Branching Processes in Biology

Marek Kimmel
Department of Statistics
Rice University

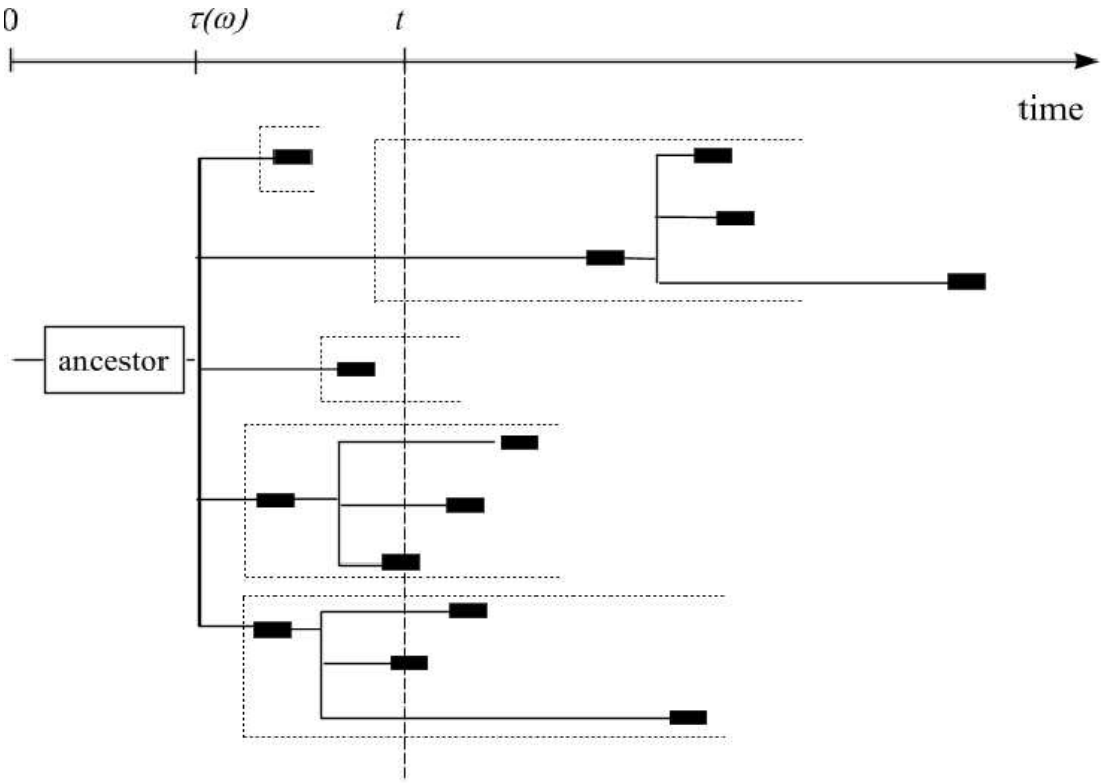
Branching Processes in Biology

Marek Kimmel

Based on Kimmel and Axelrod "Branching Processes in Biology"
Springer 2002

What are branching processes?

- Family $Z(t,\omega),\ t \geq 0\}$ of nonnegative integer-valued random variables defined on a common probability space Ω
- Process is initiated at time $t = 0$ by birth of a single ancestor particle with random lifelength $\tau(\omega)$
- Ancestor produces $X(\omega)$ progeny at death
- Each progeny is the ancestor of its own *independent* process, which is a component of our branching process.
- The number of individuals present in the process at time t is equal to the sum of numbers of the individuals present in all these subprocesses:


0
τ(ω)
t
time
ancestor

$$Z(t,\omega) = \begin{cases} \sum_{i=1}^{X(\omega)} Z^{(i)}(t, \tau(\omega), \omega), & t \geq \tau(\omega), \\ 1, & t < \tau(\omega), \end{cases}$$

where $Z^{(i)}(t, \tau(\omega), \omega)$ denotes the number of individuals time t in the i-th independent identically distributed subprocess started by the a single ancestor born at time $\tau(\omega)$

- The self-recurrence (or branching) property,

$$Z^{(i)}(t, \tau(\omega), \omega) \stackrel{d}{=} Z^{(i)}(t - \tau(\omega), \omega). \tag{1}$$

- Substitution leads to a recurrent relation

$$Z(t,\omega) = \begin{cases} \sum_{i=1}^{X(\omega)} Z^{(i)}(t - \tau(\omega), \omega), & t \geq \tau(\omega), \\ 1, & t < \tau(\omega), \end{cases}$$

  which we will use repeatedly.

Motivating example

Polymerase Chain Reaction and branching processes

- Polymerase Chain Reaction (PCR) is one of the most important tools of molecular biology.
- An experimental system for producing large amounts of genetic material from a small initial sample.
- Repeated cycles of DNA replication in a test tube that contains free nucleotides, DNA replication enzymes and template DNA molecules.
- Operates by harnessing the natural replication scheme of DNA molecules
- The result is a vast amplification of a particular DNA locus from a small initial number of molecules.

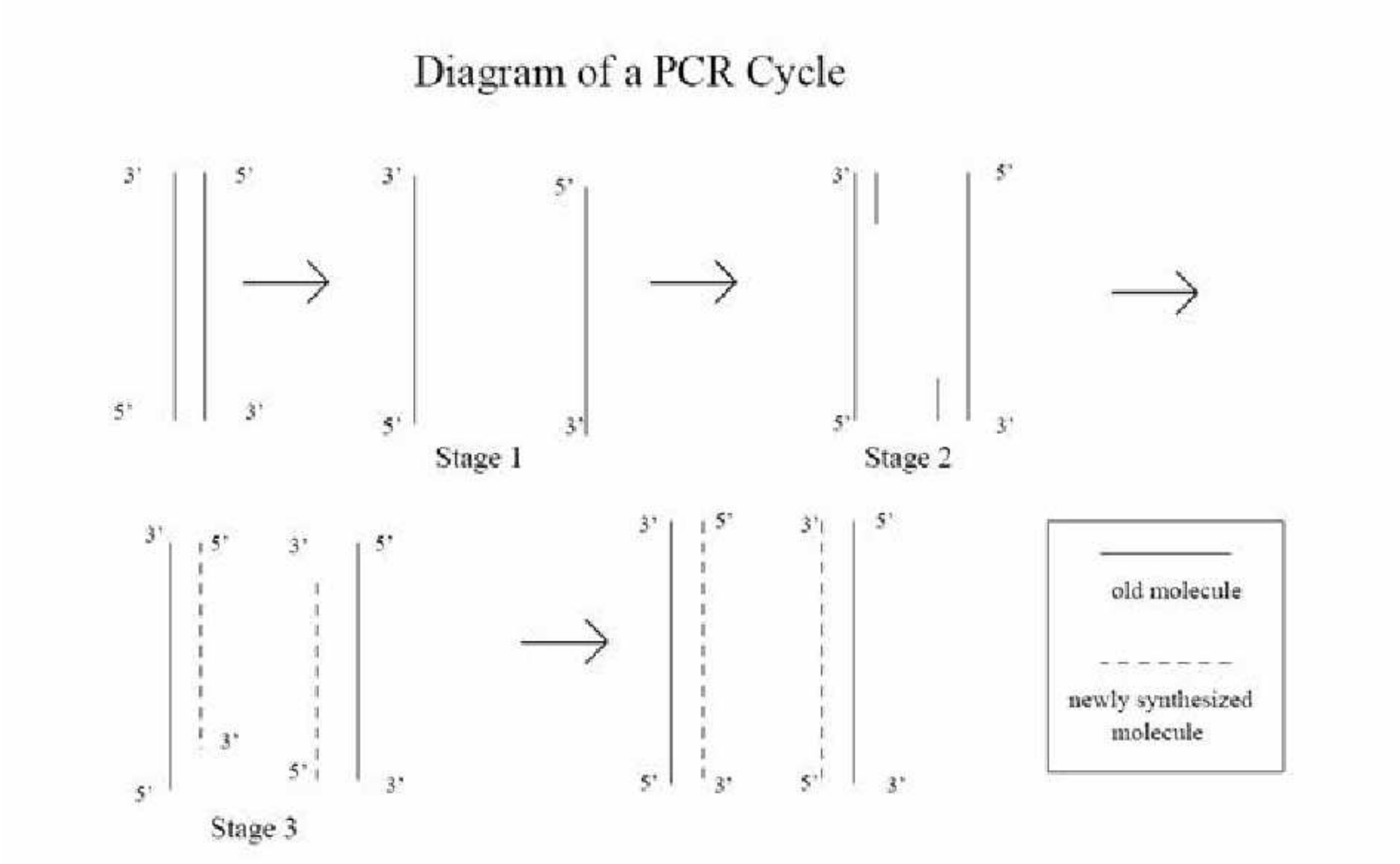


Diagram of a PCR Cycle
3'
5'
5'
3'
Stage 1
Stage 2
Stage 3
old molecule
newly synthesized molecule

Stochasticity of PCR amplification

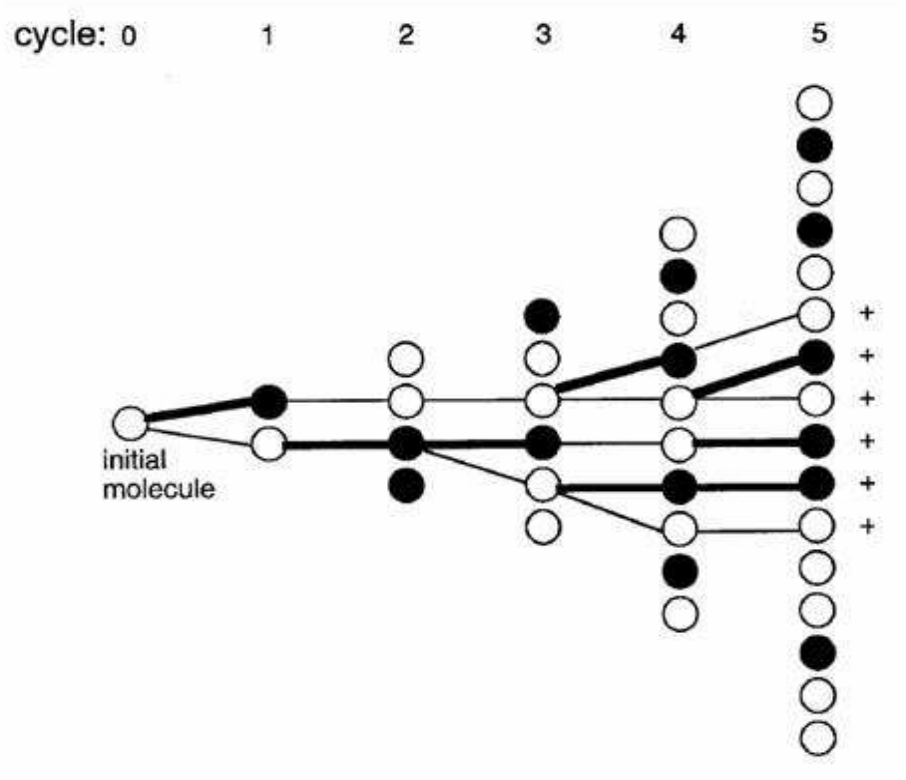
- Not every existing molecule is successfully replicated in every reaction cycle.
- Even the most highly efficient reactions operate at an efficiency around 0.8, i.e., each double stranded molecule produces an average of 0.8 new molecules in a given reaction cycle.
- Under ideal conditions in each clone the molecules are identical or complementary to the ancestral molecule of the clone (molecules in the initial samples may not be identical).

- However, random alterations of nucleotides in DNA sequences, known as mutations, also occur during PCR amplification.
- In many PCR applications (e.g., forensic), mutations hinder analysis of the initial sample
- In other settings, however, PCR mutations are desirable, as is the case in site-directed mutagenesis studies and artificial evolution experiments. (Joyce 1992).
- In order to assess the *genealogy of the molecules in a sample*, one must model the PCR and the structure of DNA replication.

