ch 1
- exponential growth
- probability simplex / Regular polygon formed by unit vector .
- HW - principle. 2allele 列出 ODE 的 stable 是否 phenotype 变了 种群到稳定
- With mutation

ch 2
- what is Quasispecies . • IBV 的研究
- sequence space
- 列出 Quasispecies equation
  $\dot{x} = x\underset{fQ}{W} - \phi x$ $Q = [\quad]$ Q_{ij} 从i到j 概率
- Adaptivitiele . / Def / Example .
- mutation meltdown / antivirus strategy . 推导 $0<1$
  $\log f_Q \approx (f_0 \approx 1)$
  $(1-u)^L$

ch3
Some MC propoties .
- Moran Def . / neutral / general case .
- Fixation Probabilty . • competing exponential 推到 $1-P[\lambda]$
- Mean fixation time . • Poisson process — IAT 推到 $P[\text{时间}]$

ch4 → 给了 GG
- Def cancer . How were
- TSG . Oncogenes . → 有了 GG
- Prevent of cancer / linear 1 compartmentazion . $P = \frac{r}{N}$
- 2 - stage model for TSG why 2 hits / tunnelling .
  $\square \xrightarrow{u_1} \square \xrightarrow{Nu_2}$ (why)

CIN model ↗ Independent mutation

ch5 . Multi-stage theory → each stage $WT \sim \text{Exp}(\lambda)$ / (Mutation Pathway) 推导这些推导
- Wg Fisher Process Binomial according to the number
  Fixed — Expeted Fixation Time
  - MC with Binorial as transition kernel
  - Example / length Is there a analytic solution ?
  这是 Extension 到了 binary sequence of d $P[\text{Sampling 到了 } i]$

ch6 . Diffusion Theory .
Δ Kolmogorov forward/Backward .
Fixation prob .
Mean Fix time .

ch7 Pay-off matrix. / 各种平衡
ESS unbeatable , strict Nash .
- Determine linear stability by det(A) why ?
- TFT .
  $f_A(1-\varepsilon) > f_B(\varepsilon)$ 被invaded by A
  $a(1-\varepsilon) + b\varepsilon > c(1-\varepsilon) + d\varepsilon$

- Intensity of selection . How does selection favors A .
- Fixation Prob . selection opposes invasion .
  1/3 law
- Risk dominance

ch 9 . spatial Moran .
- How to jump from i → j cell
- Eder growth { Particle / Bond / cell Pene-based .

ch 10 G W Process $\xi^{(i)}_{n,j}$ Branching process
(Prob generating function) its relation to Variance / Expectation
$f_{n+1} = f^{n+1}(s)$
Extinction Prob . and the theorem
- 解释 $P_m(k, j)$

ch 11 • Poset : set with " order operator "
- Distributed latter • chain
- genotype lattice 还有多
- Fitness / Mutational neighbourhood .
- Mutation matrix and its relationship with the
- Multiple Branching / 相关信到了 Poisson dist .

ch 12
- Def of coalescent • what genology tree comes from
- How does the $T(j)$ written
- How is $T(j)$ distributed / Derivation $N \to \infty$
- How the Expectation
- Infinite Sites Model / Assumption ,
- Tajima's D , what does it do ? null hypothesis

For neutral
$X_0 \xrightarrow{u_1} X_1 \xrightarrow{Nu_2} X_2$
$X_0 \to X_1$ Avg $\frac{1}{Nu_2}$
$X_0 \to X_1$ $\frac{1}{N}$

Rate of mutation

Is WF Process a death-birth process ? No analytic sol.

$P_k = P[Z = k]$ set by distribution
P_{ij} transition probabilty . $P_{1k} = P_k$
$P_{2j} = P_0 P_{j-1} \cdots P_j P_0$
$P_{ij} = \sum_{k_1 \cdots k_i = j} P_{k_1} \cdots P_{k_i} \ \forall i$

五福 拼手气

一

集齐五福，拼手气

二

集齐五福，拼手气分5亿

$\frac{1-r^{-1}}{1-r^{-N}}$

五福，拼手气

Chp 2 Evolution

Def. changes in the frequency of types from 1 gen. to the next.
↳ from mutation, replication and selection.

Def (Malthusian Law) Exponential Growth
$x' = rx \Rightarrow x(t) = x_0 e^{rt}$ r : rate of cell division : average gen. time : $\frac{1}{r}$
with death $\frac{1}{d}$ on average
$x' = (r-d)x \Rightarrow x(t) = x_0 e^{(r-d)t}$ $R_0 = r/d$:= basic reproductive ratio. ≥ 1 ↑ < 1 extinct $= 1$

Logistic growth. population has a carrying capacity K
$x' = rx(1-\frac{x}{K}) \Rightarrow x(t) = \frac{K x_0 e^{rt}}{K + x_0(e^{rt}-1)} = 1 - \frac{\square}{\square}$ $\lim_{t\to\infty} x(t) = K$

Rescaly $x' = \alpha x(1-x)$ α is growth rate.

Bifurcation

Selection : 2 exp. growing types. type A : a $x(t)$ $x' = ax$ 两个都不extinct
B : b $y(t)$ $y' = by$
with cap. let $x(t)+y(t)=1$ $\forall t \geq 0$ 有限

$\begin{cases} x' = x(a-\phi) \\ y' = y(b-\phi) \end{cases}$ $\phi = ax+by$: averge fitness of the population. 整个系统[世界]的 relative fitness
$\Leftrightarrow$
$x' = x(1-x)(a-b)$

if $a>b$ → All a $x=1$
$a<b$ → All b $x=0$

obability Simplex

$S_n = \{(x_1 \dots x_n), \sum_{i=1}^{n} x_i = 1 \quad x_i \geq 0\}$ 0-1 S_2 S_3 S_4 Regular
every pt. in S_i is a prob. dist.
最fit的总会变成 (0.... 1 ....)

ore type $x_1(t) + \dots + x_n(t) = 1$ in S_n. average fitness: $\phi = x_1 f_1 + \dots + x_n f_n$
$x_i' = x_i(f_i - \phi) \Rightarrow$ a single equilibrium pt.

ub/super exp. growth.
$\begin{cases} x' = ax^c - \phi x \\ y' = by^c - \phi y \end{cases}$ $\phi = ax^c + by^c$ $\Rightarrow$ $\begin{cases} c=0 & \text{linear (immigration)} \\ c=1 & \text{exp.} \\ c<1 & \text{sub.} \\ c>1 & \text{super.} \end{cases}$
$\Leftrightarrow$
$x' = x(1-x)f(x)$, $f(x) = ax^{c-1} - b(1-x)^{c-1}$
Fixed pt: $\begin{cases} x=0 \\ x=1 \\ x^* = 1/(1+(a/b)^{1/(c-1)}) \end{cases}$ if $c \neq 1$ why $f(x)=0 \Rightarrow$ $\{0,1\}$ 不stable.
hen $c<1$: survival of all x^* is globally stable. Can invade to stable pt, 即使a>b, 也有可能extinct
$c>1$: unstable x^* : If $x > x^*$ A out compete B Can NOT invade.
$x < x^*$ B out-compete A
$\{0,1\}$ stable

ation u_1 Prob(A→B) u_2 Prob(B→A) durig reproduction
$\begin{cases} x' = x(1-u_1) + y u_2 - \phi x \\ y' = y(1-u_2) + x u_1 - \phi y \end{cases}$ $\phi = 1$ why $x+y=1$ 区别 growth difference 是一样的.
$\Rightarrow x' = u_2 - x(u_1+u_2)$ $\Rightarrow$ stable equilibrium $x^* = \frac{u_2}{u_1+u_2}$.
If $u_2 = 0$. $x' = x u_1$ $\Rightarrow$ A extinct

tation of n types. q_{ij} : Prob(i→j) Q stochastic matrix.
$x_i' = \sum_{j=1}^{n} x_j q_{ji} - \phi x_i$ $x' = xQ - \phi x$

y-Weinberg Principle aa aA AA : x y z a: p A: q
$x+y+z = p+q = 1$ Random mating $\Rightarrow$ $\begin{cases} x = p^2 \\ y = 2pq \\ z = q^2 \end{cases}$ $p' = p$ $q' = q$ constant!

