Branching Processes
Dem260 Math Demog
Spring 2020
Lecture 8

Joshua R. Goldstein

March 12, 2020

Agenda for today

1. A viral example, simulating a branching process, our questions
2. Generating functions and extinction
3. Next time, finish up BP and get started on Fisher-Wright

Stochastic modeling

- Until now, we've focused on the hidden structures of heterogeneity.
- Now, we're wwitching gears:
  - Stochastic not deterministic
  - In small populations, randomness matters. (Even when risks are homoteneous.)
  - Todayn: branching processes ("parents producing children"), next Fisher-Wright ("children choosing parents"), and then historical reconstruction from contemporary diversity ("coalescent").

Very brief history of Branching Processes

- Bienayme's lost notes
- Galton and Watson's (extinction of families)
- Genetics (survival of a mutant)
- The bomb (chain reactions)
- Anywhere "incipient dynamics" matter.

Applicability to the Coronavirus? Yes and no.

- Perhaps the beginning, with first few cases.
- But once scale gets large, we'll see that deterministic dynamics take over.
- One lesson: beyond R_0.

Our viral example

Here are the chances that the first carrier passes on the virus to k people?

k	p_k	digits
0	.3	0-2
1	.4	3-5
2	.3	6-9

- What is R_0, (aka m)? Calculate.
- Let's diagram one chance outcome, using my number "(xxx) xxx-9056"

9, 0, 5, 6

k	p_k	digits
0	.3	0-2
1	.4	3-5
2	.3	6-9

Let's repeat

k	p_k	digits
0	.3	0-2
1	.4	3-5
2	.3	6-9

1. Let's try another diagram as a group. First name alphabetically (last cell digit)
2. Everyone do their last 4 digits

What is a (Bienaymé-)Galton-Watson branching process?

p_k	Each individual in each generation reproduces independently, following same offspring distribution, with p_k as the probability of having k offspring.
Z_n	The si$Z$e of the n'th generation Z_n. ($Z_1 \equiv 1$)
$p_0 > 0$	Some non-zero probability of no children.
Variance	None of the p_k are 1

Some questions

- What is the chance d of eventual extinction (no “outbreak”)?
- Or, what is the distribution of surviving family sizes?
- What are the aggregate properties of many branching processes? (Mean growth, variance, time-paths, eventual size)?

Galton's original question

300 THE EDUCATIONAL TIMES. [Mar. 1, '73.

4001. (Proposed by Francis Galton.)—A large nation, of whom we will only concern ourselves with the adult males, N in number, and who each bear separate surnames, colonise a district. Their law of population is such that, in each generation, a_0 per cent. of the adult males have no male children who reach adult life; a_1 have one such male child; a_2 have two; and so on, up to a_5 who have five. Find (1) what proportion of the surnames will have become extinct after r generations; and (2) how many instances there will be of the same surname being held by m persons.

[The Proposer remarks that a general solution of this problem would be of much aid in certain rather important statistical enquiries, and that he finds it a laborious matter to work it out numerically, in even the simplest special cases, and to only a few generations. In reality, the generations would overlap and mix, but it is not necessary to suppose them otherwise than as occurring in successive steps.]

A simulation

```
k = 0:2
p0 = .3; p1 = .3; p2 = .4;
p_k = c(p0, p1, p2)
Z1 = 1
set.seed(9)
(kids.of.Z1 = sample(x = k, size = Z1, replace = T, prob = p_k))

## [1] 2

(Z2 = sum(kids.of.Z1))

## [1] 2

(kids.of.Z2 = sample(x = k, size = Z2, replace = T, prob = p_k))

## [1] 2 2

(Z3 = sum(kids.of.Z2))

## [1] 4

(kids.of.Z3 = sample(x = k, size = Z3, replace = T, prob = p_k))

## [1] 2 1 2 2

(Z4 = sum(kids.of.Z3))

## [1] 7
```

Let's draw the tree.

