Evolutionary Dynamics Homework 4

Minghang Li

October 24, 2024 02:10 (America/New_York)

This solution is organized in R Markdown.

Problem 2: TSG inactivation for medium and large populations

> Consider two models for TSG inactivation: One for a medium-sized population with population size $1/\sqrt{u_2} \ll N \ll 1/u_1$ and one for a large-sized population with population size $N \gg 1/u_1$. Here u_1 and u_2 denote the mutation rates for inactivating the first and second allele.

(a)

> $P(t)$ is the probability that at least one cell with two hits has arisen before time t. Calculate the time $t_{0.5}$ such that $P(t_{0.5}) = 0.5$ for the cases $N = 10^5$ and $N = 10^{11}$ using a suitable model for each population size. The mutation rates are $u_1 = 10^{-8}$ and $u_2 = 10^{-4}$.

In the problem setting, we have:

- $1/\sqrt{u_2} = 1/\sqrt{10^{-4}} = 10^2$
- $1/u_1 = 1/10^{-8} = 10^8$

$N_1 = 10^5$

Here we have $1/\sqrt{u_2} = 10^2 \ll N_1 \ll 1/u_1 = 10^8$.

In this parameter region, we have

$$P(t) = 1 - \exp\left(-Nu_1\sqrt{u_2}t\right)$$

We want to find $t_{0.5}$ such that $P(t_{0.5}) = 0.5$.

$$0.5 = 1 - \exp\left(-Nu_1\sqrt{u_2}t_{0.5}\right) \Longrightarrow t_{0.5} = \frac{\log(2)}{Nu_1\sqrt{u_2}}$$

$N_2 = 10^{11}$

Here we have $N_2 \gg 1/u_1 = 10^8$.

In this parameter region, we have

$$P(t) = 1 - \exp\left(-\frac{1}{2}Nu_1u_2t^2\right)$$

Using similar trick:

$$0.5 = 1 - \exp\left(-\frac{1}{2} N u_1 u_2 t_{0.5}^2\right) \Longrightarrow t_{0.5} = \sqrt{\frac{2\log(2)}{N u_1 u_2}}$$

(b)

In this context of TSG inactivation briefly explain the concept of rate limiting events for overall cancer progression and why the number of these events depends on the population size.

In the context of TSG inactivation, the rate limiting event translates to "mutation in the TSG allele and the propagation of mutation in the population". Whether the mutation accumulation is rate limiting depends on the population size.

The number of events changed because $P(t)$ (probability that at least one cell with two hits has arisen before time t) has different formula for different population size.

For small population size, type 1 cell (with 1 allele mutated) reaches fixation before type 2 (two alleles both mutated) cell arises, indicating we need to wait for two event: 1. type 1 mutation to occur, 2. type 1 mutation reach fixation.

For intermediate size, type 2 cel is generated before fixation of the type 1 lineage, but the average waiting time for a type 1 cell to occur ($1/(Nu_1)$) is still rather long. Hence we have 1 rate-limiting event.

For large population size, type 1 cells are generated "immediately" (no "wait" before any type 1 cell occurs) and grow according to $X_1(t) = Nu_1 t$ and the type 2 cell is produced before it reaches fixation. Hence we have no rate-limiting event.

Problem 3: Chromosomal instability

For TSG inactivation with chromosomal instability (CIN) we have distinguished three cases: *neutral CIN*, *costly CIN in small compartments*, and *costly CIN in large compartments*. Cancer can thus either arise from two subsequent "normal" mutations, or by a normal mutation followed by CIN.

Without CIN, the rate of the second mutation can be expressed as

$$u_2 = u + p_0 \approx 2u,$$

where $u \approx 10^{-7}$ is the normal point mutation rate per cell division and p_0 the rate of LOH. With CIN, the rate of the second mutation is approximately $u_3 = 0.01$.

Assuming there are n_1 class I and n_2 class II CIN genes, the rate u_c of producing a CIN mutant can be expressed as

$$u_c = 2n_1(u + p_0) + 2n_2 u \approx 2(2n_1 + n_2)u.$$

(a)

Calculate the ratio C of cancers that are initiated with CIN to those initiated without CIN for the three CIN scenarios and show that C is independent of time.

For all derivations below, X_i denotes the *probability* of i state *without* CIN, and Y_i denotes the *probability* of i state *with* CIN.

Neutral CIN

The ODE system of state probabilities for X_i, $i \in 0, 1, 2$:

$$\begin{aligned}
\dot{X}_0 &= -(u_1 + u_c)X_0 \\
\dot{X}_1 &= u_1 X_0 - (u_c + Nu_2)X_1 \\
\dot{X}_2 &= Nu_2 X_1
\end{aligned}$$

For Y_i, $i \in 0, 1, 2$:

$$\begin{aligned}
\dot{Y}_0 &= u_c X_0 - u_1 Y_0 \\
\dot{Y}_1 &= u_c X_1 + u_1 Y_0 - Nu_3 Y_1 \\
\dot{Y}_2 &= Nu_3 Y_1
\end{aligned}$$

On the relevant time scale, the approximate solution of the system (we're only interested in X_2 and Y_2 as they are the tumor states) is:

$$X_2(t) \approx \frac{Nu_1u_2t^2}{2}, \qquad Y_2(t) = u_1u_ct^2$$

Then the ratio C is calculated as:

$$C = \frac{NY_2(t)}{NX_2(t)} = \frac{2u_c}{Nu_2} = \frac{2 \cdot 2(2n_1 + n_2)u}{2Nu} = \frac{2(2n_1 + n_2)}{N}$$

Costly CIN in small compartments

By costly we define CIN cells as having fitness $r < 1$, and their fixation probability ρ in a Moran process is:

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

the non-CIN-to-CIN transition rate is $N\rho u_c$.

