Gamblers Ruin

$$\boxed{0} \leftarrow \boxed{1} \rightleftarrows \cdots \rightarrow \boxed{N}$$

Random Walk on the integers.

$$\cdots \; \boxed{0} \rightleftarrows \boxed{1} \rightleftarrows \cdots \rightleftarrows \boxed{N} \rightleftarrows \cdots$$

never stops.

all states "equal".

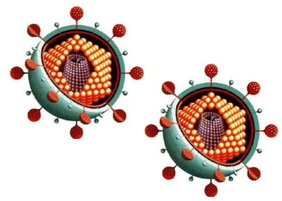
$$S^T = \begin{cases} S^{1+T'} \\ S^1 \\ S^{1+T''+T'''} \end{cases}$$

Defective?

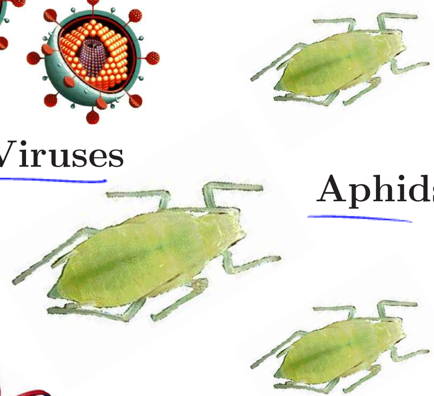
Chapter 6: Branching Processes: The Theory of Reproduction

Henry VIII

Royalty



Viruses

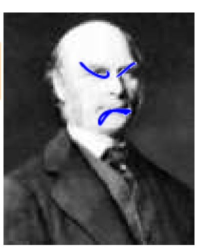


Aphids

DNA

Although the early development of Probability Theory was motivated by problems in gambling, probabilists soon realised that, if they were to continue as a breed, they must also study reproduction.

Galton



Watson

Reproduction is a complicated business, but considerable insights into population growth can be gained from simplified models. The **Branching Process** is a simple but elegant model of population growth. It is also called the **Galton-Watson Process,** because some of the early theoretical results about the process derive from a correspondence between Sir Francis Galton and the Reverend Henry William Watson in 1873. Francis Galton was a cousin of Charles Darwin. In later life, he developed some less elegant ideas about reproduction — namely eugenics, or selective breeding of humans. Luckily he is better remembered for branching processes.

$H(s) = \mathbb{E}(s^T)$

T defective? $\longrightarrow$ $H(1) = 1 \Rightarrow$ no

Guaranteed to reach your goal

T: # steps to reach goal.

$H(1) < 1 \Rightarrow$ yes

Not guaranteed to reach goal.

$$\mathbb{P}(\text{never get there}) = 1 - H(1)$$
$$= \mathbb{P}(T = \infty)$$

$$\therefore \mathbb{P}(\text{get there eventually}) = H(1).$$

$\overset{q}{\longleftarrow} \qquad \overset{p}{\longrightarrow}$

$\boxed{\text{Branching Processes}}$ $\qquad$ $\boxed{\text{Markov Chains}}$

6.1 Branching Processes

Consider some sort of ***population*** consisting of reproducing individuals.

Examples: living things (animals, plants, bacteria, royal families);
diseases; computer viruses;
rumours, gossip, lies (one lie always leads to another!)

Start conditions: Start at time $n=0$, with a single individual.

Each individual: lives for exactly 1 unit of time. At time $n=1$ it produces a family of offspring, and immediately dies.

How many offspring? Could be 0, 1, 2, ... This is the "family size", Y. (Y stands for # Young.)

Each offspring: lives for 1 unit of time. At time $n=2$, it produces its own family of offspring, and immediately dies.

and so on... Time 0 1 2 3 4

Assumptions

1. All individuals reproduce independently of each other.
2. The family sizes of different individuals are independent, identically distributed r.v.s. (Denote by $Y_1, Y_2, \ldots$)

Family size distribution, Y

y	0	1	2	3	4	...
$P(Y=y)$	p_0	p_1	p_2	p_3	p_4	...

Definition: A **branching process** is defined as follows.

- Single individual at time $n = 0$. $Z_0 = 1$.
- Every individual lives exactly one unit of time, then produces Y offspring, and dies. – parents & offspring never coexist.
- The number of offspring, Y, takes values $0, 1, 2, \ldots$, and the probability of producing k offspring is $P(Y=k) = p_k$.
- All individuals reproduce independently. Individuals $1, 2, \ldots, n$ have family sizes $Y_1, Y_2, \ldots Y_n$, where each Y_i has the same distribution as Y.
- Let Z_n be the number of individuals born at time n, for $n = 0, 1, 2, \ldots$. Interpret Z_n as the size of generation n. (# indivs alive at time n)
- Then the branching process is $\{Z_0, Z_1, Z_2, \ldots\}$. $= \{Z_n : n \in \mathbb{N}\}$.