$\dot{x}_2 = f - x - \beta = 0$

HIV Infection Some biological bias on HIV

stats: short genome, high mutation rate ↑, short gen. time, large population.

Sequence space

Def space $A^L = \{(\ \ ) \mid a_i \in A\}$, $A^* = \cup_{L \geq 0} A^L$

Evolution is a trajectory of a population in sequence space

Def. Hamming distance

Def Manhattan dist. no. of mismatches

Def. Fitness $f: G \to \mathbb{R}$ or $\{G \to P \to \mathbb{R}$. hard to measure.

(G: Genotype, P: penotype.)

Fitness is a MAP from genotype to

Quasispecies equation

Ex. $i = 0, 1, 2 \ldots n$

$x(t) = (x_0(t) \ldots x_n(t))$ genotypes frequencies

$Q = q_{ij}$:= mutation matrix $= I$ (error free)

$f = (f_0 \ \ f_n)$ fitness $\phi = x \cdot f$ average fitness

$\dot{x}_i = \sum_{j=0}^{n} x_j f_j \underline{q_{ji}}$ (mutation) $i = 0, \ldots n$ $\iff$ $\dot{x}_i = x f Q - \phi x_i$

If Q is irreducible, global x^* exist.

Solving

$X_i(t) = x_i(t) e^{\varphi(t)}$ $\varphi(t) = \int_0^t \phi(s) ds$

$\Rightarrow \dot{X}_i = \sum_{j=0}^{n} X_j f_j q_{ji}$ $\Rightarrow X = \sum_{j=0}^{n} X_j = e^{\varphi}$ thus $x_i = \frac{X_i}{X}$

$\Rightarrow \dot{X} = \dot{\varphi} e^{\varphi} = \phi X$ (X grow exponentially)

$xW = \phi x$ ϕ is the largest eigenvalue of W

This can be rewritten as $\dot{x} = xW - \phi x$ with $W = (w_{ij}) = (f_j q_{ji})$

x^* is the solution of $xW = \phi x$ [Eigenvalue problem

$\Rightarrow$ ϕ is the largest eigenvalue of W and x^* is left eigenvector.

Adaption

Def. Location of maximal fitness in sequence space.

{ ~~mutation: generates "peaks"~~
{ selection: drive to max peak

> balance x^*

If mutation removes "fit" types adaption is NOT possible.

Ex. Binary seq. of L (0: f_0) (1: 1) 两分中

x_0, x_1 be the frequency of Type 0 Type 1

Type 0 copy without error $q_L = (1-u)^L$

$\begin{cases} \dot{x}_0 = x_0 f_0 q_L - \phi x_0 \\ \dot{x}_1 = x_0 f_0 (1 - q_L) + x_1 - \phi x_1 \end{cases}$ $\iff$ (应介乏有代前 [?]) $\dot{x}_0 = x_0 [f_0 q_L - 1 - x_0 (f_0 - 1)]$

$x_0^* = \begin{cases} \frac{f_0 q_L - 1}{f_0 - 1} & f_0 q_L > 1 \\ 0 & \text{else}. \end{cases}$

What is mutation meltdown?

To get the bound for mutation rate

$\log f_0 \approx 1$

$\log(f_0 q_L) = 1 + L \log(1-u) \approx 1 - Lu > \log 1 \iff u < \frac{1}{L}$

$\Rightarrow u_c = \frac{1}{L}$ is the error threshold.

$u > \frac{1}{L}$ "mutation meltdown"

⇑ Strategy to anti viral

It's like the learning rate in some way

Ch 3. Stochastic Model in Finite Population

Basis exponential. pdf $f(x) = \lambda e^{-\lambda x}$ $E[X] = \frac{1}{\lambda}$ $Var[X] = \frac{1}{\lambda^2}$

$F(x) = 1 - e^{-\lambda x}$

$X \sim Exp(\lambda)$ $Y \sim Exp(\mu)$ $\min(X, Y) \sim Exp(\lambda + \mu)$

MC transition matrix, homogenous. absorbing / e.

Ergodic { i) aperiodic ii) irreducible. iii) positive recurrent

↓

π_i stationary dist.

唯一 fixation

从某一个数开始 一直到 N

Moran simplest A, B 2 type fixed N population. , state from A $i = 0, 1, 2, \ldots N$

$P_{i\,i+1} = p(1-p)$ $P_{i\,i-1} = (1-p)p$ $P_{ii} = p^2 + (1-p)^2$ $p = i/N$ frequency of A

! Each step only change 1 subjects inside population.

Def Birth-death process is a process with. P (tri-aylor) and 1 subjects change in one step. 每一次只能改变

Fixation Probability x_i : prob. of getting to N starting from i

$x_i = \frac{i}{N}$ since each allele has the same chance of being fix

More general cases $P_{i\,i+1} = \alpha_i$ $P_{i\,i-1} = \beta_i$ $\alpha_0 = \beta_N = 0$

let $\gamma_i = \frac{\beta_i}{\alpha_i}$ $x_i = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^{j} r_k}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} r_k}$ Prob. of ending in N starting from i

Absorption

Mean fixation time

$-N^2[(1-p)\log(1-p) + p\log p]$ generation

Moran w/ constant selection A waiting time $\sim exp(\lambda_A = r)$ $B \sim exp(1)$

r is the relative fitness of A over B

T_A T_B be the waiting times till birth.

$T_A \sim \min\{Exp(\lambda_A) \ldots Exp(\lambda_A)\} = exp\{\lambda_A i\}$ $T_B \sim Exp((N-i)\lambda_B)$

$P(T_A < T_B) = \frac{ri}{ri + N - i}$ $P(T_A > T_B) = \frac{N-i}{ri+N-i}$ competeing exponential

Transt. Prob { $P_{i\,i+1} = \frac{ri}{ri+N-i} \frac{N-i}{N}$ $P_{i\,i-1} = \frac{N-i}{ri+N-i} \frac{i}{N}$ $P_{ii} = 1 - \ldots$

$P(T_A < T_B)$

Fixation Prob $x_i = \frac{1 - 1/r^i}{1 - 1/r^N}$

Cancer . imbalance of cellular cooperation
. sometic (Non-germ cells) evolution : selfish, uncontrolled replication
. accumulation of mutations. [cellular cooperation]

Some hallmarks of cancer: disruption in i) integrity of genome ii) correct division iii) cell status iv) mobility.

