

3. Random genetic drift

- 3.1 genetic sampling
- 3.2 Wright-Fisher model
- 3.3 effective population size
- 3.4 diffusion approximation
- 3.5 RGD-mutation equilibrium
- 3.6 RGD-migration equilibrium

3.1 Genetic sampling

random inheritance of alleles in a finite population

RGD = random change in allele frequencies
due to genetic sampling

Ex 1: *D. melanogaster* experiment

Fig 3.4 p101 (7.4 p273): $K = 107$ experimental populations

population size $N = 16$

8 males + 8 females chosen at random

One gene with two alleles of equal fitness

brown eye allele $bw = a$

yellow eye allele $bw^{75} = A$

X_t = total number of alleles A in generation t

107 observed scenarios for 20 generations

20 histograms for X_t , $t = 0, 1, \dots, 19$

Properties of RGD

has no direction, accumulates with time

causes the loss of gen. variability within a population

causes an increase of gen. var. between populations

average allele freq. across populations remains constant

3.2 Wright-Fisher model

WFM is a simple population model on the allele level
assuming random mating and finite population size

Constant generation size

N diploids at each generation

$2N$ haploids (gene copies)

Backward description of the reproduction law

every gene copy picks its parent at random

from $2N$ gene copies in the previous generation

Offspring number distribution $\nu \sim \text{Bin}(2N, \frac{1}{2N})$

Pois(1) approximation: $E(\nu) = 1, \text{Var}(\nu) = 1$

k		0	1	2	3	4	5
$P(\nu = k)$		0.37	0.37	0.18	0.06	0.02	0.00

Ex 2: RGD simulation

WFM of size $N = 3, X_0 = 3$

random numbers produced by students (or dice)

six numbers between 1 and 6 per generation

Allele frequency dynamics

$$p_t = \frac{1}{2N} X_t \text{ frequency of allele } A \text{ in generation } t$$

Fixed initial frequency p_0

independent trajectories of p_t for different populations

Conditional distribution

$$X_t \sim \text{Bin}(2N, p_{t-1})$$

$$E(p_t | p_{t-1}) = p_{t-1}, \text{Var}(p_t | p_{t-1}) = \frac{p_{t-1}q_{t-1}}{2N}$$

$$\text{Var}(\Delta p) = \frac{pq}{2N}$$

Average p_t across populations remains constant

$$E(p_t) = E(p_{t-1}) = \dots = E(p_1) = p_0$$

Variation in p_t among populations increases with t

$$\sigma_t^2 = \frac{p_0 q_0}{2N} + \left(1 - \frac{1}{2N}\right) \sigma_{t-1}^2$$

Fixation index of RGD

Metapopulation of isolated populations under RGD

$$\text{fixation index } F_t = \frac{\sigma_t^2}{\bar{p}_t \bar{q}_t} \approx \frac{\sigma_t^2}{p_0 q_0} = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1}$$

$$\boxed{1 - F_t = \left(1 - \frac{1}{2N}\right) (1 - F_{t-1})}$$

Average heterozygosity across populations (unlinked loci)

$$\bar{H}_t = 2\bar{p}_t \bar{q}_t (1 - F_t) \approx 2p_0 q_0 (1 - F_t)$$

In terms of the pedigree inbreeding coefficient

$$F_t = \text{probab. for two random gene copies at gener. } t$$

to descend from the same gene copy at gener. 0

The rate of RGD in the WFM

$$1 - F_t = \left(1 - \frac{1}{2N}\right)^t$$

$$\text{average heterozygosity } \bar{H}_t \approx 2p_0 q_0 \left(1 - \frac{1}{2N}\right)^t$$

3.3 Effective population size

for a randomly mating population different from WFM

EPS N_e is the size of WFM with the same RGD rate

$$\boxed{1 - F_t = \left(1 - \frac{1}{2N_e}\right)^t}$$

Actual size $N \neq N_e$ (usually $N_e < N$)

due to various deviations from WFM assumptions

Ex 1: D. melanogaster experiment

Fig 3.12 p119 (7.12 p288): $N = 16$ and $N_e = 9$

$$\text{average heterozygosity } \bar{H}_t \approx 0.5 \left(1 - \frac{1}{18}\right)^t$$

$$\boxed{\text{Exchangeable reproduction law } N_e = \frac{N}{\text{Var}(\nu)}}$$

