ETH zürich D BSSE Computational Biology Group

# Evolutionary Dynamics

Exercises 10

Prof. Dr. Niko Beerenwinkel
Johannes Gawron
Michael Schneider
Norio Zimmermann

November 20, 2024

**Problem 1: The Galton-Watson process (tutorial exercise)**

(a) Consider the following four cell lines. Every 24 hours, each cell of these cell lines either divides into two daughter cells, dies, or survives into the next generation without producing offspring (which is equivalent to the cell creating a single offspring). The probabilities of each event are fixed for each cell line as follows:

(A) $P(X=0)=0.1, P(X=1)=0.8, P(X=2)=0.1, P(X=i)=0\ \forall i>2$

(B) $P(X=0)=0.4, P(X=1)=0.4, P(X=2)=0.2, P(X=i)=0\ \forall i>2$

(C) $P(X=0)=0.2, P(X=1)=0.2, P(X=2)=0.6, P(X=i)=0\ \forall i>2$

(D) $P(X=0)=0.4, P(X=1)=0.2, P(X=2)=0.4, P(X=i)=0\ \forall i>2$

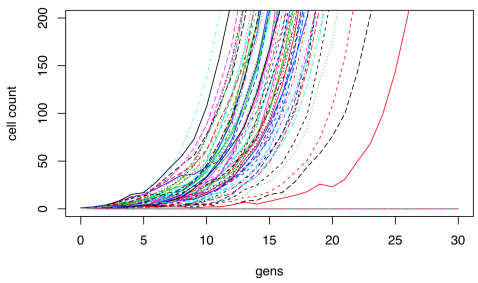
Calculate for each distribution analytically the expected number of offspring a single cell produces in the first iteration. Use your findings to decide for each distribution which type of Galton-Watson process (subcritical, critical, or supercritical) it gives rise to.

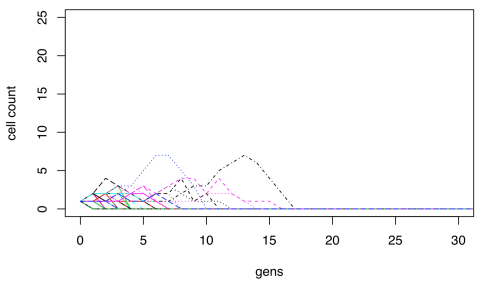
(b) The plots given below show the results of simulating the population growth for each offspring probability distribution given in (a) starting with a single cell.

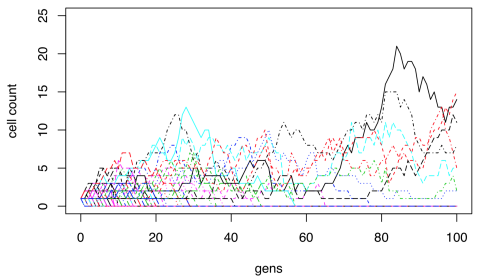
The following R function was used for the simulation:

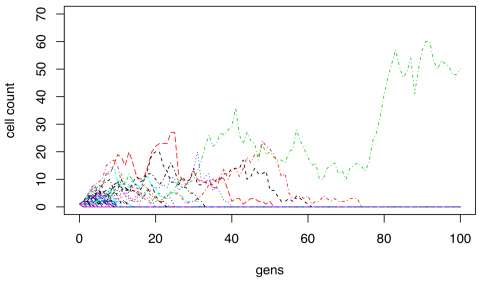
```
gw <- function(gens, init, distr) {
  counter <- rep(0, gens+1)
  counter[1] <- init
  i <- 1
  while (i <= gens && counter[i] > 0){
    nextGen <- sample(c(0,1,2), counter[i], replace = TRUE, distr)
    counter[i+1] <- sum(nextGen)
    i <- i+1
  }
  return(counter)
}
```

Match each plot to the corresponding offspring probability distribution from (a).






(c) Your wet lab colleagues lost the labels of their four cell lines with the properties described above. They ask you for help to figure out which cell line is which.

