A Markov-chain model of chromosomal instability

Sergi Elizalde

Dartmouth College

Joint with Sam Bakhoum and Ashley Laughney,

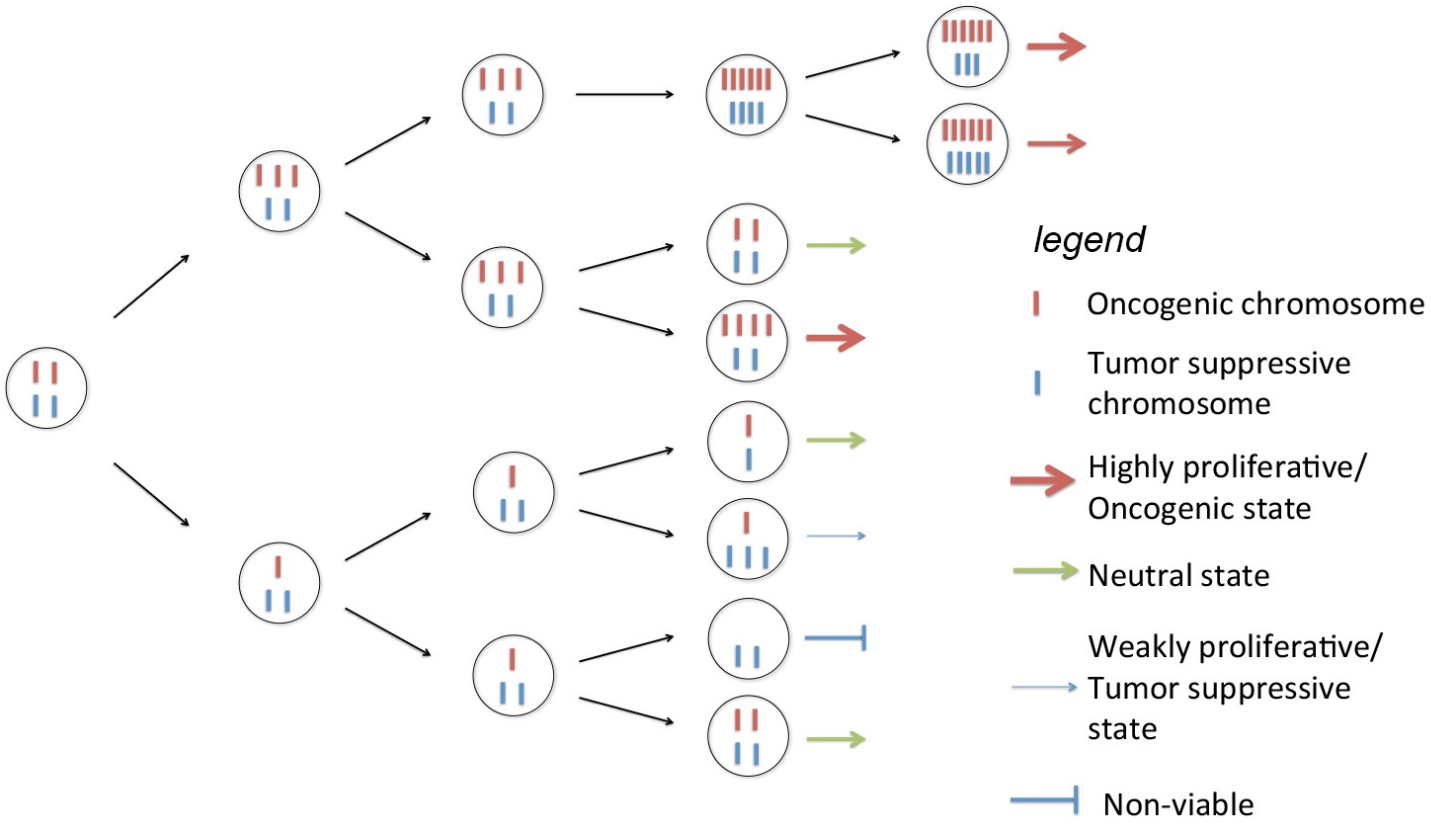
Memorial Sloan Kettering Cancer Center

Dagstuhl Seminar on Genomics, Pattern Avoidance, and Statistical Mechanics

November 2018

Missegregation

During mitosis, cancer cells undergo chromosome missegregation events, causing one of the two daughter cells to inherit more copies of a chromosome than the other.



Advantages of genomic instability

A cell dies if it loses all copies of a chromosome or ends up with too many. In addition, each cell has some probability of spontaneously dying at any time.

Advantages of genomic instability

A cell dies if it loses all copies of a chromosome or ends up with too many. In addition, each cell has some probability of spontaneously dying at any time.

It has been observed that

- more copies of oncogenic chromosomes (with proliferative genes) increase the cell's chances of surviving, while
- more copies of tumor supressive chromosomes (with anti-profiferative genes) increase its chances of dying.

Advantages of genomic instability

A cell dies if it loses all copies of a chromosome or ends up with too many. In addition, each cell has some probability of spontaneously dying at any time.

It has been observed that

- more copies of oncogenic chromosomes (with proliferative genes) increase the cell's chances of surviving, while
- more copies of tumor supressive chromosomes (with anti-profiferative genes) increase its chances of dying.

A recent genomic analysis by Davoli et al. assigned scores to individual chromosomes based on the presence of such genes.

Since the karyotype of a cell affects its fitness level, genomic instability allows for Darwinian selection to occur.

History

The first stochastic model of missegregation was developed by Gusev, Kagansky and Dooley in 2000. It has a few disadvantages:

1. Simulations are very slow.
2. It can't be analyzed mathematically to find long-term behavior.
3. It doesn't account for chromosome scores, and consequently its predictions are unrealistic.

History

The first stochastic model of missegregation was developed by Gusev, Kagansky and Dooley in 2000. It has a few disadvantages:

1. Simulations are very slow.
2. It can't be analyzed mathematically to find long-term behavior.
3. It doesn't account for chromosome scores, and consequently its predictions are unrealistic.

To address 1 and 2, we will build a Markov chain model to study how the probability distribution of karyotypes evolves over time.

History

The first stochastic model of missegregation was developed by Gusev, Kagansky and Dooley in 2000. It has a few disadvantages:

1. Simulations are very slow.
2. It can't be analyzed mathematically to find long-term behavior.
3. It doesn't account for chromosome scores, and consequently its predictions are unrealistic.

To address 1 and 2, we will build a Markov chain model to study how the probability distribution of karyotypes evolves over time.

To address 3, we will assign a survival probability to each cell based on the chromosome scores computed by Davoli et al.

Assumptions of our model

- Each chromosome copy has probability p (typically $p \approx 0.0025$) of missegregating at a given cell division, independent from other copies.

Assumptions of our model

- Each chromosome copy has probability p (typically $p \approx 0.0025$) of missegregating at a given cell division, independent from other copies.
- If the number of copies of any chromosome reaches 0 or goes above N (typically $N = 8$), the cell dies.

Assumptions of our model

- Each chromosome copy has probability p (typically $p \approx 0.0025$) of missegregating at a given cell division, independent from other copies.
- If the number of copies of any chromosome reaches 0 or goes above N (typically $N = 8$), the cell dies.
- Starting from a single founder cell, all the cells in the colony divide simultaneously at each generation.

Assumptions of our model

- Each chromosome copy has probability p (typically $p \approx 0.0025$) of missegregating at a given cell division, independent from other copies.
- If the number of copies of any chromosome reaches 0 or goes above N (typically $N = 8$), the cell dies.
- Starting from a single founder cell, all the cells in the colony divide simultaneously at each generation.

The karyotype of a cell is the vector $(i_1, i_2, \ldots, i_{23})$ where

$$i_k = \#\text{ copies of chromosome } k.$$

A live cell has $1 \leq i_k \leq N$ for all k.

Simulations vs. Markov chain

One can implement this algorithm and run a forward simulation, keeping track of the karyotypes of all the cells in the colony. Unfortunately, that is extremely show.

Simulations vs. Markov chain

One can implement this algorithm and run a forward simulation, keeping track of the karyotypes of all the cells in the colony. Unfortunately, that is extremely show.

Instead, we will build a Markov chain that describes the distribution of karyotypes probabilistically. The advantages are:

- Computations are much faster, since they amount to taking powers of matrices.
- We can analyze the Markov chain mathematically to predict long-term behavior.

A key simplification

Since missegregations of different chromosomes are independent, we focus on one type of chromosome at a time.

A key simplification

Since missegregations of different chromosomes are independent, we focus on one type of chromosome at a time.

For each of the 23 chromosomes, we build a Markov chain with states $0, 1, 2, \ldots, N$. State i corresponds to cells with i copies of the chromosome. State 0 corresponds to (dead) cells with 0 or $> N$ copies.

A key simplification

Since missegregations of different chromosomes are independent, we focus on one type of chromosome at a time.

For each of the 23 chromosomes, we build a Markov chain with states $0, 1, 2, \ldots, N$. State i corresponds to cells with i copies of the chromosome. State 0 corresponds to (dead) cells with 0 or $> N$ copies.

The probability of a given karyoptype $(i_1, \ldots, i_{23})$ is obtained by multiplying the probability that the Markov chain corresponding to chromosome k is in state i_k for $1 \leq k \leq 23$.