Now we have:

$$X_2(t) = \frac{Nu_1u_2t^2}{2}, \qquad Y_2(t) = N\rho u_1u_ct^2$$

Then the ratio C is calculated as:

$$C = \frac{NY_2(t)}{NX_2(t)} = \frac{2\rho u_c}{u_2} = \frac{2 \cdot 2(2n_1 + n_2)u \cdot \rho}{2u} = 2\rho(2n_1 + n_2)$$

Costly CIN in large compartments

For large N and $r < 1$ the fixation will not be reached. The population will "tunnel" from X_1 to Y_2 at a rate of:

$$R = \frac{Nu_cru_3}{1 - r}$$

Then we have:

$$X_2(t) = \frac{Nu_1u_2t^2}{2}, \qquad Y_2(t) = Ru_1t^2/2$$

Then the ratio C is calculated as:

$$C = \frac{NY_2(t)}{NX_2(t)} = \frac{R}{Nu_2} = \frac{u_c r u_3}{u_2(1-r)} = \frac{2(2n_1+n_2)u_3 r}{2(1-r)}$$

Are they independent of time?

Well of course they are independent of time as the t^2 always cancels out in the ratio.

(b)

> Explicitly compute C for $n_1 = 6$, $n_2 = 4$ and $N = 10$ cells, in the case of neutral CIN and costly CIN in small compartments, and $N = 200$ for large compartments. Assume a relative fitness disadvantage of $r = 0.9$ in the costly cases. Compare the ratios in a few sentences.

I'm so bad at calculations so I'll use code blocks.

```
n1 <- 6
n2 <- 4
N  <- 10
N_l<- 200
r  <- 0.9
u3 <- 0.01
```

Neutral CIN

```
C.neutral <- 2 * (2 * n1 + n2) / N
```

The ratio C for neutral CIN is 3.2.

Costly CIN in small compartments

```
rho <- (1 - 1 / r) / (1 - 1 / r^N)
C.small <- 2 * rho * (2 * n1 + n2)
```

The ratio C for costly CIN in small compartments is 1.903.

Costly CIN in large compartments

```
C.large <- (2 * (2 * n1 + n2) * u3 * r) / (2 * (1 - r))
```

The ratio C for costly CIN in large compartments is 1.44.

Comparison

Generally we have $C_{\text{neutral}} > C_{\text{small}} > C_{\text{large}}$.

C_{small} is smaller than C_{neutral} because of the fitness cost decreases the transition rate from u_c to $N\rho u_c$.

C_{large} is smaller than C_{small} might be because the mutation generates immediately at the begginning of the process, and therefore the population already has a lot of cells with one mutation ready to initiate a non-CIN type of cancer (increase in $X_2(t)$).

Problem 4: Linear process of colonic crypt transformation

The colon consists of approximately $M = 10^7$ crypts, each consisting of $N = 10^3$ cells. The mutation rate per gene is of order of $u = 10^{-8}$ per cell division. We assume a generation time of 1 day. Consider an oncogenic mutation causing a selective advantage $s = 10\%$ in the following three scenarios:

(a)

Homogeneous tissue: Use the Moran model to calculate the expected number of neoplastic crypts (i.e. crypts in which all cells are mutated) at age 50.

```
s <- 0.1   # selective advantage
r <- 1 + s # fitness
N <- 10^3
M <- 10^7
rho <- (1 - 1 / r) / (1 - 1 / r^N) # fixation probability
t <- 50 * 365 # age 50 in days, as we assume a generation time of 1 day
u <- 10^-8 # mutation rate
prob <- 1 - exp(-N * u * rho * t)
neoplastic <- M * prob
```

The expected number of neoplastic crypts at age 50 is 1.6454×10^5.

(b)

Single stem cell: By a similar calculation, compute the number of transformed crypts at age 50, if the cells in each crypt originate from a single stem cell. Stem cells replicate more slowly with a generation time of approx. 10 days.

```
# rho, u, M are the same as previously defined
N.stem <- 1
t.stem <- 365 * 50 / 10 # 10 day generation
rho.stem <- (1 - 1 / r) / (1 - 1 / r^N.stem)
prob.stem <- 1 - exp(- N.stem * u * rho.stem * t.stem)
neoplastic.stem <- M * prob.stem
```

The expected number of neoplastic crypts at age 50 is 182, if the cells in each crypt originate from a single stem cell.

(c)

Multiple stem cells: Use the Moran model to compute the number of neoplastic crypts after 50 years, assuming that each crypt is maintained by a pool of $N_S = 4$ stem cells. Assume the same fitness advantage $s = 10\%$.

```
N.stem.multi <- 4
rho.stem.multi <- (1 - 1 / r) / (1 - 1 / r^N.stem.multi)
prob.stem.multi <- 1 - exp(- N.stem.multi * u * rho.stem.multi * t.stem)
neoplastic.stem.multi <- M * prob.stem.multi
```

The expected number of neoplastic crypts at age 50 is 209, if the cells in each crypt originate from a pool of $N_S = 4$ stem cells.

(d)

Discuss which tissue architecture prevents best the initiation of cancer.

The Single Stem Cell architecture best prevents cancer initiation since it reduces the expected number of neoplastic crypts after 50 years down to 182, which is smaller than the other two architectures.