ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D-BSSE
Department of Biosystems
Science and Engineering

Evolutionary escape

Niko Beerenwinkel

Outline

- Posets and distributive lattices
- The risk polynomial
- A multi-type branching process
- The risk of escape
- Example: TSG inactivation

Binary sequence space

- We consider n loci in a genome, each of which can be unmutated (“0”) or mutated (“1”). Mutations are irreversible.
- Thus, each genotype is a binary string of length n.
- The string

$$\mathbf{0} = \hat{0} = 000\ldots0$$

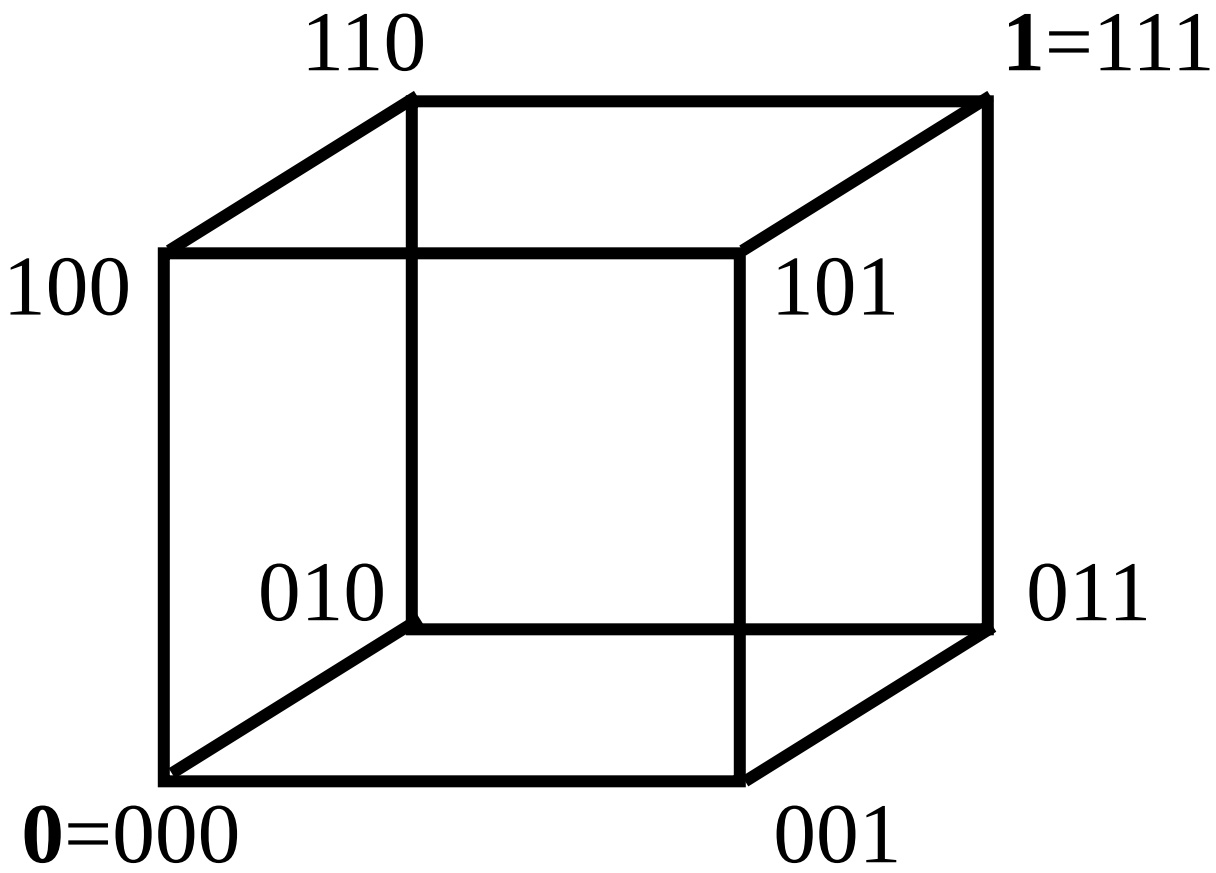
is called the *wild type*.

- The string

$$\mathbf{1} = \hat{1} = 111\ldots1$$

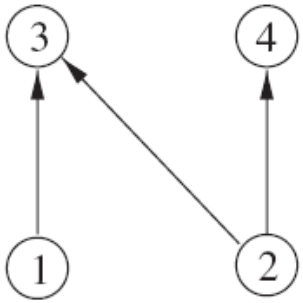
is called the *ecape type*.