Your colleagues perform the following wet lab experiment: (i) Sample 1000 cells from each cell line and put each in a separate dish. (ii) Let the cells grow out into cell colonies (or go extinct) over the course of 20 days. (iii) Then end the experiment and count the cells in each dish (the counts are reported in the spreadsheet).

Propose a strategy to find out which dataset belongs to which cell line.

**Problem 2: Probability generating function**

Let $Z$ be a random variable such that $Z \in \mathbb{Z}^+$ ($Z$ is a non-negative integer), and $p_i$ its distribution, *i.e.* $\text{Prob}[Z = i] = p_i$. The probability generating function (*pgf*) of $Z$ is a function of a symbolic argument $s$ defined as the expected value $\text{E}[s^Z] = \sum_{i=0}^{\infty} p_i s^i$ and denoted by $f_Z(s)$. We assume that all probability generating functions are absolutely convergent on the interval $[0,1]$. *Note: This is a technical requirement to assure that summand-wise operations are permitted.*

Prove the following statements:

(a) the expectation of $Z$ is given by $\text{E}[Z] = f'_Z(1)$; **($\frac{1}{2}$ point)**

(b) the variance of $Z$ is $\text{Var}[Z] = f'_Z(1) + f''_Z(1) - f'_Z(1)^2$ **($\frac{1}{2}$ point)**

(c) $\frac{d^k f_Z}{ds^k}|_{s=0} = k!\, p_k$; **(1 points)**

(d) if $Z$ and $Y$ are two independent random variables in $\mathbb{Z}^+$, then $f_{Z+Y}(s) = f_Z(s) f_Y(s)$; **(1 points)**

(e) if $Y$ is a $\mathbb{Z}^+$-valued random variable and $\{Z^{(i)}, i \geq 1\}$ a sequence of independent identically distributed random variables in $\mathbb{Z}^+$ independent of $Y$, then $V = \sum_{i=1}^{Y} Z^{(i)}$ has the pgf $f_V(s) = f_Y[f_{Z^{(1)}}(s)]$. Hint: Use d) and the law of total expectations. **(1 points)**

**Problem 3: The Luria-Delbrück experiment**

The Luria–Delbrück experiment (Salvador E Luria and Max Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28(6):491, 1943) tests two hypotheses for how bacteria acquire resistance to the virus. The first hypothesis (adaptive immunity) states that the mutations leading to resistance to the virus were caused by an induced activation (exposure to the virus). The second hypothesis (random mutation) states that the mutations to resistance may occur any time prior to the addition of the virus. The experiment demonstrated that in bacteria, genetic mutations arise in the absence of selective pressure rather than being a response to it.

Experiment setup (see the figure below): Several bacterial cultures are grown from a single cell into separate culture tubes. After a period of growth, the cultures are exposed to the virus. If the resistance to the virus was caused by adaptive immunity, then each plate should contain roughly the same number of resistant colonies. Otherwise (the random mutation case) the number of resistant colonies on each plate should vary (the variance greater than the mean).

