Lecture 9. Effective population size. Diffusion approximation.

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

$$\boxed{1 - F_t = (1 - \tfrac{1}{2N_e})^t}$$

Actual size $N \neq N_e$ (usually $N_e < N$)
due to various deviations from WFM assumptions

Ex 1: D. melanogaster experiment

Fig 7.12, p. 288: $N = 16$ and $N_e = 9$
average heterozygosity $\bar{H}_t \approx 0.5(1 - \frac{1}{18})^t$

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

Variable population size

Assuming large historical population sizes $N_1, N_2, \ldots, N_t$

$$1 - F_t = (1 - \tfrac{1}{2N_1})(1 - \tfrac{1}{2N_2}) \ldots (1 - \tfrac{1}{2N_t})$$
$$\approx 1 - \tfrac{1}{2N_1} - \tfrac{1}{2N_2} - \ldots - \tfrac{1}{2N_t}$$
$$1 - F_t = (1 - \tfrac{1}{2N_e})^t \approx 1 - \tfrac{t}{2N_e}$$

$$\boxed{\text{Harmonic mean formula } \tfrac{1}{N_e} = \tfrac{1}{t}(\tfrac{1}{N_1} + \tfrac{1}{N_2} + \ldots + \tfrac{1}{N_t})}$$

Bottleneck effect

$N_1 = \ldots = N_5 = 1000$
$N_6 = N_7 = 100$, $N_8 = \ldots = N_{12} = 1000$
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:
$N_1 = 10, N_2 = 20, N_3 = 40, N_4 = 80, \ldots, N_{10} = 5120$
compare $N_e = 50$, $\bar{N} = 1023$, and current $N = 5120$

Uniform population dispersion

Density parameter δ = the number
of breeding individuals per hectare ($= 10^4$ 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

> 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

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
$\mathrm{E}(\Delta\xi_t | \xi_t = p) = M(p)\Delta t \qquad M(p) = -ap + bq + cpq$
$\mathrm{Var}(\Delta\xi_t | \xi_t = p) = V(p)\Delta t \qquad V(p) = pq$

> 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
$u(p) = \mathrm{P}(T < \infty | \xi_0 = p)$ probability of fixation
$T = \infty$ means that allele A is never fixed i.e. lost
$u(p)$ satisfies stationary BKE with $u(0) = 0$, $u(1) = 1$

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

WFM with selection and without mutation
solution of the stationary BKE $\qquad u(p) = \frac{1-e^{-2cp}}{1-e^{-2c}}$
in particular, if no selection $\qquad u(p) = p$

Fixation of a new mutation

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

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

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

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

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 $\mathrm{E}(T|T < \infty) \approx \frac{2}{s}\ln 2N$

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

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

If neutral mutation, then

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

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

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

WFM with reversible mutation, no selection:

Beta$(2b, 2a)$ 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

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

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

Literature:

1. D.L.Hartl, A.G.Clarc. Principle of population genetics. Sinauer Associates, 2007.

2. R.Nielson, M. Statkin. An introduction to population genetics: theory and applications, Sinauer Associates. 2013.