S&DS 241 Lecture 25

Basics of Branching processes (not in final exam)

B-H: Example 2.7.2, Sec 6.7

Extinction of a species

[Francis Galton '1873] For simplicity, suppose:

- Start with a single individual as the ancestor
- The number of children of each individual is independent and identically distributed with PMF $p_0, p_1, p_2, \ldots$
  - For example: zero child with probability $p_0 = \frac{1}{3}$, one child with probability $p_1 = \frac{1}{3}$, two children with probability $p_2 = \frac{1}{3}$

Extinction of a species

[Francis Galton '1873] For simplicity, suppose:

- Start with a single individual as the ancestor
- The number of children of each individual is independent and identically distributed with PMF $p_0, p_1, p_2, \ldots$
  - For example: zero child with probability $p_0 = \frac{1}{3}$, one child with probability $p_1 = \frac{1}{3}$, two children with probability $p_2 = \frac{1}{3}$

Question

- What is the probability of eventual extinction?

Extinction of a species

[Francis Galton '1873] For simplicity, suppose:

- Start with a single individual as the ancestor
- The number of children of each individual is independent and identically distributed with PMF $p_0, p_1, p_2, \ldots$
  - For example: zero child with probability $p_0 = \frac{1}{3}$, one child with probability $p_1 = \frac{1}{3}$, two children with probability $p_2 = \frac{1}{3}$

Question

- What is the probability of eventual extinction?
- What is the average number of k-generation descendents?

History

There was concern amongst the Victorians that aristocratic surnames were becoming extinct. Galton originally posed a mathematical question regarding the distribution of surnames in an idealized population in an 1873 issue of *The Educational Times*, and the Reverend Henry William Watson replied with a solution.

https://en.wikipedia.org/wiki/Galton%E2%80%93Watson_process#History

History

There was concern amongst the Victorians that aristocratic surnames were becoming extinct. Galton originally posed a mathematical question regarding the distribution of surnames in an idealized population in an 1873 issue of *The Educational Times*, and the Reverend Henry William Watson replied with a solution.

138 WATSON and GALTON.—*Extinction of Families.*

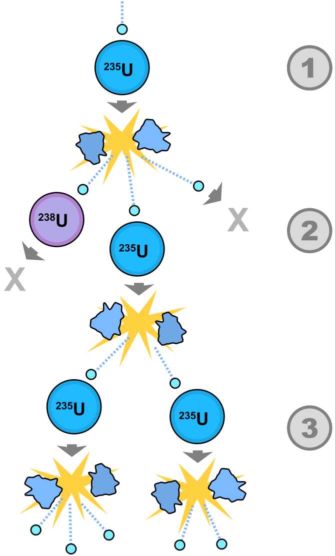
Mr. Galton then read the following paper by the Rev. H. W. Watson and himself:

On the PROBABILITY *of the* EXTINCTION *of* FAMILIES. By the Rev. H. W. WATSON. With PREFATORY REMARKS, by FRANCIS GALTON, F.R.S.

https://en.wikipedia.org/wiki/Galton%E2%80%93Watson_process#History

Branching process (Galton-Watson tree)

Nuclear fission chain reaction



Other applications

- Spread of disease/rumor
- Evolution/Genetic mutation
- Population growth/Extinction of species
- etc

Simple example

- Consider $p_0 = p_1 = p_2 = \frac{1}{3}$. Find $p = P(\text{extinction})$.

Simple example

- Consider $p_0 = p_1 = p_2 = \frac{1}{3}$. Find $p = P(\text{extinction})$.
- Call the ancestor Adam. By LOTP (Lec 4):

$$\begin{aligned} p = \ & \underbrace{P(\text{extinction}|\text{Adam has 0 child})}_{1} \times \frac{1}{3} \\ & + \underbrace{P(\text{extinction}|\text{Adam has 1 child})}_{p} \times \frac{1}{3} \\ & + \underbrace{P(\text{extinction}|\text{Adam has 2 children})}_{p^2} \times \frac{1}{3} \end{aligned}$$

- We arrive at an equation $p = (1 + p + p^2)/3$, which has a unique solution $p = 1$. Thus

$$P(\text{extinction}) = 1$$

Simple example

- Consider $p_0 = \frac{1}{3}, p_1 = \frac{1}{6}, p_2 = \frac{1}{2}$. Find $p = P(\text{extinction})$.