A simulation (2)

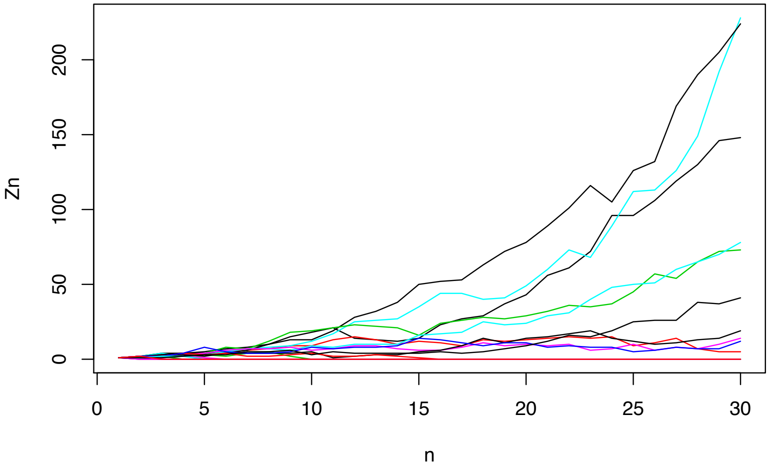
A function

```
branch <- function(n_max = 30, pk = c(p0, p1, p2), Z1 = 1)
{
    Z.vec <- rep(NA, n_max)
    Z.vec[1] <- Z1
    for (i in 1:(n_max-1))
    {
        Z.vec[i+1] <- sum(sample(x = k,
                                 size = Z.vec[i],
                                 replace = T,
                                 prob = p_k))
    }
    return(Z.vec)
}
set.seed(19); branch()
## [1] 1 2 2 4 5 2 2 3 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
set.seed(99);  branch()
## [1] 1 1 2 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

A simulation (3)

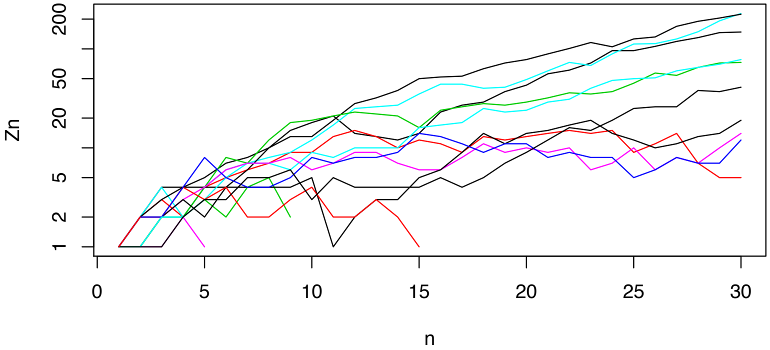
Let's see what happens with 20 trials (up to 30 generations)

```
n_trials = 20; n_gen = 30
Z.mat <- matrix(NA, n_trials, n_gen)
set.seed(131)
for (i in 1:n_trials)
    Z.mat[i,] <- branch(n_max = n_gen)
matplot(t(Z.mat),
        type = "l", lty = 1, ylab = "Zn", xlab = "n")
```



How many survive (out of 20)?

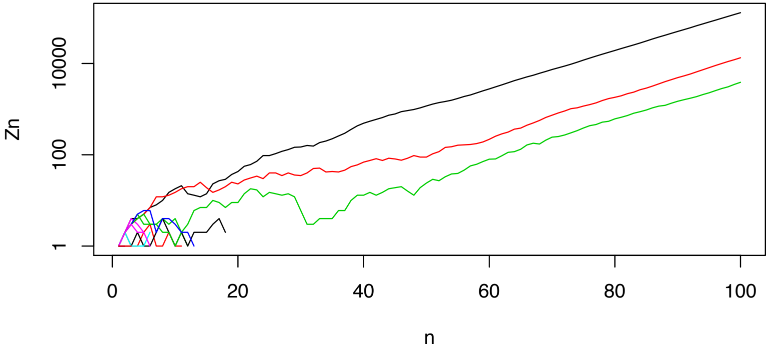
log-scale