Mutation type : i) point ii) chromosomal iii) indel iv) mitotic

Mutations accumulates if the rate of replication is high → Ex. colon cancer.

Tumor suppresor genes

Oncogenes : ↑ fitness if one allele is mutated or inappropriately expressed

Fixation of oncogenes : Moran w/ N being the number of cells. $\rho = x_i = \frac{1-r^{-i}}{1-r^{-N}}$ $i=1$ if one mutation start

$P(t) = 1 - e^{-Nu\rho t}$: Prob. of a mutant is fixed by time t

$P(t)$ ↑ in N if $r > 1$ else decreasing.

Large N increases the fixation probability so ⇒ slow it down by small cellular compartment.

Linear process ↑ Have to look at video . Eg. crypos in colon . Compartmentalization helps

1-D linear dimension process. divide asymmetry ⇒ apoptosis. shifted $P = 1/N$

TSG: w/o it, you would get cancer ⇒ inactivation of both allele leads to fitness advantage ↑

type 0 —u_1→ 1 —u_2→ 2 fitness advantage. 2 pts mutation 1 mutation LOH

Fixation: $\frac{1}{\rho} = N \frac{1}{Nu_2}$. So type 1 cells reach fixation before type 2. if $N << \frac{1}{\sqrt{u_2}}$.

Let $x_i(t)$ be probability of staying in $\begin{cases} 0 : \text{type } 0 \\ 1 \quad -\,-\,1 \\ 2 \quad -\,-\,2 \end{cases}$

ODE
$\dot{x}_0 = -u_1 x_0$
$\dot{x}_1 = u_1 x_0 - Nu_2 x_1$
$\dot{x}_2 = Nu_2 x_1$

$X(0) = (1,0,0)$ at first "all healthy"

$t \to 1$ $X(t) = (0,0,1)$

$P(t) = X_2(t) = 1 - \frac{Nu_2 e^{-u_1 t} - u_1 e^{-Nu_2 t}}{Nu_2 - u_1}$

Time scales short: $t << 1/Nu_2$ $P(t) \approx Nu_1 u_2 t^2/2$ [2 rate limiting events]

long: $1/Nu_2 < t < 1/u_1$ $P(t) \approx 1 - \exp(-u_2 t)$ [first hit is rate limiting]

Very long: $t >> 1/u_1$ No rate limiting

? "Tunnelly" of type 1 $1/\sqrt{u_2} \le N \le 1/u_1$

$P(t) \doteq 1 - e^{-Nu_1\sqrt{u_2}t}$ (The probability at least one cell with two hits has arisen before t)

Large population size ($N \ge 1/u_1$)

Type 1 are generated as soon as possible. $x_1(t) = Nu_1 t$ [Why]

Prob'y of Type 2 growth durin type 1 growth

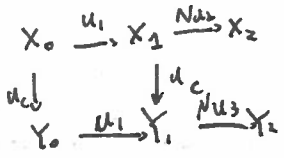
$P(t) = 1 - \exp\left\{-u_2 \int_0^t x_1(t)\,dt\right\} = 1 - \exp\left\{-\frac{1}{2} Nu_1 u_2 t^2\right\}$

D\wangyong (82.130.107.167)

Genetic instability — MIN microsatellite — CIN Chromosomal

- I one is mutated or lost (oncogene)
- II one is mutated
- III both alleles (suppressor)

Neutral CIN and $A^{+/-}$ are neutral $P = 1/N$. $A^{-/-}$ will be fixed immediately.



ODE

$$\begin{cases} \dot{x}_0 = -(u_1 + u_c)x_0 \\ \dot{x}_1 = u_1 x_0 - (u_c + Nu_2)x_1 \\ \dot{x}_2 = Nu_2 x_1 \end{cases}$$

$$\begin{aligned} \dot{Y}_0 &= u_c x_0 - u_1 Y_0 \\ \dot{Y}_1 &= u_c x_1 + u_1 Y_0 - Nu_3 Y_1 \\ \dot{Y}_2 &= Nu_3 Y_1 \end{aligned}$$

$X_2(t) \approx N u_1 u_2 t^2/2$ $\quad Y_2(t) \approx u_1 u_c t^2$ $\quad Y_1 \approx Y_2$ if $u_3 \approx 10^{-2}$ is much higher

$Y_1 \to Y_2$ is negligible.

Costly CIN. in small aparenet. CIN with fim $r < 1$ $\quad P = \frac{1-r^{-1}}{1-r^{-N}}$

non CIN → CIN with Npu_c

large aparenet complicated.

$X_2(t) = N u_1 u_2 t^2/2$

$Y_2(t) = Np u_1 u_c t^2$.

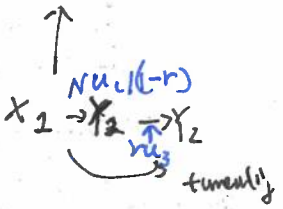
Why the difference b/w large and small compartment in CIN?

Np become so small such tha A^{++} CIN (Y_0) will NOT reach A^{++} CIN (Y_1) fixation.

$Y_2(t) = N u_1 u_2 t^2/2$

$Y_2(t) = R u_1 t^2/2$.

$R = N u_c r u_3 / 1-r$.



ch 5. Cancer progression.

In the last chapter, we know that small compart ($N\downarrow$) helps decrease the fixation probability (?) why

In colon, epithelial cells replicates in a rapid rate. Cypto combined with linear evolution

Progression A few specific mutation ⟶ adenoma ⟶ cancer

Multistage theory cancer progress in a discrete state space in a linear fashion.

$0 \xrightarrow{u_1} 1 \xrightarrow{u_2} 2 \xrightarrow{u_3} \dots \xrightarrow{u_k} k$ u_i are small ⇐ transition are rare.

$T_{01}, T_{ij} \sim Exp(u_i)$ τ_k: time to get to final state k. $\tau_k \overset{d}{=} Exp(u_1) + \dots + Exp(u_k)$ independent

$$E[\tau_k] = \sum_{j=1}^{k} E[Exp(u_j)] = \sum_{j=1}^{k} \frac{1}{u_j}$$

why? Since the independance of each transition and $\tau_k = \min\{T_{j_1} \dots T_{j_k}\}$ waiting time for k out of d.