Binary sequences of length n = 3



Partially ordered sets (posets)

- A *partially ordered set* (or *poset*) is a set $\mathcal{E}$ together with a binary relation, denoted "$\leq$", which is
  - reflexive: for all $e \in \mathcal{E}$, $e \leq e$
  - antisymmetric: for all $e_1, e_2 \in \mathcal{E}$, $(e_1 \leq e_2$ and $e_2 \leq e_1) \Rightarrow e_1 = e_2$
  - transitive: for all $e_1, e_2, e_3 \in \mathcal{E}$, $(e_1 \leq e_2$ and $e_2 \leq e_3) \Rightarrow e_1 \leq e_3$
- We write $e_1 < e_2$, if $e_1 \neq e_2$ and $e_1 \leq e_2$.
- $e_1 < e_2$ is a *cover relation*, if there is no e' with $e_1 < e' < e_2$.
- In the *Hasse diagram*, $e_1 \rightarrow e_2$, if $e_1 < e_2$ is a cover relation.
- Example:
  - $\mathcal{E} = \{1, 2, 3, 4\}$
  - relation: $1 \leq 3$, $2 \leq 3$, $2 \leq 4$, and $i \leq i$ for all i = 1, 2, 3, 4.

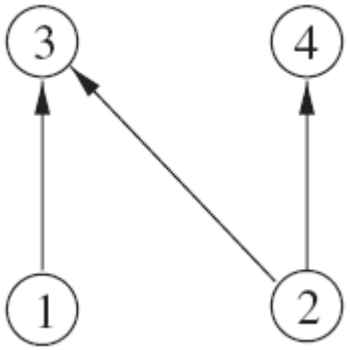


Distributive lattice of order ideals

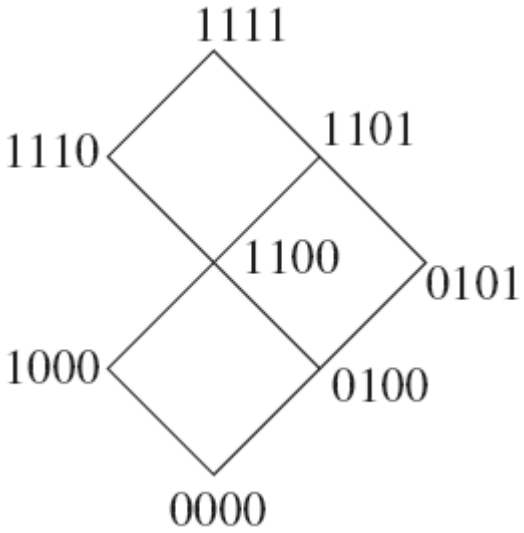
- An *order ideal* g in a poset $\mathcal{E}$ is a subset of $\mathcal{E}$ that is closed downward, i.e., if $e_2 \in g$ and $e_1 \leq e_2$, then $e_1 \in g$.
- The set of all order ideals of $\mathcal{E}$ forms a *distributive lattice* $J(\mathcal{E})$ under inclusion, that is,
  - $(J(\mathcal{E}), \subseteq)$ forms a poset, and
  - every pair of order ideals (g_1, g_2) has a unique supremum (namely the *join* $g_1 \cup g_2$) and an infimum (namely the *meet* $g_1 \cap g_2$).

- All distributive lattices have the form $J(\mathcal{E})$ for a poset $\mathcal{E}$ (Birkhoff's Representation Theorem).

Example



poset, $\mathcal{E}$

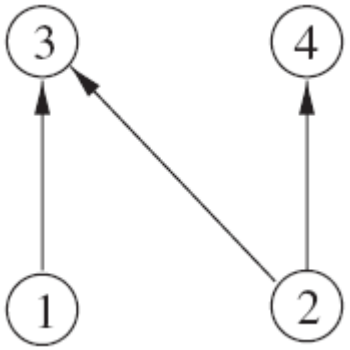


distributive lattice
of order ideals, J($\mathcal{E}$)

