

Goal

Build a rejection algorithm to simulate exactly from coupled Wright-Fisher diffusion models.

Coupled Wright-Fisher diffusion model

- Proposed by Timo Koski and Erik Aurell, the coupled Wright-Fisher diffusion models the evolution of frequencies of L different loci with M_L different types of alleles each, in an haploid asexually reproducing population of constant size N .
- The model characterizes possible pairwise interactions between frequencies of allele types across different loci.

Expressed as a system of stochastic differential equations, the coupled Wright-Fisher diffusion, in its most general form, reads

$$d\mathbf{X}_t = \mu(\mathbf{X}_t)dt + \mathbf{G}(\mathbf{X}_t)dt + \mathbf{D}^{\frac{1}{2}}(\mathbf{X}_t)d\mathbf{B}_t,$$

where \mathbf{X}_t is a vector of allele types frequencies at each locus, \mathbf{B}_t a vector of independent Brownian motions and $\mathbf{D}(\mathbf{x})$ a diffusion block-diagonal matrix with entries

$$d_{ij}^l = \begin{cases} x_i^l(1-x_i^l), & i=j, \\ -x_i^l x_j^l, & i \neq j, \end{cases} \quad \text{where } i, j \in \{1, \dots, M_l - 1\}, l \in \{1, \dots, L\}.$$

- The drift function μ governs mutation of allele types at each locus separately.
- The coupling function \mathbf{G} parametrizes the pairwise interactions and has general form

$$\mathbf{G}(\mathbf{x}) = \mathbf{D}(\mathbf{x})\nabla_{\mathbf{x}}\mathbf{V}(\mathbf{x}),$$

where $\mathbf{V}(\mathbf{x}) = \mathbf{x}^T \sigma + \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x}$, with σ a vector of selection parameters and \mathbf{A} a matrix of pairwise interactions parameters.

Two-dimensional example

- Consider the following coupled Wright Fisher diffusion model for 2 loci and 2 allele types with law $\mathbb{CWF}_{\alpha, G, x_0}$ (target distribution)

$$\begin{aligned} dX_t^1 &= \alpha(X_t^1)dt + hX_t^1(1-X_t^1)X_t^2dt + \sqrt{X_t^1(1-X_t^1)}dB_t^1 \\ dX_t^2 &= \alpha(X_t^2)dt + hX_t^2(1-X_t^2)X_t^1dt + \sqrt{X_t^2(1-X_t^2)}dB_t^2, \end{aligned} \quad (1)$$

where the drift $\alpha(x) = \theta_1(1-x) - \theta_2x$ describes recurrent mutation between allele types governed by parameters $\theta_1, \theta_2 > 0$ and the coupling term describes allele types interaction governed by parameter h .

- Consider the following 2 neutral (independent) Wright-Fisher diffusions with joint law $\mathbb{WF2}_{\alpha, x_0}$ (candidate distribution)

$$\begin{aligned} dX_t^1 &= \alpha(X_t^1)dt + \sqrt{X_t^1(1-X_t^1)}dB_t^1 \\ dX_t^2 &= \alpha(X_t^2)dt + \sqrt{X_t^2(1-X_t^2)}dB_t^2, \end{aligned}$$

that can be simulated exactly using the algorithm presented in [1].

The strategy is to use samples from the candidate distribution $\mathbb{WF2}_{\alpha, x_0}$ in a rejection algorithm to sample from our target distribution $\mathbb{CWF}_{\alpha, G, x_0}$.

Rejection probability

- The Radon-Nikodým derivative of $\mathbb{CWF}_{\alpha, G, x_0}$ w.r.t. $\mathbb{WF2}_{\alpha, x_0}$, exists and is given by Girsanov's formula

$$\frac{d\mathbb{CWF}_{\alpha, G, x_0}}{d\mathbb{WF2}_{\alpha, x_0}} = \exp \left\{ \int_0^T h[X_t^2 dX_t^1 + X_t^1 dX_t^2] - \frac{1}{2} \int_0^T h^2 [(X_t^2)^2 X_t^1(1-X_t^1) + (X_t^1)^2 X_t^2(1-X_t^2)] + h[X_t^2 \alpha(X_t^1) + X_t^1 \alpha(X_t^2)] dt \right\}. \quad (2)$$