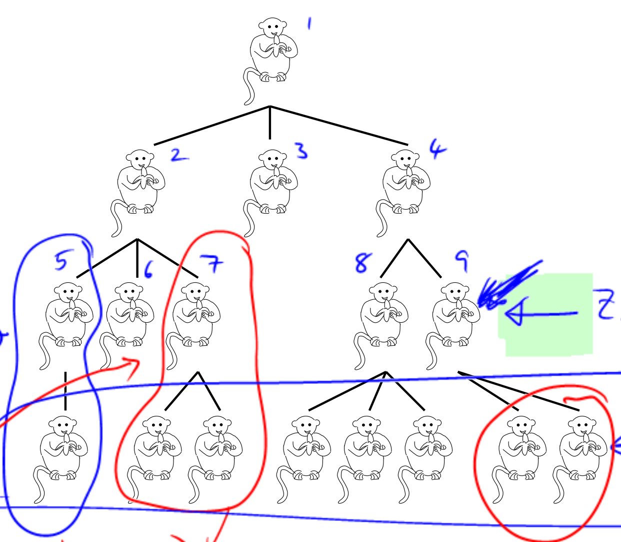
Definition: The **state** of the branching process at time n is z_n, where each z_n can take values $0, 1, 2, \ldots$.
Note that $z_0 = 1$ always.
z_n = # indivs alive @ time n = population size at time n.

Note: When we want to say that two random variables X and Y have the same distribution, we write: $X \sim Y$.
For example: $Y_i \sim Y$, where Y_i = family size of any individual i.
$Z_1 \sim Y$ use this a lot! Because $Z_0 = 1$.

$Z_0 = 1$, Z_1

Note: The definition of the branching process is easily generalized to start with more than one individual at time $n = 0$.

Branching Process



(p_1) $Y_5 = 1$

(p_0) $Y_6 = 0$

$Y_7 = 2$ (prob $= p_2$)
etc.

Z_{n-1}

Z_n

Generation	Size
0	$Z_0 = 1$
1	$Z_1 = 3$
2	$Z_2 = 5$
3	$Z_3 = 8$ G_3
⋮	⋮

6.2 Questions about the Branching Process

When we have a situation that can be modelled by a branching process, there are several questions we might want to answer.

If the branching process is just beginning, what will happen in the future?

1. What can we find out about the distribution of Z_n (the population siZe at generation n)?

   - can we find the mean and variance of Z_n?
     — *yes, using the probability generating function of family size, Y;*

   - can we find the whole distribution of Z_n?
     — *for special cases of the family size distribution Y, we can find the PGF of Z_n explicitly;* $Y \sim$ Geometric is only well-known non-trivial case.

   - can we find the probability that the population has become **extinct** by generation n, $\mathbb{P}(Z_n = 0)$?
     — *for special cases where we can find the PGF of Z_n (as above).*

2. What can we find out about eventual extinction?

   - can we find the probability of eventual extinction, $\mathbb{P}(\lim_{n\to\infty} Z_n = 0)$?
     — Yes, always, using the PGF of Y.

   - can we find general conditions for eventual extinction?
     — yes: we can find (easy) conditions that guarantee extinction with probability 1: $\mathbb{E}(Y) \leq 1$. (essentially)

   - if eventual extinction is definite, can we find the distribution of the time to extinction?
     — for special cases only.

Example: Modelling cancerous growths. Will a colony of cancerous cells become extinct before it is sufficiently large to overgrow the surrounding tissue?

If the branching process is already in progress, what happened in the past?

1. How long has the process been running?
   - *how many generations do we have to go back to get to the single common ancestor?*

2. What has been the distribution of family size over the generations?

3. What is the total number of individuals (over all generations) up to the present day?

Example: It is believed that all humans are descended from a single female ancestor, who lived in Africa. How long ago? ~200,000 years ago.