A key simplification

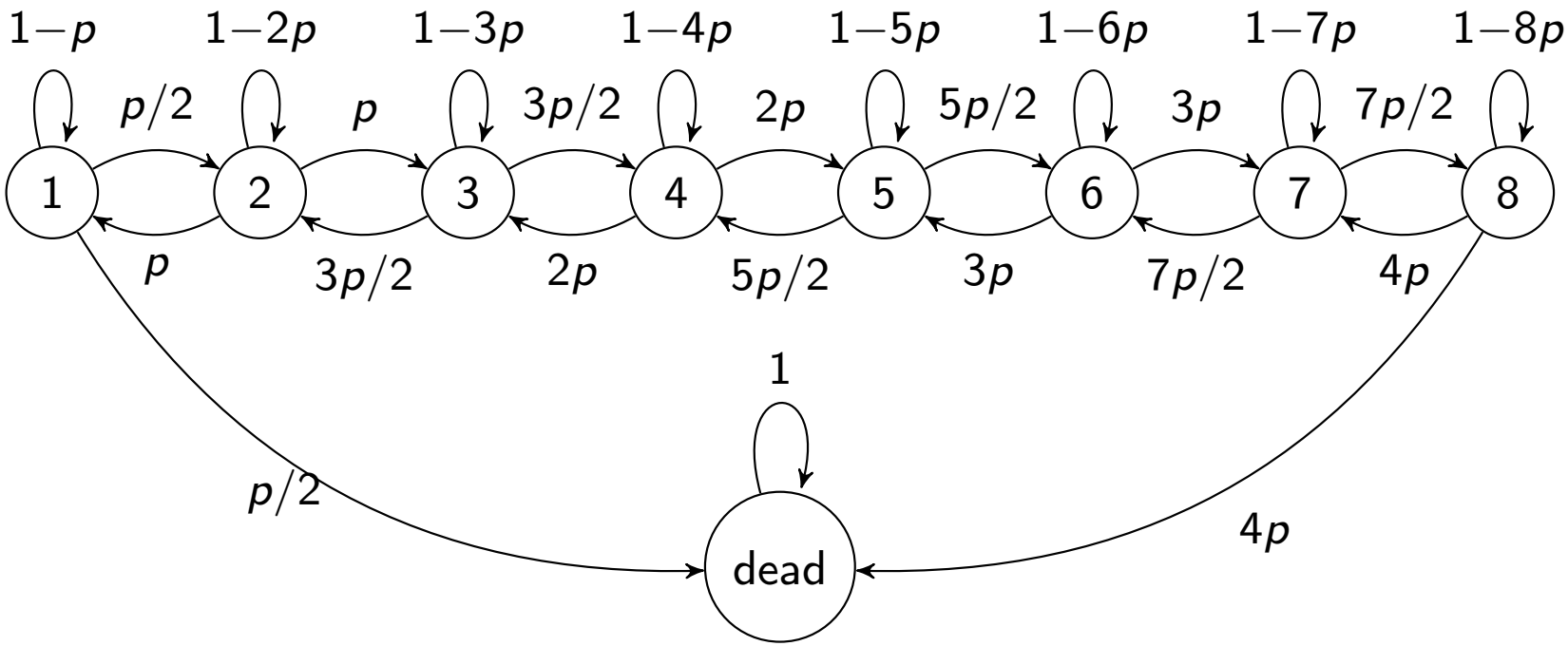
Since missegregations of different chromosomes are independent, we focus on one type of chromosome at a time.

For each of the 23 chromosomes, we build a Markov chain with states $0, 1, 2, \ldots, N$. State i corresponds to cells with i copies of the chromosome. State 0 corresponds to (dead) cells with 0 or $> N$ copies.

The probability of a given karyoptype $(i_1, \ldots, i_{23})$ is obtained by multiplying the probability that the Markov chain corresponding to chromosome k is in state i_k for $1 \leq k \leq 23$.

Each step is a cell division, and one of the 2 daughter cells is chosen with probability 1/2.

The Markov chain for the basic model without scores



(we ignore quadratic terms in p for simplicity)

The transition matrix

Ignoring quadratic terms in p, the transition matrix for the basic model (without chromosome scores) is

$$M_{ij} = \begin{cases} 1 - ip & \text{if } i = j, \\ ip/2 & \text{if } |i - j| = 1, \\ 0 & \text{if } |i - j| \geq 2, \end{cases}$$

for $1 \leq i, j \leq N$.

The transition matrix

Ignoring quadratic terms in p, the transition matrix for the basic model (without chromosome scores) is

$$M_{ij} = \begin{cases} 1 - ip & \text{if } i = j, \\ ip/2 & \text{if } |i-j| = 1, \\ 0 & \text{if } |i-j| \geq 2, \end{cases}$$

for $1 \leq i, j \leq N$.

Being more precise, M_{ij} is the coefficient of x^j in

$$\left(\frac{p}{2} + (1-p)x + \frac{p}{2}x^2\right)^i \approx \frac{ip}{2}x^{i-1} + (1-ip)x^i + \frac{ip}{2}x^{i+1} + [\text{terms involving } p^2].$$

The transition matrix

For example, for $N = 8$, we get

$$\left[\begin{array}{c|cccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline p/2 & 1-p & p/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p & 1-2p & p & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3p/2 & 1-3p & 3p/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2p & 1-4p & 2p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5p/2 & 1-5p & 5p/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3p & 1-6p & 3p & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7p/2 & 1-7p & 7p/2 \\ 4p & 0 & 0 & 0 & 0 & 0 & 0 & 4p & 1-8p \end{array}\right].$$

The transition matrix

For example, for $N = 8$, we get

$$\left[\begin{array}{c|cccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline p/2 & 1-p & p/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p & 1-2p & p & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3p/2 & 1-3p & 3p/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2p & 1-4p & 2p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5p/2 & 1-5p & 5p/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3p & 1-6p & 3p & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7p/2 & 1-7p & 7p/2 \\ 4p & 0 & 0 & 0 & 0 & 0 & 0 & 4p & 1-8p \end{array}\right].$$

Let $\mathbf{M}$ be the matrix obtained by removing the row and column corresponding to state 0.

The transition matrix

For example, for $N = 8$, we get

$$\mathbf{M} = \begin{bmatrix} 1-p & p/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ p & 1-2p & p & 0 & 0 & 0 & 0 & 0 \\ 0 & 3p/2 & 1-3p & 3p/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2p & 1-4p & 2p & 0 & 0 & 0 \\ 0 & 0 & 0 & 5p/2 & 1-5p & 5p/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3p & 1-6p & 3p & 0 \\ 0 & 0 & 0 & 0 & 0 & 7p/2 & 1-7p & 7p/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4p & 1-8p \end{bmatrix}.$$

Let $\mathbf{M}$ be the matrix obtained by removing the row and column corresponding to state 0.

Properties of the transition matrix

- State 0 is an absorbing state, and so a high proportion of the 2^g potential cells after g generations are dead. Still, we are interested in the distribution of copy numbers among live cells.

Properties of the transition matrix

- State 0 is an absorbing state, and so a high proportion of the 2^g potential cells after g generations are dead. Still, we are interested in the distribution of copy numbers among live cells.
- If $\mathbf{v}$ is a vector describing the initial distribution of the number of copies of a given chromosome, the vector $\mathbf{v}\mathbf{M}^g$, normalized so its entries sum to one, is the distribution of copy numbers among live cells after g generations.

Properties of the transition matrix

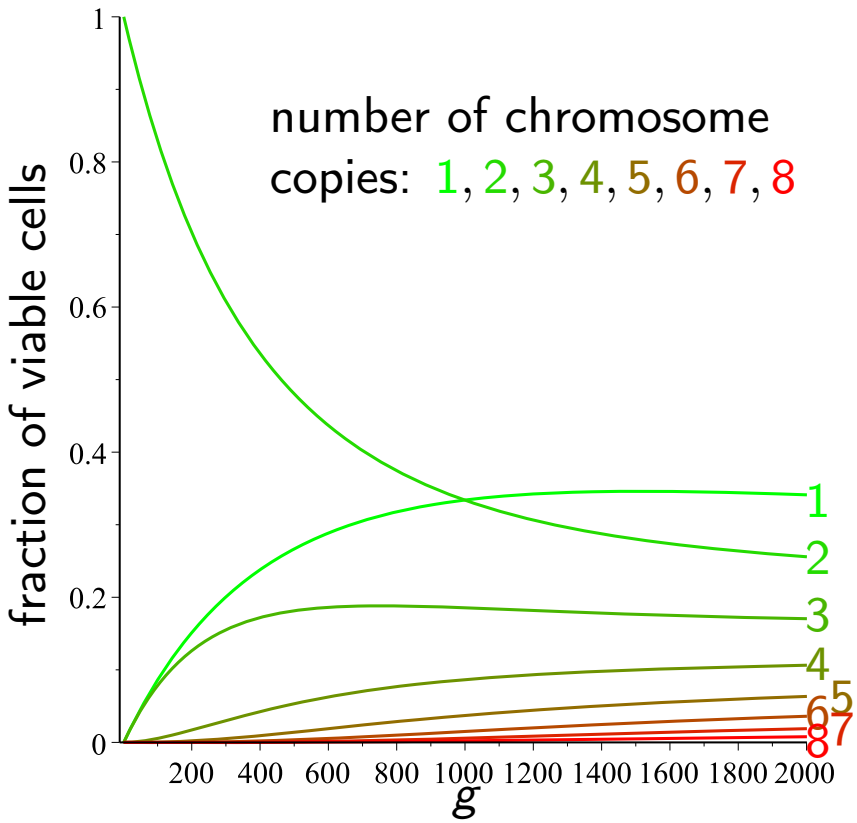
- State 0 is an absorbing state, and so a high proportion of the 2^g potential cells after g generations are dead. Still, we are interested in the distribution of copy numbers among live cells.
- If $\mathbf{v}$ is a vector describing the initial distribution of the number of copies of a given chromosome, the vector $\mathbf{v}\mathbf{M}^g$, normalized so its entries sum to one, is the distribution of copy numbers among live cells after g generations.
- Letting $s_g(i) =$ sum of the entries of the ith row of $\mathbf{M}^g$,

$$2^g \prod_{k=1}^{23} s_g(i_k)$$

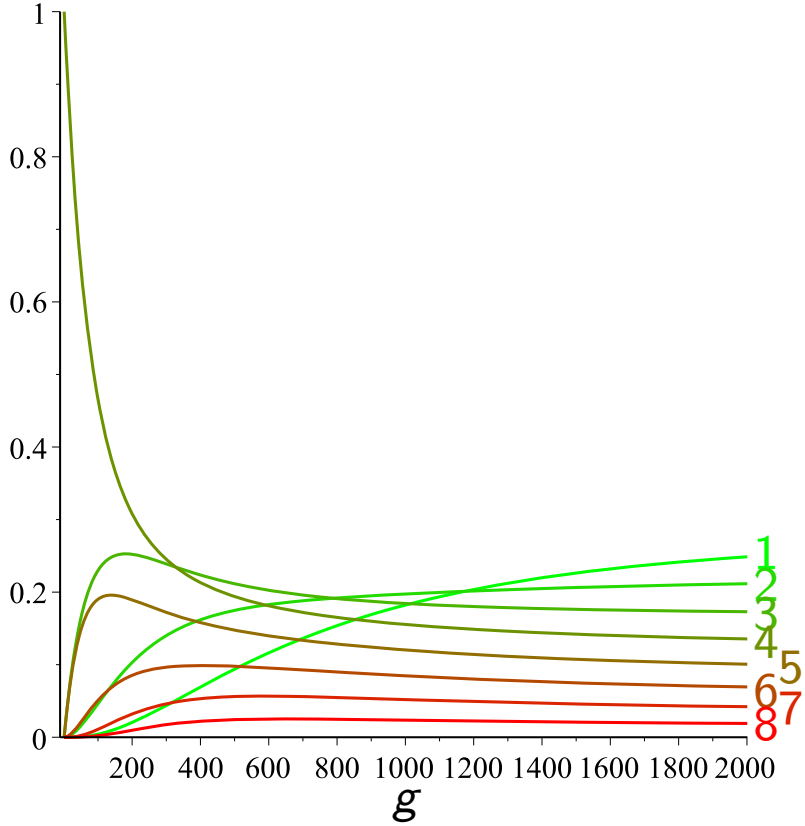
is the expected number of live cells after g generations when the founder cell has i_k copies of chromosome k for each k.