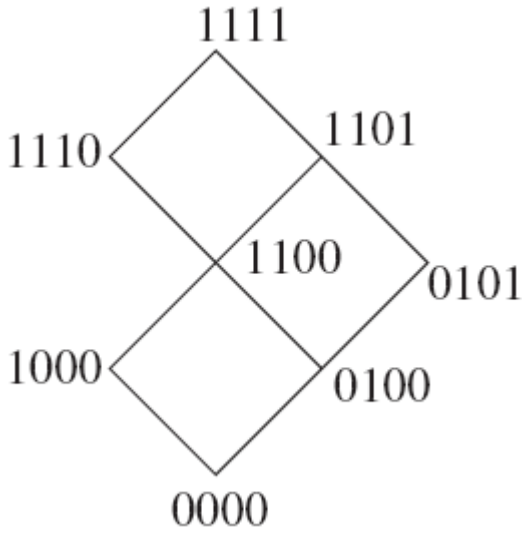
The genotype lattice

- Let $\mathcal{E}$ be a set of n = $|\mathcal{E}|$ irreversible genetic events (e.g., point mutations at different loci, indels, genomic rearrangements, etc.).
- The poset $(\mathcal{E}, \leq)$ encodes constraints on the order in which mutations can accumulate. (For example, some intermediate types might not be viable).
- The order ideals g of J($\mathcal{E}$) are precisely the subsets of $\mathcal{E}$ that are compatible with the partial order, i.e., they are the genotypes that can evolve subject to the order constraints.
- We call $\mathcal{G}$ = J($\mathcal{E}$) the genotype lattice.

Example



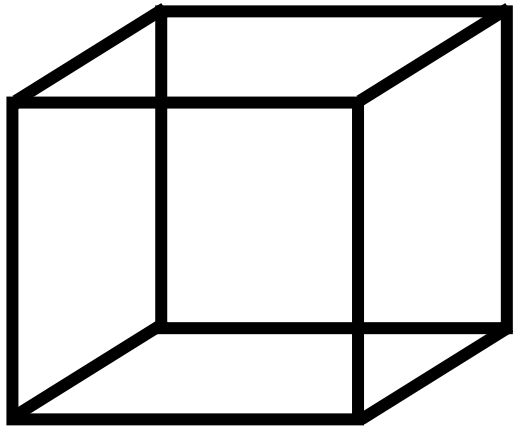
poset of genetic events (mutations), $\mathcal{E}$



genotype lattice, $\mathcal{G} = J(\mathcal{E})$

The empty poset

- $\mathcal{E}$ = {1, 2, …, n}, no relations
- $\mathcal{G}$ = J($\mathcal{E}$) is the set of all subsets of $\mathcal{E}$, or the Boolean lattice $\{0, 1\}^n$, or the space of all binary sequences of length n.
- All genotypes can occur in any order; there are no order constraints.



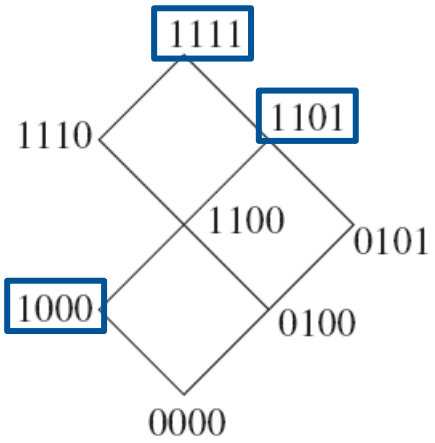
- We will regard a genotype g interchangeably as a binary string, $g \in \{0,1\}^n$, or as a subset, $g \subseteq \mathcal{E}$.

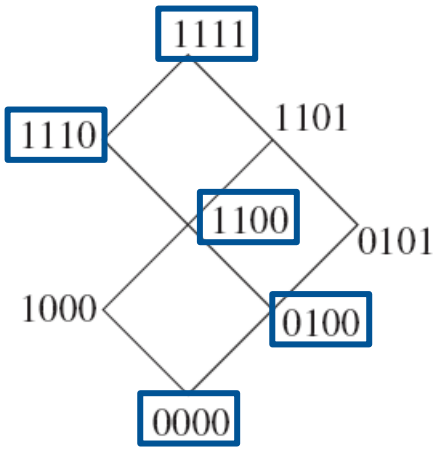
Chains

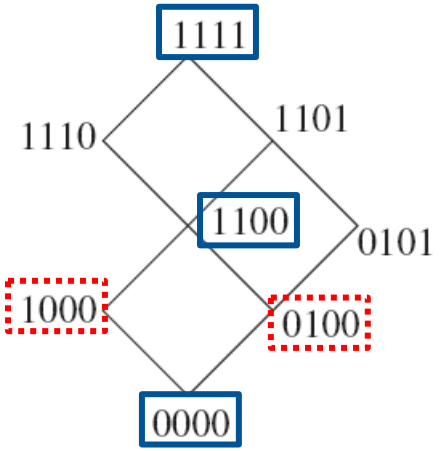
- A *chain* in $\mathcal{G}$ = (J($\mathcal{E}$), $\subseteq$) of length k is a totally ordered subset

$$g_1 \subset g_2 \subset \cdots \subset g_k$$

- The chains in $\mathcal{G}$ are the mutational pathways consistent with the poset $\mathcal{E}$.