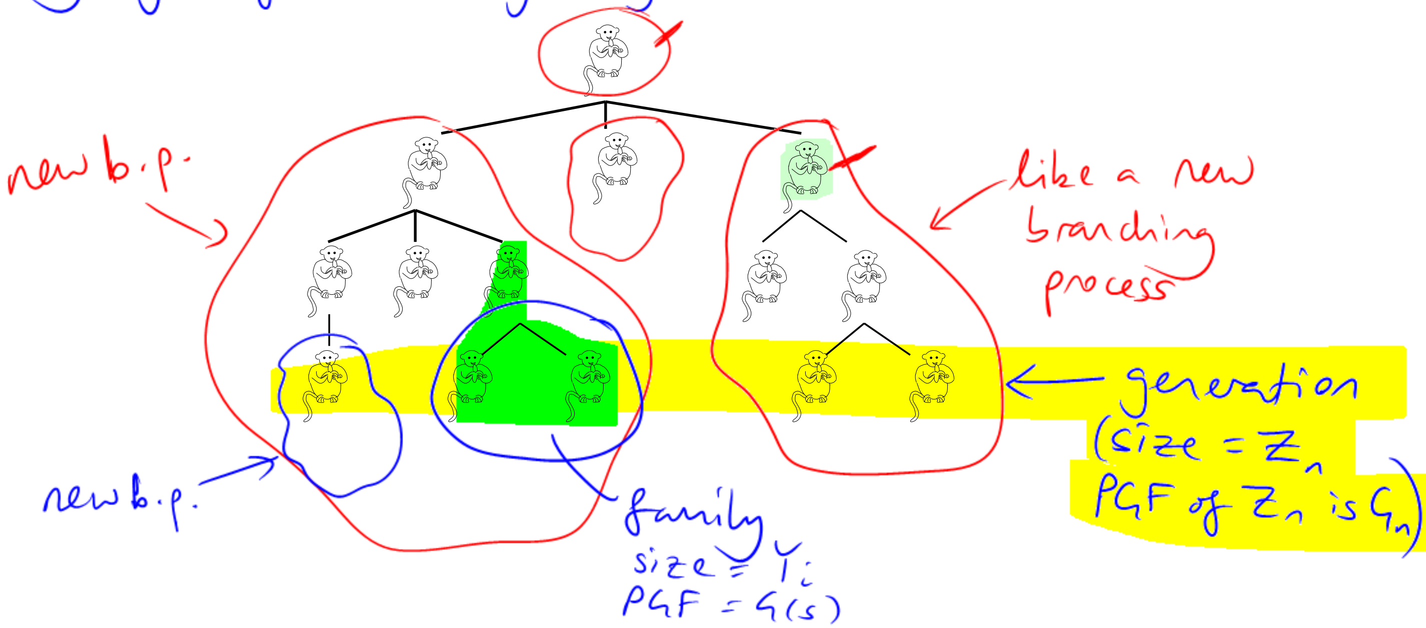
What has been the mean family size over that period?
Probably very close to 1 female child per female adult: estimate is 1.002.

6.3 Analysing the Branching Process

$$A \xrightarrow{} B, \quad A \xrightarrow{p} A$$

$$m_A = 1 + m_A p + m_B(1-p)$$

Key Observation: Every individual in every generation starts a new, independent, branching process, as if the whole process were starting again from the beginning.



Z_n as a randomly stopped sum

Most of the interesting properties of the branching process centre on the distribution of Z_n (the population size at time n). Using the Key Observation from overleaf, we can find an expression for the probability generating function of Z_n.

Consider the following.

- The pop^n size at time $n-1$ is Z_{n-1}. (parents)
- Label these parents as $1, 2, 3, \ldots, Z_{n-1}$. (unknown random number)
- Each individual $1, 2, \ldots, Z_{n-1}$ starts a new branching process.
  Let $Y_1, Y_2, \ldots, Y_{Z_{n-1}}$ be the #children of parents $1, 2, \ldots, Z_{n-1}$
- Then total pop^n size at time n, Z_n, is

$$Z_n = \underbrace{Y_1}_{\substack{\#\text{kids of}\\ \text{parent } 1}} + \underbrace{Y_2}_{\substack{\#\text{kids of}\\ \text{parent } 2}} + \ldots + \underbrace{Y_{Z_{n-1}}}_{\substack{\#\text{kids of}\\ \text{parent } Z_{n-1}\\ \text{(last one)}}} = \sum_{i=1}^{Z_{n-1}} Y_i .$$

So Z_n is a randomly stopped sum: a sum of $Y_1, Y_2, \ldots$, randomly stopped by the r.v. Z_{n-1} = #parents.

Note: 1. Each $Y_i \sim Y$.