Modeling PCR

- Ideally, PCR is a binary fission process with discrete time
- PCR starts with S_0 identical copies of single-stranded sequences
- Let S_i be the number of sequences present after the i-th cycle.
- In cycle i each of the S_{i-1} template molecules is amplified independently of the others with probability λ_i (the efficiency in cycle i). More precisely, the efficiency does not simply depend on the cycle number, but on the number of amplifications in the previous cycles and on PCR conditions.

- If $\lambda_i = \lambda$ independently of the cycle number, then the the sequence $S_0, S_1, \ldots, S_i, \ldots$(accumulation of PCR product) is a Galton-Watson branching process.
- We add a mutation process to the model: we assume that a new mutation occurs at a position that has not mutated in any other sequence before.
- We model the process of nucleotide substitution as a Poisson process with parameter μ, which is the error rate (mutation rate) of PCR per target sequence and per replication.


cycle: 0
1
2
3
4
5
initial
molecule
+
+
+
+
+
+

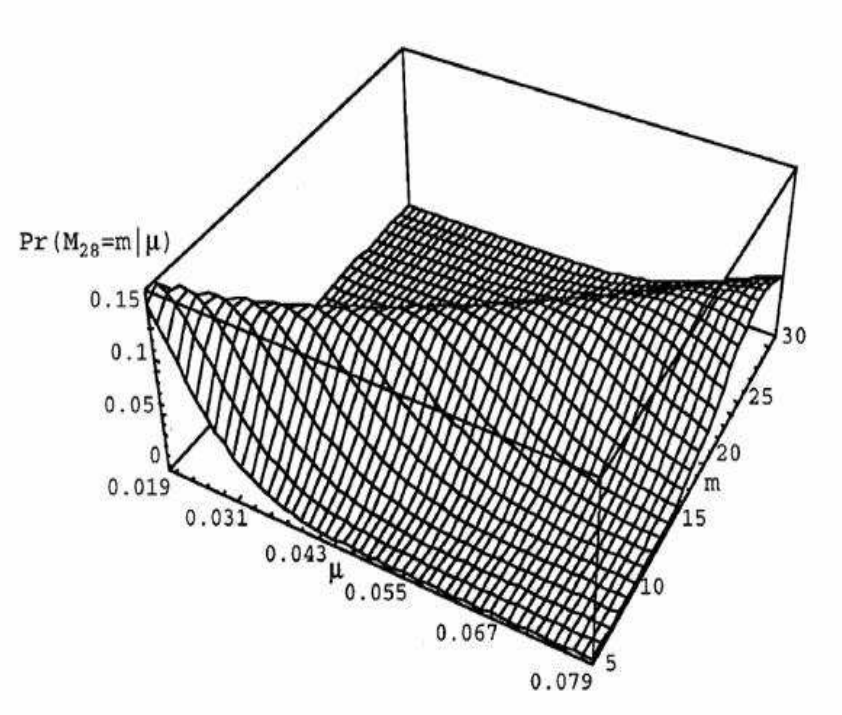
Statistical estimation of the efficiency and mutation rate

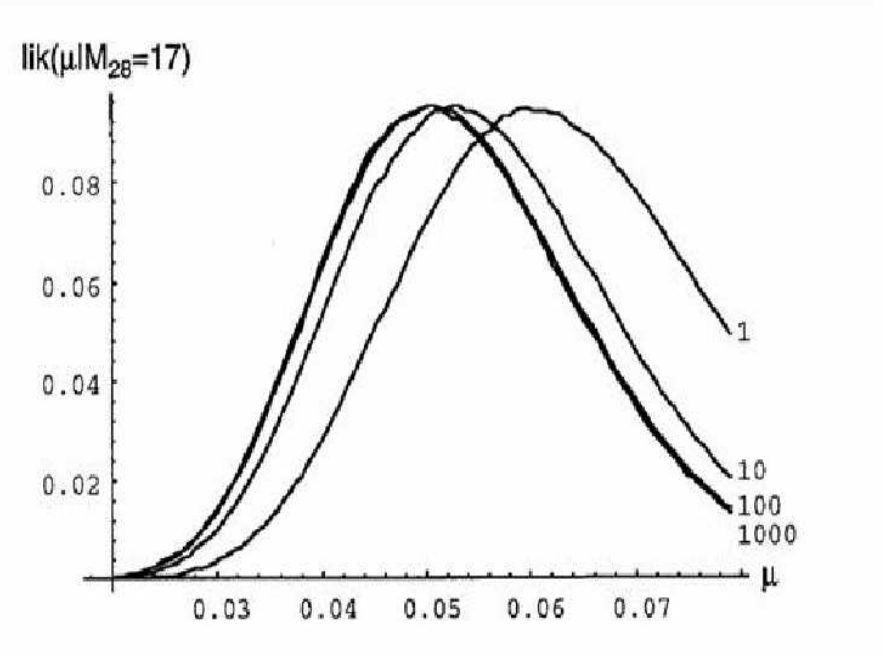
$$\frac{\mathrm{E}(S_i)}{\mathrm{E}(S_{i-j})} = (1+\lambda_i)^j,\ i \geq j.$$

From simulation, the probabilities $\Pr(M_n = m|\mu)$ that there are m mutants in the n-th generation can be computed

$$lik(\mu|M_n = m) = \Pr(M_n = m|\mu)$$

and the MLE of μ can be determined by maximization.


Pr(M28=m|μ)
0.15
0.1
0.05
0
0.019
0.031
0.043
μ
0.055
0.067
0.079
5
10
15
m
20
25
30


lik(μ|M28=17)
0.08
0.06
0.04
0.02
0.03
0.04
0.05
0.06
0.07
μ
1
10
100
1000

The Galton-Watson process

- A single ancestor particle lives for exactly one unit of time and at the moment of death produces a random number Z_1 of progeny according to a prescribed probability distribution

$$\{p_k;\ k = 0, 1, 2, \ldots\},\ p_k = P[Z_1 = k]$$

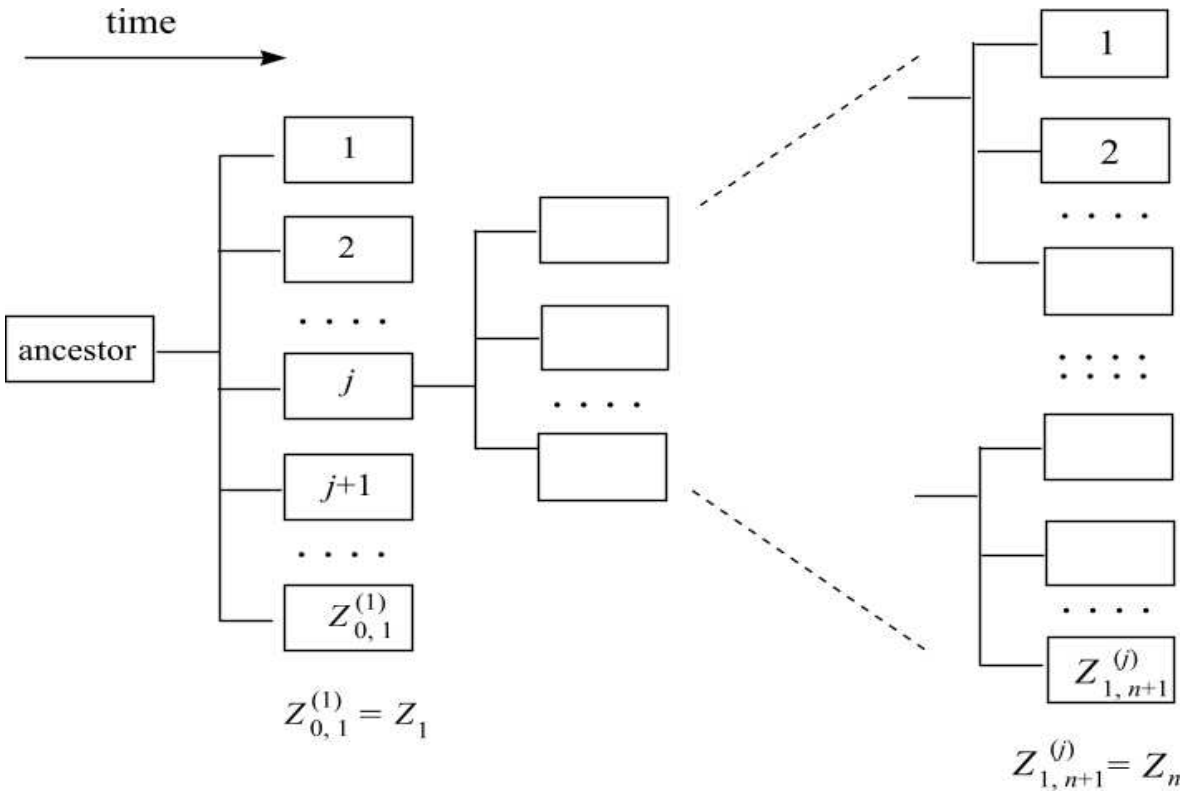
- Each of the first generation progeny behaves, independently of each other, as the initial particle did.
- The particle counts Z_n in the successive generations $n = 0, 1, 2, \ldots$ (where generation 0 is composed of the single initial particle) form a sequence of random variables

We have

$$Z_{n+1} = \sum_{j=1}^{Z_1} Z_{1,n+1}^{(j)}.$$

where random variables $Z_{1,n+1}^{(j)}$ are *independent identically distributed (iid)* copies of processes initiated by ancestor's progeny with distributions identical to that of Z_n, so

$$Z_{n+1} = \sum_{j=1}^{Z_1} Z_n^{(j)}$$


time
ancestor
1
2
j
j+1
$Z_{0,1}^{(1)}$
$Z_{0,1}^{(1)} = Z_1$
1
2
$Z_{1,n+1}^{(j)}$
$Z_{1,n+1}^{(j)} = Z_n$

Definition 1. *The pgf f_X of a Z_+-valued rv X is a function $f_X(s) = E\big(s^X\big) = \sum_{i=0}^{\infty} p_i s^i$, of a symbolic argument $s \in U \equiv [0,1]$. With some abuse of notation, we write $X \sim f_X(s)$. The following are the basic properties of the pgf's.*

Theorem 1. *Suppose X is a Z_+-valued rv with pgf $f_X(s)$ which may not be proper. Let us denote (N) the nontriviality condition $p_0 + p_1 < 1$.*

1. *f_X is nonnegative and continuous with all derivatives on $[0, 1)$. Under (N), f_X is increasing and convex.*
2. *If X is proper, $f_X(1) = 1$, otherwise $f_X(1) = P[X < \infty]$.*
3. *$d^k f_X(0)/ds^k = k! p_k$.*
4. *If X is proper, the k-th factorial moment of X, $\mu_k = E[X\,(X-1)\,(X-1)\,...\,(X-k+1)]$, is finite if and only if $f_X^{(k)}(1-) = \lim_{s\uparrow 1} f_X^{(k)}(s)$ is finite. In such case, $\mu_k = f_X^{(k)}(1-)$.*

5. *If X and Y are two independent Z_+-valued rv's, then* $f_{X+Y}(s) = f_X(s) f_Y(s)$.

6. *If Y is a Z_+ -valued rv and $\left\{X^{(i)},\ i \geq 1\right\}$ is a sequence of iid Z_+-valued rv's independent of Y, then $V = \sum_{i=1}^{Y} X^{(i)}$ has pgf* $f_V(s) = f_Y\left[f_{X^{(1)}}(s)\right]$.