Consider

- $\phi : [0, 1]^2 \mapsto \mathbb{R}$ bounded by $\phi^- \leq \phi \leq \phi^+$ with

$$\phi(x^1, x^2) := \frac{1}{2} h^2 [(x^2)^2 x^1(1-x^1) + (x^1)^2 x^2(1-x^2)] + 2h [x^2 \alpha(x^1) + x^1 \alpha(x^2)],$$

and

- $A : [0, 1]^2 \mapsto \mathbb{R}$ bounded above by $A^+ \geq A$ with

$$A(X_T^1, X_T^2) := \int_0^T h[X_t^2 dX_t^1 + X_t^1 dX_t^2] = h \int_0^T d(X_t^1 X_t^2) = h[X_T^1 X_T^2 - X_0^1 X_0^2].$$

Then, (2) can be rewritten as

$$\frac{d\mathbb{CWF}_{\alpha, G, x_0}}{d\mathbb{WF2}_{\alpha, x_0}} \propto \exp \left\{ A(X_T^1, X_T^2) - A^+ \right\} \exp \left\{ - \int_0^T [\phi(X_t^1, X_t^2) - \phi^-] dt \right\} \quad (3)$$

References

- [1] P. A. Jenkins and D. Spanò. Exact simulation of the Wright-Fisher diffusion. *Ann. Appl. Probab.*, 27(3):1478-1509, 2017.

Decision event

Is it possible to construct an event that occurs exactly with probability (3)?

- Let Φ be a Poisson process with unit intensity on $D := [0, T] \times [0, \phi^+ - \phi^-] \subset \mathbb{R}^2$.
- Let N be the number of points of Φ that lie in $B = \{(x, y) \in D : y \leq (\phi(X_t^1, X_t^2) - \phi^-)\}$.

Then, given a realization $\mathbf{x}_t = (x_t^1, x_t^2)$ of $\mathbf{X}_t = (X_t^1, X_t^2) \sim \mathbb{WF2}_{\alpha, x_0}$

$$\Pr(N=0 | \mathbf{X}_t = \mathbf{x}_t) = \exp \left\{ - \int_0^T [\phi(x_t^1, x_t^2) - \phi^-] dt \right\},$$

is the probability that no points from Φ lie in B , i.e., that no points from Φ lie below the graph of $t \mapsto \phi(X_t^1, X_t^2) - \phi^-$.

Observe that such event can be exactly evaluated given a realization of \mathbf{X}_t at only a finite collection of time points (t_1, \dots, t_J) , i.e., given only a skeleton of \mathbf{X}_t .

Exact rejection algorithm

Algorithm 1 returns an exactly simulated skeleton of the solution $(X_t^1, X_t^2) \sim \mathbb{CWF}_{\alpha, G, x_0}$ of (1) at times (t_1, \dots, t_J) .

Algorithm 1: Exact algorithm for simulating skeletons of the paths (X_t^1, X_t^2) for $t \in [0, T]$ of a coupled Wright-Fisher diffusion process with law $\mathbb{CWF}_{\alpha, G, x_0}$.

- repeat**
- Simulate Φ , a Poisson process on D .
- Simulate $U \sim \text{Uniform}(0, 1)$
- Given $\Phi = \{(t_j, \psi_j) : j = 1, \dots, J\}$, simulate $(X_t^1, X_t^2) \sim \mathbb{WF2}_{\alpha, x_0}$ at times $\{t_1, \dots, t_J\}$ and time T .
- if** $\phi(X_{t_j}^1, X_{t_j}^2) - \phi^- \leq \psi_j$, $\forall j$ and $U \leq \exp\{A(X_t^1, X_t^2) - A^+\}$ **then**
- return** $\{(t_j, X_{t_j}^1, X_{t_j}^2), \forall j\} \cup \{(T, X_T^1, X_T^2)\}$
- end if**
- until** false

Once a skeleton is accepted, the entire path of (X_t^1, X_t^2) can be filled by simulating the corresponding independent Wright-Fisher bridges, as proposed in [1].

Simulation results

We simulate 10,000 skeletons of $(X_t^1, X_t^2) \sim \mathbb{CWF}_{\alpha, G, x_0}$ on $t \in [0.2, 1]$ with parameters $\theta_1 = \theta_2 = h = 0.5$ and bound constants $\phi^- = 0$, $\phi^+ = 0.375$ and $A^+ = 0.375$.