Evolution of the number of chromosome copies over time

Proportion of live cells having each number of copies, for the Markov chain model with $N = 8$ and a founder cell with f copies:



$p = 0.001,\ f = 2$



$p = 0.0025,\ f = 4$

The limiting behavior

We are interested in the limiting distribution of the copy numbers among live cells as $g \to \infty$.

The limiting behavior

We are interested in the limiting distribution of the copy numbers among live cells as $g \to \infty$.

Since the Markov chain has an absorbing state, its stationary distribution is trivial and unhelpful.

The limiting behavior

We are interested in the limiting distribution of the copy numbers among live cells as $g \to \infty$.

Since the Markov chain has an absorbing state, its stationary distribution is trivial and unhelpful.

However, we can use a result from probability theory to restrict to non-absorbing states (equivalently, live cells):

Theorem

Let ρ be the largest eigenvalue of $\mathbf{M}$. The limiting distribution conditional on the non-absorbing states is given by the vector $\mathbf{v}$ satisfying $\mathbf{vM} = \rho\mathbf{v}$ and $\sum_{i=1}^{N} v_i = 1$.

The limiting behavior

In particular, this limiting distribution does not depend on the karyotype of the founder cell.

The limiting behavior

In particular, this limiting distribution does not depend on the karyotype of the founder cell.

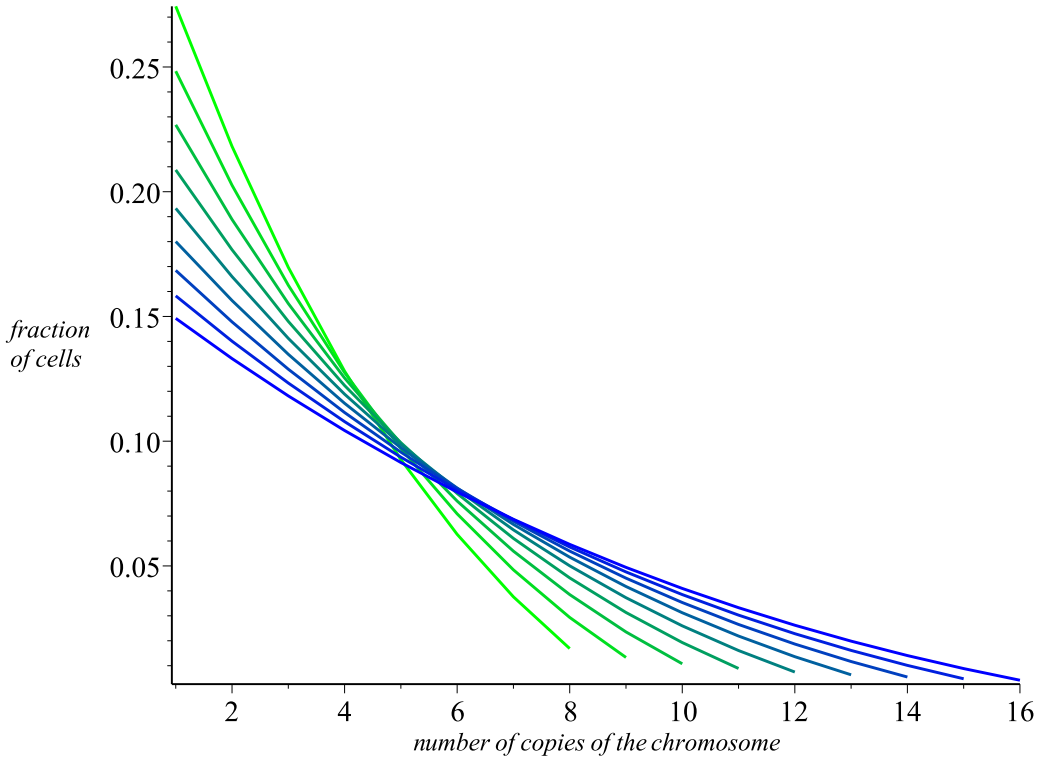
Surprisingly, it does not depend on the missegregation rate either:

Theorem

The limiting distribution of the above basic model conditional on the non-absorbing states is independent of p.

The limiting distribution

Limiting distributions for $N = 8, 9, 10, 11, 12, 13, 14, 15, 16$:



The most frequent copy number is always 1, which is not very realistic. This will change once we incorporate chromosome scores.

Chromosome scores and survival probability

Based on experiments by Davoli et al., we assign a score s_k to each chromosome k. The total score of a cell with karyotype $(i_1, \ldots, i_{23})$ is:

$$S = \sum_{k=1}^{23} s_k i_k,$$

Chromosome scores and survival probability

Based on experiments by Davoli et al., we assign a score s_k to each chromosome k. The total score of a cell with karyotype $(i_1, \ldots, i_{23})$ is:

$$S = \sum_{k=1}^{23} s_k i_k,$$

and its survival probability at a given generation is

$$Q_{\text{surv}} = e^{c+dS}$$

for some parameters c and $d > 0$.

Chromosome scores and survival probability

Based on experiments by Davoli et al., we assign a score s_k to each chromosome k. The total score of a cell with karyotype $(i_1, \ldots, i_{23})$ is:

$$S = \sum_{k=1}^{23} s_k i_k,$$

and its survival probability at a given generation is

$$Q_{\text{surv}} = e^{c+dS}$$

for some parameters c and $d > 0$.

Again, we can implement this algorithm and run lengthy simulations.

Instead, we'll incorporate the chromosome scores into the Markov chain, and use it to run fast computations and to determine limiting behavior.

Decomposing the survival probability

$$Q_{\text{surv}} = e^{c+dS} = e^{c+d\sum_k s_k i_k} = e^c \prod_{k=1}^{23} \underbrace{e^{ds_k i_k}}.$$

Decomposing the survival probability

$$Q_{\text{surv}} = e^{c+dS} = e^{c+d\sum_k s_k i_k} = e^c \prod_{k=1}^{23} \underbrace{e^{ds_k i_k}}_{q_k(i_k)}.$$

Let

$$q_k(i) = e^{ds_k i} = \mu^i$$

denote the contribution of chromosome k to the survival probability, where $\mu = e^{ds_k}$.

Decomposing the survival probability

$$Q_{\text{surv}} = e^{c+dS} = e^{c+d\sum_k s_k i_k} = e^c \prod_{k=1}^{23} \underbrace{e^{ds_k i_k}}_{q_k(i_k)}.$$

Let

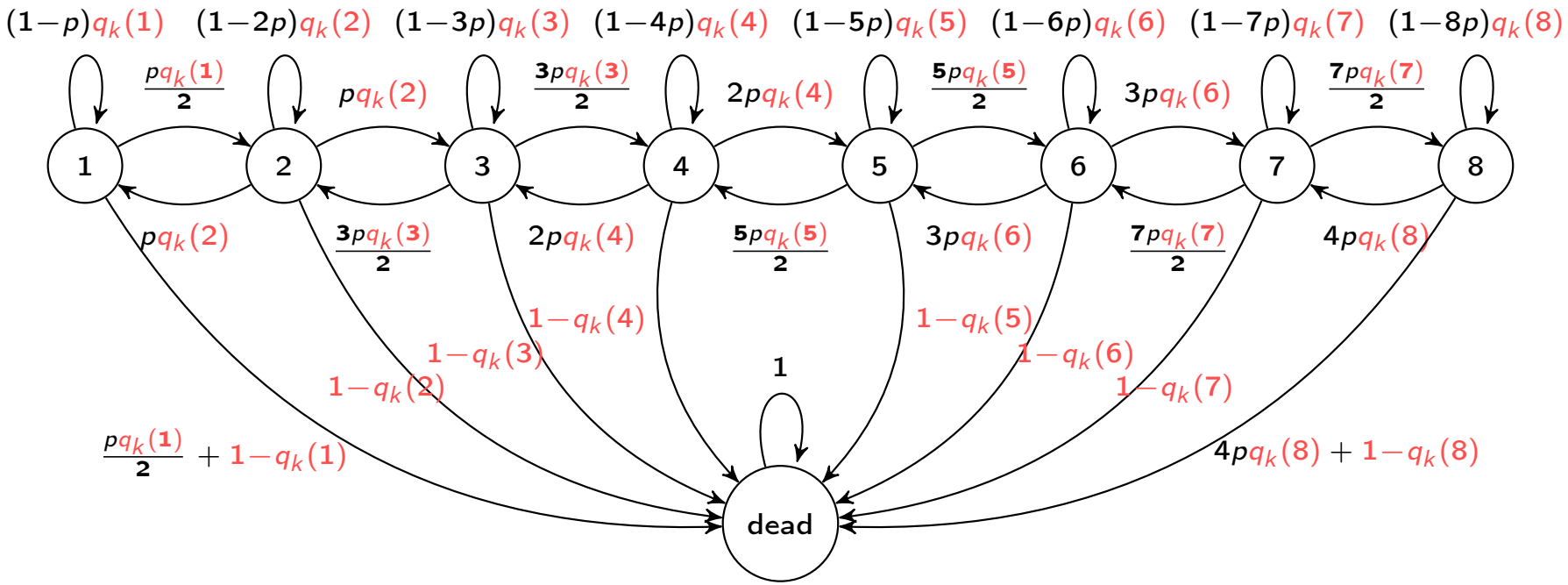
$$q_k(i) = e^{ds_k i} = \mu^i$$

denote the contribution of chromosome k to the survival probability, where $\mu = e^{ds_k}$.