$k=1$ $\tau_1 = \min\{T_1 \dots T_d\} \sim Exp(\lambda_1 + \dots \lambda_d) \overset{\text{indep}}{=} Exp(d\lambda)$

Assume $\lambda_i = \lambda \ \forall i$ $\tau_j \sim \tau_{j+} + Exp((d-j+1)\lambda)$ τ $\Rightarrow E[\tau_j] = \frac{1}{\lambda} \sum_{j=1}^{k} \frac{1}{d-j+1}$

Mutation pathway

Exit i : set of all possible mutations in step i Exit $1 = \{1,2,3\}$

Probability of a mutation pathway: $P = j_1 \to \dots \to j_k$

$$Prob(P) = \prod_{i=1}^{k} \frac{\lambda_{j_i}}{\sum_{j \in Exit\, i} \lambda_j}$$ competing exponentials

Expected time for P:

$$E[\tau_P] = \sum_{i=1}^{k} \frac{1}{\sum_{j \in Exit\, i} \lambda_j}$$

① → ② → ③

① ② ③ ④ ⑤ ⑥

The Wright-Fisher Process

- Haploid population of A/B of size N.
- Reproduction occurs in discrete, non-overlappty generations. ?
- $X(t)$:= number of A in size N $\in \{1, \dots, N\}$

$X(t)$ is a MC. with $X(t+1) \mid X(t) \sim Bin(N, i/N)$ Markov chain

$$P_{ij} = Prob[X(t+1)=j \mid X(t)=i] = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$$

Properties $E[X(t)] = E[X(0)]$ Variance $N i (1-p)[1-(1-p)^t]$

Absorbing state $X(t)=0$, $X(t)=N$ let $x_i = \lim_{t\to\infty} P[X(t)=N \mid X(0)=i]$ Probability of absorbed starting from i

$$i = \lim E[X(t)] = 0(1-x_i) + N x_i \Rightarrow x_i = \frac{i}{N}$$

Prob of Not fixed. Prob. of fix

Expected Fixation Time

k_i := no. of generation $k_2 \approx 2[\log N - 1 + \sigma]$

Extension of Wright-Fisher

accumulating mutation

genome
binary of length d $\{0,1\}^d$

$X_j(t)$ be the number of cells with j^{th} mutation at the t. $x_j(t) = X_j(t)/N$

At first $X_0(0) = N$ 野生型没有 mutation

Assumed fitness $(1+s)^i$

Ignore backward.

~~Probab~~

Probably of getting a j-cell [with j mutation]

$$\theta_j(t) = \sum_{i=0}^{j} P[i\text{-cell} \to j\text{-cell}] = \sum_{i=0}^{j} \underbrace{P[i\text{-cell}]}_{\downarrow} P_{i\to j}$$

$$= \sum_{i=0}^{j} \frac{(1+s)^i x_i(t)}{\sum_\ell (1+s)^\ell x_\ell(t)} \Rightarrow (1-u)^{d-j} \binom{d-i}{j-i} u^{j-i}$$

多种可能性的关系

Prob of getting a i

Binomal

Transition from $X(t+1) \leftarrow X(t)$

$P_{m,n} = P[X(t+1) = n \mid X(t) = m]$, Multi-normal dist.

generalization of binomial

mutation waves travel's at constant speed

waiting time for j-mutated cells are fixed in the population.

Assumed: wave is Gaussian. $(1+s)^i \approx 1+si$ $s \ll 1$

$$X_{(j)}(t) = A \exp\left\{-\left(\frac{j-vt}{2\sigma^2}\right)^2\right\}$$ Normal.

$A \approx \frac{1}{\sqrt{2\pi}\sigma}$, v determined by decoupling clonal expansion & generations of new types.

$$\dot{x}_j = s x_j \left[j - \sum_{i=0}^{d} i\, x_i(t)\right] \Rightarrow v = s\sigma^2$$

Adaptation

- generating a new mutant τ: average time it takes to produce a new mutant

$v = 1/\tau$ is solved by

$$X_{j+1}(\tau) = u d \int_0^\tau x_j(t)\,dt$$

- speed of adaptation

$$v \approx \frac{2s \log N}{\log[s/(ud)]^2}$$ approximated of asexual reproduction

$\propto s$

- waiting time to cancer [The for the first cells w/ k mutations]

$$\tau_k \propto k/s$$

$$\approx \frac{k \log[s/(ud)]^2}{2s \log N}$$

cnt'd Diffusion theory

Selection → dominate in large pop Drift in small pop.

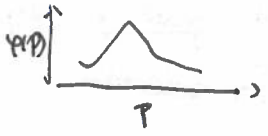
Probability of a population over time / instead of paths.

$\varphi(p,t)$:= a certain allele is at frequency p at time t

Evolutionary Process

Directional process $M(p)$: Non-zero expected change in allele frequency. selection / migration

Non-directional process $V(p)$: Zero expected change. $\varphi(p;t)$ is random



Ex. $V(p) = E[\mathrm{Var}\ p(t+1)/p(t)] = \frac{2p(1-p)}{N_e}$

let g: probability that frequency changes from $p \to p+\varepsilon$ in time interval dt ← drift / Expected variance in the next gen. distn.

$g(p,\varepsilon;dt)$

$$\varphi(p,t+dt) = \int \varphi(p-\varepsilon,t)\, g(p-\varepsilon,\varepsilon;dt)\, d\varepsilon$$

$$\text{(Taylor expansion)} = \int \left[\varphi(p,t)\, g(p,\varepsilon;dt) - \varepsilon \frac{\partial(\varphi g)}{\partial p} + \frac{\varepsilon^2}{2} \frac{\partial^2}{\partial p}(\varphi g) \right] d\varepsilon$$

Approximation

$$\varphi(p,t+dt) = \varphi(p,t) \underbrace{\int g(p,\varepsilon;dt)\, d\varepsilon}_{\Downarrow 1} - \frac{\partial}{\partial p}\varphi \underbrace{\int g\varepsilon\, d\varepsilon}_{\Downarrow} + \frac{1}{2}\frac{\partial^2}{\partial p^2}\varphi \int g\varepsilon^2 d\varepsilon$$

$$= \varphi - \frac{\partial}{\partial p}\varphi \underbrace{\int g\varepsilon\, d\varepsilon}_{\Downarrow} + \frac{1}{2}\frac{\partial}{\partial p^2} \underbrace{\int g\varepsilon^2 d\varepsilon}$$

$\Downarrow$ $E(\varepsilon)$ $E(\varepsilon^2) \propto \mathrm{Var}(\varepsilon)$

$$\Rightarrow \varphi(p,t+dt) = \varphi(p,t) - \frac{\partial}{\partial p}[\varphi(p,t) M_p]\, dt + \frac{1}{2}\frac{\partial^2}{\partial p^2}[p_p(t) V(p)\, dt)$$

$$\Rightarrow \frac{\partial \varphi_{p(t+dt)}}{\partial t} = -\frac{\partial}{\partial p}[\varphi\ M] + \frac{1}{2}\frac{\partial^2}{\partial p^2}[\varphi\ V]$$

Change in allele frequency in terms of { shape of dist. M, V

Equilibrium $\frac{\partial}{\partial t} = 0 = -\frac{\partial}{\partial p}\varphi M + \frac{1}{2}\frac{\partial^2}{\partial p}\varphi V$

ODE $\frac{1}{2}\frac{\partial}{\partial p}\varphi V - \varphi M = 0$ after integrate over p

$$\Rightarrow \varphi^*(p^*) = \frac{c}{V(p)} \exp\left[\int_0^p \frac{2M(q)}{V(q)}\, dq. \right]$$

↳ accounts for Non-directional forces

Selective competition b/w 2 alleles. A_1 A_2 p / $(1-p)$ fitness w_1 / w_2.