2. $Y_1, Y_2, \ldots, Y_{Z_{n-1}}$ are indept of each other and of Z_{n-1}.

$$\overset{T}{Z_n} = \sum_{i=1}^{Z_{n-1} \leftarrow N} Y_i \leftarrow X$$

$T = X_1 + \cdots + X_N$ N random

$G_T(s) = G_N(G_X(s))$ Thm 4.6.

Probability Generating Function of Z_n

Let $G_Y(s) = \mathbb{E}(s^Y)$ be the probability generating function of Y.
(Recall that Y is the # Young of any individual: the family size.

Now Z_n is a randomly stopped sum: it is the sum of $Y_1, Y_2, \ldots$, stopped by the random variable Z_{n-1}. So we can use Theorem 4.6 (Chapter 4) to express the PGF of Z_n directly in terms of the PGFs of Y and Z_{n-1}.

By Theorem 4.6, if $Z_n = Y_1 + Y_2 + \ldots + Y_{Z_{n-1}}$, and Z_{n-1} is itself random, then the PGF of Z_n is given by:

$G_{Z_n}(s) = \mathbb{E}(s^{Z_n})$

$G_{Z_{n-1}}(s) = \mathbb{E}(s^{Z_{n-1}})$

$$(*) \quad G_{Z_n}(s) = G_{Z_{n-1}}(G_Y(s)).$$

For ease of notation, we can write:

$$G_{Z_n}(s) = G_n(s), \qquad G_{Z_{n-1}}(s) = G_{n-1}(s), \text{ and so on.}$$

Note that $Z_1 \sim Y$ (# offspring of a single parent, as $Z_0 = 1$)
so we can also write:

$$G_Y(s) = G_{Z_1}(s) = G_1(s) = \underline{G(s)} \text{ for simplicity.}$$

Thus, from $(*)$:

$$\boxed{G_n(s) = G_{n-1}(G(s)) \quad \text{Branching Process Recursion Formula.}}$$

Note:

1. $G_n(s) = \mathbb{E}(s^{Z_n})$ = PGF of family size at time n.

2. $G_{n-1}(s) = \mathbb{E}(s^{Z_{n-1}})$ = ______________ $n-1$

3. $G(s) = \mathbb{E}(s^Y) = \mathbb{E}(s^{Z_1}) = G_1(s)$ = PGF of single family size, Y.

We are trying to find the PGF of Z_n, the population size at time n.

So far, we have: $G_n(s) = G_{n-1}\Big(G(s)\Big)$. $(\star)$

But by the same argument,

$$G_{n-1}(r) = G_{n-2}\big(G(r)\big) \qquad (**)$$

Substituting in $(\star)$,

$$\begin{aligned} G_n(s) &= G_{n-1}\big(G(s)\big) \\ &= G_{n-1}(r) \text{ where } r = G(s) \\ &= G_{n-2}\big(G(r)\big) \text{ by } (**) \end{aligned}$$

$$\therefore \; G_n(s) = G_{n-2}\Big(G\big(G(s)\big)\Big) \text{ replacing } r = G(s).$$

By the same reasoning, we will obtain:

$$G_n(s) = G_{\underbrace{n-3}_{n-3}}\Big(\underbrace{G(G(G(s)))}_{3 \text{ times}}\Big)$$

and so on, until we finally get:

$$G_n(s) = G_{\underbrace{n-(n-1)}_{n-1}}\Big(\underbrace{G(G(G(\cdots G(s)\cdots)))}_{n-1 \text{ times}}\Big)$$

$n-(n-1) \to 1$

$$= \underbrace{G_1}_{=G}\Big(\underbrace{G(G(G(\cdots\cdots G(s)\cdots)))}_{n-1 \; G\text{'s}}\Big)$$

$$G_n(s) = \underbrace{G(G(G(G(\cdots\cdots G(s)\cdots))))}_{n \text{ times}}.$$

We have therefore proved the following Theorem.

PGF of Z_n = $G_n(s) = \mathbb{E}(s^{Z_n})$
PGF of Y = $G(s) = \mathbb{E}(s^Y)$

Theorem 6.3: Let $G(s) = \mathbb{E}(s^Y) = \sum_{y=0}^{\infty} p_y s^y$ be the PGF of the family size distribution, Y. Let $Z_0 = 1$ (start from a single individual at time 0), and let Z_n be the population size at time n ($n = 0, 1, 2, \ldots$). Let $G_n(s)$ be the PGF of the random variable Z_n. Then

$$\rightarrow \quad G_n(s) = \underbrace{G\Big(G\Big(G\Big(\ldots G}_{n \text{ times}}(s)\ldots\Big)\Big)\Big). \qquad \square$$

Note: $G_n(s) = \underbrace{G\Big(G\Big(G\Big(\ldots G}_{n \text{ times}}(s)\ldots\Big)\Big)\Big)$ is called the n-fold iterate of G.