Oncogenic $\Leftrightarrow s_k > 0 \Leftrightarrow \mu > 1$.
Tumor-suppressive $\Leftrightarrow s_k < 0 \Leftrightarrow \mu < 1$.

This equation allows us to break up the model into 23 independent Markov chains, one for each type of chromosome.

The Markov chain for chromosome k



A cell with i copies of the chromosome has probability $1-q_k(i)$ of dying, and probability $q_k(i)$ of surviving and dividing as in the basic model.

The transition matrix

The transition matrix $\mathbf{A}^{(k)}$ restricted to live cells is:

$$A_{ij}^{(k)} = \begin{cases} (1-ip)\, q_k(i) & \text{if } i = j, \\ ip\, q_k(i)/2 & \text{if } |i-j| = 1, \\ 0 & \text{if } |i-j| \geq 2, \end{cases}$$

for $1 \leq i, j \leq N$.

The transition matrix

The transition matrix $\mathbf{A}^{(k)}$ restricted to live cells is:

$$A_{ij}^{(k)} = \begin{cases} (1-ip)\, q_k(i) & \text{if } i = j, \\ ip\, q_k(i)/2 & \text{if } |i-j| = 1, \\ 0 & \text{if } |i-j| \geq 2, \end{cases}$$

for $1 \leq i, j \leq N$.

Letting $s_g^{(k)}(i) =$ sum of the entries of the ith row of $(\mathbf{A}^{(k)})^g$,

$$2^g \prod_{k=1}^{23} s_g^{(k)}(i_k)$$

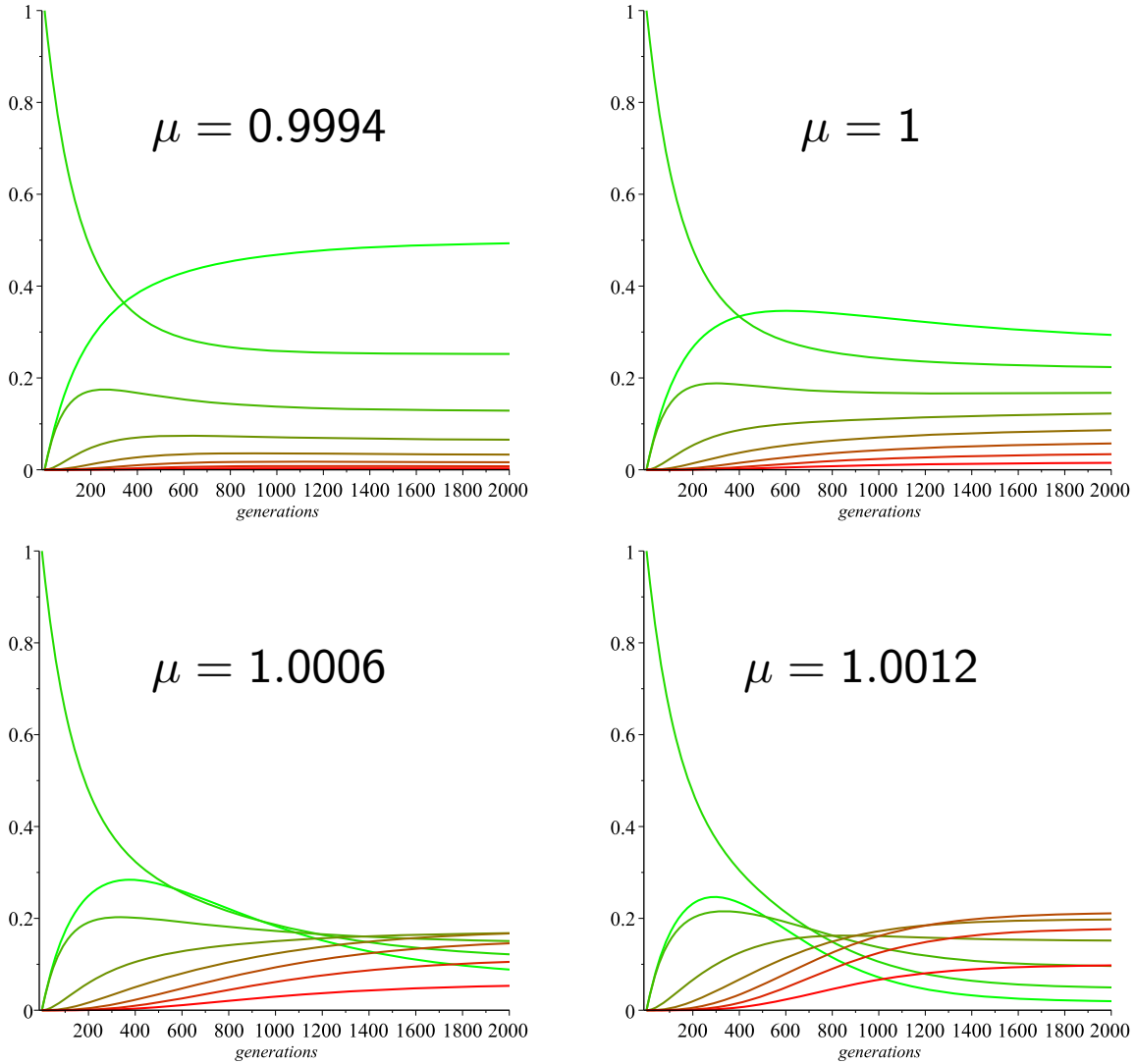
is the expected number of live cells after g generations when the founder cell has i_k copies of chromosome k for each k.

Distribution of the number of chromosome copies over time

In human chromosomes, $\mu \in [0.9994, 1.0012]$.

Fix $p = 0.0025$ and a founder cell with 2 copies.

Each curve represents a number of copies: 1, 2, 3, 4, 5, 6, 7, 8.

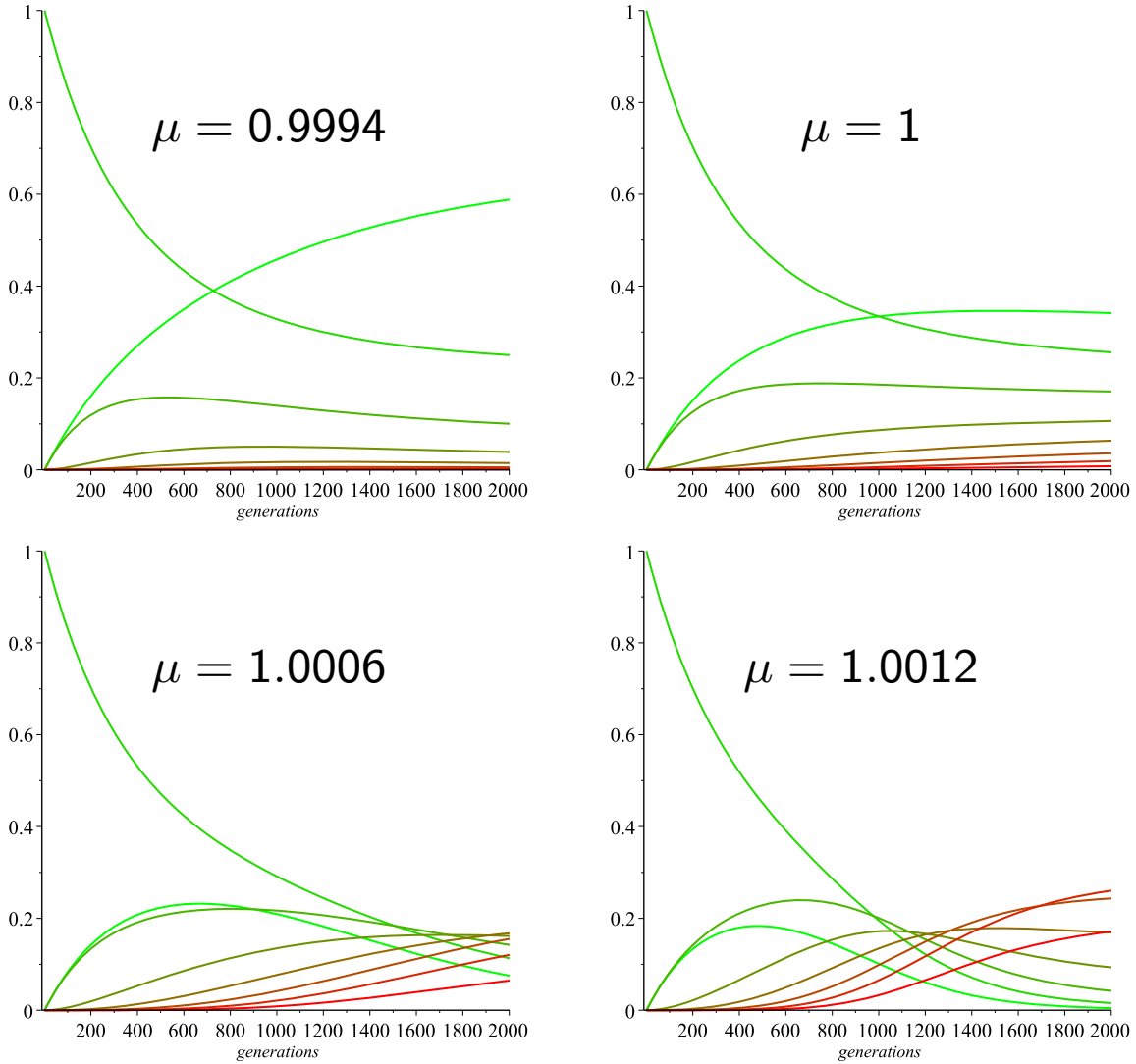


Distribution of the number of copies over time

In human chromosomes, $\mu \in [0.9994, 1.0012]$.

Fix $p = 0.001$ and a founder cell with 2 copies.

Each curve represents a number of copies: 1, 2, 3, 4, 5, 6, 7, 8.



The limiting behavior

As before, if ρ is the largest eigenvalue of $\mathbf{A}^{(k)}$, the limiting distribution conditional on the non-absorbing states is given by the vector $\mathbf{v}$ satisfying $\mathbf{v}\mathbf{A}^{(k)} = \rho\mathbf{v}$ and $\sum_{i=1}^{N} v_i = 1$.