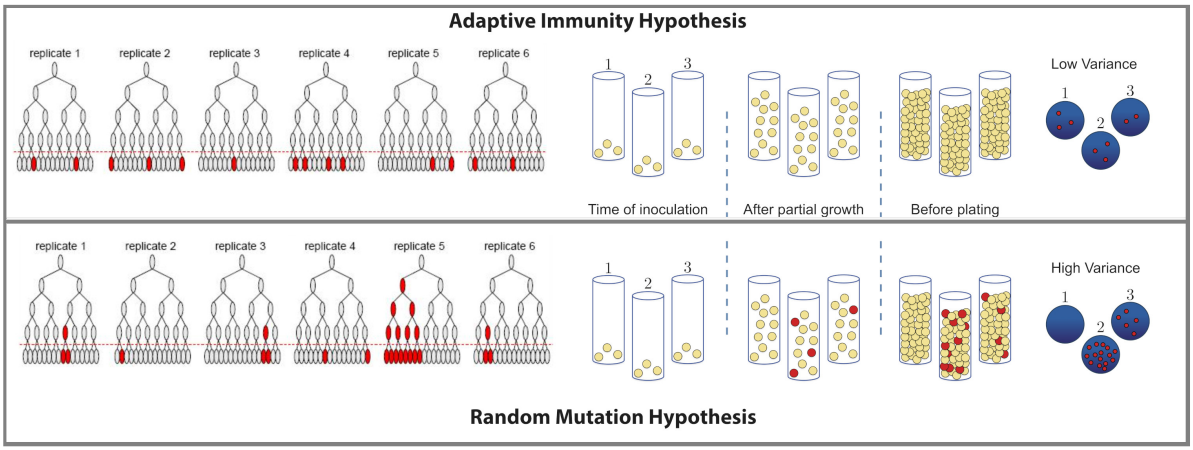



Figure adapted from:
https://evilutionarybiologist.blogspot.com/2008/07/this-weeks-citation-classic-fluctuation.html
http://justinbois.github.io/bootcamp/2016/lessons/ld_hypotheses.png

The number of cells grows at a rate $\beta$, such that the total number of cells at time point $t$ is $N(t) = N(0)e^{\beta t}$, and they give stochastically rise to a mutant (resistant) offspring with rate $\alpha$. Hence the number of cells that directly arise through mutation are a non-homogeneous Poisson process with time dependent rate $\lambda(t) = \alpha N(t)$. Thus, the distribution of the number of mutations that occur in $[0,t]$ is Poissonian with parameter $\Lambda(t) = \int_0^t \lambda(\tau)\mathrm{d}\tau$. In the absence of virus, mutant cells grow at the same rate as normal bacteria.

(a) Compute the probability $P_0(t)$ that no mutations have occurred at time $t$. Show that the mutation rate $\alpha$ can be estimated as $\alpha = \frac{\beta \ln \rho}{1-e^{\beta t}}$, where $\rho$ is the ratio of experiments in which resistance was not found (estimator for $P_0(t)$). Assume $N(0) = 1$.

*Hint:* Assume that in a small time interval interval $[t, t+\Delta t]$ the number of mutants is poisson distributed at rate $\alpha N(t)\Delta t$ to show that

$$\begin{aligned} P_0(t) &= P(0 \text{ mutants in } [0,\Delta t])P(0 \text{ mutants in } [\Delta t, 2\Delta t]) \cdots P(0 \text{ mutants in } [t-\Delta t, t]) \\ &\approx e^{-\alpha N(0)\Delta t} \cdots e^{-\alpha N(t-\Delta t)\Delta t} \end{aligned}$$

and let $\Delta t \to 0$. Explain the assumptions made in this calculation. **(2 points)**

(b) Derive $\alpha$ from the expression for $P_0$ derived in the lecture for the Galton-Watson process and explain the differences. **(1 point)**

(c) Compute the expected number of mutant cells at time $t$, $m(t)$, and their variance $\sigma^2(t)$ under the Luria-Delbrück model.
*Hint:* The expected number of new mutant cells that arise in the interval $[\tau, \tau + \mathrm{d}\tau]$ is:

$$\nu(\tau + \mathrm{d}\tau) = \lambda(\tau)\mathrm{d}\tau.$$

Compute to which size these newly generated subclones have grown to at time $t$ and express $m(t)$ by integrating out the time point of mutations.
Consider a similar strategy for the variance. Use that the new mutant cells arising in $[\tau, \tau + \mathrm{d}\tau]$ are Poissonian variables and remember that $\mathrm{Var}[aX] = a^2\,\mathrm{Var}[X]$. **(2 points)**

(d) Luria and Delbrück used the mean and variance to distinguish the proposed mechanism of mutations stochastically accumulating prior to viral infection from an active adaptation scenario. Suppose that in the adaptation case, bacteria have no resistance, but stochastically acquire resistance upon infection with high rate $\delta$. In this short period of time, the population size can be considered constant. What would be the resulting relation between the expected number of resistant cells and their variance? Compare this with your results from the accumulation scenario, part (c).
**(1 point)**