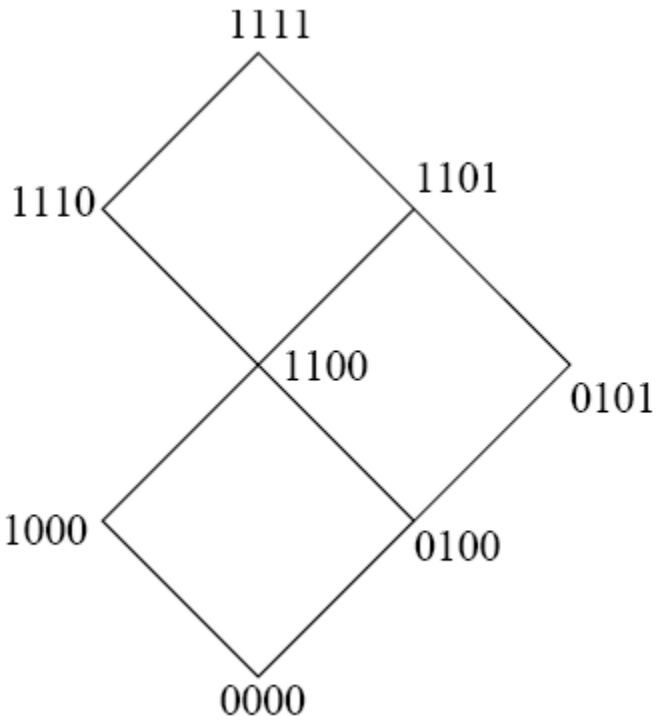


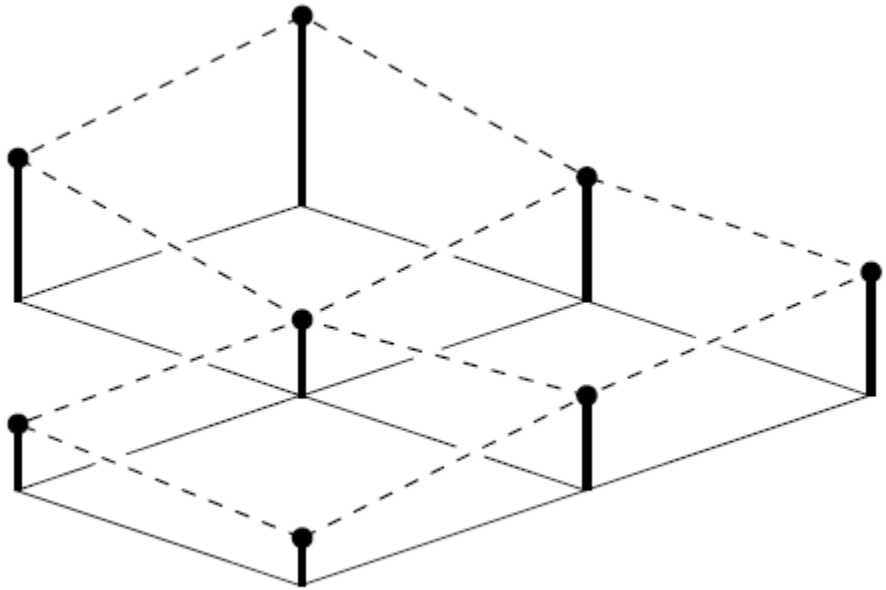




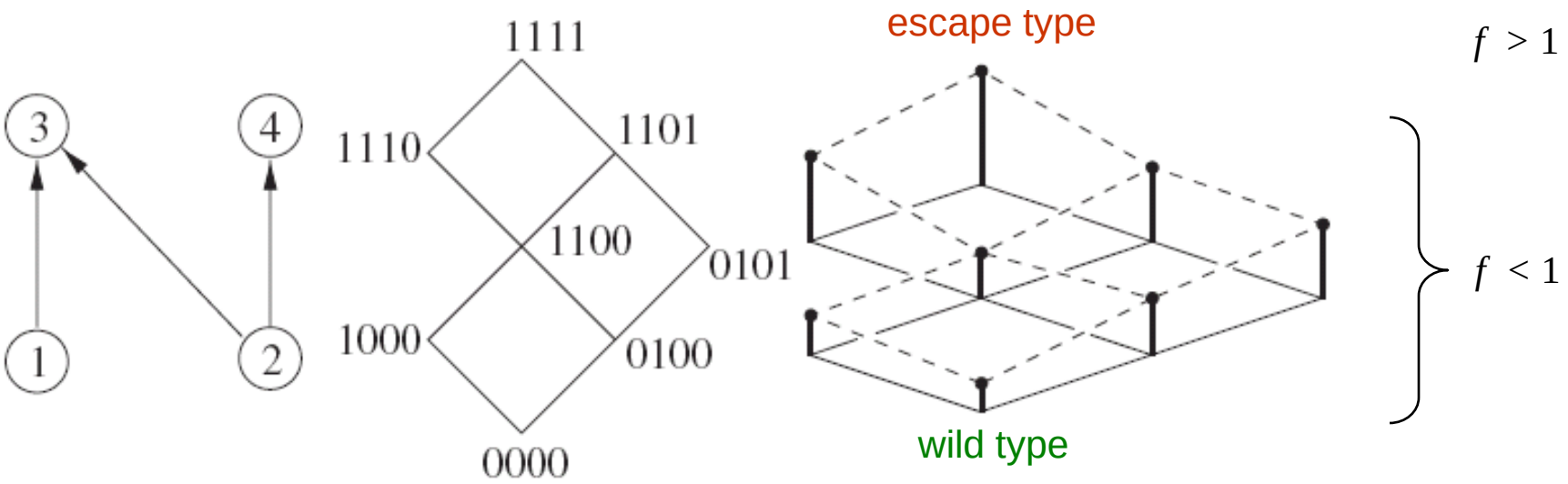
Fitness landscapes

- A fitness landscape is a mapping $\mathbf{f} : \mathcal{G} \rightarrow \mathbb{R}$.





Evolutionary escape



event poset genotype lattice fitness landscape

- Does the wild type population reach the escape state before extinction? (infection of a new host, escape from therapy)

Mutational neighborhood

- The *mutational neighborhood* of a genotype g $\in \mathcal{G}$ is the set of genotypes h $\in \mathcal{G}$ that can be reached by mutation:

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

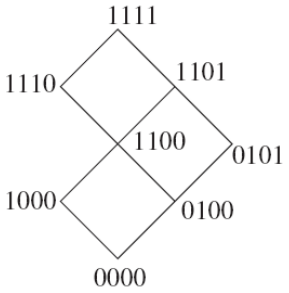
- We model a directed evolutionary process that proceeds upward the genotype lattice from the wild type **0** to the escape type **1**.

Mutation

- Let m = $|\mathcal{G}|$. We fix a total order (or linear extension) of $\mathcal{G}$.
- Let μ_e be the mutation rate of event e $\in \mathcal{E}$. We assume that mutations are independent of each other.
- Define the (m $\times$ m) mutation matrix **U** = $(u_{gh})_{g,h\in\mathcal{G}}$ by

$$u_{gh} = \begin{cases} \prod_{e\in h\setminus g} \mu_e & \text{if } h \in N(g) \\ 0 & \text{otherwise} \end{cases}$$

Example



$$
\begin{array}{c}
 \\
0000 \\
1000 \\
0100 \\
1100 \\
0101 \\
1110 \\