7. *Suppose that $\{X_i,\ i \geq 1\}$ is a sequence of Z_+-valued rv's.* $\lim_{i\to\infty} f_{X_i}(s) = f_X(s)$ *exists for each $s \in [0,1)$ if and only if the sequence $\{X_i,\ i \geq 1\}$ converges in distribution to a rv X, i.e., if limits $\lim_{i\to\infty} P[X_i = k]$ exist for all k and are equal to $P[X_i = k]$, respectively. Then $f_X(s)$ is the pgf of the limit rv X.*

Application of the pgf Theorem to the Galton-Watson process

$$Z_{n+1} = \sum_{j=1}^{Z_1} Z_n^{(j)}$$

- Denote $f(s) = f_1(s)$, the pgf of Z_1,and by $f_n(s)$, the pgf of Z_n,
- The pgf Theorem (part 6) yields the following pgf identity

$$f_{n+1}(s) = f_1[f_n(s)] = f[f_n(s)].$$

- If we note that $Z_0 = 1$ implies $f_0(s) = s$, this yields

$$f_n(s) = f^{(n)}(s) = \underbrace{f\{\cdots f[f}_{n \text{ times}}(s)] \cdots \},$$

i.e., the pgf of Z_n is the n-th functional iterate of the progeny pgf $f(s)$.

Moments of the GW process

- The moments of the process are expressed in the terms of the derivatives of $f(s)$ at $s = 1$,e.g.,

$$\mathrm{E}(Z_1) = f'(1-) \equiv m,$$

  where m is the mean number of progeny of a particle. From the chain rule

$$\mathrm{E}(Z_n) = f_n'(1-) = f_{n-1}'(1-)f'(1-) = \ldots = m^n.$$

- GW process is subcritical if $m < 1$, critical if $m = 1$, supercritical if $m > 1$.

- Similarly, using the chain rule for the second derivative, one concludes that

$$\mathrm{Var}(Z_n) = \begin{cases} \frac{\sigma^2 m^{n-1}(m^n-1)}{m-1}; & m \neq 1, \\ n\sigma^2; & m = 1, \end{cases}$$

  where $\sigma^2 = \mathrm{Var}(Z_1)$ is the variance of the progeny count.

- The linear growth of variance in the critical case ($m = 1$) is consistent with the "heavy tails" of the distribution of Z_n in the critical case.

Linear fractional case of the GW process

- After several iterations, the functional form of the iterates $f_n(s)$ becomes intractable.
- The linear fractional case is the only non-trivial counterexample. Suppose

$$p_0 = \frac{1-b-p}{1-p},\ p_k = bp^{k-1},\ k = 1, 2, \ldots,\ p \in (0,1).$$

$$f(s) = 1 - \frac{b}{1-p} + \frac{bs}{1-ps},$$

and $m = \frac{b}{(1-p)^2}$.

- Equation $f(s) = s$ has roots q and 1.

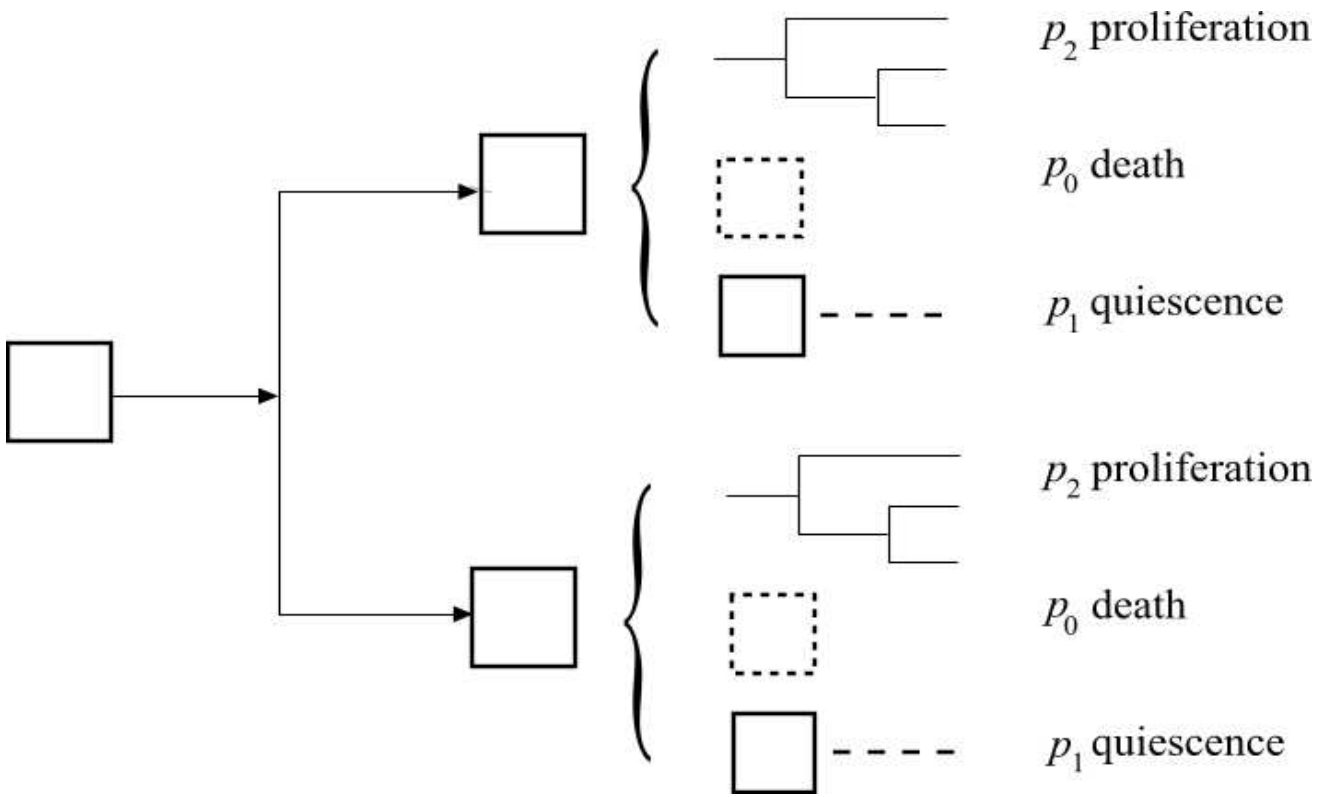
$$f_n(s) = 1 - m^n\left(\frac{1-q}{m^n-q}\right) + \frac{m^n\left(\frac{1-q}{m^n-q}\right)^2 s}{1-\left(\frac{m^n-1}{m^n-q}\right)s}; \quad m \neq 1,$$

$$f_n(s) = \frac{np-(np+p-1)s}{1-p+np-nps}; \quad m = 1.$$

- The linear functional pgf corresponds to the geometric distribution with a rescaled term at zero.

Application : Cell cycle model with death and quiescence

- Fundamental step in the proliferation of a population of cells is the *division* of one cell into two cells.
- After completing its life cycle each cell approximately doubles in size and then divides into two progeny cells of approximately equal sizes.
- Populations derived from single cells are referred to as *clones or colonies*.
- Founder cells may not yield colonies with the same number of cells after the same time. This may be due to various factors, like the *randomness of cell death and quiescence*.


p_2 proliferation
p_0 death
p_1 quiescence
p_2 proliferation
p_0 death
p_1 quiescence

Mathematical description of the cell cycle model

- Process more general than the standard Galton-Watson process, initiated by a single proliferating cell, which divides and each of its progeny, independently, may
  - become proliferative with probability (wp) p_2,
  - become quiescent wp p_1, or
  - die wp p_0.

(Z_1, Q_1)	Probability	(Z_{n+1}, Q_{n+1})	$f_{n+1}(s, w)$
$(0, 0)$	p_0^2	$(0, 0)$	1
$(0, 1)$	$2p_0p_1$	$(0, 1)$	w
$(0, 2)$	p_2^2	$(0, 2)$	w^2
$(1, 0)$	$2p_2p_0$	$(Z^{(1)}_{1,n+1}, Q^{(1)}_{1,n+1})$	$f_n(s, w)$
$(1, 1)$	$2p_2p_1$	$(Z^{(1)}_{1,n+1}, Q^{(1)}_{1,n+1} + 1)$	$f_n(s, w)\ w$
$(2, 0)$	p_2^2	$(Z^{(1)}_{1,n+1} + Z^{(2)}_{1,n+1}, Q^{(1)}_{1,n+1} + Q^{(2)}_{1,n+1})$	$f_n(s, w)^2$

Pgf equations for the cell cycle model

$$Z_{n+1} = \sum_{j=1}^{Z_1} Z_{1,n+1}^{(j)}.$$

$$Q_{n+1} = \sum_{j=1}^{Z_1} Q_{1,n+1}^{(j)} + Q_1.$$

$f_n(s,w)$ the joint pgf of random variables Z_n and Q_n. We obtain the pgf recurrence:

$$f_{n+1}(s,w) = [p_2 f_n(s,w) + p_1 w + p_0]^2.$$

Modeling biological cell cycle data

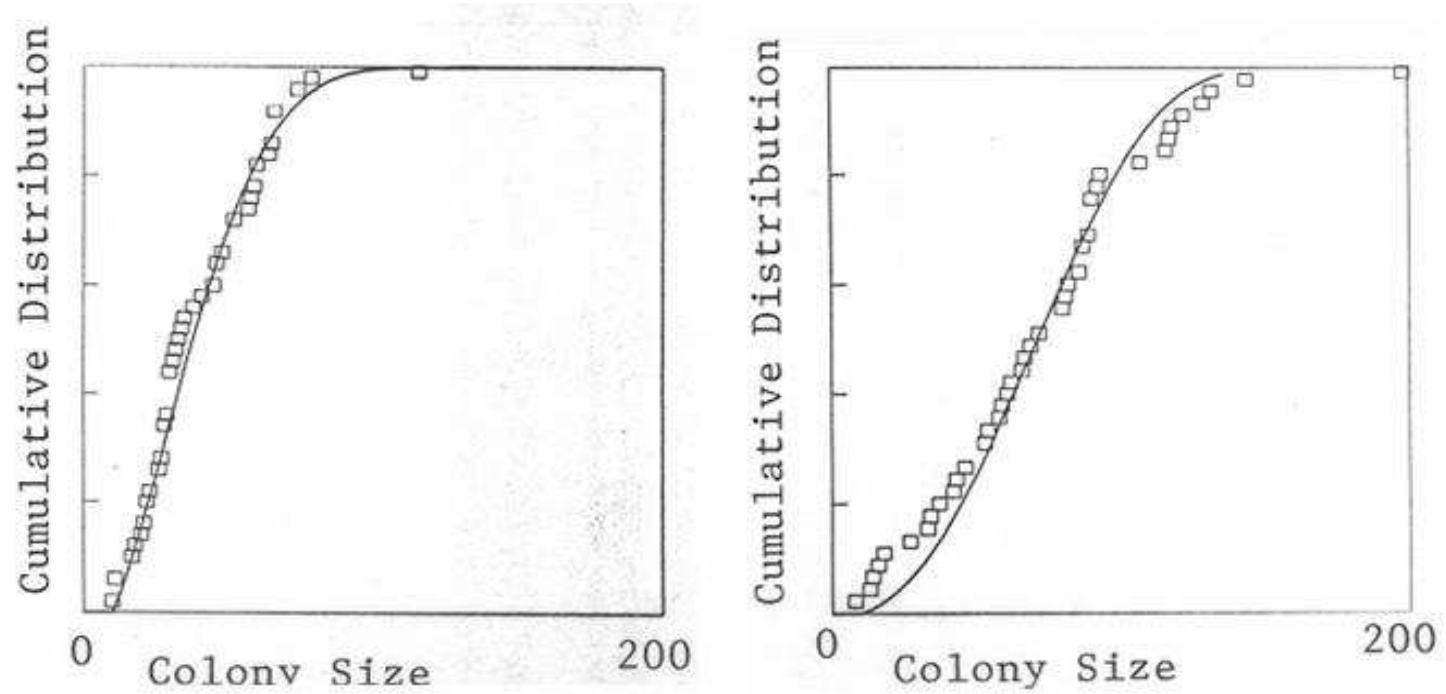
- In the experiment, it is impossible to discern proliferative cells from quiescent cells. Therefore, distributions of $Z_n + Q_n$ are used.
- The pgf of this sum is equal to $g_n(s) = f_n(s, s)$, and therefore

$$g_{n+1}(s) = [p_2 g_n(s) + p_1 s + p_0]^2.$$

- This pgf is equal to $g_n(s) = \sum_j \ \pi_n(j) s^j$, where $\pi_n(j) = \mathrm{P}\{Z_n + Q_n = j\}$ and so

$$\{\pi_{n+1}(j)\} = \{p_2\pi_n(j) + p_1\delta_{j1} + p_0\delta_{j0}\} \ * \ \{p_2\pi_n(j) + p_1\delta_{j1} + p_0\delta_{j0}\},$$

- During the time of the experiment, about $n = 8$ divisions occurred. Distributions $\{\pi_8(j),\ j \geq 0\}$ of colony size can be computed using different values of probabilities p_0 and p_1.