Simple example

- Consider $p_0 = \frac{1}{3}, p_1 = \frac{1}{6}, p_2 = \frac{1}{2}$. Find $p = P(\text{extinction})$.
- Call the ancestor Adam. By LOTP:

$$\begin{aligned} p = & \underbrace{P(\text{extinction}|\text{Adam has 0 child})}_{1} \times \frac{1}{3} \\ & + \underbrace{P(\text{extinction}|\text{Adam has 1 child})}_{p} \times \frac{1}{6} \\ & + \underbrace{P(\text{extinction}|\text{Adam has 2 children})}_{p^2} \times \frac{1}{2} \end{aligned}$$

- Solving

$$p = 1/3 + p/6 + p^2/2 \implies p = 1 \text{ or } p = 2/3$$

Which solution should I pick?

Systematic solution

Probability generating function (B-H Sec 6.7)

Let X be a random non-negative integer (e.g. Bin, Pois, Geom)

- Probability Generating function (PGF) of a random variable

$$G_X(z) = E(z^X), \quad |z| \le 1$$

Probability generating function (B-H Sec 6.7)

Let X be a random non-negative integer (e.g. Bin, Pois, Geom)

- Probability Generating function (PGF) of a random variable

$$G_X(z) = E(z^X), \quad |z| \le 1$$

- PGF vs MGF: just a change of variable

$$\begin{cases} G_X(z) = M_X(\log z) \\ M_X(t) = G_X(e^t) \end{cases}$$

Probability generating function (B-H Sec 6.7)

Let X be a random non-negative integer (e.g. $\text{Bin}, \text{Pois}, \text{Geom}$)

- Probability Generating function (PGF) of a random variable

$$G_X(z) = E(z^X), \quad |z| \leq 1$$

- PGF vs MGF: just a change of variable

$$\begin{cases} G_X(z) = M_X(\log z) \\ M_X(t) = G_X(e^t) \end{cases}$$

- Let X denote the number of children of a given individual, with PMF $P(X = i) = p_i$, $i = 0, 1, 2, \ldots$. It turns out the PGF of X determines the chance of extinction:

$$G(z) = E(z^X) \overset{\text{LOTUS}}{=} p_0 + p_1 z + p_2 z^2 + \cdots$$

e.g. $p_0 = p_1 = p_2 = \frac{1}{3} \implies G(z) = (1 + z + z^2)/3$.

Probability of extinction

Let $e_k = P(\text{extinction by the } k\text{th generation})$.

- Then $e_0 = 0, e_1 = p_0, \ldots$.
- Find $P(\text{eventual extinction})$: $e_\infty = \lim_{k\to\infty} e_k$

Probability of extinction

Let $e_k = P(\text{extinction by the } k\text{th generation})$.

- Then $e_0 = 0, e_1 = p_0, \ldots$.
- Find $P(\text{eventual extinction})$: $e_\infty = \lim_{k\to\infty} e_k$
- Recursion: by LOTP, conditioned on the first generation,

$$e_{k+1} = p_0 + p_1 e_k + p_2 e_k^2 + \cdots$$

Probability of extinction

Let $e_k = P(\text{extinction by the } k\text{th generation})$.

- Then $e_0 = 0, e_1 = p_0, \ldots$.
- Find $P(\text{eventual extinction})$: $e_\infty = \lim_{k\to\infty} e_k$
- Recursion: by LOTP, conditioned on the first generation,

$$e_{k+1} = p_0 + p_1 e_k + p_2 e_k^2 + \cdots$$

that is

$$\boxed{\begin{cases} e_{k+1} = G(e_k) \\ e_0 = 0 \end{cases}}$$

Probability of extinction

Let $e_k = P(\text{extinction by the } k\text{th generation})$.