1101 \\
1111
\end{array}
\begin{array}{c}
\begin{array}{cccccccc}
0000 & 1000 & 0100 & 1100 & 0101 & 1110 & 1101 & 1111
\end{array} \\
\begin{pmatrix}
0 & \mu_1 & \mu_2 & \mu_1\mu_2 & \mu_2\mu_4 & \mu_1\mu_2\mu_3 & \mu_1\mu_2\mu_4 & \mu_1\mu_2\mu_3\mu_4 \\
0 & 0 & 0 & \mu_2 & 0 & \mu_2\mu_3 & \mu_2\mu_4 & \mu_2\mu_3\mu_4 \\
0 & 0 & 0 & \mu_1 & \mu_4 & \mu_1\mu_3 & \mu_1\mu_4 & \mu_1\mu_3\mu_4 \\
0 & 0 & 0 & 0 & 0 & \mu_3 & \mu_4 & \mu_3\mu_4 \\
0 & 0 & 0 & 0 & 0 & 0 & \mu_1 & \mu_1\mu_3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu_4 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu_3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
$$

k-step offspring

- Let **f** : $\mathcal{G} \to \mathbb{R}$ be a fitness landscape and set **F** = diag(f), the (m $\times$ m) diagonal matrix.
- The entry (g,h) of the matrix **UF** is the probability of genotype g producing offspring of type h in one step.
- The entry (g,h) of the matrix $(\mathbf{UF})^k$ is the probability of genotype g producing offspring of type h along any mutational pathway of length k in the genotype lattice $\mathcal{G}$.
- Let $\mathbf{I}$ be the identity matrix. We consider the matrix

$$\begin{aligned} \mathbf{B} &= (\mathbf{I} - \mathbf{UF})^{-1} - \mathbf{I} \\ &= \mathbf{UF} + (\mathbf{UF})^2 + \cdots + (\mathbf{UF})^n \end{aligned}$$