The limiting behavior

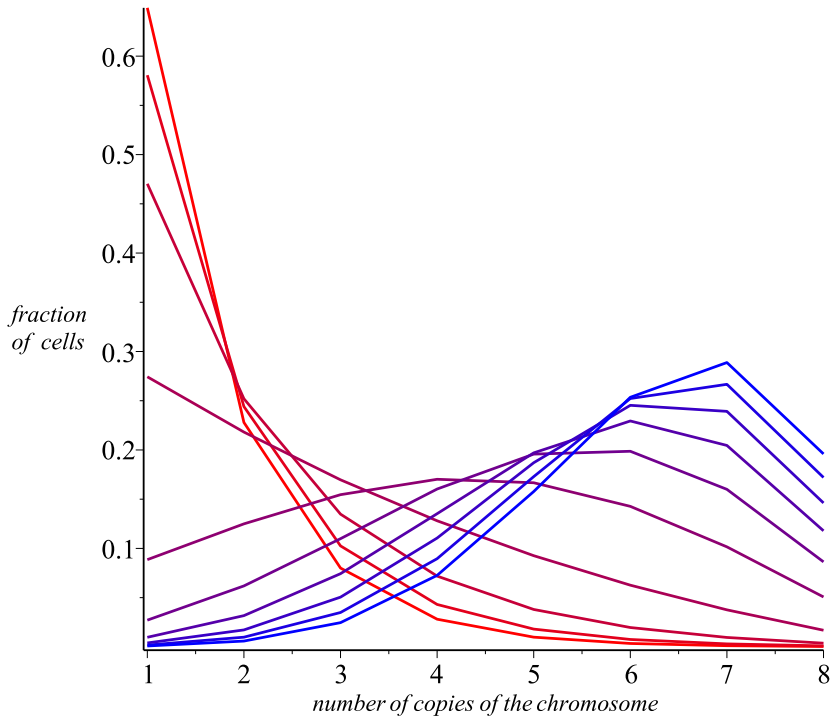
As before, if ρ is the largest eigenvalue of $\mathbf{A}^{(k)}$, the limiting distribution conditional on the non-absorbing states is given by the vector $\mathbf{v}$ satisfying $\mathbf{v}\mathbf{A}^{(k)} = \rho\mathbf{v}$ and $\sum_{i=1}^{N} v_i = 1$.

Again, this limiting distribution does not depend on the number of copies of the founder cell.

However, unlike for the model without scores, it now depends on p and on μ (i.e., on the chromosome score).

The limiting distribution

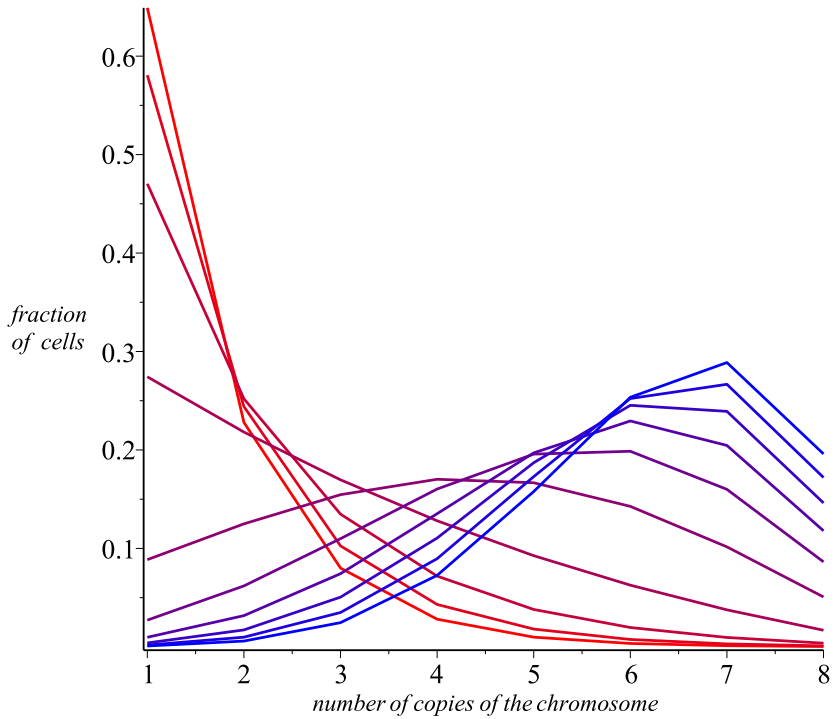
Liming distributions for $\mu = 0.9994$, 0.9996, 0.9998, 1.0000, 1.0002, 1.0004, 1.0006, 1.0008, 1.0010, 1.0012.



$p = 0.001$

The limiting distribution

Liming distributions for $\mu = 0.9994$, 0.9996, 0.9998, 1.0000, 1.0002, 1.0004, 1.0006, 1.0008, 1.0010, 1.0012.



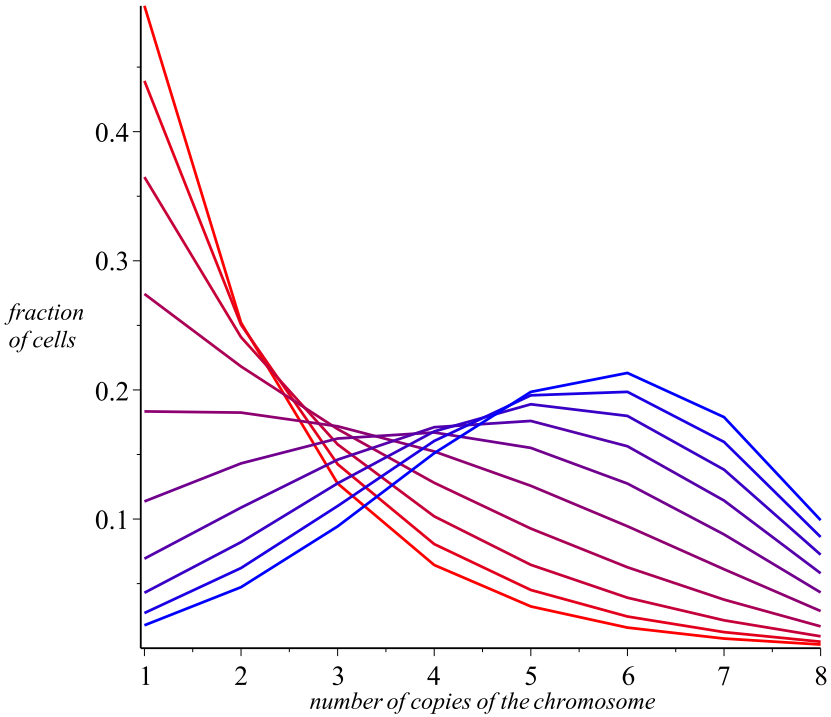
$p = 0.001$

For higher chromosome scores, the limiting distribution favors higher copy numbers.

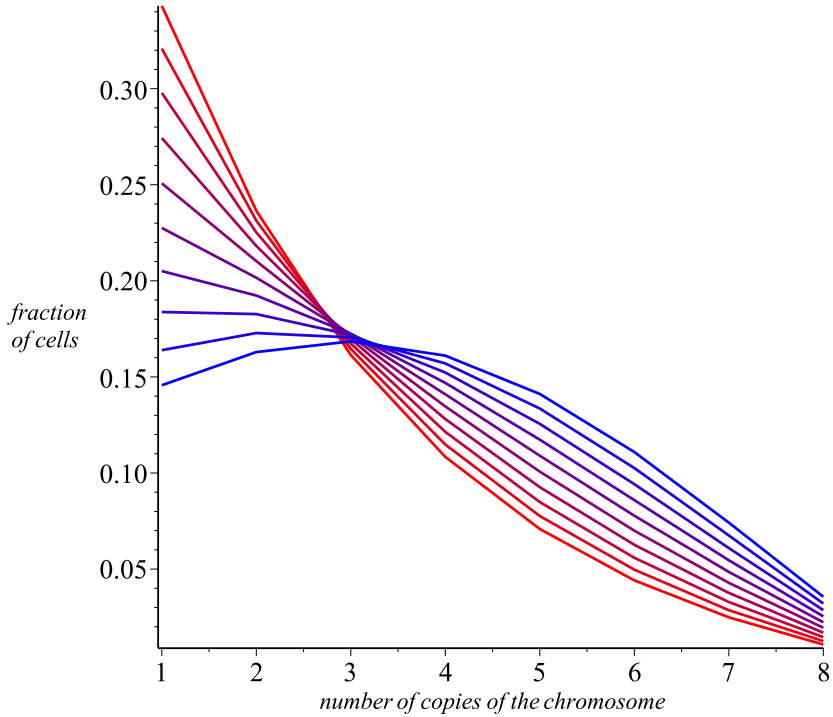
For positive chromosome scores ($\mu > 1$), the most frequent copy number is no longer 1, making this model more realistic than the basic model without scores.

The limiting distribution

Liming distributions for $\mu = 0.9994$, 0.9996, 0.9998, 1.0000, 1.0002, 1.0004, 1.0006, 1.0008, 1.0010, 1.0012.