Cumulative Distribution
0
Colonv Size
200
Cumulative Distribution
0
Colony Size
200

Extinction and criticality

- Asymptotic behavior of $\{f_n(s)\}$ provides insight into the limit theorems for the $\{Z_n\}$ process.
- $f(s)$ is a power series with nonnegative coefficients $\{p_k\}$ and with $p_0 + p_1 < 1$, so that
- $f(s)$ is strictly convex and increasing in $[0, 1]$;
- $f(0) = p_0$; $f(1) = 1$;
- if $m \leq 1$ then $f(s) > s$ for $s \in [0, 1)$;
- if $m > 1$ then $f(s) = s$ has a unique root in $[0, 1)$.
- Let q be the smallest root of $f(s) = s$ for $s \in [0, 1]$.

Lemma 1. *If $m \leq 1$ then $q = 1$; if $m > 1$ then $q < 1$.*

Lemma 2. *If $s \in [0, q)$ then $f_n(s) \uparrow q$ as $n \to \infty$. If $s \in (q, 1)$ then $f_n(s) \downarrow q$ as $n \to \infty$. If $s = q$ or 1 then $f_n(s) = s$ for all n.*

As a special case of the Lemma 2 we note that $f_n(0) \uparrow q$.

$$\begin{aligned} \lim_{n\to\infty} f_n(0) &= \lim_n P\{Z_n = 0\} = \lim_n P\{Z_i = 0, \text{ for some } 1 \leq i \leq n\} \\ &= P\{Z_i = 0, \text{ for some } i \geq 1\} = P\{\lim_{n\to\infty} Z_n = 0\}, \end{aligned}$$

Therefore q is the *extinction probability*

Critical process becomes extinct wp. $q = 1$ although its mean stays constant (!)

Application : Complexity threshold in the evolution of early life

- Consider a polymeric chain of ν nucleotides
- If p is the probability that a single nucleotide is correctly copied, then the probability that a copy of the whole chain is correct is p^ν.
- During one generation step the molecule either survives (with probability w) and produces a copy, which is accurate with probability p^ν, or it is destroyed with probability $1 - w$.

- A given molecule yields 0, 1 or 2 molecules of the same type after one unit of time: the probabilities are $1-w$, $w(1-p^\nu)$ and wp^ν, respectively.

$$f(s) = (1-w) + w(1-p^\nu)s + wp^\nu s^2$$

- This biomolecule is indefinitely preserved with a positive probability only if the process is supercritical, i.e., if $m = w(1+p^\nu) > 1$, which yields

$$p^\nu > \frac{1-w}{w}.$$

Asymptotics of the supercritical GW process

If we set

$$W_n \equiv Z_n/m^n,$$

then $\mathrm{E}(W_n) = 1$ and

$$\mathrm{E}(W_{n+1}|W_n) = m^{-(n+1)}\mathrm{E}(Z_{n+1}|Z_n) = m^{-(n+1)}mZ_n = W_n.$$

Since the Galton-Watson process is a Markov chain, this means that W_n is a martingale. By Doob's Theorem

Theorem 2. *If* $0 < m \equiv f'(1-) < \infty$, *then there exists a random variable* W *such that*

$$\lim_{n\to\infty} W_n = W, \quad \text{wp } 1.$$

W might be nondegenerate only if $m > 1$.

Theorem 3. *If* $m > 1$, $\sigma^2 < \infty$, *and* $Z_0 \equiv 1$, *then*

1. $\lim_{n\to\infty} E(W_n - W)^2 = 0$
2. $E(W) = 1$, $Var(W) = \sigma^2/(m^2 - m)$
3. $P\{W = 0\} = q = P\{Z_n = 0 \text{ for some } n\}$.

Subcritical process

In the subcritical case, the process becomes extinct with probability 1. What can be said about the asymptotic behavior?

Example: Linear Fractional Case. The probability of nonextinction is now equal to

$$1 - f_n(0) = m^n(\frac{1-q}{m^n - q}),$$

$$\mathrm{E}(Z_n | Z_n > 0) = \frac{\mathrm{E}(Z_n)}{1 - f_n(0)} = \frac{m^n - q}{1-q} \to \frac{q}{q-1}; \quad n \to \infty.$$

Conditioning on nonextinction sufficient to obtain a limit law.

Theorem 4. *(Yaglom's) . If* $m < 1$ *then* $P\{Z_n = j | Z_n > 0\}$ *converges, as* $n \to \infty$*, to a probability function whose generating function* $\mathcal{B}(s)$ *satisfies equation*

$$\mathcal{B}[f(s)] = m\mathcal{B}(s) + (1 - m). \tag{2}$$

Also,

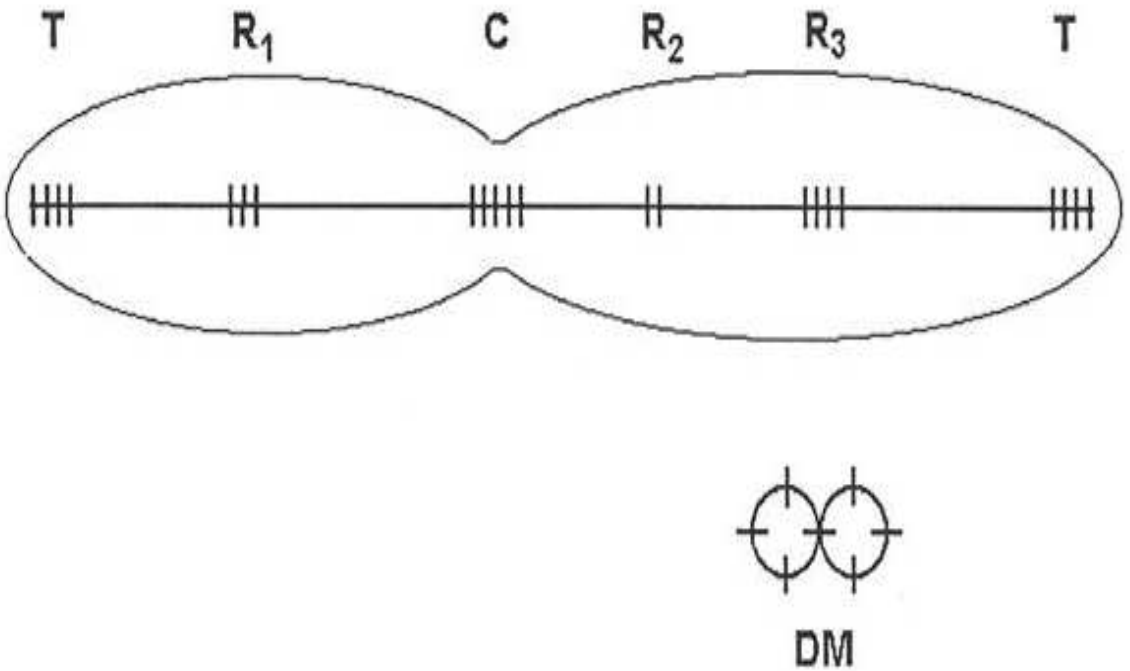
$$1 - f_n(0) \sim \frac{m^n}{\mathcal{B}'(1-)}; \quad n \to \infty. \tag{3}$$

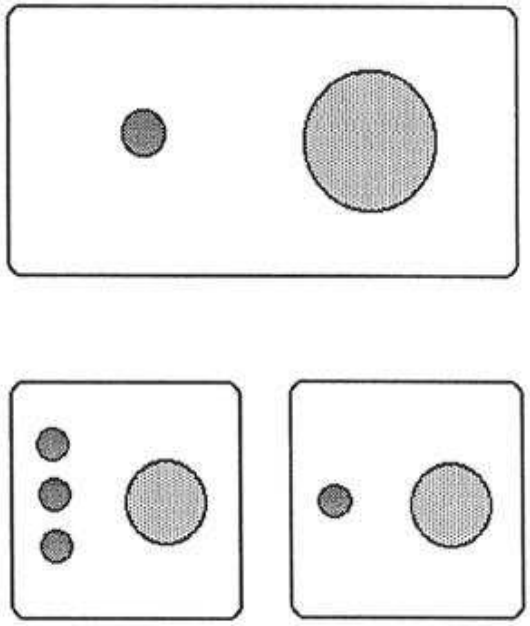
Convergence to a limit distribution conditional on nonabsorption is known as *quasistationarity*

Application **: Gene amplification and drug resistance**

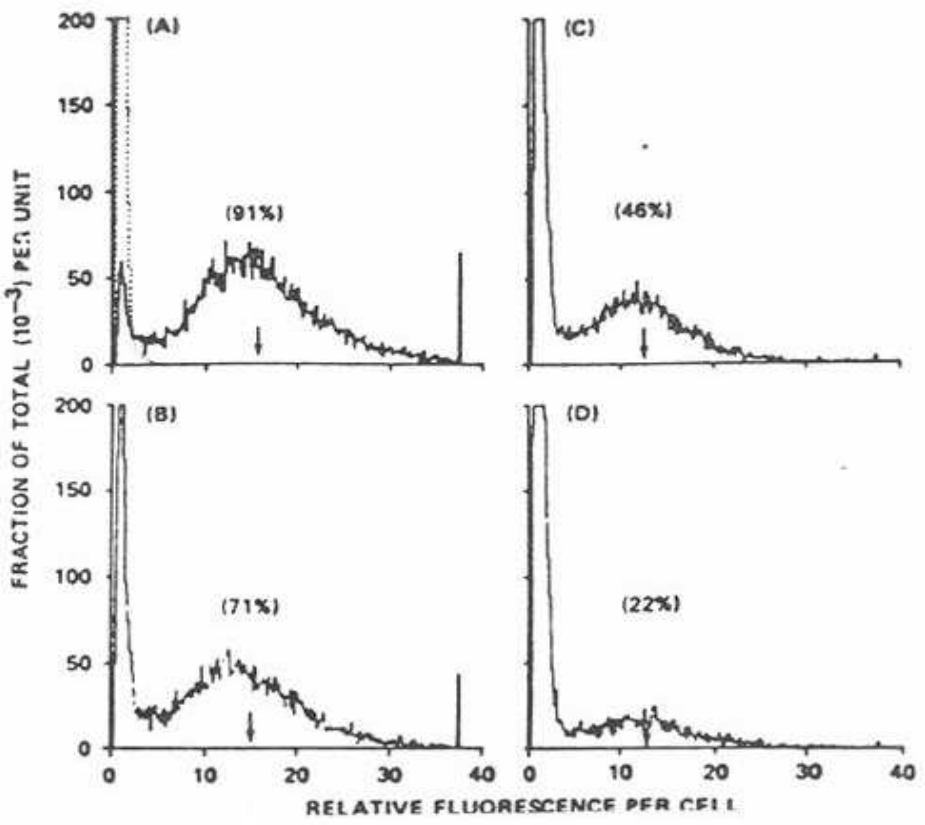
- *Amplification of a gene* is an increase of the number of copies of that gene in a cell.
- Amplification of genes coding for the *enzyme dihydrofolate reductase* (DHFR) has been associated with *cellular resistance to the anticancer drug* methotrexate (MTX).
- A resistant population with an increased number of DHFR gene copies per cell can be obtained after a sensitive population is grown in increasing concentrations of the drug.