$\bar{w} = p w_1 + (1-p) w_2$ $\quad \frac{d\bar{w}}{dp} = w_1 - w_2$

In the next gen. $p' = \frac{p w_1}{\bar{w}}$

Wright's equation for adaptive landscape $\quad \Delta p_{sel} = p' - p = \frac{p(w_1 - \bar{w})}{\bar{w}} = \frac{p(1-p)(w_1 - w_2)}{\bar{w}} = p(1-p)\frac{d \ln \bar{w}}{dp}$

For selection

Mutation u_1 : rate from A_1 to A_2 $\quad u_2$: rate from $A_2 \to A_1$

per-generation change due to mutation

Ch7. Evolutionary Game Theory

※ Fitness may change according to the dynamics of the population. [Frequency-dependent]

$$\begin{cases} \dot{x}_A = x_A[f_A(x) - \phi(x)] \\ \dot{x}_B = x_B[f_B(x) - \phi(x)] \end{cases} \iff \dot{x} = x(1-x)[f_A(x) - f_B(x)]$$

x is the frequency

$x^* = 0$ if $f_A(0) < f_B(0)$

$x^* = 1$ if $f_A(1) > f_B(1)$

a third equilibrium is possible if $\frac{df_A}{dx}(x^*) < \frac{df_B}{dx}(x^*)$

Evolutionary Games

Pay-off matrix

	A	B
A	a	b
B	c	d

against A

row: expected pay-off

- fitness as expected pay-off $f_A(x_A, x_B) = a x_A + b x_B$

$f_B(\quad) = c \cdot + d \cdot$

let $x_B = 1 - x \Rightarrow$

$$\dot{x} = x(1-x)[(a-b-c+d)x + b - d]$$

See page 11

Nash equilibrium

If two players play the same and no one increase by changing strategy.

→ Nash equilibrium. [best reply to itself] 对手选择自己就会最好的.

A is strict NE if $a > c$ [选A更好] Similarly B is NE if $d > b$

NE $a \geq c$ $d \geq b$

Evolutionary stable strategy

If ε of B invades a All-A population

selection will oppose

If $\begin{cases} f_A(1-\varepsilon) > f_B(\varepsilon) \\ a(1-\varepsilon) + b\varepsilon > c(1-\varepsilon) + d\varepsilon \end{cases}$ → 对手选择一点点 B(ε) 但总选择 A 一定会更好.

if $\varepsilon \to 0$

① $a > c$ or ($a = c$, $b > d$)

② ESS

Hierarchy of Nash

for all $i \neq k$.

Unbeatable	Strict	ESS	Nash
$a_{kk} > a_{ik}$ 一样好 $a_{kk} > a_{ik}$.	$a_{kk} > a_{ik}$	$a_{kk} > a_{ik}$	$a_{kk} \geq a_{ik}$
$a_{ki} > a_{ii}$	or	or	
选下家一样都赢	$a_{kk} = a_{ik}$, $a_{ki} > a_{ii}$		

$\phi(x) = \sum f_i x_i$

$f_i = \sum_j x_j a_{ij}$

Replicator equation

$\dot{x}_i = x_i[f_i(x) - \phi(x)]$

kind of like the RPC game

$$A = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \quad \phi() = 0$$

Case 1: $\det(A) > 0 \Rightarrow$ unique stable equilibrium

Case 2: $\det(A) < 0 \Rightarrow$ unique unstable equilibrium

$= 0$ neutral oscillation

通过 A 的 determinant 判断是否稳定

The prisoner's dilemma

silent / confess $\begin{pmatrix} S & C \\ 0 & -7 \end{pmatrix}$

example $\begin{matrix} & C & D \\ C & 3 & 0 \\ D & 5 & 1 \end{matrix}$ $\quad f_C < f_D \Rightarrow$ D dominating

$\begin{matrix} & C & D \\ C & R & S \\ D & T & P \end{matrix}$ $\quad T > R > P > S \quad R > \frac{T+P}{2}$

R reward P punishment

TFT Repeat opponent last round: (m : number of Rounds)

TFT vs ALLD. $\begin{matrix} & TFT \\ ALLD & \begin{pmatrix} mR & S+(m-1)P \\ T+(m-1)P & mP \end{pmatrix} \end{matrix}$

TFT : never first to defect

Not greedy.

If $m > (T-P)(R-P)$ resist invasion

GRIM : Cooperate at first as long as opponent does not defect

$\begin{matrix} & GRIM \\ & \begin{pmatrix} mR & S+(m-1)P \\ T+(m-1)P & mP \end{pmatrix} \end{matrix}$

Noise in TFT

long run is as long as choosing randomly b/c C and D

TFT vs ALLC $\begin{pmatrix} mR & mR \\ mR & mR \end{pmatrix}$ TFT Not evolutionary stable $\Rightarrow$ in the long run TFT can become ALLC

PD as MC b/w S_1 S_2 is a Markov chain {CC, DD, DC, DD}.

M is the transition matrix for it

$X(t)$ be the prob. {CC, . . DD} at the t^{th} games $\quad X(t+1) = X(t) \cdot M$

May reach stationary.

$E[S_1, S_2] = RS_1S_2 + S\,S_1(1-S_2) + T(1-S_1)S_2 + P\,(1-S_1)(1-S_2)$

CTFT $S(1, 1/3)$ $f_D \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}$

Memory one : affected by Opponent's and own last move

TFT = S(1,0,1,0)

WSLS = S(1,0,0,1) [cooperate after CC or DD, defect after CD or DC]

ch 8 Evolutionary Games in finite populations.

Moran Process

$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ A B

A: i individuals.

Pay-off. A: $F_i = \frac{(i-1)a + (N-i)b}{N-1}$ B: $G_i = \frac{ic + (N-i-1)d}{N-1}$

selection opposes A invading B if $F_i < G_i \iff b(N-1) < c + d(N-2)$ why "<"

拒绝 invade selection

Intensity of selection

$f_i = 1 - w + wF_i$ → Expected pay-off

$g_i = 1 - w + wG_i$

$w = 0$ 也叫 weak selection Game does not contribute.

$w = 1$, 全取决 Game 收益

Fixation Probability

For a moran $P_{i,i+1}$ $\frac{if_i}{if_i + (N-i)g_i}$ $\frac{N-i}{N}$.

Fixation let $\rho_A = \frac{1}{1 + \sum_{k=1}^{N-1} \prod_{i=1}^{k} (g_i/f_i)}$ 得 $\frac{P_{i,i-1}}{P_{i,i+1}} = \frac{g_i}{f_i}$.

weak selection limit.

$\rho_A \approx \frac{1}{N} \frac{1}{1 - (\alpha N - \beta) w/6}$

$\alpha = a + 2b - c - 2d$

$\beta = 2a + b - c - 4d$.

If $\rho_A > \frac{1}{N}$ selection favors the fixation of A

$\alpha N > \beta$

$\alpha(N-2) + b(2N-1) > c(N+1) + d(2N-4)$

当 $N = 2$ $b > c$

$N \to \infty$ ↑

$a + 2b > c + 2d$

Equilibrium $F_i = G_i$

$x^* = \frac{d-b}{a-b-c+d}$.

如果 $a > c$ $d > b$,

A, B, ⇓ ⇓ 是最好的策略.