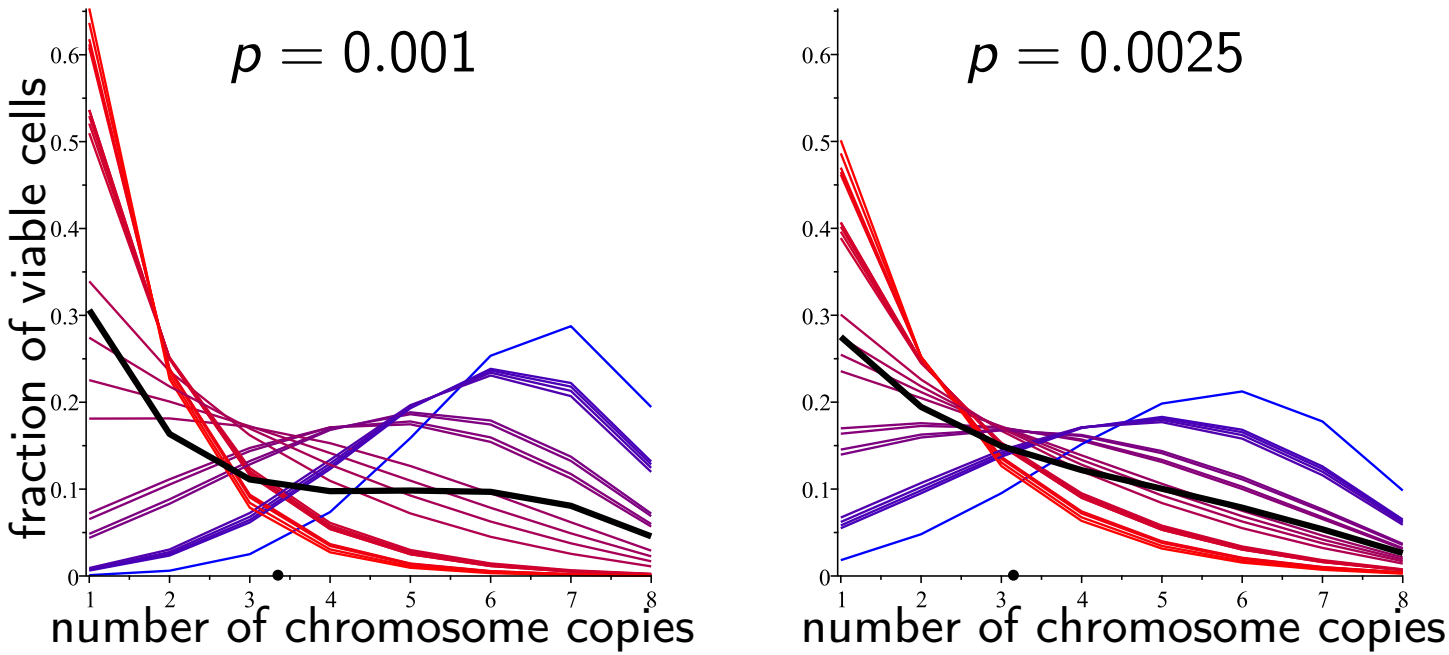
$p = 0.0025$



$p = 0.01$

The limiting distribution

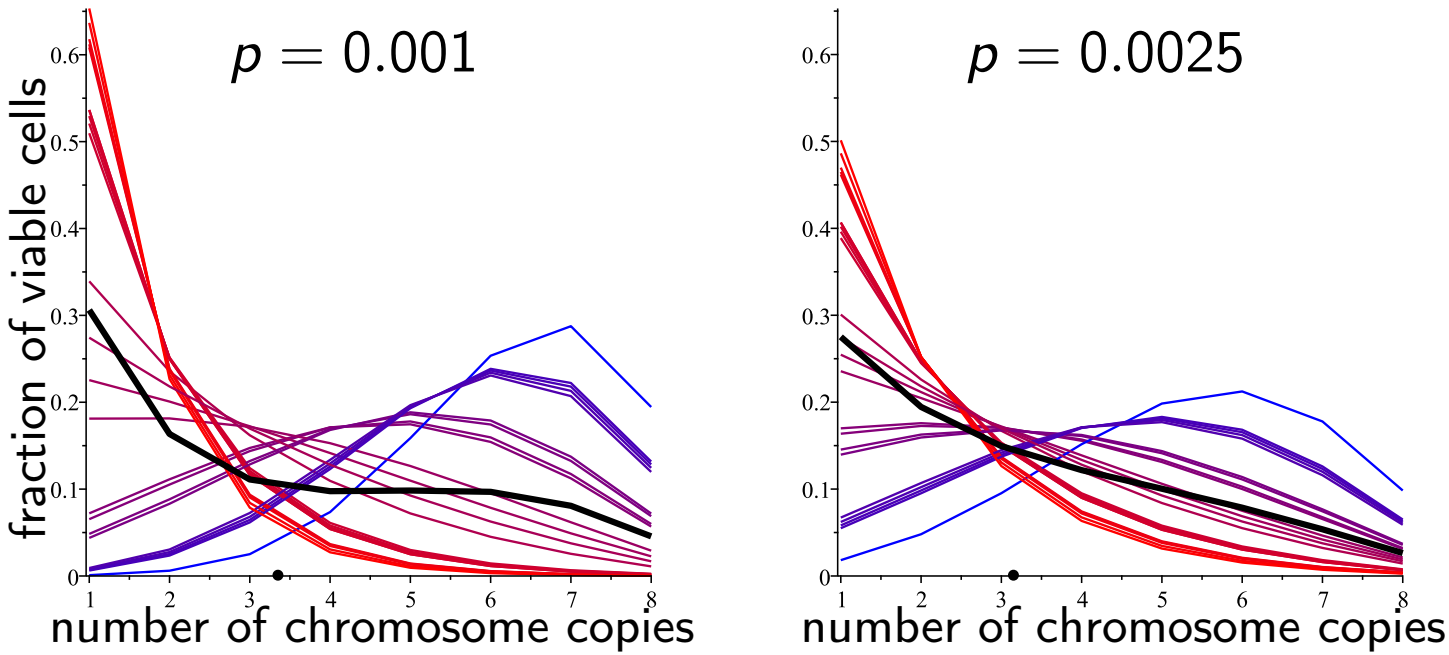
Limiting distributions for the experimentally found values of μ corresponding to the 23 human chromosomes:



The black curve is the average of the 23 limiting distributions, and the black dot on the x-axis is the average number of chromosome copies.

The limiting distribution

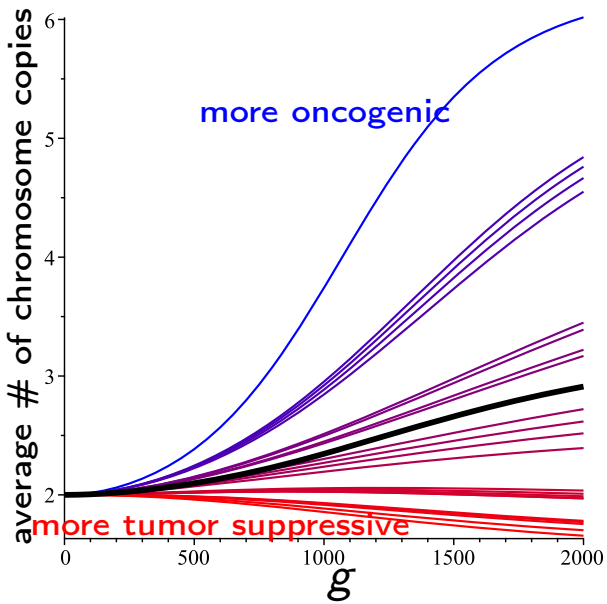
Limiting distributions for the experimentally found values of μ corresponding to the 23 human chromosomes:



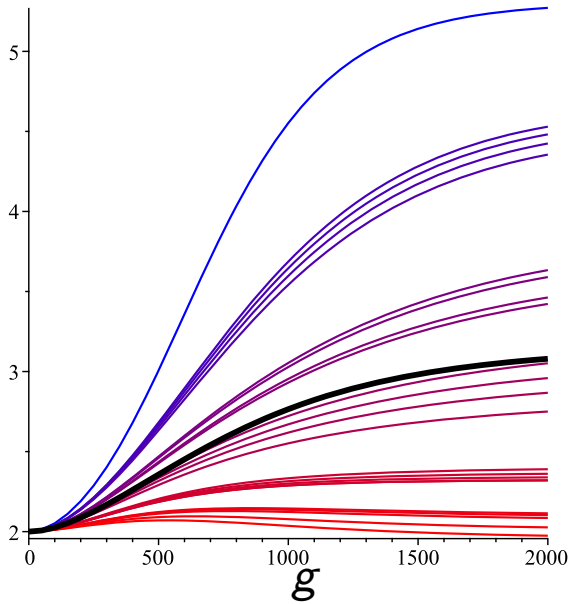
The black curve is the average of the 23 limiting distributions, and the black dot on the x-axis is the average number of chromosome copies. This average of about 3 copies agrees with observations.

The average number of copies over time

The evolution of the average number of copies of the 23 human chromosomes, starting with 2 copies of each.
The average of the 23 averages is shown in black.



$p = 0.001$

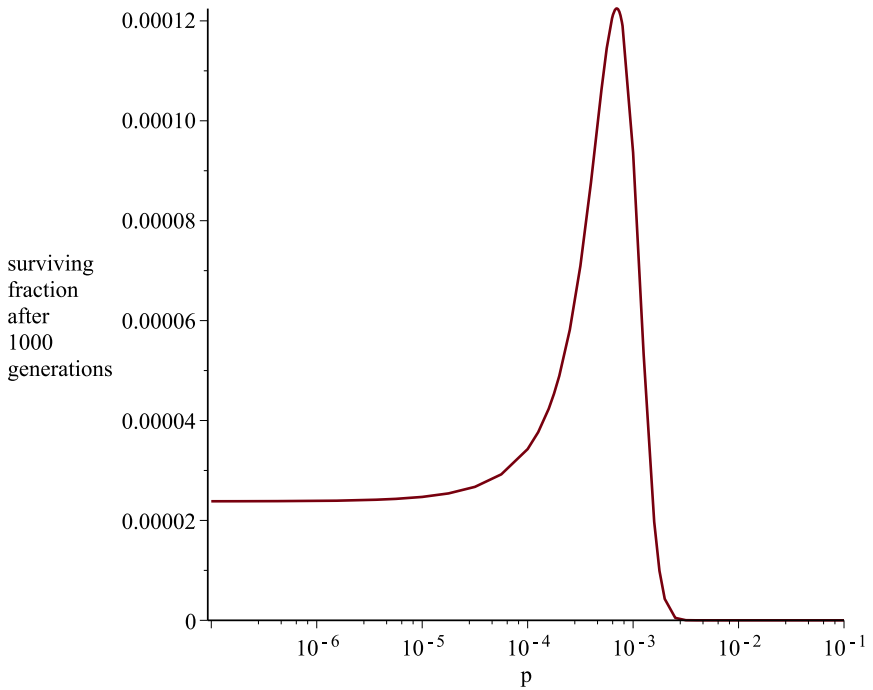


$p = 0.0025$

The convergence to ≈ 3 copies is observed in experiments.