- Increased resistance is correlated with increased numbers of gene copies on small extrachromosomal DNA elements called *double minute chromosomes.*
- The number of DHFR genes on double minutes in a cell may increase or decrease at each cell division. This is because double minutes do not have *centromeres*, which are required to faithfully segregate chromosomes into progeny cells.
- Increased drug resistance and the increase in number of gene copies are *reversible*. Cells grown in the absence of the drug gradually lose resistance to the drug, by losing extra gene copies.


T
R1
C
R2
R3
T
DM



Loss of resistance visualized by flow cytometry

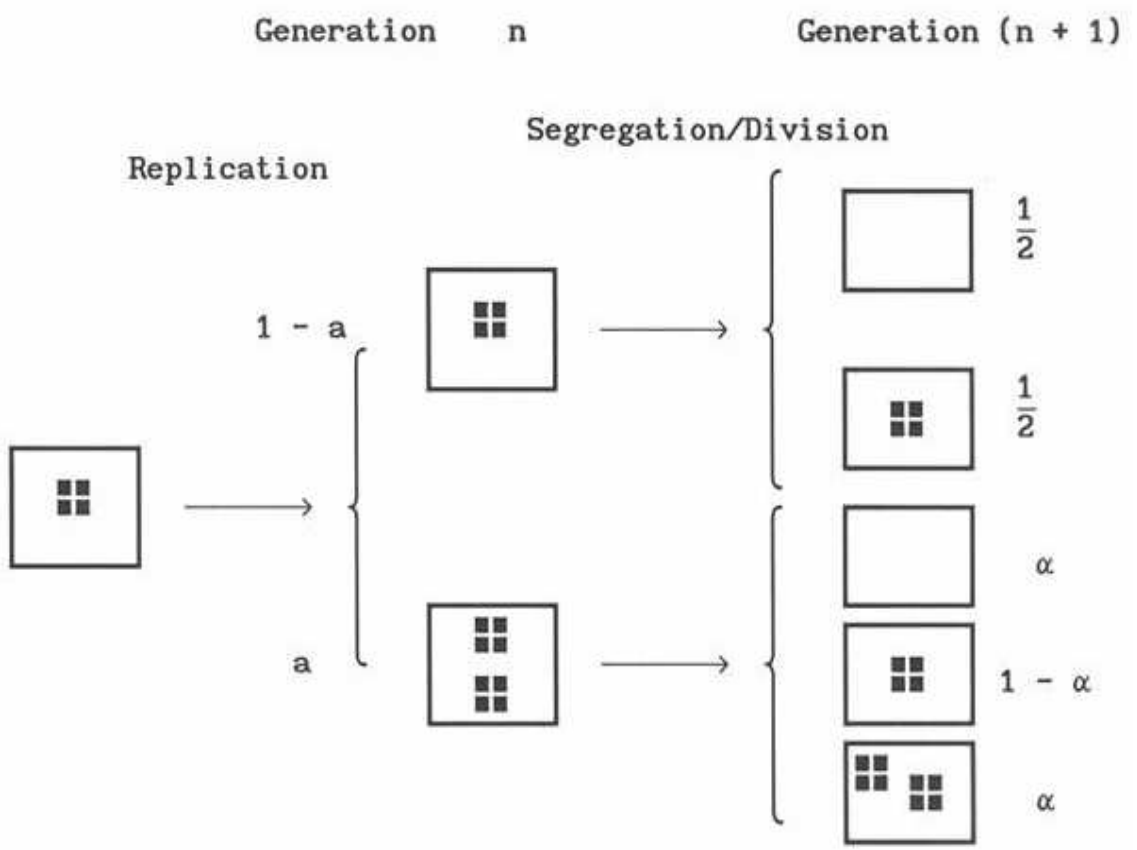


- Population distribution of numbers of copies per cell can be estimated by the experimental technique called *flow cytometry*.
- Proportion of cells with amplified genes decreases with time.
- The shape of the distribution of gene copy number within the subpopulation of cells with amplified genes appears stable as resistance is being lost.

Galton-Watson process model of gene amplification and deamplification

- Consider a cell, one of its progeny (randomly selected), one of the progeny of that progeny (randomly selected) and so forth.
- Cell of the n-th generation contains Z_n double minutes.
- During cell's life, each double minute is either replicated with probability a, or not replicated, with probability $1 - a$, independently of the other double minutes.

- At the time of cell division, the double minutes are segregated to progeny cells.
  - If the double minute has not been replicated, it is assigned to one of the progeny cells with probability $\frac{1}{2}$.
  - If it has been replicated, then either both copies are assigned to progeny 1 (wp $\alpha/2$), or to progeny 2 (wp $\alpha/2$), or they are divided evenly between both progeny (wp $1-\alpha$).


Generation n
Generation (n + 1)
Segregation/Division
Replication
1 - a
a
1/2
1/2
α
1 - α
α

Galton-Watson process model of gene amplification and deamplification

Randomly selected progeny in the line of descent contains

- no replicas of the original double minute (wp $(1-a)/2 + a\alpha/2$), or
- one replica of the original double minute (wp $(1-a)/2 + a(1-\alpha)$), or
- both replicas of the original double minute (wp $a\alpha/2$).

- The count of double minutes in the n-th generation of the cell lineage is a Galton-Watson process with

$$f(s) = d + (1 - b - d)s + bs^2,$$

where $b = a\alpha/2$ and $d = (1 - a)/2 + a\alpha/2$ are the probabilities of *gene amplification and deamplification*

- The process is subcritical, $b < d$ and $m = f'(1-) = 1 + b - d < 1$.

Mathematical model of the loss of resistance

- Cell is *resistant* if it carries at least one double minute chromosome with the DHFR gene.
- Otherwise it is called *sensitive*.
- Population of cells resistant to MTX, previously cultured for N generations in medium containing MTX, initially consists only of cells with at least one DHFR gene copy, i.e., $Z_N > 0$.

- By Yaglom Theorem, if N is large, distribution of $\{Z_N \mid Z_N > 0\}$ has pgf $\mathcal{B}(s)$

$$\mathcal{B}[f(s)] = m\mathcal{B}(s) + (1 - m)$$

- For $n > N$, when the cell clone has been transferred to the MTX-free medium, based on Yaglom Theorem, the fraction of resistant cells decreases roughly geometrically

$$1 - f_n(0) \ \sim \ m^{n-N}/\mathcal{B}'(1-)$$

  while $\{Z_n \mid Z_n > 0\}$ stays unchanged.

- This behavior is consistent with the flow cytometry experimental data.

Application : Iterated Galton-Watson (IGW) process and dynamic mutations

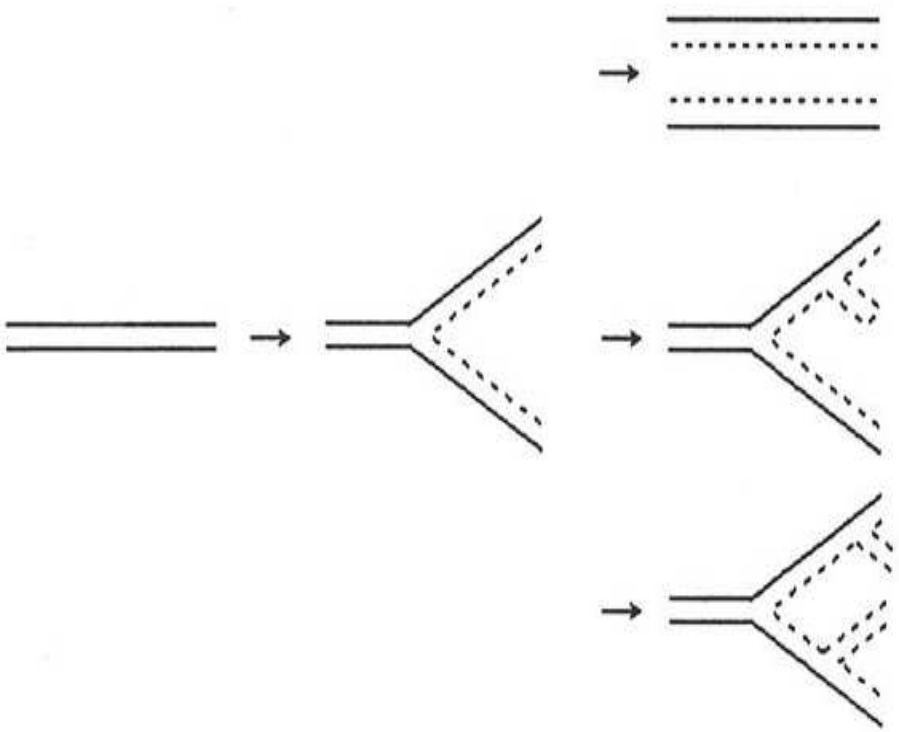
- Several heritable disorders have been associated with dynamic increases of the number of repeats of DNA-triplets in certain regions of human genome. In two to three subsequent generations, the transitions from normal individuals to non-affected or mildly-affected carriers, and then to full-blown disease, occur. Examples:
  - The fragile X syndrome, caused by a mutation of the FMR-1 gene characterized by expansion of the $(CCG)_n$ repeats (normal 6-60, carrier, 60-200, affected >200 repeats).
  - Myotonic dystrophy, caused by a mutation of the DM-1 autosomal gene characterized by expansion of the $(AGC)_n$ repeats (normal 5-27, affected >50 repeats).

Important questions that have not been fully answered are:

1. What is the mechanism of fluctuation of the number of repeat sequences in normal people (not in affected families)?
2. What is the mechanism of the modest increase in repeat sequences in unaffected carriers?
3. What is the mechanism of the rapid expansion of the number of repeat sequences in affected progeny within one or two generations?

Proposed biological mechanism to answer questions 1, 2 and 3

Formation of branching hairpin structures during DNA replication, facilitated by GC-rich repeats.