```
## surviving
## extinct survive
##     0.5     0.5
```

- How would you discribe the time path of the surviving lines?

Long term



What does this remind you of? (Hint: "Leslie"). (See Harris figure)

“Extinction” vs “breakout”

- We see that in a super-critical ($m > 1$) branching process, if a line can survive a few generations and reach a large enough size, it will grow exponentially.
- What happens if $m < 1$, if $m = 1$? Discuss.

Mathematical analysis

The Probability Generating Function: Our mathematical tool

$$h(z) = p_0 + p_1 z + p_2 z^2 + \dots$$

The PGF "keeps book" on the probabilities. The chance of k is the coefficient on z^k.

Some interesting properties

$h(0) =$

$h(1) =$

$h'(1) =$

But the magic is next.

The story of two brothers

A father has two sons. The probability generating function of their children combined is:

$$[h(z)]^2 = (p_0 + p_1 z + p_2 z^2) \times (p_0 + p_1 z + p_2 z^2)$$

Multiply out, and tell me the coefficients on $z^0, z^1, \ldots$.

Even more amazing

What is the probability generating function for the distribution of grandsons?

- A man has two sons, with probability p_2, so PGF in that case is $p_2[h(z)]^2$.
- But let's sum over all possible numbers of sons.

$$p_0 + p_1 h(z) + p_2[h(z)]^2 + p_3[h(z)]^3 + \dots$$

- Which is? (Hint: write a new argument for PGF)

Even more amazing

What is the probability generating function for the distribution of grandsons?

- A man has two sons, with probability p_2, so PGF in that case is $p_2[h(z)]^2$.
- But let's sum over all possible numbers of sons.

$$p_0 + p_1 h(z) + p_2[h(z)]^2 + p_3[h(z)]^3 + \dots$$

- Which is? (Hint: write a new argument for PGF)

$$h(h(z))$$

Even more amazing

What is the probability generating function for the distribution of grandsons?

- A man has two sons, with probability p_2, so PGF in that case is $p_2[h(z)]^2$.
- But let's sum over all possible numbers of sons.

$$p_0 + p_1 h(z) + p_2[h(z)]^2 + p_3[h(z)]^3 + \ldots$$

- Which is? (Hint: write a new argument for PGF)

$$h(h(z))$$

- Can show PGF for the n'th generation is

$$h(h(h ... n \text{ times} h(z))) = h_n(z)$$

In-class exercise: write out $h_2(z) = h(h(z))$ for

$$h(z) = p_0 + p_1 z + p_2 z^2.$$

Extinction

Extinction: some generalities

"Extinction is forever." So, the probability d_n of extinction *by* generation n can never decline over time. (Must it always rise?)

Extinction: some generalities

"Extinction is forever." So, the probability d_n of extinction *by* generation n can never decline over time. (Must it always rise?)

Is non-extinction "forever"? If $\lim_{n\to\infty} = d(\infty) < 1$, then this says there's a chance $1 - d(\infty)$ of eternal persistence. We'll try to figure out more about what this means.

Extinction: a recursive trick

If the probability of a female line going extinct in n generations is d_n, then this is equivalent to her daughter(s) line(s) going extinct in $n-1$ generations. With p_k chance of having k daughters, we have

$$d_n = p_0 + p_1 d_{n_1} + \text{What is next term in series?}$$

Recursive extinction, continued

What can we do with

$$d_n = h(d_{n-1})?$$

Well, remember that d_n is non-decreasing, and that it's maximum can be no greater than 1.0. When d_n reaches it's limit, say d, we won't need generational subscripts, d will be constant, and will obey

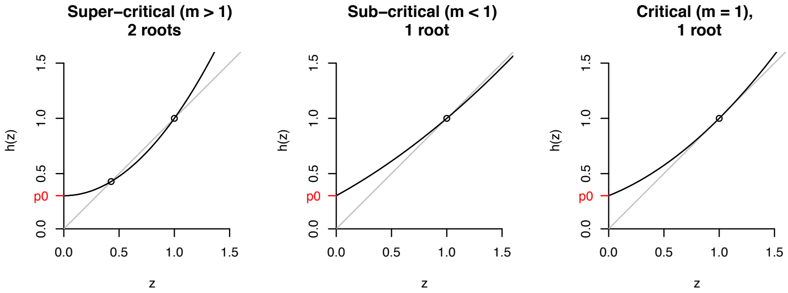
$$d = h(d)$$

Thus, an amazing result: the probability of ultimate extinction is when the argument equals the PGF of the argument.

Can $d = 1$, can $d < 1$

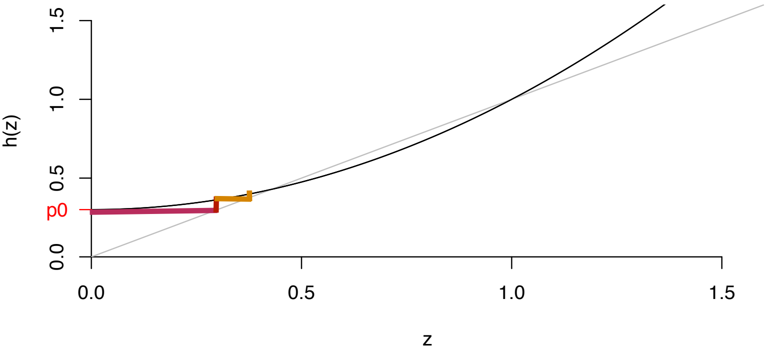
1. Try $d = 1$. What happens?
2. If we were to find a solution less than 1.0, how would we interpret that?

Three cases



We can prove by answering: What is $h'(1)$? What is $h(0)$? Is $h''(z) > 0$?

A cobweb diagram (here a “staircase”)



Where is $h(p_0)$, $h(h(p_0))$, $h(h(h(p_0)))$, ...?

So how do we actually get d?

Take the case where $p_0 = .3$, $p_1 = 0$, and $p_3 = .7$ (the one I just plotted).

So how do we actually get d?

Take the case where $p_0 = .3$, $p_1 = 0$, and $p_3 = .7$ (the one I just plotted).

- Can do some algebra

So how do we actually get d?

Take the case where $p_0 = .3$, $p_1 = 0$, and $p_3 = .7$ (the one I just plotted).

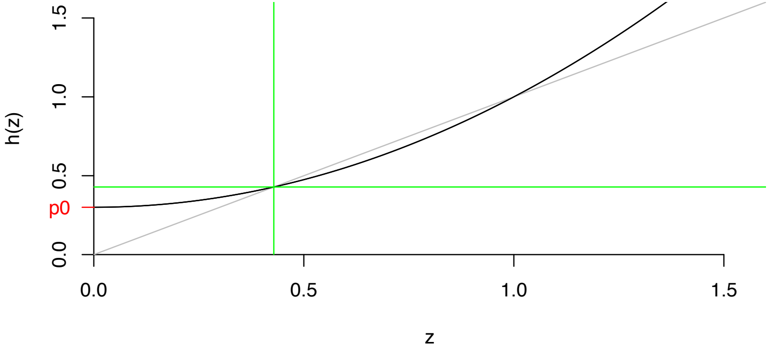
- Can do some algebra
- Or we can recursively iterate on the computer.

Numerical recursion