We have therefore found an expression for the PGF of the population size at generation n, although there is no guarantee that it is possible to write it down or manipulate it very easily for large n. For example, if Y has a Poisson(λ) distribution, then $G(s) = e^{\lambda(s-1)}$, and already by generation $n = 3$ we have the following fearsome expression for $G_3(s)$:

$$G_3(s) = e^{\lambda\left(e^{\lambda\left(e^{\lambda(s-1)}-1\right)}-1\right)}. \qquad \text{(Or something like that!)}$$

However, in some circumstances we can find quite reasonable closed-form expressions for $G_n(s)$, notably when Y has a Geometric distribution. In addition, for any distribution of Y we can use the expression $G_n(s) = G_{n-1}\Big(G(s)\Big)$ to derive properties such as the mean and variance of Z_n, and the probability of eventual extinction ($\mathbb{P}(Z_n = 0)$ for some n).

6.4 What does the distribution of Z_n look like?

Before deriving the mean and the variance of Z_n, it is helpful to get some intuitive idea of how the branching process behaves. For example, it seems reasonable to calculate the mean, $\mathbb{E}(Z_n)$, to find out what we expect the population size to be in n generations time, but why are we interested in $\text{Var}(Z_n)$?

The answer is that Z_n usually has a "boom-or-bust" distribution: either the population will take off (boom), and the population size grows quickly, or the population will fail altogether (bust). In fact, if the population fails, it is likely to do so very quickly, within the first few generations. This explains why we are

AIDS
HIV : $\mathbb{E}(Y) = 1+\varepsilon$
$1-\varepsilon$

interested in $\text{Var}(Z_n)$. A huge variance will alert us to the fact that the process does not cluster closely around its mean values. In fact, the mean might be almost useless as a measure of what to expect from the process.

Simulation 1: $Y \sim \text{Geometric}(p = 0.3)$

The following table shows the results from 10 simulations of a branching process, where the family size distribution is $Y \sim \text{Geometric}(p = 0.3)$.

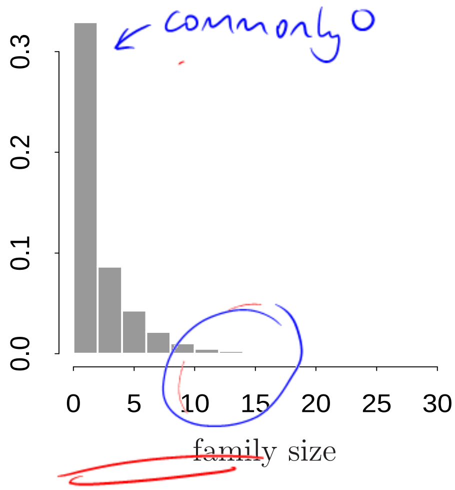
$\mathbb{E}Y = \frac{0.7}{0.3} > 1$

Simulation	Z_0	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
1	1	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0	0
3	1	4	19	42	81	181	433	964	2276	5383	12428
4	1	3	3	5	3	15	29	86	207	435	952
5	1	0	0	0	0	0	0	0	0	0	0
6	1	1	0	0	0	0	0	0	0	0	0
7	1	2	8	26	68	162	360	845	2039	4746	10941
8	1	1	0	0	0	0	0	0	0	0	0
9	1	1	0	0	0	0	0	0	0	0	0
10	1	1	4	13	18	39	104	294	690	1566	3534

Often, the population is extinct by generation 10. However, when it is not extinct, it can take enormous values (12428, 10941, ...).

The same simulation was repeated 5000 times to find the empirical distribution of the population size at generation 10 (Z_{10}). The figures below show the distribution of family size, Y, and the distribution of Z_{10} from the 5000 simulations.

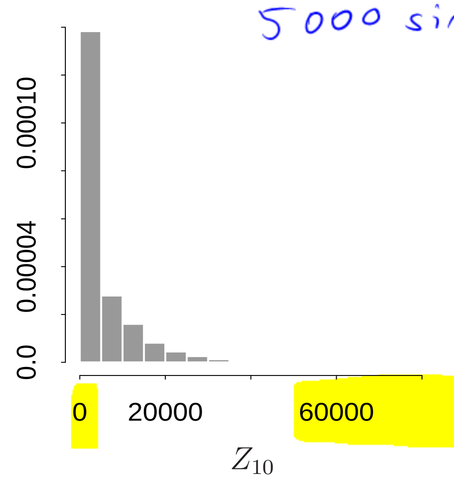
Family Size, Y



commonly 0

Z_{10}

5000 simulations.