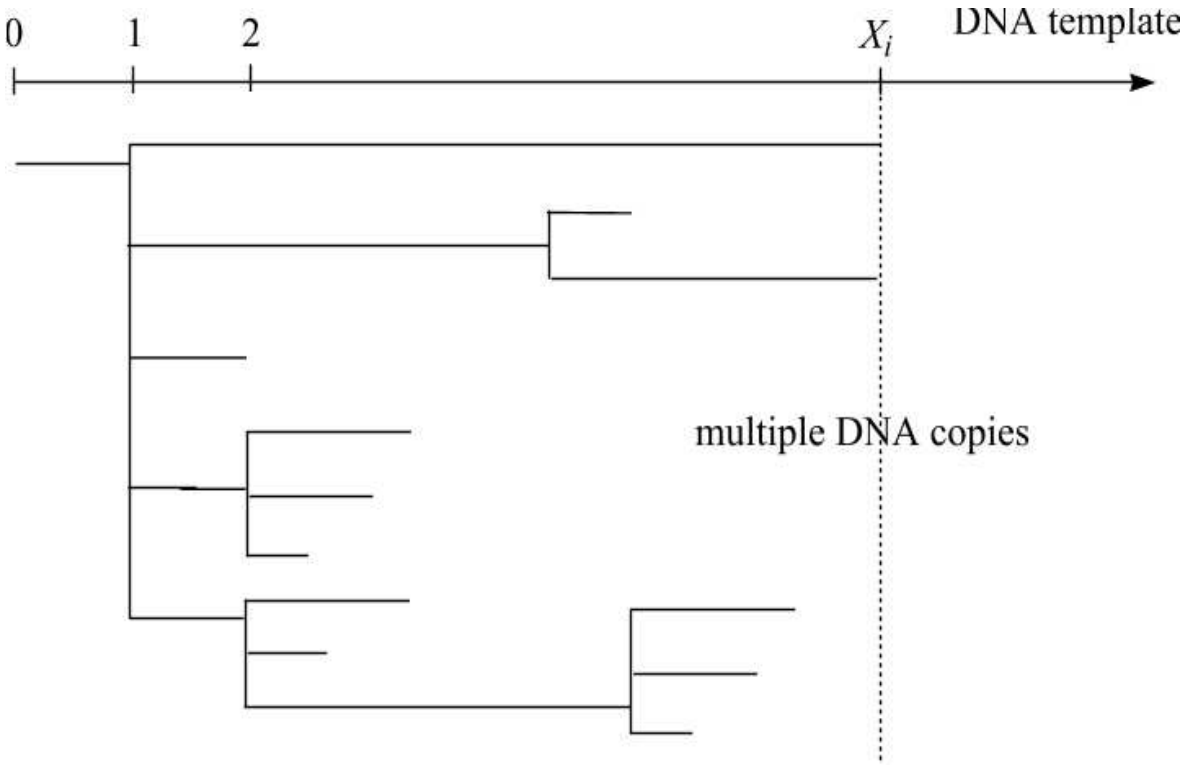
Definition of the IGW model

Assume a specific scenario of expansion of repeats:

- In the initial, 0-th, replication round, the number of repeats is n.
- In each new DNA replication round, a random number of new branching events (i.e., "initiation without termination of replication" events) occur at the endpoint of each repeat (this random number is characterized by the pgf $f(s)$), and

- All resulting branches become resolved and reintegrated into the linear DNA structure, which becomes the template for the successive replication round.
- A precedent exists in the replication of the T4 bacteriophage.
  - Virus induces production, in the host cell, of branched networks of concatenated DNA
  - Branched DNA subsequently is resolved into unbranched phage genomes

Attention **: Branching along DNA repeat sequence, not in time (!)**



$$X_{i+1} = \sum \{\text{lengths of branches in a tree of height } X_i\} = Y_{X_i - 1}$$

Definition of the IGW process

- Let

$$\{Z_k, k \geq 0\}$$

be the ordinary Galton-Watson process with pgf $f(s)$, and let

$$\{Y_k, k \geq 0\}, \text{ where } Y_k = Z_0 + Z_1 + \ldots + Z_k, \ , \ k = 0, 1, 2, \ldots$$

be the total progeny process.

- Let $\{Z_k^{(i)}, k \geq 0\}, \ i \geq 0$ be a sequence of iid copies of $\{Z_k\}$ with total progeny processes $\{Y_k^{(i)}, k \geq 0\}, \ i \geq 0$. Generic $\{Z_k\}$ is called the underlying GW process.

- Process $\{X_i, i \geq 0\}$ is defined in a recursive manner,

$$X_0 = n, \ X_{i+1} = Y_{X_i - 1}^{(i)}, \ i \geq 0.$$

Sequence $\{X_i\}$ is Markov and, since $Y_0^{(i)} = 1$, state 1 is absorbing.

Plausible example of IGW

- Suppose that at the end of each repeat a new branching event occurs with small probability p, so that

$$f(s) = (1-p)s + ps^2.$$

- The number of branches stemming from each ramification point is 1 or 2, the latter less likely.
  - This leads to a "sparse" tree $\Longrightarrow$ for a number of generations the growth will be slow.

Binomial thinning of IGW

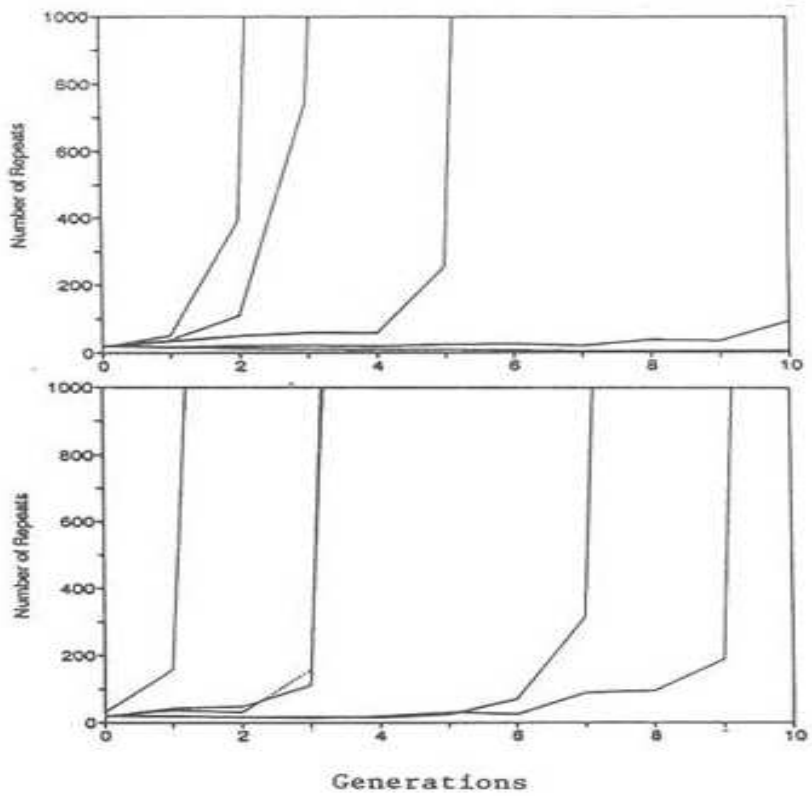
- Fluctuations of the number of triplets in *unaffected individuals* can be explained by coexistence of triplet increase and triplet loss.
- Accordingly, we assume that the process of resolution and reincorporation of repeats into the linear chromosomal structure has a limited efficiency $u < 1$.
- The new process $\{\tilde{X}_i, i \geq 0\}$ including the imperfect efficiency is defined as

$$\tilde{X}_0 = n,$$

$$\tilde{X}_{i+1} = B(u, Y^{(i)}_{\tilde{X}_i - 1} - 1) + 1, \ i \geq 0,$$

where, conditional on N, $B(u, N)$ is a binomial rv with parameters u and N.

This process produces runs of fluctuations, followed by explosive growth.


1000
800
600
400
200
0
Number of Repeats
0
2
4
6
8
10
Generations

Properties of the IGW process

$\mathrm{P}[\{X_i \to 1\} \cup \{X_i \to \infty\}] = 1$. Let X_∞ denote the almost sure limit of X_i and let $g(s, \nu)$ denote the pgf of Y_ν. Then

$$g(s, 0) = s \text{ and } g(s, \nu + 1) = sf[g(s, \nu)]$$

It follows from the definition that

$$\mathrm{E}(s^{X_{i+1}}) = \mathrm{E}[g(s, X_i - 1)],$$

and hence in all cases

$$\mathrm{E}(s^{X_\infty}) = \mathrm{E}[g(s, X_\infty - 1)]. \tag{4}$$

If

$$0 < p_0 < 1,$$

we may choose $s \in (0, q)$ and then $f(s) > s$. This gives
$g(s, 1) > sf(s) > s$ and hence, by induction, that $g(s, \nu - 1) > s^\nu$.

Since (4) can be written as

$$s + \mathrm{E}(s^{X_\infty}, X_\infty > 1) = s + \mathrm{E}[g(s, X_\infty - 1), X_\infty > 1],$$

it can hold if and only if $\mathrm{P}[X_\infty > 1] = 0$ and so the process is absorbed at unity when $0 < p_0 < 1$.

If $p_0 = 0$, then $Y_\nu^{(i)} > \nu + 1$ and hence definition implies $X_{i+1} \geq X_i$.
So, $X_i \uparrow \infty$ if $X_0 \geq 2$.

Theorem 5. *Let us consider the IGW process with no thinning (i.e., with $u = 1$). Then*

1. *$m < 1$ yields $E(X_i) \to 1$ and $X_i \stackrel{a.s.}{\to} 1$;*
2. *$m = 1$ yields $E(X_i) = E(X_0)$ and $X_i \stackrel{a.s.}{\to} X_\infty$ where X_∞ is a finite rv, and $X_\infty = 1$ if $p_0 < 1$;*
3. *$m > 1$ yields $E(X_i) \to \infty$, and*
   (a) *if $p_0 > 0$ then $X_i \stackrel{a.s.}{\to} 1$,*
   (b) *if $p_0 = 0$ i.e., $f(s) = p_1 s + p_2 s^2 + \cdots$, then $X_i \stackrel{p}{\to} \infty$.*

Theorem 6. *Suppose* $\{\tilde{X}_n\}$ *is the IGW process with binomial thinning.*

1. *Suppose* $m > 1$. *For each integer* $M > 0$, *there exists an integer* $N_0 > 0$ *such that*

$$E(\tilde{X}_{i+1}|\tilde{X}_i = N_0) > MN_0.$$

2. *Suppose* $u > 1/2$ *and* $p_0 = 0$. *There exist* $N_0 \geq 0$ *and* $\alpha > 1$ *such that*

$$E(\tilde{X}_{n+1}|\tilde{X}_n \geq N_0) \geq \alpha E(\tilde{X}_n - 1|\tilde{X}_n \geq N_0).$$

- Theorem shows that no matter how small the efficiency u in the process with thinning, the process will increase (in the expected value sense) by an arbitrary factor, after it exceeds certain threshold.

The age-dependent process: Markov case

- A single ancestor particle is born at $t = 0$.
- Ancestor lives for time τ which is exponentially distributed with parameter λ

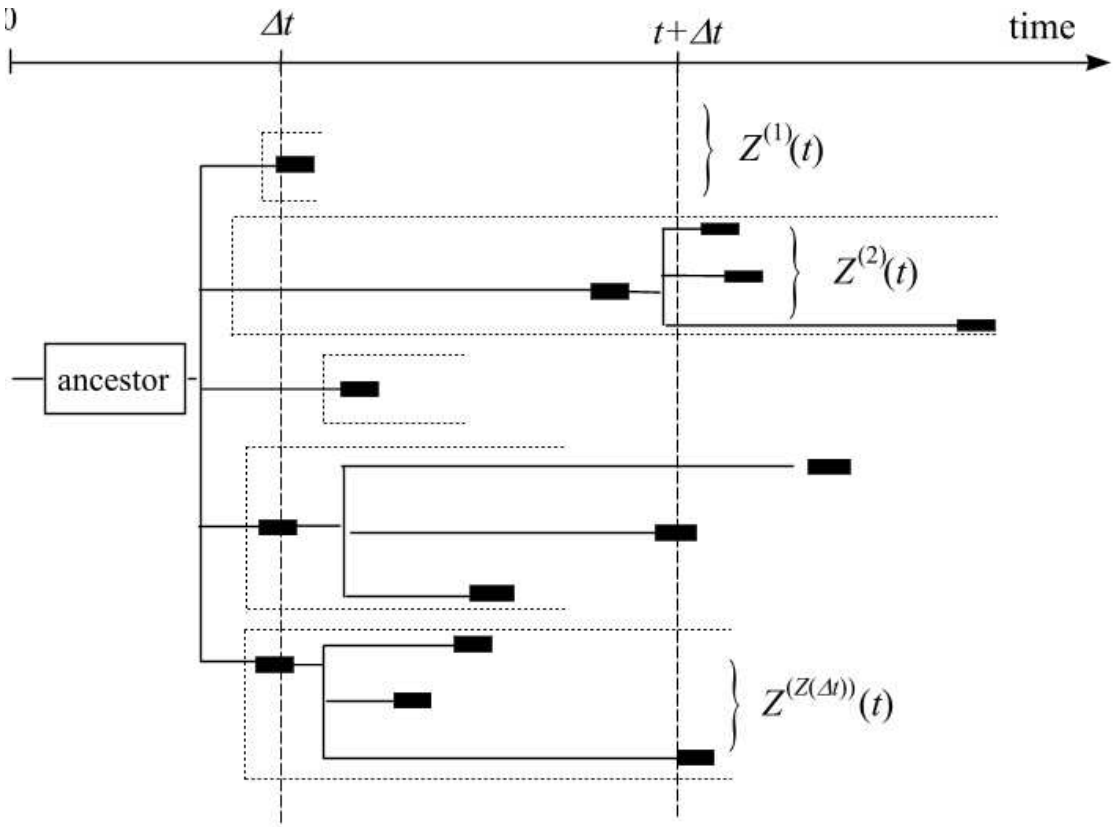
$$\tau \sim \exp(\lambda)$$

- At the moment of death the particle produces a random number of progeny according to a probability distribution with pgf $f(s)$.
- Each of the first generation progeny behaves, independently of each other, in the same way as the initial particle.

