From Wright–Fisher Diffusion to a Time-Inhomogeneous BDI Filtering Equation for Rare Alleles

Model setup and scaling

Let $X_t \in [0, 1]$ be the allele frequency at (rescaled) time t, and let the effective population size vary through

$$\rho(t) \equiv \frac{N(t)}{N_0}.$$

Define population-scaled parameters $\theta = 4N_0\mu$ (mutation; we will use a one-sided rate for rare-allele analysis) and $\gamma = 4N_0hs$ (heterozygote selection). We will be interested in the probability

$$p_k(t) \equiv \Pr\{\text{a sample of size } n \text{ contains the allele exactly } k \text{ times at time } t\}, \quad k \ll n.$$

Wright–Fisher diffusion (generator and forward PDE)

Under a standard WF diffusion with time-varying population size and genic (heterozygote) selection, an infinitesimal generator for smooth f is

$$(\mathcal{L}_{WF}f)(x,t) = \Big[\tfrac{\gamma}{2}\,x(1-x)\big(x + h(1-2x)\big) + \tfrac{\theta_1}{2}(1-x) - \tfrac{\theta_2}{2}x\Big] f'(x) + \frac{x(1-x)}{2\,\rho(t)} f''(x). \tag{1}$$

Writing $\phi(x;t)$ for the density of X_t, the forward (Fokker–Planck) equation is

$$\partial_t\phi(x;t) = \mathcal{L}^*_{WF}\phi(x;t) = -\partial_x\Big(a(x,t)\,\phi\Big) + \partial_x^2\Big(D(x,t)\,\phi\Big), \qquad D(x,t) = \frac{x(1-x)}{2\rho(t)}, \tag{2}$$

with drift $a(x,t)$ read from (1). In the no-mutation case ($\theta_1 = \theta_2 = 0$), the boundaries $0, 1$ are absorbing.

Rare-variant limit and linearized generator

For large n and rare alleles ($x \ll 1$), linearize near $x = 0$ and retain the leading terms, dropping x^2 and $\theta_i x$:

$$(\mathcal{L}f)(x,t) = \frac{1}{2}\big(\gamma h\,x + \theta\big) f'(x) + \frac{x}{2\,\rho(t)}\, f''(x). \tag{3}$$

Below we suppress h and write γ for γh (heterozygote selection).

Poissonization of sampling and definition of $p_k(t)$

Given frequency x, the sample count $K \mid x \sim \mathrm{Bin}(n, x)$. In the rare-variant regime (x small, n large), approximate $\mathrm{Bin}(n, x)$ by $\mathrm{Pois}(nx)$. Then

$$p_k(t) = \mathbb{E}\left[\frac{(nX_t)^k}{k!} e^{-nX_t}\right], \tag{4}$$

where the expectation is taken over the diffusion generated by (3).

Forward master equation for $\{p_k(t)\}$pk(t) (BDI filtering)

Applying Dynkin's formula (or acting with $\mathcal{L}$ on the Poisson basis $(nx)^k e^{-nx}/k!$ inside the expectation) yields a closed, tri-diagonal system for $\mathbf{p}(t) = (p_0(t), p_1(t), \dots)^\top$:

$$\boxed{\frac{d}{dt}p_k(t) = \big(f + (k-1)b(t)\big)\, p_{k-1}(t) - \big(f + k(b(t) + d(t))\big)\, p_k(t) + \big(k+1\big)d(t)\, p_{k+1}(t), \quad k \geq 0,} \tag{5}$$

with time-inhomogeneous rates ("birth–death with immigration"):

$$f = \frac{n\theta}{2}, \qquad b(t) = \frac{n}{2\,\rho(t)}, \qquad d(t) = \frac{n}{2\,\rho(t)} - \frac{\gamma}{2}. \tag{6}$$

Interpretation: from state k, an upward jump to $k+1$ occurs either by an *immigration* event at rate f or by *birth* of one of the k copies at rate $k\,b(t)$; a downward jump to $k-1$ occurs by *death* of one copy at rate $k\,d(t)$.

Matrix form and filtering viewpoint. Let $\mathbf{Q}(t)$ be the tri-diagonal generator with entries

$$Q_{k,k-1}(t) = f + (k-1)b(t), \quad Q_{k,k}(t) = -\big(f + k(b(t) + d(t))\big), \quad Q_{k,k+1}(t) = (k+1)d(t).$$

Then (5) is the continuous-time forward equation

$$\dot{\mathbf{p}}(t) = \mathbf{Q}(t)\,\mathbf{p}(t), \tag{7}$$

which *filters/predicts* the distribution on the discrete count states forward in time given the rate functions and $\rho(t)$.

Closed-form solution via complete Bell polynomials

The system (5) admits an analytic solution in terms of complete Bell polynomials B_k:

$$\boxed{p_k(t) = e^{-\xi_0(t)}\,\frac{B_k\big(\xi_1(t), \dots, \xi_k(t)\big)}{k!}, \qquad k \geq 0.} \tag{8}$$

Here

$$\xi_0(t) = \int_0^t f\big(1 - \alpha(t;s)\big)\,ds, \tag{9}$$

$$\xi_i(t) = i!\int_0^t f\,q_i(t;s)\,ds, \qquad i \geq 1, \tag{10}$$

where $\alpha(t;s)$ and $q_i(t;s)$ are probabilities for the *immigration-free* birth–death process started from one copy at time s:

$$\alpha(t;s) = 1 - \frac{e^{R(t;s)}}{W(t;s)}, \qquad \beta(t;s) = 1 - \frac{1}{W(t;s)}, \tag{11}$$

$$q_0(t;s) = 1 - \alpha(t;s), \qquad q_i(t;s) = \big(1 - \alpha(t;s)\big)\big(1 - \beta(t;s)\big)\,\beta(t;s)^{i-1}, \quad i \geq 1, \tag{12}$$

with

$$W(t;s) = 1 + \int_s^t e^{R(t;u)}\,b(u)\,du, \qquad R(t;s) = \int_s^t \big(b(u) - d(u)\big)\,du. \tag{13}$$

Intuitively, $\xi_i(t)$ is the expected number of independent mutational origins that are at copy count i at time t; the Bell polynomial B_k organizes all set partitions of these independent clusters into a total of k copies.

Likelihood (rare-allele truncation)

For a mutational context j with rate θ_j (shared demography ρ), one obtains

$$p_{j,k}(t) \text{ from (8) by replacing } \theta \mapsto \theta_j.$$

Because (8) is valid for rare counts, fix $K \leq n$ and renormalize

$$\widetilde{p}_{j,k} = \frac{p_{j,k}}{\sum_{r=0}^{K} p_{j,r}}, \qquad \log L = \sum_{j=1}^{J} \sum_{k=0}^{K} c_{j,k} \log \widetilde{p}_{j,k},$$

where $c_{j,k}$ is the observed number of sites in context j with sample count k.

Summary. Starting from the WF diffusion with time-varying population size, the rare-variant limit plus Poissonization yields a closed, time-inhomogeneous *birth–death with immigration* forward equation for $\{p_k(t)\}$ (*filtering in count space*). This system has an analytic Bell-polynomial solution parameterized by $\rho(t)$, θ, and γ via the rate functions (6).