- Then $e_0 = 0, e_1 = p_0, \ldots$.
- Find $P(\text{eventual extinction})$: $e_\infty = \lim_{k\to\infty} e_k$
- Recursion: by LOTP, conditioned on the first generation,

$$e_{k+1} = p_0 + p_1 e_k + p_2 e_k^2 + \cdots$$

that is

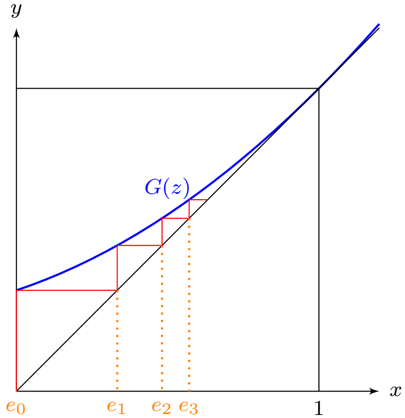
$$\boxed{\begin{cases} e_{k+1} = G(e_k) \\ e_0 = 0 \end{cases}}$$

- Trivial case: $p_0 = 0$ (always at least one child), then $e_k = 0$ for all k
- Next assume non-trivial case

$$p_0 > 0$$

Iterative maps

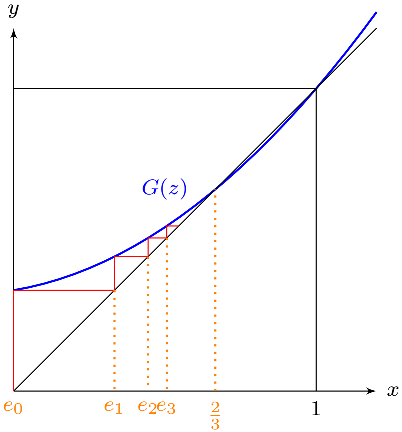
$(p_0, p_1, p_2) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \implies G(z) = (1 + z + z^2)/3$



In this case there is a unique fixed point (solution to $G(z) = z$) at 1. So e_k converges to the $e_\infty = 1$, i.e., $P(\text{extinction}) = 1$.

Iterative maps

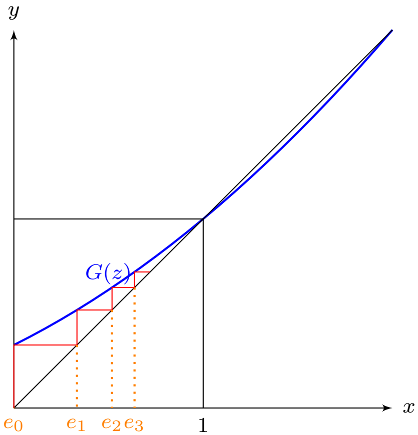
$$(p_0, p_1, p_2) = (\tfrac{1}{3}, \tfrac{1}{6}, \tfrac{1}{2}) \implies G(z) = (2 + z + 3z^2)/6$$



In this case $G(z) = z$ has two solutions $2/3$ and 1. Starting from $e_0 = 0$, e_k converges to $e_\infty = 2/3$, i.e,
$P(\text{extinction}) = 2/3$ and $P(\text{survival}) = 1/3$

Iterative maps

$(p_0, p_1, p_2) = (\frac{1}{3}, \frac{1}{2}, \frac{1}{6}) \implies G(z) = (2 + 3z + z^2)/6$



In this case $G(z) = z$ has two solutions 1 and 2. Starting from $e_0 = 0$, e_k converges to $e_\infty = 1$.

Criterion for extinction

$$\boxed{\begin{cases} e_{k+1} = G(e_k) \\ e_0 = 0 \end{cases}}$$

Note that PGF $G(z) = E(z^X)$ satisfies:

- $G(1) = 1$ so 1 is always a fixed point

Criterion for extinction

$$\boxed{\begin{cases} e_{k+1} = G(e_k) \\ e_0 = 0 \end{cases}}$$

Note that PGF $G(z) = E(z^X)$ satisfies:

- $G(1) = 1$ so 1 is always a fixed point
- $G(z) = p_0 + p_1 z + p_2 z^2 + \cdots$ is an increasing function of $z \geq 0$, $G(0) = p_0$

Criterion for extinction

$$\boxed{\begin{cases} e_{k+1} = G(e_k) \\ e_0 = 0 \end{cases}}$$

Note that PGF $G(z) = E(z^X)$ satisfies:

- $G(1) = 1$ so 1 is always a fixed point
- $G(z) = p_0 + p_1 z + p_2 z^2 + \cdots$ is an increasing function of $z \geq 0$, $G(0) = p_0$
- $G'(z) = p_1 + 2p_2 z + 3p_3 z^2 + \cdots$ is an increasing function, i.e., $G(z)$ is convex