A B

$\rho_A > 1/N$ Favors

$\iff$ $x^* < \frac{1}{3}$

B is ESS if

1. $b(N-1) < c + d(N-2)$ Expect ($F_1 < G$)

2. $p_A < 1/N$ for $w > 0$.

$$a(N-2) + b(2N-1) < c(N+1) + d(2N-4)$$

Evolutionary Graph Theory.

Moran process 동일한 Full matrix 'Every one can be another / die and reproduce
However it does not reflect the spatial connection b/w

$$\rho = \frac{1 - 1/r}{1 - 1/r^N}$$

Directed cycle

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

A: 1 B: r. Taking # of B

$$P_{m-1} = \frac{1}{N - m + rm}$$

A 1명 B 2명

$$P_{m+1} = \frac{r}{N + m + rm}$$

same as Moran

$$\rho = \frac{1 - 1/r}{1 - 1/r^r}$$

Bivalent cycle 동일

linear

$$\rho = \frac{1}{N}$$ independent of r.

G is amplifier $\rho_G > \rho_{neu}$ strongest suppress $\frac{1}{N}$ 보다 r

suppress $\rho_G < \rho_{neu}$

Isothermal Thm

$\rho_G = \rho_{neu}$ iff G is isothermal

Temp: $T_j = w_{1j} + \dots w_{Nj}$

$J_i = T_j \quad \forall i, j$

Prof $V = (v_1 \dots v_N)$ $v_i = \begin{cases} 0, & \text{occupied by A} \\ 1 & \text{occupied by B} \end{cases}$ $m = \sum v_i$

$$P_{m,m+1} = \frac{r \sum_{ij} w_{ij} v_i (1 - v_j)}{rm + N - m}$$

occupied by B, occupied A

NA → B

$$P_{m,m-1} = \frac{\sum w_{ij} (1 - v_i) v_j}{rm + N - m}$$

iff $\frac{P_{m-1}}{P_{m,m+1}} = \frac{1}{r}$ for all v. $\Rightarrow$ $\sum w_{ij} (1 - v_i) v_j = \sum w_{ij} v_i (1 - v_j)$

$\Rightarrow$ $\sum_j w_{kj} = \sum_i w_{ik} = 1$

CH7

Cellular automata

Gridspace with each grid can take one of the states

Eden growth

2 state.

S_0, S_1 occupied.

Rule. a site in the neighbor of S_1 switched to S_1.

Variants:
- Aviable: 有 S_1 的附近的 S_0 中随意选一个
- Bond: $S_0 \to S_1$ 与 no. of S_1
- Cell-for: 选一个 S_1 旁的 S_0，再随机选择 S_0

Deme-based models

每个 Deme 有很多的细胞. WF-Prob Moran process

cells can migrate from Deme to another

当 $N_{deme} = 1$ ⇓ stochastic

⇓ 0 non-spatial.

Spatial Moran

n = # of mutation vector

+1 $W_i^+(n) = \frac{\mu(1+s)}{N}(N-n_i)\left[n_i + \frac{m}{2}n_i''\right]$ 这里不是再除以一个N吗?

-1 $W_i^-(n) = \frac{\mu}{N} n_i \left[(N-n_i) - \frac{m}{2}n_i''\right]$

μ death rate

s relative diff. in fitness

$n_i'' = (n_{i-1} + n_{i+1} - 2n_i)$

Ch10. Branching process in biology.

Galton-Watson Process

After a time, produce Random no. of offspring. Z according to a fixed distribution.

$\{Z_n \mid n = 0 \dots\}$ is a MC.

Z_1 Z_2

$$Z_{n+1} = \sum_{j=1}^{Z_1} Z_n^{(j)}$$ 从 z_1 个过来的个体

Time-homogeneous

Transition Probability

$p_k = \text{Prob}(Z = k)$

P_{ij} transition from $i \to j$

$$P_{ij} = \sum_{k_1 + \dots + k_i = j} p_{k_1} \cdots p_{k_i}$$

$p_k = P(1,k)$

e.g. $P(2,j) = p_0 p_j + p_1 p_{j-1} + \dots + p_j p_0$

就是把所有的排列组合加.

Probability Generating Function Discrete RV $Z \sim \{p_k\}_{k \geq 0}$.

$$f(s) = \mathbb{E}[s^Z] = \sum_{k=0}^{\infty} p_k s^k. \quad s \in [0,1]$$

generating distribution on P 从 k 个值的到 1.

$$\frac{d^k f}{ds^k}(0) = k!\, p_k \quad k \geq 0$$

Properties: $\mathbb{E}[Z] = f'(1)$ ($\sum k p_k$.) $\quad \text{Var}[Z] = f'(1) + f''(1) - f'(1)^2$.

Iteration Def $f^{(0)}(s) = s \quad f^{(1)}(s) = f(s) \quad f^{(n+1)}(s) = f \circ f^{(n)}(s)$

$f(s) = \sum_j P(1,j) s^j$

$(f(s))^k = \sum_j P(k,j) s^j$ 从 k 开始

f_n be the pgf of Z_n.

$P_n(i,j)$ be the n-step transition Probabilities.

The CK equations $$P_{n+m}(i,j) = \sum_{k=0}^{\infty} P_n(i,k) P_m(k,j)$$

$f_n = f^{(n)}$

$$f_{n+1}(s) = \sum_j P_{n+1}(1,j) s^j = \sum_j \sum_k P_n(1,k) P(k,j) s^j = \sum_k P_n(1,k) \sum_j P(k,j) s^j$$

$$= \sum P_n(1,k)\, f(s)^k.$$

$$= f_n(f(s))$$

$$= f^{(n+1)}(s)$$ deduction

$P_{n+1}(1,j)$ means

Moment of Z_n

Assume $p_0 + p_1 < 1$ See Exercise

$$\mathbb{E}[Z_n] = m^n$$

$$\text{Var}[Z_n] = \begin{cases} \sigma^2 m^{n-1}(m^n - 1)/(m-1) & m \neq 1 \\ n\sigma^2 & m = 1 \end{cases}$$

Extinction $Z_n = 0$ Absorbing state

$P = \text{Prob}(Z_i = 0 \text{ for } i \geq 0) = \lim_{n\to\infty} \text{Prob}[Z_i = 0] = \lim_{n\to\infty} f_n(0) = \lim_{n\to\infty} f^{(n)}(0)$

Properties of pgf Power series non negative coefficients adding up to 1

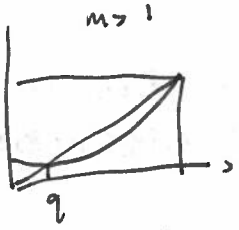
i f is strictly convex

ii $f(0) = p_0 \quad f(1) = 1$

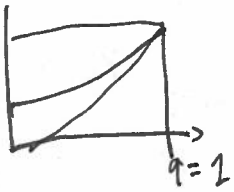
iii if $f'(1) = m \leq 1$ then $f(s) > s$ for $s \in [0,1)$ $\quad m = f'(1) < 1$

iv if $f(1) = m > 1$ $\quad f(s) = s$ has a unique root in $[0,1)$

$m > 1$



$m \leq 1$



Extinction Probability

Thm The extinction prob. of GW process is the smallest non-negative root of $f(s) = s$

