varepsilon \frac{U_{i-1}^n - 2U_i^n + U_{i+1}^n}{h^2} + b \frac{U_{i+1}^n - U_{i-1}^n}{2h} = 0, \ \ i = 1, 2, \ldots, M-1 \\
& U_0^n = 1; \quad U_M^n = 2
\end{aligned}
$$

Nonphysical oscillation



Upwind scheme

Constrain for central different scheme

To insure the stability, we must have that the *Numerical Peclet's Number*

$$Pe = \frac{h|b|}{2\varepsilon} \leq 1.$$

If we can not afford the cost, *Upwind Scheme* is the choice.

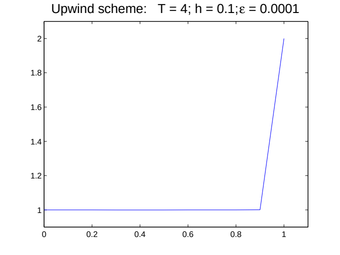
Upwind scheme for $b > 0$

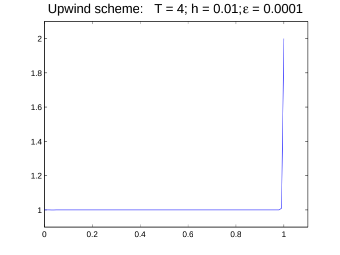
Gvien $U_i^{n-1} = u(x_i, t_{n-1})$, $i = 1, 2, \ldots, M$, wants $U_i^n = u(x_i, t_n)$ such that

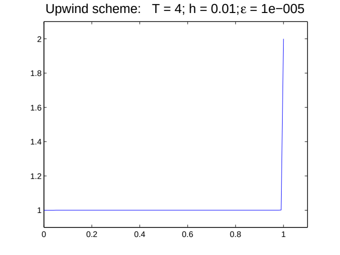
$$\frac{U_i^n - U_i^{n-1}}{\tau} - \varepsilon \frac{U_{i-1}^n - 2U_i^n + U_{i+1}^n}{h^2} + b\frac{U_i^n - U_{i-1}^n}{h} = 0, \quad i = 1, 2, \ldots, M-1$$

$$U_0^n = 1; \quad U_M^n = 2$$

Numerical results for upwind scheme







Upwind scheme for $b > 0$

Gvien $U_i^{n-1} = u(x_i, t_{n-1})$, $i = 1, 2, \dots, M$, wants $U_i^n = u(x_i, t_n)$ such that

$$\frac{U_i^n - U_i^{n-1}}{\tau} - \varepsilon \frac{U_{i-1}^n - 2U_i^n + U_{i+1}^n}{h^2} + b\frac{U_i^n - U_{i-1}^n}{h} = 0, \;\; i = 1, 2, \dots, M-1$$

$$U_0^n = 1; \quad U_M^n = 2$$

Stability of upwind different scheme

Intrinsic *numerical viscosity* introduced by upwind

$$\frac{U_i^n - U_i^{n-1}}{\tau} - (\varepsilon + \frac{|b|h}{2})\frac{U_{i-1}^n - 2U_i^n + U_{i+1}^n}{h^2} + b\frac{U_{i+1}^n - U_{i-1}^n}{2h} = 0,$$

Upwind scheme is always stable but only first order in accuracy.

Outline

Simulation for random genetic drift

Initial and boundary conditions

- Equation

$$\frac{\partial}{\partial t}f - \frac{\partial}{\partial x}[x(1-x)\frac{\partial}{\partial x}f(x,t)] + \frac{\partial}{\partial x}((2x-1)f) = 0, x \in (0,1)$$

- Initial state

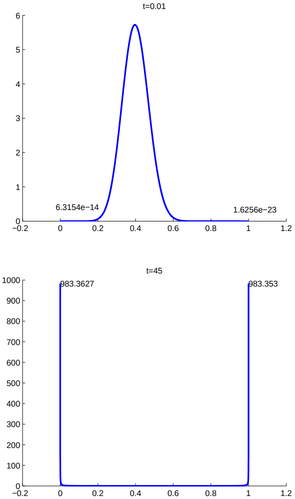
$$f(x,0) = \delta(x_0), \text{ means that the fraction of gene } A \text{ is } x_0$$