In this example, the family size is rather variable, but the variability in Z_{10} is enormous (note the range on the histogram from 0 to 60,000). Some statistics are:

```
Proportion of samples extinct by generation 10:  0.436
```

5000 sims

```
Summary of Zn:
     Min  1st Qu  Median   Mean  3rd Qu    Max
       0       0    1003   4617    6656  82486

Mean of Zn:        4617.2
Variance of Zn:   53937785.7
```

So the empirical variance is $\text{Var}(Z_{10}) = 5.39 \times 10^7$. This perhaps contains more useful information than the mean value of 4617. The distribution of Z_n has 43.6% of zeros, but (when it is non-zero) takes values up to 82,486. Is it really useful to summarize such a distribution by the single mean value 4617?

For interest, out of the 5000 simulations, there were only 35 (0.7%) that had a value for Z_{10} greater than 0 but less than 100. This emphasizes the "boom-or-bust" nature of the distribution of Z_n.

Simulation 2: $Y \sim$ Geometric($p = 0.5$)

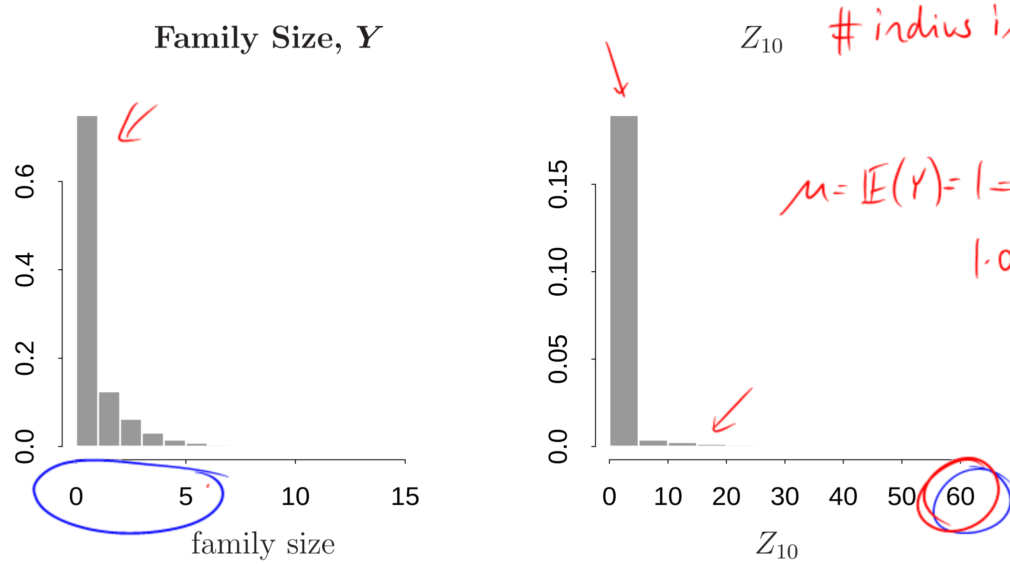
$\mathbb{E}(Y) = \frac{0.5}{0.5} = 1$.

We repeat the simulation above with a different value for p in the Geometric family size distribution: this time, $p = 0.5$. The family size distribution is therefore

Simulation	Z_0	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
1	1	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0	0	0
5	1	1	0	0	0	0	0	0	0	0	0
6	1	7	9	17	15	20	19	8	7	13	35
7	1	2	5	2	5	8	8	3	3	0	0
8	1	2	0	0	0	0	0	0	0	0	0
9	1	0	0	0	0	0	0	0	0	0	0
10	1	0	0	0	0	0	0	0	0	0	0

This time, almost all the populations become extinct. We will see later that this value of p (just) guarantees eventual extinction with probability 1.