If $m \leq 1$ $q = 1$ if $m > 1$ $q < 1$

With Prob 1 $Z_n \to 0$ or $Z_n \to \infty$ as $n \to \infty$

Instability

$\text{Prob}(\lim Z_n = 0) = q$

$\text{Prob}(\lim Z_n = \infty) = 1 - q$

supercritical	$m > 1$	$E[Z_n] \to \infty$	$q < 1$
critical	$m = 1$	$E[Z_n] = 1$	$q = 1$
subc-	$m < 1$	$E[Z_n] \to 0$	$q = 1$

Multiple Types 2 types 0 wT 1 mutant. $Z_0(t)$ $Z_1(t)$

Type 0 $\xrightarrow{\alpha}$ Type 1

irreversible

$F = F_0, F_1$ $\quad F_0(s_0, s_1; t) = E[s_0^{Z_0(t)} s_1^{Z_1(t)} \mid Z_0(0) = 1, Z_1(0) = 0)$

$\cdot F_1 \quad \mid Z_0(0) = 0, Z_1(0) = 1]$

Recurrence equations

$F_0(s;t) = [(1-\alpha) F_0(s;t-1) + \underline{\alpha} F_1(s;t-1)]^2$ mutation

$F_1(s;t) = [F_1(s;t-1)]^2$

Differentiation at $s = (1,1)$

$E[Z_0(t) \mid Z_i(0) = \delta_{0i}] = [2(1-\alpha)]^t$

$E[Z_0(t) \mid Z_i(0) = \delta_{1i}] = 2E[Z_0(t-1) \mid Z_i(0) = \delta_{1i}] = 0$

$E[Z_0(t) \mid Z_i(0) = \delta_{0i}] = 2(1-\alpha) E[Z_0(t-1) \mid \times]$

$\Rightarrow \quad N(t) = E[Z_0(t) + Z_1(t) = Z_i(0) = \delta_{0i}] = 2^t.$

$r(t) = E[Z_1(t) \mid Z_i(0) = \delta_{0i}] = 2^t - (2(1-\alpha))^t = 2^t[1 - (1-\alpha)^t]$

Prob. of a mutant type population

$$P_0(t) = F_0(1,0;t) = E\left[1^{z_0(t)} 0^{z_1(t)} \mid z_1(0) = \delta_{0,i}\right]$$

$$0^{z_1(t)} = \begin{cases} 1 & \text{if } z_1(t) = 0 \\ 0 & \end{cases}$$

Ch 11 Evolutionary escape. Reaching 1 before extinction

Simple ex in binary genome 1: mutant 0: WT

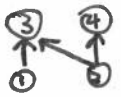
Partially ordered sets (Posets) : a set Σ together with a binary relation, denoted "$\leq$"

Properties
- reflexive : for all $e \in \Sigma$, $e \leq e$
- antisymmetric : for all $e_1, e_2 \in \Sigma$ $(e_1 \leq e_2$ and $e_2 \leq e_1) \Rightarrow e_1 = e_2$
- transitive : $e_1 \leq e_2$ and $e_2 \leq e_3 \Rightarrow e_1 \leq e_3$

If $e_1 \neq e_2 \Rightarrow e_1 \leq e_2 \Rightarrow e_1 < e_2$

$e_1 < e_2$ is a cover relation.

Example:



$1 \leq 3$
$2 \leq 3$
$2 \leq 4$
$1 \leq 5$

Distributive lattice

A order ideal g in a poset Σ is a subset of Σ that is closed downward
if $e_2 \in g$ and $e_1 \leq e_2$, then $e_1 \in g$

The distributive lattice is the set of all order ideal, under inclusion $J(\Sigma)$ 都可以找到 sup, inf 对于每一对

- $J(\Sigma)$ forms a poset
- every pair of g_1, g_2 has a unique infimum and unique supremum.

The genotype lattice

- Σ be a set of $n = |\Sigma|$ irreversible genetic events (e.g. point mutation at different loci, indels, genomics)
- The poset indicates that the mutation can be accumulate.
- the order ideals g of $J(\Sigma)$ are precisely the subset of Σ that are compatible with the partial order.
  [genotypes that can evolve subject to the order constraints]

Empty poset
- $\Sigma = \{1 \dots n\}$ (no relation)
- $G = J(\Sigma)$ is the set of all subsets of Σ / Boolean lattice of $\{0,1\}^n$
- All genotypes can occur in any order ; there are no order constraints.

Chain A chain in $G = (J(\Sigma), \subseteq)$ of length k is a totally ordered subsets $g_1 \subset g_2 \dots \subset g_k$. how to compare b/w g_i g_j

[Mutation pathways in cancer]

Fitness landscape $f : G \to \mathbb{R}$

Mutational neighbourhood : The mutational neighbourhood of a genotype $g \in G$ is the set of genotypes that can be reached by mutation.

$$N(g) = \{h \in G \mid g \subset h\}$$ can be

$g \subset h$ 表示 $g \to h$

k-steps offspring Set $F = \mathrm{diag}(f)$ f is the fitness F is a diagonal matrix $(n \times n)$

Mutation matrix $n = |G|$ mutation matrix $U = (U_{gh})_{g,h \in G}$ by $\begin{cases} \pi_{e \in h \setminus g} \mu_e & \text{if } h \in N(g) \\ 0 & \text{otherwise} \end{cases}$

Entry i,j : $i \to j$ 多个突变 mutation 的概率

the entry (g,h) of $U \cdot F$ is the probability of genotype producing in one step. $UF = F \cdot \mathrm{diag}(f)$

(g,h) of $(UF)^k$ k step. producing h along the way

$$B = (I - UF)^{-1} - I = UF + (UF)^2 + \dots + (UF)^n$$

注 $g \subset h$ 能不能在 $\{1, \dots n\}$ 当中走到

Properties The entry b_{gh} of B is zero, unless $g \subset h$ $g \to h$

If $g \subset h$ then : $b_{gh} = \underbrace{\mu_{gh}}_{\text{mutation}} \underbrace{f(h)}_{\text{fit}} \underbrace{P_{gh}(f)}_{\text{poly.}}$ where P_{gh} is a polynomial function of degree $|h \setminus g| - 1$ on $\mathbb{R}^1$

P_{gh} is the generating function for all chains from g to h in G

The risk polynomial $f(g) = f_g \quad g \in G$. case $g=0, h=1$

$$Pos(f) = \sum_{0=g_0<g_1\cdots g_k=1} f_{g_1}\cdots f_{g_{k-1}}$$ "the sum over all chains from 0 to 1 in the genotype lattice G"

从 0 -> 1 all possible combination.

$R(G;f) = Pos(f)$ is the risk polynomial [Revise chain]

Invasion Let R_g be the basic reproductive ratio of an invading pathogen.