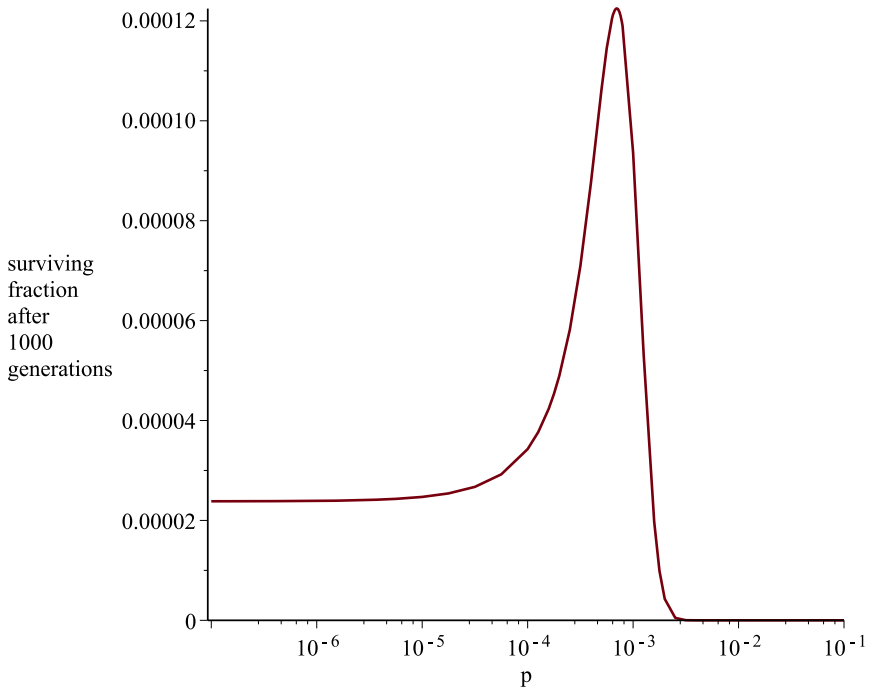
Fraction of live cells after 1000 generations

Using the experimentally found values for the chromosome scores.



Fraction of live cells after 1000 generations

Using the experimentally found values for the chromosome scores.



The number of live cells is maximized for missegregation rates around $p \approx 10^{-3}$. These are the rates observed in cancer!

Karyotypic diversity index

The *karyotype diversity index* measures the heterogeneity of the colony:

$$K = -\sum_{k=1}^{23}\sum_{i=1}^{N} a_{k,i} \ln a_{k,i},$$

where $a_{k,i}$ = fraction of live cells with i copies of chromosome k.

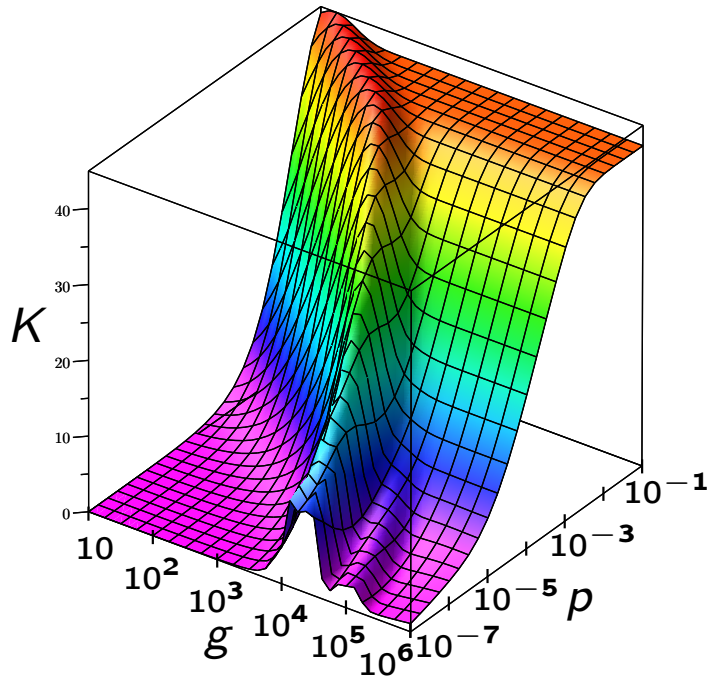
Karyotypic diversity index

The *karyotype diversity index* measures the heterogeneity of the colony:

$$K = -\sum_{k=1}^{23}\sum_{i=1}^{N} a_{k,i} \ln a_{k,i},$$

where $a_{k,i}$ = fraction of live cells with i copies of chromosome k.

Plot of K as a function of g and p (both in a logarithmic scale):



Karyotypic diversity index

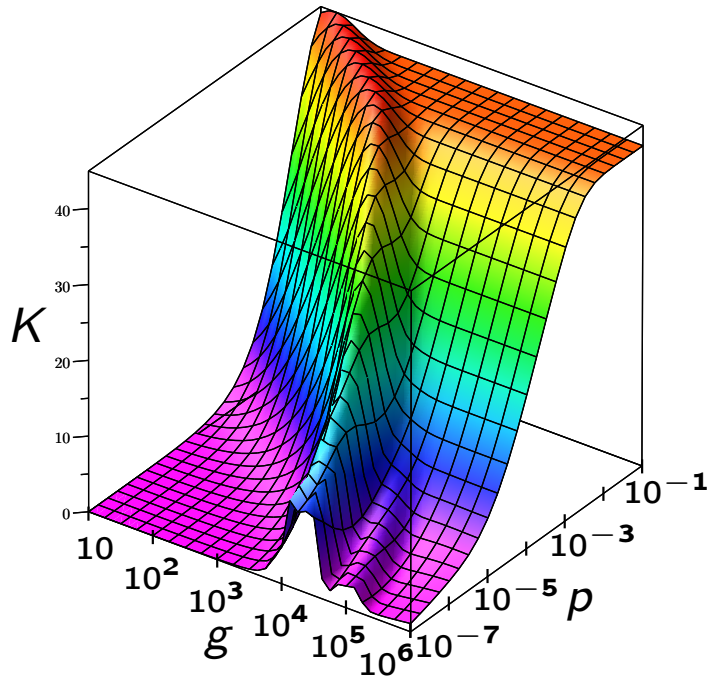
The *karyotype diversity index* measures the heterogeneity of the colony:

$$K = -\sum_{k=1}^{23}\sum_{i=1}^{N} a_{k,i} \ln a_{k,i},$$

where $a_{k,i}$ = fraction of live cells with i copies of chromosome k.

Plot of K as a function of g and p (both in a logarithmic scale):

After $g \approx 10^3$ generations, K is maximized when $p \approx 10^{-3}$ again.



Incorporating drug resistance

Multiplying the survival probability Q_{surv} by a factor < 1, we can model **targeted therapy**. A drug targets genes in a particular chromosome, decreasing the survival probability.

Incorporating drug resistance

Multiplying the survival probability Q_{surv} by a factor < 1, we can model **targeted therapy**. A drug targets genes in a particular chromosome, decreasing the survival probability.

The target genes can mutate with some probability, becoming no longer responsive to treatment. Mutated genes are inherited.

Incorporating drug resistance

Multiplying the survival probability Q_{surv} by a factor < 1, we can model **targeted therapy**. A drug targets genes in a particular chromosome, decreasing the survival probability.

The target genes can mutate with some probability, becoming no longer responsive to treatment. Mutated genes are inherited.

The survival probability of the cell depends on the number of mutated and normal copies of the treated chromosome.

Modeling mutations

We modify the Markov chain to incorporate mutations.

States are now indexed by pairs (i_1, i_2) with $1 \leq i_1 + i_2 \leq N$, representing cells having i_1 normal copies of the chromosome and i_2 mutated copies. For $N = 8$, there are 44 non-absorbing states.

Modeling mutations

We modify the Markov chain to incorporate mutations.

States are now indexed by pairs (i_1, i_2) with $1 \leq i_1 + i_2 \leq N$, representing cells having i_1 normal copies of the chromosome and i_2 mutated copies. For $N = 8$, there are 44 non-absorbing states.

In a cell division, each normal copy of the chromosome has probability $m \approx 10^{-9}$ of mutating (and becoming resistant). Each mutated copy has probability $r \approx 10^{-9}/4$ of reversing into a normal copy (amenable to treatment).

Modeling mutations

We modify the Markov chain to incorporate mutations.

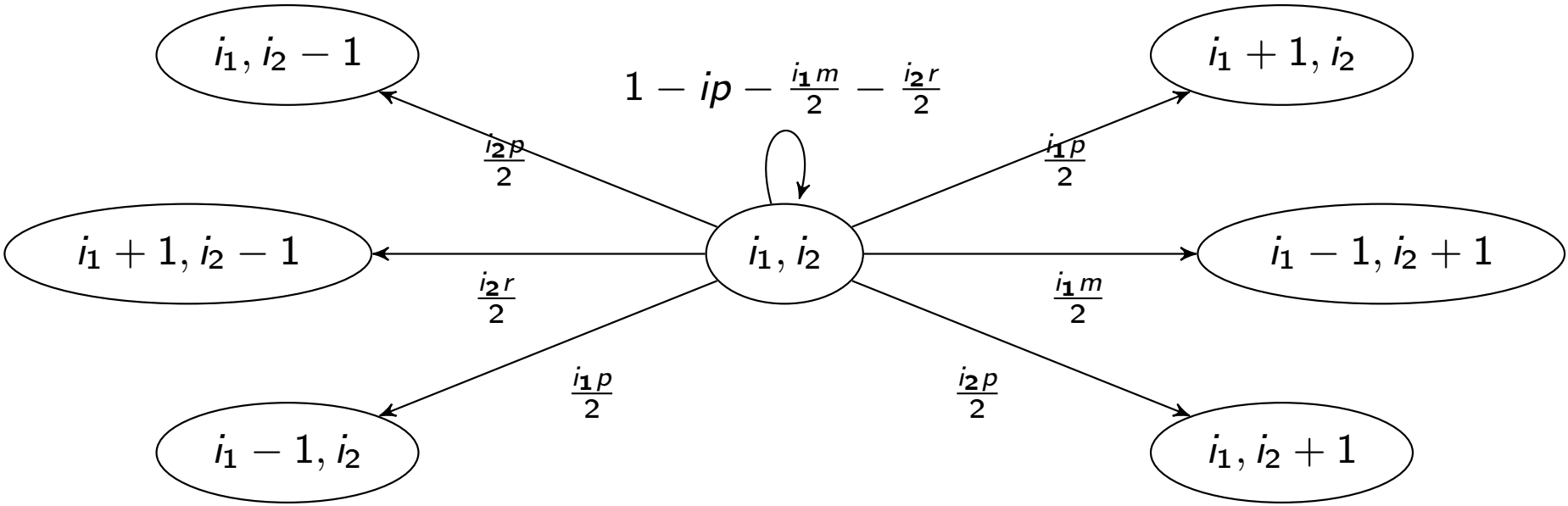
States are now indexed by pairs (i_1, i_2) with $1 \leq i_1 + i_2 \leq N$, representing cells having i_1 normal copies of the chromosome and i_2 mutated copies. For $N = 8$, there are 44 non-absorbing states.