If we denote $Z(t, \omega)$ the particle count at time t, we obtain a stochastic process

$$\{Z(t, \omega),\ t \geq 0\}.$$

Differential Equation for the pgf



Differential Equation for the pgf

From the branching property

$$Z(t+\Delta t) = \sum_{i=1}^{Z(\Delta t)} Z^{(i)}(t) \quad \Longrightarrow \quad F(s,t+\Delta t) = F[F(s,t),\Delta t].,$$

$$F(s,t+\Delta t) - F(s,t) = F[F(s,t),\Delta t] - F[F(s,t),0]. \qquad (5)$$

If Δt is small

$$F(s,\Delta t) = se^{-\lambda\Delta t} + f(s)(1-e^{-\lambda\Delta t}) + o(\Delta t), \tag{6}$$

or

$$F(s,\Delta t) - F(s,0) = [-s + f(s)](1-e^{-\lambda\Delta t}) + o(\Delta t). \tag{7}$$

Substituting (7) into (5) and dividing by Δt we obtain

$$\frac{F(s,t+\Delta t) - F(s,t)}{\Delta t} = \frac{\{-F(s,t) + f[F(s,t)]\}(1-e^{-\lambda\Delta t}) + o(\Delta t)}{\Delta t}.$$

$$\Delta t \to 0 \implies dF(s,t)/dt = -\lambda\{F(s,t) - f[F(s,t)]\},\ F(s,0) = s.$$

Application **: Clonal resistance theory of cancer cells**

- Developed by Coldman and Goldie, the *only mathematical theory* that had any *impact* on practice of cancer chemotherapy.
- The aim of *cancer chemotherapy* is to achieve remission, i.e., disappearance of clinically detectable cancers and then to prevent relapse, i.e., the regrowth of cancer.
- In many cases the failure of chemotherapy is associated with the growth of cells resistant to further treatment with the same drug.

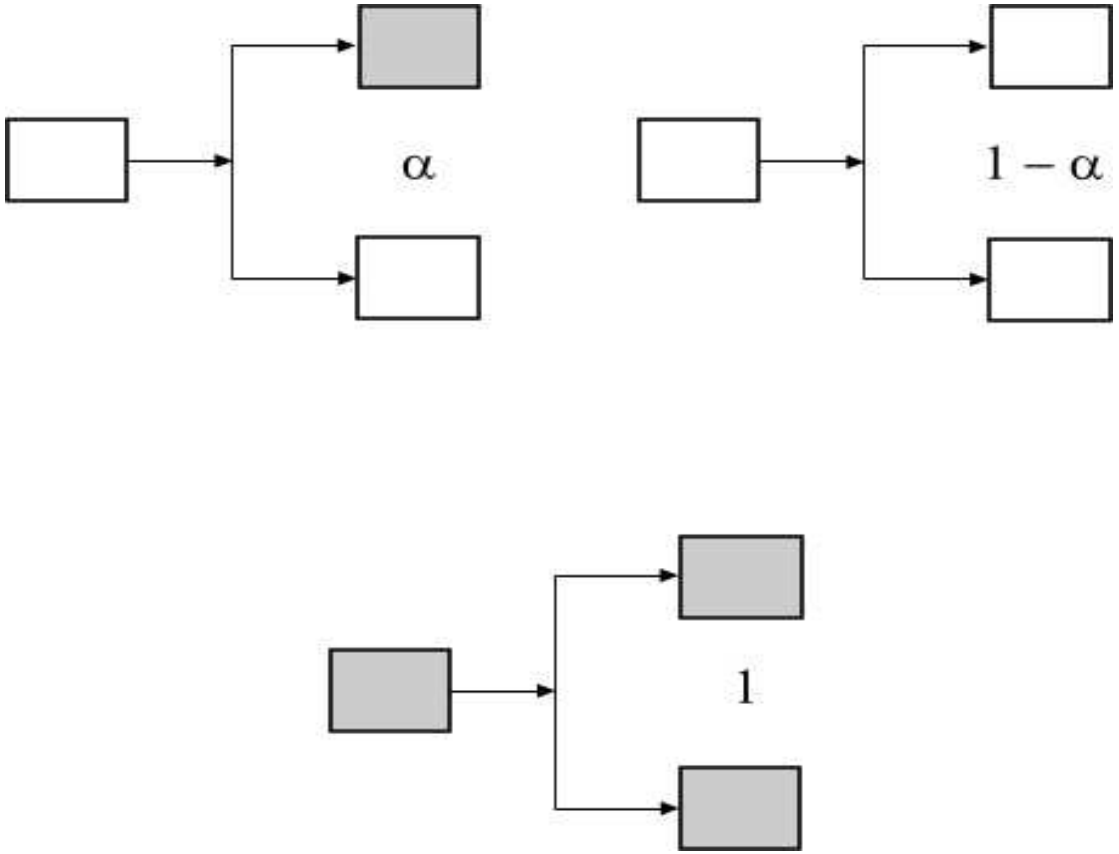
- Two modes of drug resistance:
  - Resistant cells might spontaneously arise in tumors and be selected for during treatment.
  - Alternatively, they might be induced by treatment.
- We focus on the first possibility

Assumptions of the clonal theory

1. Cancer cell population is initiated by a single cell which is sensitive to the cytotoxic (chemotherapeutic) agent. The population proliferates without losses.
2. Interdivision time of cells is a random variable with a given distribution.
3. At each division, with given probability, a single progeny cell mutates and becomes resistant to the cytotoxic agent.
4. Mutations are irreversible.

- ***Aim*** **:** Compute the probability that when the tumor is discovered, it does not contain resistant cells.
  - Only in such situation, the use of a cytotoxic agent is effective.
  - If a subpopulation of resistant cells exists, the cancer cell population will eventually re-emerge despite the therapy.

Markov branching process model: Single-mutation case



Markov branching process model: Single-mutation case

1. In the process, there exist two types of particles, labeled 0 (sensitive) and 1 (resistant).
2. The process is initiated by a single type 0 particle.
3. The lifespans of particles are independent exponential random variables with parameter λ.
4. Each particle, at death, divides into exactly two progeny particles:
   - 0-particle produces either two 0-particles, wp $1-\alpha$, or one 0− and one 1-particle, wp α.
   - 1-particle produces two 1-particles.

Notation and the ODE system

- $F_0(s_0, s_1; t)$ pgf of the numbers of cells of both types, at time t in the process initiated at time 0 by a type 0 cell.
- $F_1(s_1; t)$ is the pgf of the numbers of cells of type 1, at time t in the process initiated at time 0 by a type 1 cell.
- Progeny pgf's for cells of type 0 and 1

$$f_0(s_0, s_1) = (1-\alpha)s_0^2 + \alpha s_0 s_1, \quad f_1(s_1) = s_1^2$$

In consequence,

$$\frac{dF_0}{dt} = -\lambda F_0 + \lambda f_0(F_0, F_1) = -\lambda F_0 + \lambda[(1-\alpha)F_0^2 + \alpha F_0 F_1],$$

$$\frac{dF_1}{dt} = -\lambda F_1 + \lambda f_1(F_1) = -\lambda F_1 + \lambda F_1^2.$$

Solutions

Theorem 7. *The solution of the differential equation*

$$\frac{dF(t)}{dt} = f(t)F(t) + hF(t)^2, \tag{8}$$

where $f \in C[0,\infty)$, with initial condition $F(0)$, is a uniquely defined function $F \in C^1[0,\infty)$

$$F(t) = \frac{F(0)e^{\int_0^t f(u)du}}{1 - hF(0)\int_0^t e^{\int_0^u f(v)dv}du}. \tag{9}$$

Second equation is solved by separation of variables,

$$F_1(s;t) = \frac{s_1}{s_1 + (1 - s_1)e^{\lambda t}}.$$

Substituting this into the first equation and employing Theorem we obtain

$$F_0(s;t) = \frac{s_0 e^{-\lambda t}[e^{-\lambda t}s_1 + (1 - s_1)]^{-\alpha}}{1 + s_0\{[e^{-\lambda t}s_1 + (1 - s_1)]^{1-\alpha} - 1\}s_1^{-1}}.$$

Conclusions

Differentiating $F_0(s;t)$ with respect to s_0 and s_1 we obtain expected counts of the sensitive and resistant cells

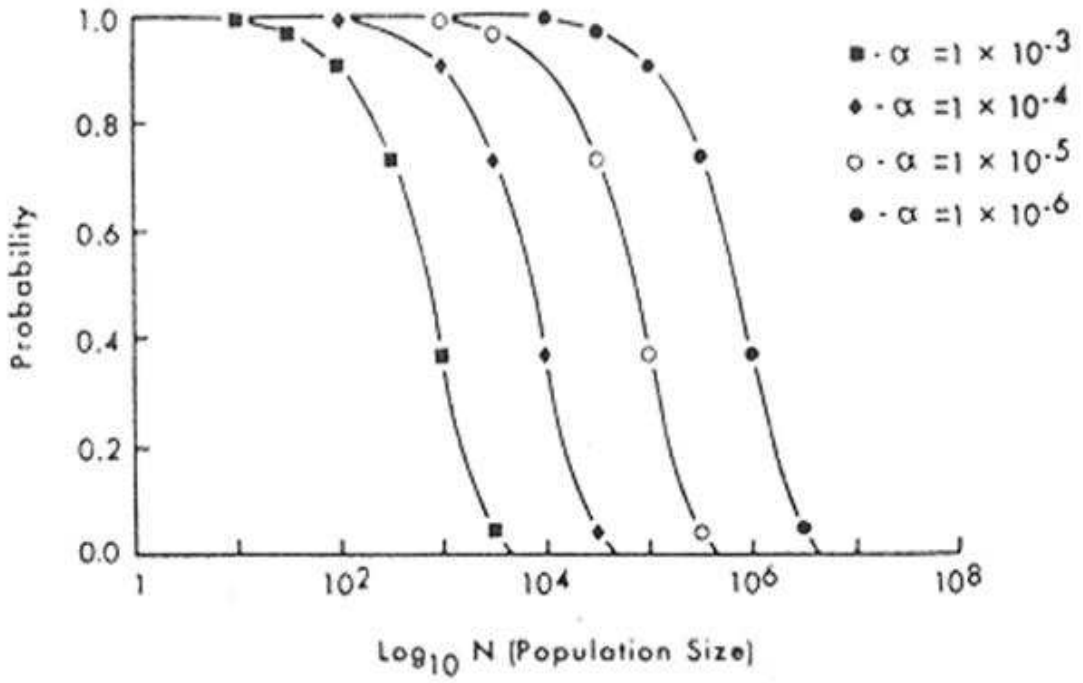
$$M_0(t) = \frac{\partial F(1,1;t)}{\partial s_0} = e^{\lambda(1-\alpha)t}, \quad t \geq 0,$$

$$M_1(t) = \frac{\partial F(1,1;t)}{\partial s_1} = e^{\lambda t} - e^{\lambda(1-\alpha)t}, \quad t \geq 0.$$

Conclusion: In absence of intervention, resistant cells eventually outgrow the sensitive ones.