Properties of $\mathbf{B} = \mathbf{UF} + (\mathbf{UF})^2 + \cdots + (\mathbf{UF})^n$

- The entry b_{gh} of **B** in row g and column h is zero, unless $g \subset h$.
- If $g \subset h$, then

$$b_{gh} \;=\; u_{gh}\, \mathbf{f}(h)\, P_{gh}(\mathbf{f})$$

  where P_{gh} is a polynomial function of degree $|h \setminus g| - 1$ on $\mathbb{R}^{\mathcal{G}}$.
- The polynomial P_{gh} is the generating function for all chains from g to h in $\mathcal{G}$.
- We now restrict to the case g = **0** and h = **1**.

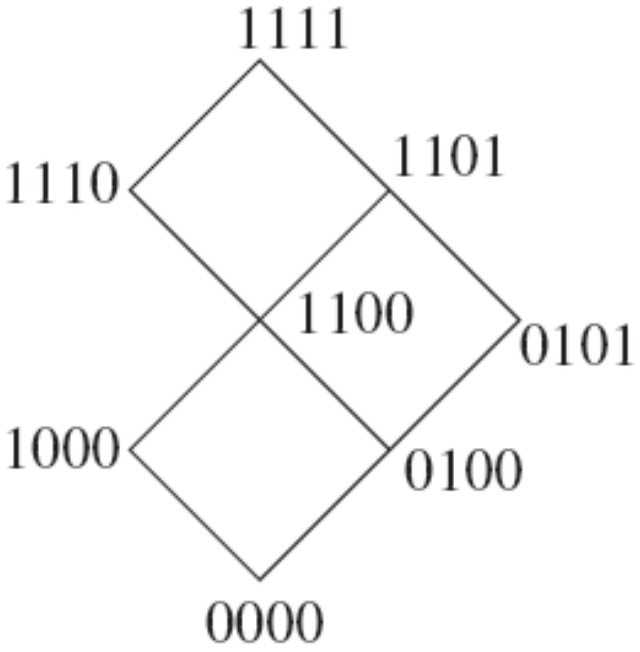
The risk polynomial

- We write the fitness values **f**(g) = f_g, g $\in \mathcal{G}$.
- The polynomial P_{01} in the upper-right entry of the matrix B equals

$$P_{01}(\mathbf{f}) = \sum_{\mathbf{0}=g_0 \subset g_1 \subset \cdots \subset g_k = \mathbf{1}} f_{g_1} f_{g_2} \cdots f_{g_{k-1}}$$

  where the sum runs over all chains from **0** to **1** in the genotype lattice $\mathcal{G}$.
- We call $\mathcal{R}(\mathcal{G}; \mathbf{f}) = P_{01}(\mathbf{f})$ the *risk polynomial*. (It is also known as the chain polynomial.)

Example



$$\mathcal{R}(\mathcal{G};\mathbf{f}) = 1 + f_{1000} + f_{0100} + f_{1100} + f_{0101} + f_{1110} + f_{1101} + f_{1000}f_{1100} + f_{0100}f_{1100} + f_{0100}f_{0101} + f_{1000}f_{1110} + f_{0100}f_{1110} + f_{1000}f_{1101} + f_{0100}f_{1101} + f_{1100}f_{1110} + f_{1100}f_{1101} + f_{0101}f_{1101} + f_{1000}f_{1100}f_{1110} + f_{0100}f_{1100}f_{1110} + f_{1000}f_{1100}f_{1101} + f_{0100}f_{1100}f_{1101} + f_{0100}f_{0101}f_{1101}.$$

Invasion

- Let R_g be the basic reproductive ratio of an invading pathogen.
- We are interested in the case where $R_1 > 1$ and $R_g < 1$ for all $g \neq \mathbf{1}$.
- We define the fitness landscape by

$$f_g = \frac{R_g}{1 - R_g} = R_g + R_g^2 + R_g^3 + \dots$$

- Then $R_g = f_g/(1 + f_g)$.
- For $g \neq 1$, we have $f_g \approx R_g$.