Interested in the case where $R_1 > 1$ and $R_g < 1$ for all $g \neq 1$

Fitness landscape $f_g = \frac{R_g}{1-R_g} = R_g + R_g^2 + R_g^3 + \cdots$

Then $R_g = \frac{f_g}{1+f_g}$ $\quad f_g \approx R_g$ for $g \neq 1$

Multitype Branching Process

Eg G w/ a Poisson offspring distribution.

$$p_{gh}^k = Pois(k; u_{gh}R_g) = \frac{(u_{gh}R_g)^k}{k!} e^{-u_{gh}R_g}.$$

p_{gh}^k $g \to h$ 有 k 个儿子

The risk of escape

Let ε_g be the probability of escape (reaching 1 before extinction) starting w/ one individual of type g.

$\varepsilon_g :=$ Probab. reach 1 [escape] $\quad 1-\varepsilon_g :=$ Probab. of extinction

$$1-\varepsilon_g = \prod_{h\geq g} \sum_{k=0}^{\infty} (1-\varepsilon_h)^k p_{gh}^k$$ 所有项都 0 u_{gh} is defined as before $u_{gg}=1$

Extinction

Recurrence equation substitutting the Poisson distr. 运用 $F(\cdot)=1$ For $g \neq 1$ $\varepsilon_g << 1$ and $R_g^2 \approx 0$

what does k mean.

$$\log(1-\varepsilon_g) = -\sum_{h\geq g} \varepsilon_h u_{gh} R_g. \qquad \varepsilon_g \approx R_g \sum_{h\geq g} \varepsilon_h u_{gh}.$$

$$\approx \frac{R_g}{1-R_g} \sum_{h>g} \varepsilon_g u_{gh} = f_g \sum_{h>g} \varepsilon_h u_{gh}$$

In particular, $\varepsilon_0 = f_0 \sum_{h\in G} \varepsilon_h u_{0h}$.

$$\Rightarrow \varepsilon_0 \approx \varepsilon_1 f_0 \prod_{e\in\varepsilon} \mu_e \quad R(G;f).$$ How to solve this recursion

The risk of escape of N WT pathogens is

$$1-(1-\varepsilon_0)^N \approx 1-e^{-N\varepsilon_0}$$ 假设 ε_0 已经算出.

The critical population size $N^* = 1/\varepsilon_0$

$\gg N^*$ escape is almost certain.

$= N^*$ risk of escape $1-1/e$

$1/e :=$ prob. of sucessful intervention

$\ll N^*$. escape is almost impossible.

Example: Anti-cancer therapy.

Cancer cells can die X chemotherapy by TSG inactivation.

We assume that at the beginning of therapy all cells have TSG.

Inactivation WT TSP^{++} —2u→ TSP^{+-} —u→ TSP^{--} esc

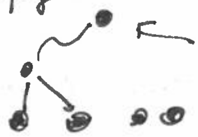
TSP^{++} —2p→ TSP^{+-} —u→ TSP^{--} esc

TSP^{+-} —p→ TSP^{--}

The coalescent theory

The coalescent — WF Process

To get into geno that gives rise to the Final offspring.



Branching time (looking back into root)

Branch time is the time where linege break off

[coalescent event]

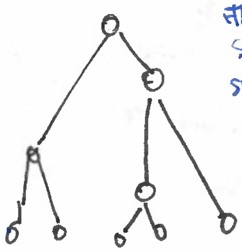
Coalescent time

Time for linege for multiple to coale. / come together.

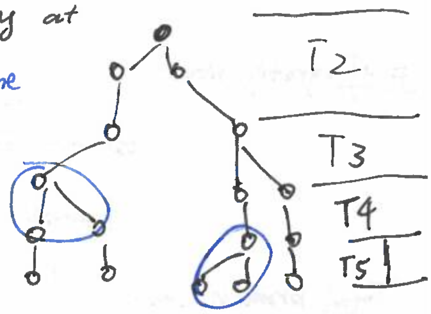
In principle you can have multiple of these events happening at the same time.

Interested in branch length

Every vector is a coalscent event.



then may can be simplified as such



The prob of j genes has no common ancestor in previous gen. 完全不同源于 n^{th} 代

上一代不相仓

$j=2 \quad (1-\frac{1}{N})$ 选 3 个 选择

$j=3 \quad (1-\frac{1}{N})(1-\frac{1}{N}-\frac{1}{N})$

$$\Rightarrow \prod_{i=1}^{j-1}(1-\frac{i}{N}) = 1-\binom{j}{2}N^{-1}+O(N^{-2})$$

Coalescent time — let $T(j)$ be coalescent time b/w j and $j-1$ genes. 记合并时间

$$\mathbb{P}[T(j)>t] = \left[\prod_{i=1}^{j-1}(1-\frac{i}{N})\right]^{Nt} \to \exp\{-\binom{j}{2}t\} \text{ as } N\to\infty$$

指数分布

只看 pairwise coalescent process 多久该有 coalsent expl dist

MRCA $\quad T_{MRCA}(5) = T(5)+\cdots+T(2) = \sum_{j=2}^{n} T(j)$

How old is the tumor.

Expectation — 指数函数

$$\mathbb{E}[T_{MRCA}(n)] = \sum_{j=2}^{n}\mathbb{E}[T(j)] = \sum_{j=2}^{n}\frac{2}{j(j-1)} = 2\sum_{j=2}^{n}\frac{1}{j-1}-\frac{1}{j}$$

$$= 2(1-\frac{1}{n})$$

$\mathbb{E}[T(2)]=1$

2 generation?

Variance

$$Var(T_{MRCA}(n)) = \sum_{j=2}^{n} Var(T(j)) = \sum_{j=2}^{n}\left(\frac{2}{j(j-1)}\right)^2 = 4\sum\left(\frac{1}{j-1}-\frac{1}{j}\right)^2$$ 没有表达式

$=$ whn $\to$ 1.16.

$Var(T(2))=1 \Rightarrow$

T_{MRCA} is dominated by $T(2)$ 等一次分支最重要

Huge Variation b/w any two coalescent event $T(2)$

Mutation Poisson Process puts down mutation at $\frac{\theta}{2} = \mu n$ 在 Branch.

Infinite sites model. ○ 有很长的 gene
- infinite # of alleles.
- Each Mutation will create a new mutation

每次 mutation 都在新的 loci

# of segrating sites. 大概每次不同

S equals the total # of mutations of the genealogy

Number of 独 colume that are not identical.

Total branch lon. Under Infinite Site Model

$$T_{tot}(n) = \sum_{j=2}^{n} j\,T(j)$$

$$E[S] = \frac{\theta}{2} E[T_{tot}(n)] = \frac{\theta}{2} \sum j \frac{1}{\binom{j}{2}} = \theta \sum_{j=2}^{n} \frac{1}{j-1} = \theta\,(C_n) \leq \ln n + 1$$

↑ mutation rate

Average pairwise nucleotide distance K

Expected value of

$$E[K] = \frac{\theta}{2} 2 E[T(2)] = \theta$$

↑ parallel lines

	2	3	4	5
1				
2				
3				
4				

Hamming distance

Detecting selection

Under neutral infinite size model.

$E[K] = \theta = C_n^{-1} E[S]$ ← sensitive to low-allele frequencies

↑ strongly affected by allele frequencies

Tanjima's D

$$D = \frac{\hat{K} - C_n^{-1}\hat{S}}{\sqrt{\hat{V}}}$$

$\hat{V}$ ← Variance

Ho: no selection

=> To check whether it's outlined.

This is essentially a Random tree generated