In a cell division, each normal copy of the chromosome has probability $m \approx 10^{-9}$ of mutating (and becoming resistant). Each mutated copy has probability $r \approx 10^{-9}/4$ of reversing into a normal copy (amenable to treatment).

For simplicity, let's disregard highly unlikely events such as a mutation and a missegregation in the same cell division.

The modified Markov chain

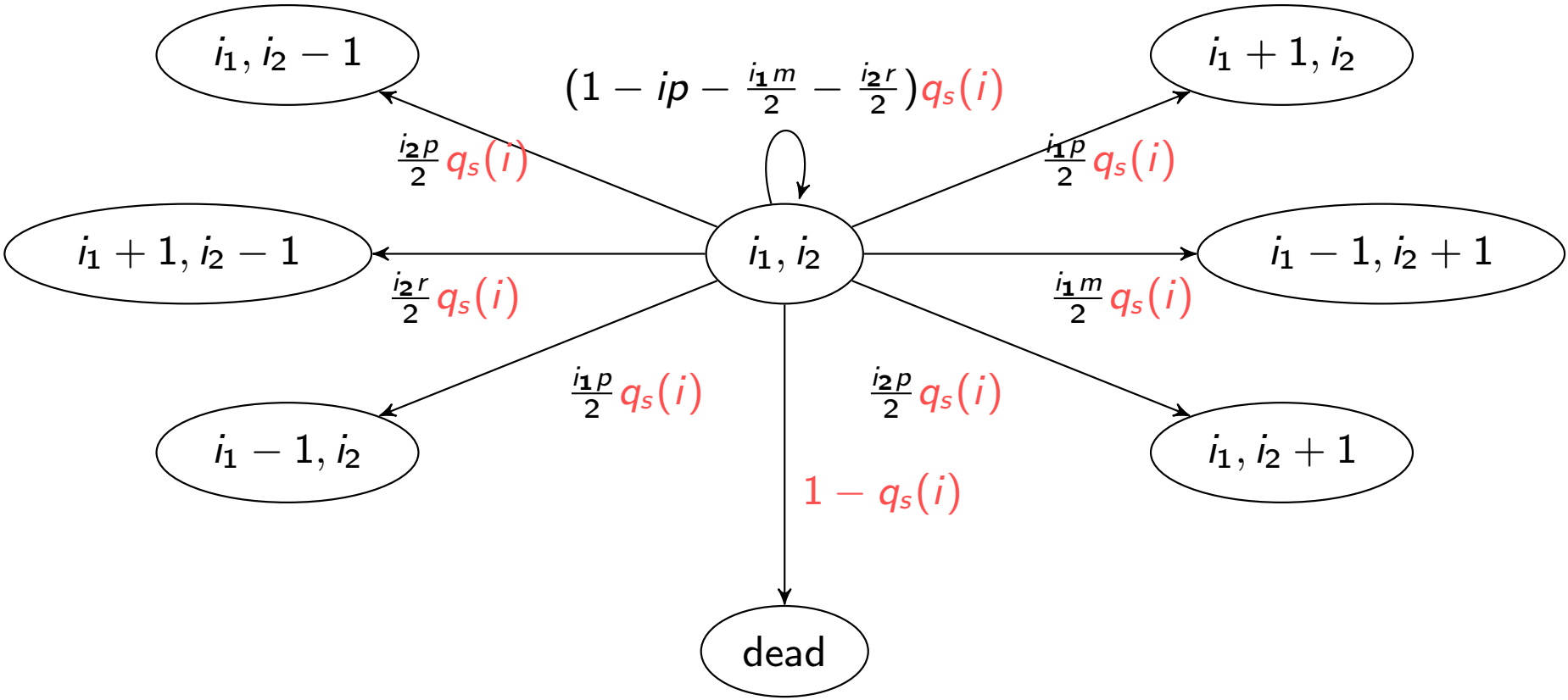
Arrows leaving a typical node (i_1, i_2): (let $i = i_1 + i_2$)



Missegregations and mutations

The modified Markov chain

Arrows leaving a typical node (i_1, i_2): (let $i = i_1 + i_2$)



Missegregations and mutations, survival probability determined by scores

Modeling drug resistance

First, we let the tumor grow until it reaches about 10^9 cells and it becomes detectable with a CT scan.

Then we apply a drug that targets a given gene, decreasing the survival probability of cells containing that gene.

Modeling drug resistance

First, we let the tumor grow until it reaches about 10^9 cells and it becomes detectable with a CT scan.

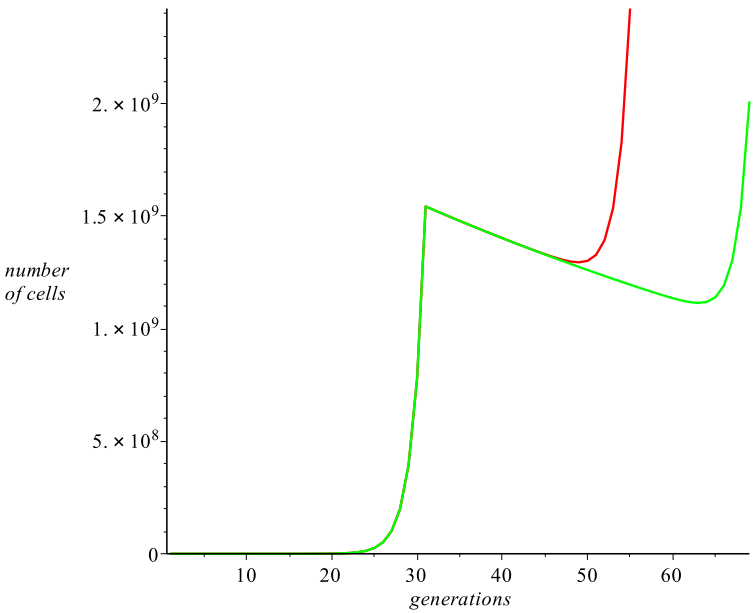
Then we apply a drug that targets a given gene, decreasing the survival probability of cells containing that gene.

Resistance to the drug can be modeled in several ways:

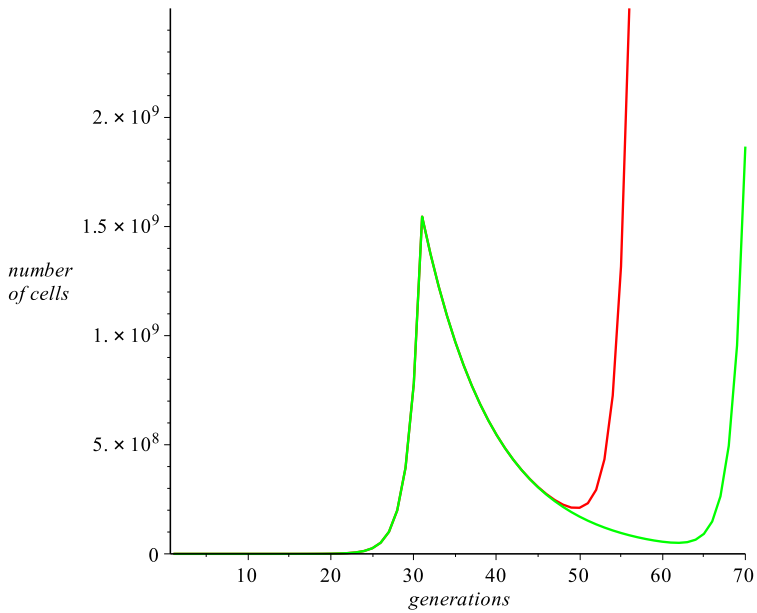
1. Binary resistance: cells with at least one mutated copy of the treated chromosome are resistant.
2. Graded resistance: the level of resistance depends on the ratio of copies of normal vs. mutated copies of the treated chromosome.

Modeling drug resistance

These plots show the evolution of the number of cells when applying targeted therapy to chromosome 1, comparing binary resistance and graded resistance.



weaker treatment



stronger treatment

Work in progress

- Study **time to resistance** for different parameters and validate with data obtained by applying different drugs on real patients.

Work in progress

- Study **time to resistance** for different parameters and validate with data obtained by applying different drugs on real patients.
- **Whole genome duplication events.** Sometimes the genome of a cell doubles but the cell does not divide. We can model this for each chromosome individually, introducing transitions from state i to state $2i$ in the Markov chain. However, we currently cannot account for correlations among different chromosomes, which in practice double simultaneously.

Work in progress

- Study **time to resistance** for different parameters and validate with data obtained by applying different drugs on real patients.
- **Whole genome duplication events.** Sometimes the genome of a cell doubles but the cell does not divide. We can model this for each chromosome individually, introducing transitions from state i to state $2i$ in the Markov chain. However, we currently cannot account for correlations among different chromosomes, which in practice double simultaneously.
- **Arm level events.** Sometimes chromosomes can split into two arms, which missegregate at high rates. Additionally, arms of different chromosomes can fuse to form neo-chromosomes.

Work in progress

- Study **time to resistance** for different parameters and validate with data obtained by applying different drugs on real patients.
- **Whole genome duplication events.** Sometimes the genome of a cell doubles but the cell does not divide. We can model this for each chromosome individually, introducing transitions from state i to state $2i$ in the Markov chain. However, we currently cannot account for correlations among different chromosomes, which in practice double simultaneously.
- **Arm level events.** Sometimes chromosomes can split into two arms, which missegregate at high rates. Additionally, arms of different chromosomes can fuse to form neo-chromosomes.
- Keep track of **maternal and paternal alleles**.

References

- A. Laughney, S.E., G. Genovese and S. Bakhoum, Dynamics of tumor heterogeneity derived from clonal karyotypic evolution, *Cell Rep.* 12 (2015), 809–820.
- S.E., A. Laughney and S. Bakhoum, A Markov chain for numerical chromosomal instability in clonally expanding populations, *PLoS Comput. Biol.* 2018; 14(9): e1006447.

Thank you!