- Boundary condition – no flow condition (to keep total probability $\int_0^1 f(x,t)dx \equiv 1$)

$$j(x,t) = -\frac{\partial}{\partial x}[x(1-x)f(x,t)] = -x(1-x)\frac{\partial}{\partial x}f(x,t) + (2x-1)f = 0, \text{ at } x = 0, 1$$

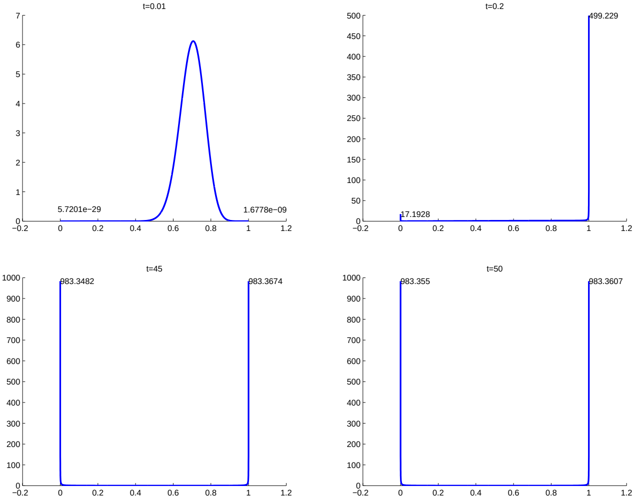
Numerical results by upwind scheme

Initial state $f(x,0) = \delta(0.4) \sim \mathcal{N}(0.4, 0.01^2)$; Step size $h = 0.0001, \quad \tau = 0.0001$



Numerical results by upwind scheme

Initial state $f(x,0) = \delta(0.7) \sim \mathcal{N}(0.7, 0.01^2)$; Step size $h = 0.0001, \quad \tau = 0.0001$



Numerical results by upwind scheme

Conclusions by upwind scheme

- The long time behavior has nothing to do with the initial state
- The steady state of the first moment of the random process $x(t)$ is always $1/2$. There is a perfect balance between genes A and B
- But . . . ?

Conservations of total probability and the first moment

- Equation with no flow boundary condition

$$\frac{\partial}{\partial t}f - \frac{\partial^2}{\partial x^2}[x(1-x)f(x,t)] = 0, x \in (0,1)$$

- Conservation of total probability: $\int_0^1 f(x,t)dx \equiv 1$
- Conservation of first moment: $\int_0^1 xf(x,t)dx \equiv \int_0^1 xf(x,0)dx = x_0$

Outline

viscosity vanishing

- The key feature of upwind scheme is its intrinsic numerical viscosity
- So we check the effect of infinitesimal artificial viscosity
- Consider the following steady state problem with no-flow boundary

$$-\frac{d^2}{dx^2}[(x(1-x)+\varepsilon)f_\varepsilon] = 0$$

- We have

$$f_\varepsilon \to \frac{1}{2}\delta(0) + \frac{1}{2}\delta(1), \ \ \varepsilon \to 0$$

- That is the reason why the upwind scheme never give a right result

Outline

Central scheme

- Denote the coef by $D(x) = x(1-x)$

$$\frac{\partial}{\partial t} f - \frac{\partial^2}{\partial x^2}[D(x)f(x,t)] = 0, x \in (0,1)$$

- For inner points $i = 1, 2, \ldots, M-1$,

$$\frac{f_i^n - f_i^{n-1}}{\tau} - \frac{D_{i+1} f_{i+1}^n - 2D_i f_i^n + D_{i-1} f_{i-1}^n}{h^2} = 0$$