Criterion for extinction

$$\boxed{\begin{cases} e_{k+1} = G(e_k) \\ e_0 = 0 \end{cases}}$$

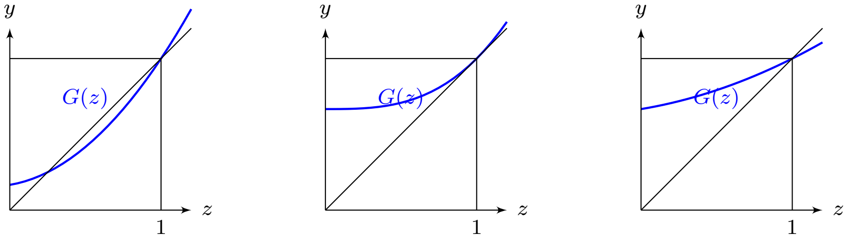
Note that PGF $G(z) = E(z^X)$ satisfies:

- $G(1) = 1$ so 1 is always a fixed point
- $G(z) = p_0 + p_1 z + p_2 z^2 + \cdots$ is an increasing function of $z \geq 0$, $G(0) = p_0$
- $G'(z) = p_1 + 2p_2 z + 3p_3 z^2 + \cdots$ is an increasing function, i.e., $G(z)$ is convex
- Key quantity:

$$G'(1) = E(X) = \text{ average number of children per individual}$$

This also follows from MGF: $G'(1) = M'_X(0) = E(X)$.

Criterion for extinction

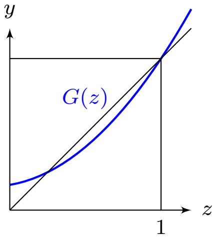


Supercritical: $E(X) > 1 \Rightarrow e_\infty < 1$

Critical: $E(X) = 1 \Rightarrow e_\infty = 1$

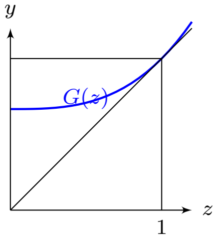
Subcritical: $E(X) < 1 \Rightarrow e_\infty = 1$

Criterion for extinction



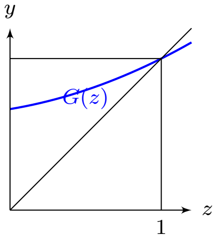
Supercritical:

$E(X) > 1 \Rightarrow e_\infty < 1$



Critical:

$E(X) = 1 \Rightarrow e_\infty = 1$



Subcritical:

$E(X) < 1 \Rightarrow e_\infty = 1$

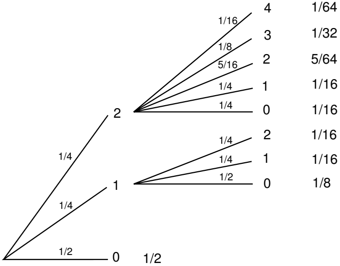
In terms of expected number of children:

- $E(X) \leq 1 \Leftrightarrow P(\text{extinction}) = 1$
- $E(X) > 1 \Leftrightarrow P(\text{extinction}) < 1$, given by the smaller solution to $G(z) = z$

Number of descendents

Number of descendents

- Let N_k be the number of descendents at kth generation, with $N_0 = 1$.
- Find the PMF is very involved:



So N_2 takes values $0, 1, 2, 3, 4$ with probabilities $\frac{11}{16}, \frac{1}{8}, \frac{9}{64}, \frac{1}{32}, \frac{1}{64}$.

- Let's find the expected number of kth-gen descendents $E(N_k)$

Strategy: Generation functions

Denote the PGF of N_k by:

$$G_k(z) \equiv E(z^{N_k})$$

Let's compute $G_k(z)$ and then differentiate:

$$G'_k(1) = E(N_k)$$

Recursions between PGFs

- $N_0 = 1$: $G_0(z) = z$.
- N_1 has PMF $(p_0, p_1, \ldots)$: $G_1(z) = G(z)$
- What about N_2?

Recursions between PGFs

Relation between N_1 and N_2:

- Suppose $N_1 = n$, then

$$N_2 = X_1 + X_2 + \cdots + X_n$$

  where the numbers of 2nd-generation childern X_i's are iid with common PMF $(p_0, p_1, \ldots)$.