Probability of no resistant cells at time t

$$P(t) = \lim_{s_0 \uparrow 1} \lim_{s_1 \downarrow 0} F_0(s;t) = \frac{1}{(1-\alpha) + \alpha e^{\lambda t}} = \frac{1}{(1-\alpha) + \alpha[M_0(t) + M_1(t)]}.$$


1.0
0.8
0.6
0.4
0.2
0.0
Probability
1
10^2
10^4
10^6
10^8
$\log_{10}$ N (Population Size)
■ - α =1 × 10^{-3}
♦ - α =1 × 10^{-4}
○ - α =1 × 10^{-5}
● - α =1 × 10^{-6}

Conclusions

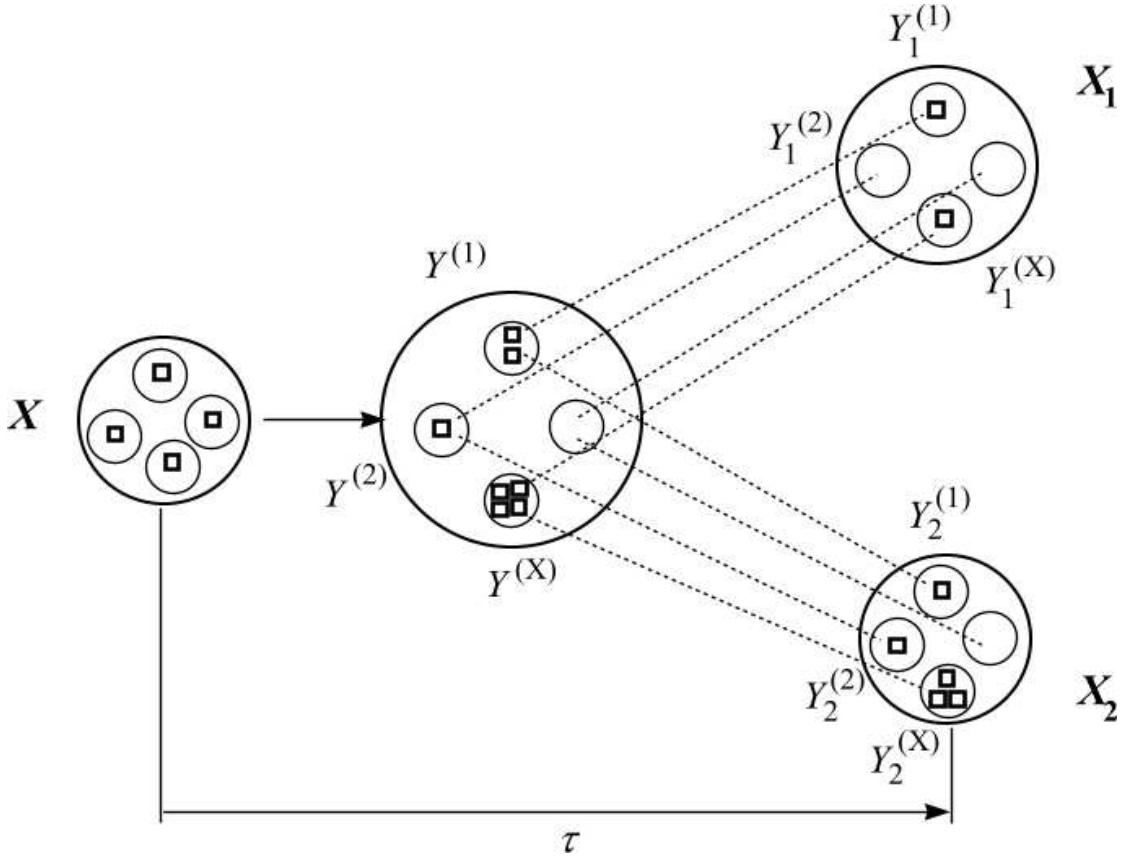
Based on the model, the following observations can be made:

- The probability that there are no resistant cells at time t is inversely related to the total number of cells.
- For different mutation rates α, if α's are small, the plots of $P(t)$ are approximately shifted, with respect to each other, along the t axis.
- The time interval in which the resistant clone is likely to emerge, i.e., in which $P(t)$ falls from near 1 to near 0, for example from 0.95 to 0.05, constitutes a relatively short "window" (Fig. 4.3).
- Therefore, the therapy should be prompt and radical to decrease cell number and probability $(1 - P(t))$ of emerging resistance.

Quasistationarity in a branching model of division-within-division

- Examples of *branching-within-branching* occurs in various settings in cell and molecular biology.
  - Gene amplification in cancer cells
  - Plasmid dynamics in bacteria
  - Proliferation of viral particles in host cells.
- *General motivating idea:* Stability arising from selection superimposed on a random mechanism.
- Consider a set of large particles (biological cells), following a binary fission process.
- Each of the large particles is born containing a number of small particles, which multiply or decay during the large particle's lifetime.

- Arising population of small particles is then split between the two progeny of the large particle
- Suppose the presence of at least one small particle is necessary to ensure the viability of the large particle. This can be due to a selection factor existing in the environment.
- We are interested in the behavior of the population of large particles surviving selection, i.e., large particles having at least one small particle inside.


$Y_1^{(1)}$
X_1
$Y_1^{(2)}$
$Y^{(1)}$
$Y_1^{(X)}$
X
$Y^{(2)}$
$Y_2^{(1)}$
$Y^{(X)}$
$Y_2^{(2)}$
X_2
$Y_2^{(X)}$
τ

Definition of the process

1. Population of large particles evolves according to a binary-fission time-continuous Markov branching process, i.e., each particle lives for a random time $\tau \sim \exp(\lambda)$, and then splits into two progeny, each of which independently follows the same scenario.

2. Each large particle contains X small particles at its birth. Each of these proliferates producing

$$Y^{(1)}, Y^{(2)}, \ldots, Y^{(X)},$$

small particle progeny at the end of the large particle's lifetime.

3. Each of the $Y^{(k)}$ progeny of the initial $k-$th small particle is independently split between the progeny of the large particle, so that large progeny 1 and 2 receive correspondingly $Y_1^{(k)}$ and $Y_2^{(k)}$ small progeny.

4. The joint distributions of the pairs $(Y_1^{(k)}, Y_2^{(k)})$ are identical, independent for all (k), and symmetric in $Y_1^{(k)}$ and $Y_2^{(k)}$. They are described by the pgf

$$f_{12}(s_1, s_2) = \mathrm{E}[s_1^{Y_1^{(1)}} s_2^{Y_2^{(1)}}].$$

5. As a result, each of the large progeny receives the total of

$$X_1 = \sum_{k=1}^{X} Y_1^{(k)} \quad \text{and} \quad X_2 = \sum_{k=1}^{X} Y_2^{(k)}$$

small progeny particles.

Resulting branching process and its properties

- Markov time-continuous process with denumerable infinity of types of large particles. The large particle is *of type* i if it contains i copies of small particles at its birth.
- Denote the vector of counts of large particles of all types at time t, by

$$Z(t) = [Z_0(t), Z_1(t), Z_2(t), \ldots],$$

and the infinite matrix of expected values $M(t) = [M_{ij}(t)]$ by

$$M_{ij}(t) = E[Z_j(t)|Z_i(0) = 1,\ Z_k(0) = 0,\ k \neq i].$$

- Let us define coefficients $a_{nm}(i)$ using the expansion of the pgf of the sums of numbers of small particles, given $X = i$,

$$[f_{12}(s_1, s_2)]^i = \sum_{n,m \geq 0} a_{nm}(i) s_1^n s_2^m.$$

  $a_{nm}(i)$ is equal to the probability that among the progeny of the i small particles present at birth of the large particle, n will end in large progeny 1 and m will end in large progeny 2.

- The expected value equations

$$\frac{d}{dt}M(t) = \lambda(2A - I)M(t),\ M(0) = I,$$

where $A = [A_{ij}] = [a_j(i)]$ is the matrix of coefficients of the marginal pgf of X_1 given $X = i$

$$[f(s_1)]^i = [f_{12}(s_1, 1)]^i = \sum_{j,l\geq 0} a_{jl}(i)s_1^j = \sum_{j\geq 0} a_j(i)s_1^j,$$

$a_j(i)$ is equal to the probability that of the i small particles present in the large particle at its birth, j will end in large progeny 1 (or in large progeny 2).

- Equations can be explicitly solved using the Laplace transform. The solution can be expressed in the form of generating function

$$M_k(u,t) = \sum_{l \geq 0} M_{kl}(t) u^l, \ u \in [0,1].$$

We obtain

$$M_k(u,t) = \sum_{j \geq 0} \frac{(2\lambda t)^j}{j!} [f_j(u)]^k e^{-\lambda t}, \ k \geq 0.$$

where $f_j(u)$ is the j-th iterate of the marginal pgf of $Y_1^{(1)}$.

Quasistationarity

Back to the Galton-Watson process with progeny pgf $f(u)$:

- If $f'(1-) < 1$ (the subcritical case) then as $j \to \infty$,

$$\frac{f_j(u) - f_j(0)}{1 - f_j(0)} \to \mathcal{B}(u),$$

  i.e., conditionally on nonextinction, the process tends to a limit distribution. This behavior is known as quasistationarity.

- Moreover, as $j \to \infty$

$$f_j(u) - 1 \ \sim \ \rho^j Q(u),$$

where $\rho = f'(1-)$ and the function $Q(u)$ satisfies

$$\frac{Q(0) - Q(u)}{Q(0)} = \mathcal{B}(u),$$

with $Q(1) = 0$, $Q'(1-) = 1$, $Q(u) \leq 0$ and $Q(u)$ increasing for $u \in [0, 1]$.

- Functions $\mathcal{B}(u)$ and $Q(u)$ are unique solutions of certain functional equations.

Theorem 8. *Let us consider the process defined in Section* **??** *started by a large ancestor of type k and let $\rho = f'(1-) < 1$. Then, as $t \to \infty$,*

$$e^{\lambda t} - M_k(u,t) \ \sim \ -kQ(u)e^{(2\rho-1)\lambda t}, \tag{10}$$

for all $k \geq 1$.

Corollary 1. *The expected frequencies $\{\mu_{kl}(t),\ l \geq 1\}$ of large particles of type l among the particles of nonzero type tend, as $t \to \infty$, to a limit distribution independent of k, characterized by the pgf $\mathcal{B}(u)$.*

Application : Gene amplification

- During cell's lifetime each extrachromosomal copy of the gene is successfully replicated with probability β, less than 1.
- The resulting two copies are segregated to the same progeny cell with probability α and to two different progeny cells with probability $1-\alpha$. α may be called *the probability of co-segregation* and has been showed to be ≈ 0.9 in one cell system.
- The above hypotheses yield

$$f_{12}(s_1, s_2) = \beta\left[(1-\alpha)s_1 s_2 + \frac{\alpha}{2}(s_1^2 + s_2^2)\right] + (1-\beta), \tag{11}$$

$$f(u) = \frac{\beta\alpha}{2}u^2 + \beta(1-\alpha)u + \left(\frac{\beta\alpha}{2} + 1 - \beta\right), \tag{12}$$

with $\rho = f'(1-) = \beta < 1$. Therefore our Theorem and its Corollary apply.

- Qualitatively, all the experimental observations above are explained by our results: The stable quasistationary distribution of gene copy count is predicted by the Corollary.
- If the type 0 cells are not removed by the drug, then the Theorem proves they dominate the population. Indeed by the Theorem the resistant cells grow as

$$\sum_{l\geq 1} M_{kl}(t) = M_k(1,t) - M_k(0,t) \sim -kQ(0)e^{(2\rho-1)\lambda t},\ \rho < 1, \quad (13)$$

  while the entire population grows as $e^{\lambda t}$.
- If $\rho > 1/2$, then in the presence of the drug, resistant cells grow as $e^{(2\rho-1)\lambda t}$, i.e., exponentially but slower than in the nonselective conditions.