Multitype branching process

- Let us consider a branching process on the type space $\mathcal{G}$ with a Poisson offspring distribution.
- The probability that a single individual of type g produces k children of type h is

$$\rho_{gh}^{k} = \mathrm{Pois}(k; u_{gh}R_g) = \frac{(u_{gh}R_g)^k}{k!}\, e^{-u_{gh}R_g}$$

The risk of escape

- Let ξ_g be the probability of escape (reaching **1** before extinction) starting with one individual of type g.
- $1 - \xi_g$ is the probability of extinction. Because for extinction, all lineages need to go extinct, we have the recursion

$$1 - \xi_g = \prod_{h \supseteq g} \sum_{k=0}^{\infty} (1 - \xi_h)^k \, \rho_{gh}^k$$

where the u_{gh} are defined as before, except $u_{gg} = 1$.

Recurrence equation

- Substituting the Poisson distribution, simplifying, and taking logarithms, we find

$$\log(1-\xi_g) = -\sum_{h \supseteq g} \xi_h u_{gh} R_g$$

- For g ≠ 1, $\xi_g \ll 1$ and $(R_g)^2 \approx 0$. Thus

$$\xi_g \approx R_g \sum_{h \supseteq g} \xi_h u_{gh}$$

$$\approx \frac{R_g}{1-R_g} \sum_{h \supset g} \xi_h u_{gh} = f_g \sum_{h \supset g} \xi_h u_{gh}$$

Risk polynomial and risk of escape

- In particular,

$$\xi_0 = f_0 \sum_{h \in \mathcal{G}} \xi_h u_{0h}$$

- This recursion can be solved and it yields

$$\xi_0 \approx \xi_1 f_0 \prod_{e \in \mathcal{E}} \mu_e \mathcal{R}(\mathcal{G}; \mathbf{f})$$

where $\mathcal{R}(\mathcal{G}; \mathbf{f})$ is the risk polynomial.

- The risk of escape of N wild type pathogens is

$$1 - (1 - \xi_0)^N \approx 1 - e^{-N\xi_0}$$

The critical population size

- We define the critical population size

$$N^* = 1/\xi_0$$

- If $N \gg N^*$, then escape is almost certain.
- If $N = N^*$, then the risk of escape is $1 - 1/e$ and the probability of successful intervention is $1/e$.
- If $N \ll N^*$, then escape is almost impossible.