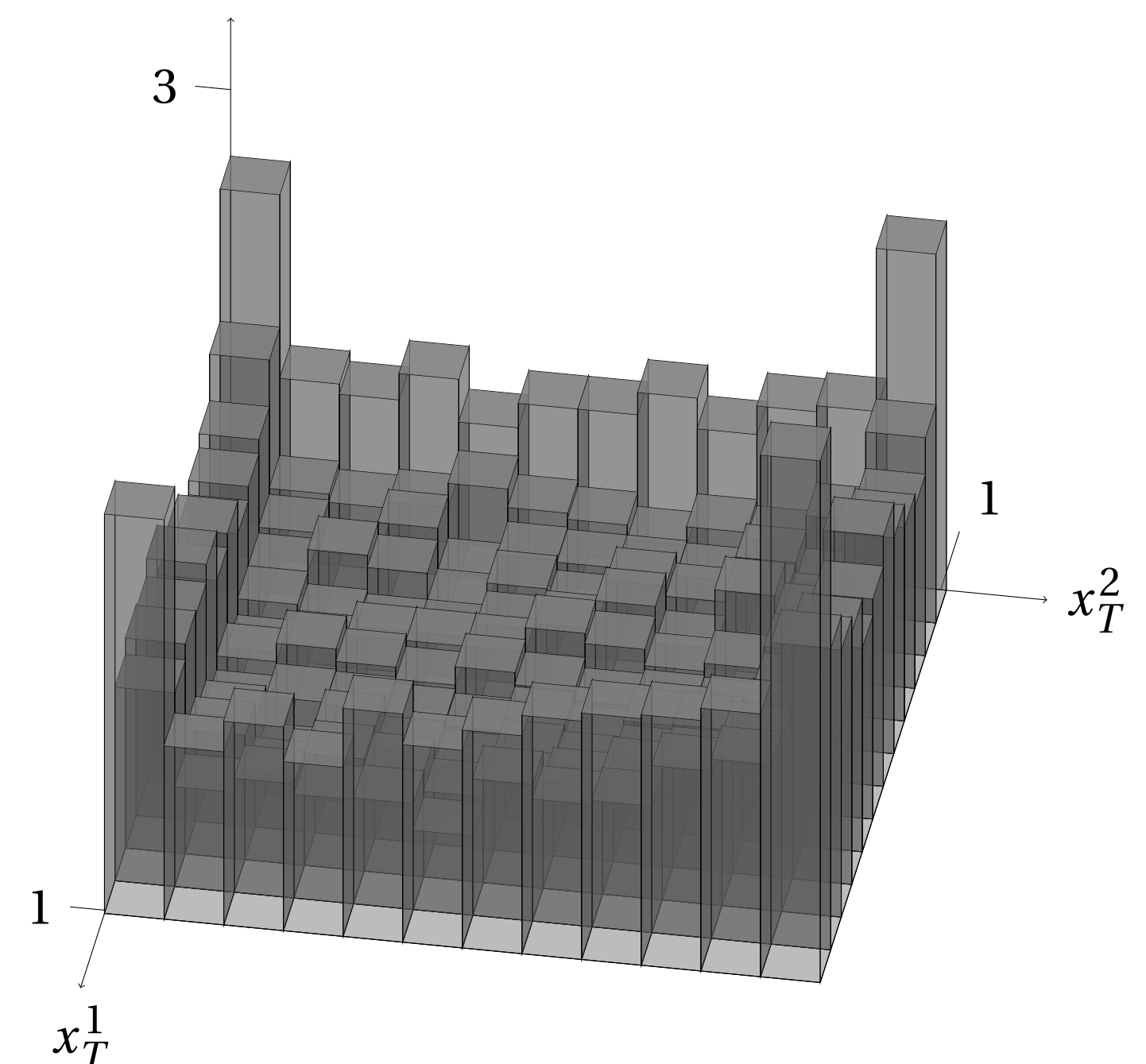


Figure 1: Histogram of 10,000 samples from (X_T^1, X_T^2) at $T=1$.

We obtained an estimated acceptance probability of 0.69. The average number of Poisson points needed to accept a skeleton was 0.14 and the maximum number of rejected skeletons before acceptance was 8 (average 1.44).