The family size distribution, $Y \sim \text{Geometric}(p = 0.5)$, and the results for Z_{10} from 5000 simulations, are shown below. Family sizes are often zero, but families of size 2 and 3 are not uncommon. It seems that this is not enough to save the process from extinction. This time, the maximum population size observed for Z_{10} from 5000 simulations was only 56, and the mean and variance of Z_{10} are much smaller than before.



indivs in generation 10

$\mu = \mathbb{E}(Y) = 1 \Rightarrow$ certain ext

$1.0001 \Rightarrow$ likely ext but not certain

```
Proportion of samples extinct by generation 10:  0.9108

Summary of Zn:
     Min  1st Qu  Median    Mean  3rd Qu    Max
       0       0       0   0.965       0     56

Mean of Zn:       0.965
Variance of Zn:  19.497
```

What happens for larger values of p?

It was mentioned above that $Y \sim \text{Geometric}(p = 0.5)$ just guarantees eventual extinction with probability 1. For $p > 0.5$, extinction is also guaranteed, and tends to happen quickly. For example, when $p = 0.55$, over 97% of simulated populations are already extinct by generation 10.

6.5 Mean and variance of Z_n

The previous section has given us a good idea of the significance and interpretation of $\mathbb{E}(Z_n)$ and $\text{Var}(Z_n)$. We now proceed to calculate them. Both $\mathbb{E}(Z_n)$ and $\text{Var}(Z_n)$ can be expressed in terms of the mean and variance of the family size distribution, Y.

Thus, let $\mathbb{E}(Y) = \mu$ and let $\text{Var}(Y) = \sigma^2$.

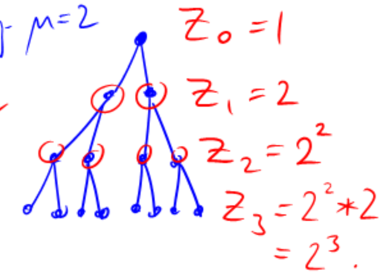
These are mean & variance of # offspring of a single individual.

Theorem 6.5: Let $\{Z_0, Z_1, Z_2, \ldots\}$ be a branching process with $Z_0 = 1$ (start with a single individual). Let Y denote the family size distribution, and suppose that $\mathbb{E}(Y) = \mu$. Then

$$\boxed{\mathbb{E}(Z_n) = \mu^n}.$$

e.g. $\mu = 2$

"ideal" version of the B.P.



Proof:

By p. 111, $Z_n = Y_1 + Y_2 + \ldots + Y_{Z_{n-1}}$ is a randomly stopped sum:

$$Z_n = \sum_{i=1}^{Z_{n-1}} Y_i .$$

Z_{n-1} ← random

Thus by Section 3.4 (page 60),

$$\mathbb{E}(Z_n) = \underbrace{\mathbb{E}(Z_{n-1})} * \underbrace{\mathbb{E}(Y_i)}_{\mu} \qquad \text{Wald's eqn.}$$

$$= \mu\, \mathbb{E}(Z_{n-1})$$

$$= \mu \{ \mu\, \mathbb{E}(Z_{n-2}) \} \quad \text{by same argument}$$

$$= \mu^2\, \mathbb{E}(Z_{n-2})$$

$$\vdots$$

$$= \mu^{n-1}\, \mathbb{E}(Z_{n-(n-1)})$$

$$= \mu^{n-1}\, \mathbb{E}(Z_1)$$

$$= \mu^{n-1} * \mu$$

$$= \mu^n \text{ as required.} \qquad \square$$

Examples: Consider the simulations of Section 6.4.

1. Family size $Y \sim \text{Geometric}(p = 0.3)$. So $\mu = \mathbb{E}(Y) = \frac{q}{p} = \frac{0.7}{0.3} = 2.33.$

   Expected population size by generation $n = 10$ is:

   $$\mathbb{E}(Z_{10}) = \mu^{10} = (2.33)^{10} = 4784.$$

   The theoretical value, 4784, compares well with the sample mean from 5000 sims, 4617 (p. 116).

2. Family size $Y \sim \text{Geometric}(p = 0.5)$. So $\mu = \mathbb{E}(Y) = \frac{q}{p} = \frac{0.5}{0.5} = 1.$

   So $\mathbb{E}(Z_{10}) = \mu^{10} = 1^{10} = 1.$

   Compares well with sample mean of 0.965 (p. 117).

Variance of Z_n

$\mathbb{E}(Z_n)$ R.S.S L.o.T.E.

$\text{Var}(Z_n)$ RSS L.o.T.V.