Variable population size

Assuming large historical population sizes N_1, N_2, \dots, N_t

$$\begin{aligned}1 - F_t &= \left(1 - \frac{1}{2N_1}\right)\left(1 - \frac{1}{2N_2}\right) \dots \left(1 - \frac{1}{2N_t}\right) \\ &\approx 1 - \frac{1}{2N_1} - \frac{1}{2N_2} - \dots - \frac{1}{2N_t} \\ 1 - F_t &= \left(1 - \frac{1}{2N_e}\right)^t \approx 1 - \frac{t}{2N_e}\end{aligned}$$

Harmonic mean formula $\frac{1}{N_e} = \frac{1}{t}\left(\frac{1}{N_1} + \frac{1}{N_2} + \dots + \frac{1}{N_t}\right)$

Bottleneck effect

$$N_1 = \dots = N_5 = 1000$$

$$N_6 = N_7 = 100, N_8 = \dots = N_{12} = 1000$$

Compare $N_e = 400$, $\bar{N} = 850$, and current $N = 1000$

Ex 3: northern elephant seals

hunted down to $N = 20$ in 1890's

now $N = 30,000$ and no genetic variation in 24 genes

Southern elephant seals

to the contrary have wide genetic variation

Founder effect

Small number of founders and exponential growth:

$$N_1 = 10, N_2 = 20, N_3 = 40, N_4 = 80, \dots, N_{10} = 5120$$

compare $N_e = 50$, $\bar{N} = 1023$, and current $N = 5120$

Ex 4: Amish in Pennsylvania

A = recessive allele causing Dwarfism

in Switzerland $p_{AA} = 1/1000$

among 12,000 Amish in Pennsylvania $p_{AA} = 1/14$

descendants of 30 people migrated from Switz in 1720

Ex 5: 50/500 conservation rule

begin a conservation program with at least $N_e = 50$
 and release the animals when $N_e = 500$

Example: $N_t = 5120, 2560, \dots, 40, 20, 10$

$N_e = 50$, the species might be saved even though
 the actual population size is only $N = 10$

Unequal sex ratio

Unequal number of breeding males N_m and females N_f
 two random gene copies descend from the same
 parental gene with probability $\frac{1}{2N_e} = \frac{1}{4} \cdot \frac{1}{2N_m} + \frac{1}{4} \cdot \frac{1}{2N_f}$

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

Ex 6: buffalo population

5 bulls mating 100 cows $N_e = \frac{4 \cdot 5 \cdot 100}{5 + 100} = 19.05$

Uniform population dispersion

Density parameter $\delta =$ the number

of breeding individuals per hectare ($= 10^4 \text{ m}^2$)

Mobility parameter σ measured in 100 m

offspring birthplaces are IID with $N(\mu_1, \mu_2, \sigma^2, \sigma^2, 0)$

parent's birthplace (μ_1, μ_2) measured in 100 m

$$\text{Neighborhood size: } N_e = 4\pi\delta\sigma^2$$

Non-random mating: $1 - F_t = (1 - \frac{1}{2N_e})^t$, where

$F_t =$ probab. for two gene copies in an individual
 at generation t to descend from the same
 ancestral gene copy at generation 0

Ex 7: deer mouse, Michigan

$\delta = 6$ breeding individuals per ha, $\sigma = 114$ m
neighborhood size $N_e = 4 \cdot 3.14 \cdot 6 \cdot (1.14)^2 = 98$
low density and low mobility

EPS of mt DNA

Nuclear gene EPS = N_e

X-linked gene EPS = $\frac{3}{4}N_e$

Y-linked gene EPS = mt DNA EPS = $\frac{1}{4}N_e$

3.4 Diffusion approximation

Diffusion approximation $\xi_t = p_{[2Nt]}$

allele frequency with time running faster by factor $2N$
in WFM with add. selection and reversible mutation