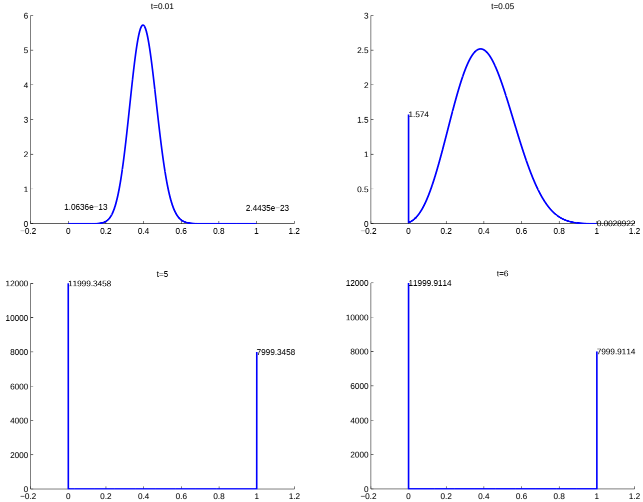
- For boundary points

$$\frac{f_0^n - f_0^{n-1}}{\tau} - \frac{2D_1 f_1^n}{h^2} = 0, \quad \frac{f_M^n - f_M^{n-1}}{\tau} - \frac{2D_{M-1} f_{M-1}^n}{h^2} = 0$$

- The inner system is decoupled with the boundary conditions! The solution is naturally decomposed into regular part (inner) and singular part (boundary)

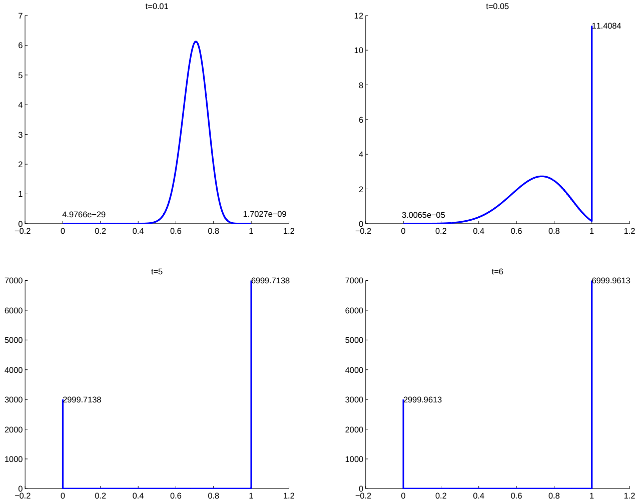
Numerical results by central scheme

Initial state $f(x,0) = \delta(0.4) \sim \mathcal{N}(0.4, 0.01^2)$; Step size $h = 0.0001$, $\tau = 0.0001$



Numerical results by central scheme

Initial state $f(x,0) = \delta(0.7) \sim \mathcal{N}(0.7, 0.01^2)$; Step size $h = 0.0001, \quad \tau = 0.0001$



Numerical results by central scheme

Conclusions by central scheme

- The long time behavior determined by the initial state
- Two spikes develop at the boundary points – that is the fixation behavior.
- For the second simulation with initial state $f(x,0) = \delta(0.7)$, with a probability of 70 percent, Gene A will survive and while Gene B will totally lost; with a probability of 30 percent, Gene B will survive and while Gene A will totally lost
- Different treatments for intrinsic convection and external convection:

  Central scheme works well for intrinsic convection no matter whether it is dominated or not.

  For external convection, central scheme will lead to non-physical solution if the resolution is not fine enough.

- The solution has a splitting as $f(x,t) = f_c + m_0(t)\delta(0) + m_1(t)\delta(1)$

Outline

Future Work

Discussions and future work

1. Resolution for the singularity – h scale now
2. Variational particle methods may have a resolution of machine precision.
3. 3 genes problems – really challenge

$$\frac{\partial}{\partial t} f - \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}[(\delta_{i,j} x_i - x_i x_j) f(x,t)] = 0,$$

$$x \in \{x_1 > 0, x_2 > 0, x_1 + x_2 < 1\}$$

谢 谢 !