Theorem 6.5: Let $\{Z_0, Z_1, Z_2, \ldots\}$ be a branching process with $Z_0 = 1$ (start with a single individual). Let Y denote the family size distribution, and suppose that $\mathbb{E}(Y) = \mu$ and $\text{Var}(Y) = \sigma^2$. Then

721 extra exam Q 2009 prove by induction?

$$\text{Var}(Z_n) = \begin{cases} \sigma^2 n & \text{if } \mu = 1, \\ \sigma^2 \mu^{n-1} \left(\dfrac{1 - \mu^n}{1 - \mu} \right) & \text{if } \mu \neq 1 \quad (> 1 \text{ or } < 1). \end{cases}$$

Proof:

Write $V_n = \text{Var}(Z_n)$. The proof works by finding a recursive formula for V_n.

V_n i.t.o. V_{n-1}

$\to V_1 = \sigma^2$

Same idea useful for Ass Q3. (different quantity).

Using the Law of Total Variance for randomly stopped sums from Section 3.4 (page 60),

$$Z_n = \sum_{i=1}^{Z_{n-1}} Y_i$$

R.S.S. for Z_n

$$\begin{aligned}
\Rightarrow \quad \text{Var}(Z_n) &= \{\mathbb{E}(Y_i)\}^2 \times \text{Var}(Z_{n-1}) + \text{Var}(Y_i) \times \mathbb{E}(Z_{n-1}) \\
\Rightarrow \quad V_n &= \mu^2 V_{n-1} + \sigma^2 \mathbb{E}(Z_{n-1}) \\
\Rightarrow \quad V_n &= \mu^2 V_{n-1} + \sigma^2 \mu^{n-1},
\end{aligned}$$

L.o.T.V. for R.S.S.

using $\mathbb{E}(Z_{n-1}) = \mu^{n-1}$ as above.

Also,

$$V_1 = \text{Var}(Z_1) = \text{Var}(Y) = \sigma^2.$$

Find V_n by repeated substitution:

$$V_1 = \sigma^2$$

$$V_2 = \mu^2 V_1 + \sigma^2\mu = \mu^2\sigma^2 + \mu\sigma^2 = \mu\sigma^2(1+\mu)$$

$$V_3 = \mu^2 V_2 + \sigma^2\mu^2 = \mu^2\sigma^2\left(1+\mu+\mu^2\right)$$

$$V_4 = \mu^2 V_3 + \sigma^2\mu^3 = \mu^3\sigma^2\left(1+\mu+\mu^2+\mu^3\right)$$

$$\vdots \quad \text{etc.}$$

Completing the pattern,

$$\begin{aligned}
V_n &= \mu^{n-1}\sigma^2\left(1+\mu+\mu^2+\ldots+\mu^{n-1}\right) \\
&= \mu^{n-1}\sigma^2 \sum_{r=0}^{n-1} \mu^r \\
&= \mu^{n-1}\sigma^2\left(\frac{1-\mu^n}{1-\mu}\right).
\end{aligned}$$

always true.

Valid for $\mu \neq 1$.

(sum of first n terms of Geometric series)

When $\mu = 1$:

$$V_n = 1^{n-1}\sigma^2 \underbrace{\left(1^0 + 1^1 + \ldots + 1^{n-1}\right)}_{\text{n times}} = \sigma^2 n.$$

Hence the result:

$$\text{Var}(Z_n) = \begin{cases} \sigma^2 n & \text{if } \mu = 1, \\ \sigma^2 \mu^{n-1} \left(\dfrac{1-\mu^n}{1-\mu}\right) & \text{if } \mu \neq 1. \end{cases} \qquad \square$$

Examples: Again consider the simulations of Section 6.4.

1. Family size $Y \sim \text{Geometric}(p = 0.3)$. So $\mu = \mathbb{E}(Y) = \dfrac{q}{p} = \dfrac{0.7}{0.3} = 2.33.$

$\sigma^2 = Var(Y) = \dfrac{q}{p^2} = \dfrac{0.7}{(0.3)^2} = 7.78.$

$Var(Z_{10}) = \sigma^2 \mu^9 \left(\dfrac{1-\mu^{10}}{1-\mu}\right) = 5.72 * 10^7.$ Huge variance!

Sample var from 5000 sims was $5.39 * 10^7$.

2. Family size $Y \sim \text{Geometric}(p = 0.5)$. So $\mu = \mathbb{E}(Y) = \dfrac{q}{p} = \dfrac{0.5}{0.5} = 1.$

$\sigma^2 = Var(Y) = \dfrac{q}{p^2} = \dfrac{0.5}{0.5^2} = 2.$

Use special formula for $\mu = 1$:

$Var(Z_{10}) = \sigma^2 n = 2 * 10 = 20.$

Sample var $= 19.5$ (p. 117).