1 generation corresponds to $\Delta t = \frac{1}{2N}$ of diffusion time

Conditional expectation and variance

$$E(\Delta \xi_t | \xi_t = p) = M(p) \Delta t \quad M(p) = -ap + bq + cpq$$

$$\text{Var}(\Delta \xi_t | \xi_t = p) = V(p) \Delta t \quad V(p) = pq$$

Infinitesimal mutation rates $a = 2N\mu$, $b = 2N\nu$ and selection coefficient $c = 2Ns$

Backward and Forward Kolmogorov Equations for
conditional pdf $\phi(p, x, t)$ of $\xi_t = x$ given $\xi_0 = p$

BKE: first generation change $\phi'_t = M(p)\phi'_p + \frac{1}{2}V(p)\phi''_p$ FKE: last gen. change $\phi'_t = -[M(x)\phi]'_x + \frac{1}{2}[V(x)\phi]''_x$
--

Allele fixation

T = time to fixation of allele A at frequency p

$u(p) = P(T < \infty | \xi_0 = p)$ probability of fixation

$T = \infty$ means that allele A is never fixed i.e. lost

$u(p)$ satisfies stationary BKE with $u(0) = 0$, $u(1) = 1$

$$\boxed{\text{Stationary BKE } M(p)u' + \frac{1}{2}V(p)u'' = 0}$$

WFM with selection and without mutation

solution of the stationary BKE $u(p) = \frac{1-e^{-2cp}}{1-e^{-2c}}$

in particular, if no selection $u(p) = p$

Fixation of a new mutation

New mutation $p = \frac{1}{2N}$ fixation prob. $u(\frac{1}{2N}) = \frac{1-e^{-2s}}{1-e^{-4sN}}$

$$\boxed{u(\frac{1}{2N}) \approx \frac{2s}{1-e^{-4sN}} \text{ if } |s| \ll 1}$$

Neutral selection: if $|s| \ll \frac{1}{4N}$, then $u(\frac{1}{2N}) = \frac{1}{2N}$

mean time to fixation $E(T|T < \infty) \approx 4N$

average time to loss $\approx 2 \ln(2N)$

Positive selection

if $\frac{1}{4N} \ll s \ll 1$, then $u(\frac{1}{2N}) \approx 2s$

mean time to fixation $E(T|T < \infty) \approx \frac{2}{s} \ln 2N$

Negative selection

if $\frac{1}{4N} \ll -s \ll 1$, then $u(\frac{1}{2N}) \approx -2s \cdot e^{4sN}$

Ex 8: numerical example

population size $N = 10^6$

If $s = 0.05$, then

$u(\frac{1}{2N}) = 0.1$ or 90% probability of loss

mean time to fixation 580 generations

If $s = -0.01$, then

$$u\left(\frac{1}{2N}\right) = 0.02 \cdot e^{-40000} = 0 \text{ fixation is impossible}$$

If neutral mutation, then

$$u\left(\frac{1}{2N}\right) = 0.5 \cdot 10^{-6}$$

mean time to fixation ≈ 4000000 generations

average time to loss ≈ 29 generations

Stable distribution of the allele frequency

Under unchanged circumstances distribution $\phi(p, x, t)$

becomes a stationary distribution: $\phi(p, x, \infty) \equiv f(x)$

independent of t and p , so that initial state is forgotten

$$\text{Stationary FKE: } [M(x)f(x)]' = \frac{1}{2}[V(x)f(x)]''$$

WFM with reversible mutation, no selection:

$$\text{Beta}(2b, 2a) \text{ pdf } f(x) = \frac{\Gamma(2a)\Gamma(2b)}{\Gamma(2a+2b)} x^{2b-1}(1-x)^{2a-1}$$

Mean value and variance of Beta($2b, 2a$) distribution

$$\text{mean } \hat{p} = \frac{2b}{2a+2b} = \frac{\nu}{\nu+\mu} \text{ equilibrium frequency}$$

$$\text{variance} = \frac{ab}{(a+b)^2(2a+2b+1)} \text{ strength of RGD}$$

Beta($2b, 2a$) curve shapes

U curve, $2a < 1$ and $2b < 1$, weak SF strong RGD

bell curve, $2a > 1$ and $2b > 1$, strong SF weak RGD

L curve, $2a < 1$ and $2b > 1$, strong SF against fixation