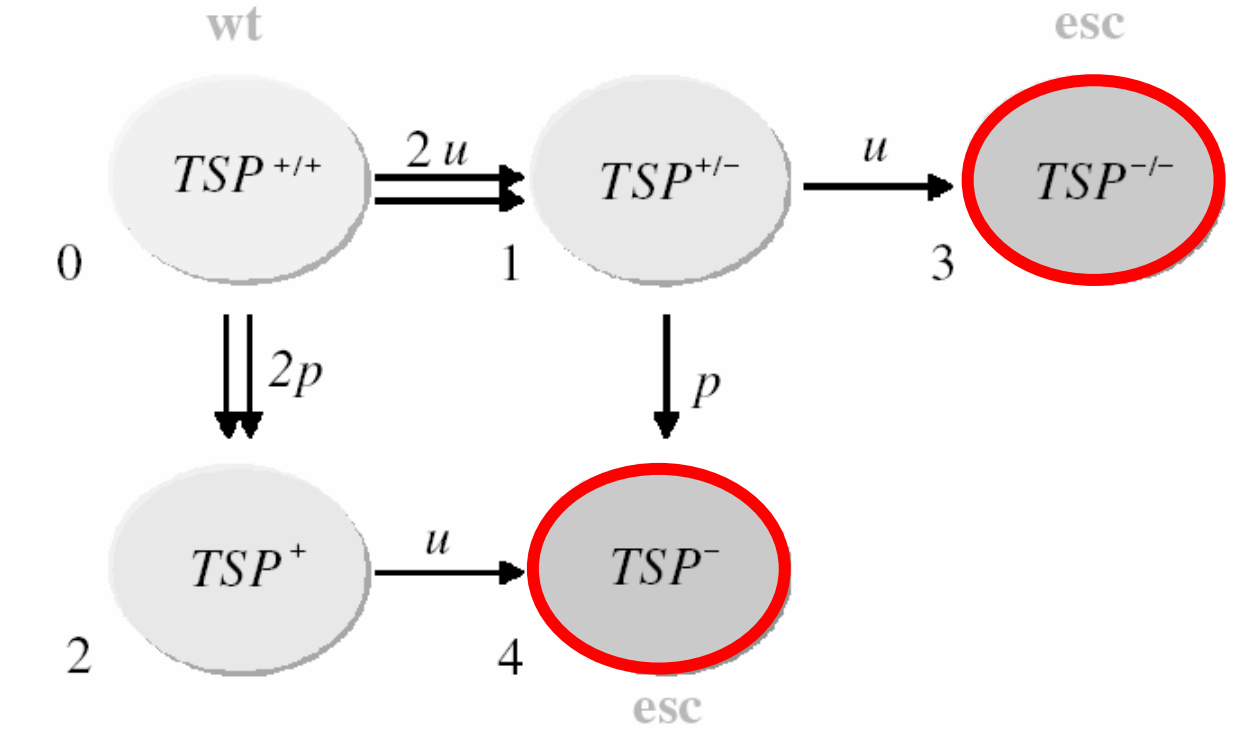
Example: Anti-cancer therapy

- Cancer cells can evolve resistance to chemotherapy by inactivation of tumor suppressor genes (TSGs).
- We assume that at the beginning of therapy all cells have both copies of a certain TSG.
- Inactivation occurs by point mutations at rate u, and/or by LOH at rate p.
- Cells with at least one wild type allele are sensitive to therapy and have the same basic reproductive ratio $R < 1$.
- Cells with no wild type allele escape with probability 1.

- Which tumor size can be controlled with chemotherapy?

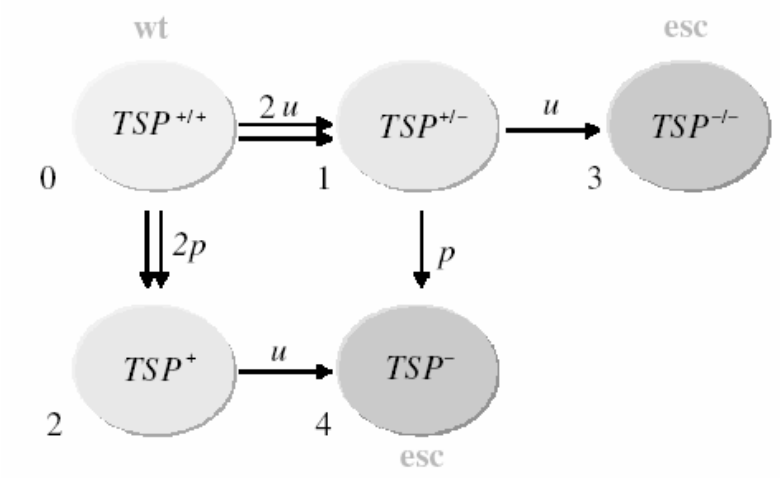
TSG inactivation

$$f_g = \frac{R_g}{1 - R_g} = a \ll 1 \quad \text{for all } g \neq \text{esc}$$



Recurrence equations

- Remember: $\xi_g = f_g \sum_{h \supset g} \xi_h u_{gh}$



$$\xi_- = \xi_{-/-} = 1$$

$$\xi_+ = f_+ u \xi_- = au$$

$$\xi_{+/-} = f_{+/-}(u\xi_{-/-} + p\xi_-) = a(u+p)$$

$$\xi_{+/+} = f_{+/+}(2p\xi_+ + 2u\xi_{+/-} + 2pu\xi_- + u^2\xi_{-/-})$$
$$= a(1+2a) \cdot u(u+2p)$$

Critical tumor size for successful chemotherapy

- The critical population size is

$$N^* = 1/[a(1+2a) \cdot u(u+2p)]$$

- For example, if R = 0.9, u = 10^{-7}, and p = 10^{-6}, then $N^* = 3 \cdot 10^{10}$.
- With chromosomal instability, p = 10^{-2}, and hence $N^* = 3 \cdot 10^{6}$.

Summary

- Mutational escape pathways can be described by order constraints on the occurrence of mutations.
- The order constraints define the genotype lattice, i.e., the set of viable genotypes to consider.
- The risk polynomial encodes the set of escape pathways in the genotype lattice.
- A multi-type branching process can be used to model the population dynamics during evolutionary escape.
- Under the branching process model, the risk of escape can be computed. It depends on the mutation rates and the fitness values, and on the structure of the genotype lattice through the risk polynomial.

Further reading

- Iwasa Y, Michor F, Nowak MA (2003). Evolutionary dynamics of escape from biomedical intervention. Proc Biol Sci 270:2573−2578.
- Beerenwinkel N, Eriksson N, Sturmfels B (2006). Evolution on distributive lattices. J Theor Biol 242:409−420.