- Therefore N_2 is an iid sum with N_1 (random) number of terms:

$$N_2 = \sum_{i=1}^{N_1} X_i$$

Recursions between PGFs

Find the PGF of N_2:

$$E(z^{N_2}|N_1 = n) = E(z^{X_1+\cdots+X_n}) \overset{\text{iid}}{=} \prod_{i=1}^{n} E(z^{X_i}) = G(z)^n$$

By Law of Total Expectation (Lec 22):

$$G_2(z) = E(z^{N_2}) = E(E(z^{N_2}|N_1)) = E(G(z)^{N_1}) = G(G(z))$$

Recursions between PGFs

- Relation between N_k and N_{k+1}:

$$N_{k+1} = \sum_{i=1}^{N_k} X_{k,i}$$

where the number of kth-gen children $X_{k,i}$'s are iid with common PMF $(p_0, p_1, \ldots)$.

- Find PGF of N_{k+1}: Entirely analogous to the previous slide,

$$G_{k+1}(z) = E(z^{N_{k+1}}) = E(E(z^{N_{k+1}}|N_k)),$$

where

$$E(z^{N_{k+1}}|N_k = n) = E(z^{\sum_{i=1}^{n} X_{k,i}}) \overset{\text{iid}}{=} \prod_{i=1}^{n} E(z^{X_{k,i}}) = G(z)^n$$

Thus

$$G_{k+1}(z) = E(G(z)^{N_k}) = G_k(G(z))$$

Expression of PGF

- Generating function of N_k:

$$G_k(z) = \underbrace{G \circ G \cdots \circ G}_{k \text{ times}}(z)$$

- Taking derivative yields $G'_k(1) = E(N_k)$: let $m = E(X) = G'(1)$ be the average number of offsprings per individual.
  - 0th gen: $E(N_0) = G'_0(1) = 1$

Expression of PGF

- Generating function of N_k:

$$G_k(z) = \underbrace{G \circ G \cdots \circ G}_{k \text{ times}}(z)$$

- Taking derivative yields $G_k'(1) = E(N_k)$: let $m = E(X) = G'(1)$ be the average number of offsprings per individual.
  - 0th gen: $E(N_0) = G_0'(1) = 1$
  - 1st gen: $E(N_1) = G_1'(1) = m$

Expression of PGF

- Generating function of N_k:

$$G_k(z) = \underbrace{G \circ G \cdots \circ G}_{k \text{ times}}(z)$$

- Taking derivative yields $G'_k(1) = E(N_k)$: let $m = E(X) = G'(1)$ be the average number of offsprings per individual.
  - 0th gen: $E(N_0) = G'_0(1) = 1$
  - 1st gen: $E(N_1) = G'_1(1) = m$
  - 2nd gen:

$$E(N_2) = \frac{d}{dz} G(G(z))\Big|_{z=1} = G'(\underbrace{G(1)}_{1})G'(1) = G'(1)^2 = m^2$$

Expression of PGF

- Generating function of N_k:

$$G_k(z) = \underbrace{G \circ G \cdots \circ G}_{k \text{ times}}(z)$$

- Taking derivative yields $G_k'(1) = E(N_k)$: let $m = E(X) = G'(1)$ be the average number of offsprings per individual.
  - 0th gen: $E(N_0) = G_0'(1) = 1$
  - 1st gen: $E(N_1) = G_1'(1) = m$
  - 2nd gen:

$$E(N_2) = \frac{d}{dz} G(G(z))\Big|_{z=1} = G'(\underbrace{G(1)}_{1})G'(1) = G'(1)^2 = m^2$$

  - kth gen: by induction,

$$E(N_k) = \underbrace{G'(G_{k-1}(1))}_{=G'(1)=m} \underbrace{G_{k-1}'(1)}_{=m^{k-1}} = m^k$$

Average number of descendents

Let $m = E(X) =$ average number of offsprings per individual. Then

$$E(N_k) = m^k$$

Then

- $m > 1$: $P(\text{survival}) > 0$ and average population grows exponentially
- $m = 1$: $P(\text{survival}) = 0$ and average population stays one
- $m < 1$: $P(\text{survival}) = 0$ and average population dies exponentially

Average number of descendents

Let $m = E(X) =$ average number of offsprings per individual. Then

$$E(N_k) = m^k$$

Then

- $m > 1$: $P(\text{survival}) > 0$ and average population grows exponentially
- $m = 1$: $P(\text{survival}) = 0$ and average population stays one
- $m < 1$: $P(\text{survival}) = 0$ and average population dies exponentially

For more, see

- Grinstead-Snell §10.2
- Athreya, Krishna B.; Ney, Peter E. (1972). *Branching Processes*. Berlin: Springer-Verlag.