```r
pk = c(.3, .0, .7); names(pk) <- 0:2 ## our example
d <- pk["0"] # initial value
for (i in 1:20)
{
    d <- pk["0"]  + pk["1"]*d + pk["2"]*d^2
    if (i %in% c(1,2,19,20))
       print(paste(i, d))
}
## [1] "1 0.363"
## [1] "2 0.3922383"
## [1] "19 0.428565882081349"
## [1] "20 0.428568100698915"
```

Did we get the right value?

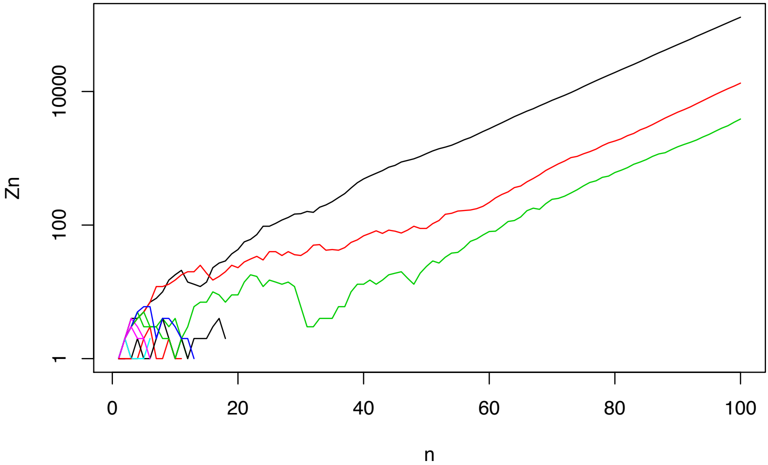
Did we get the right value?



Extinction and non-extinction revisited

- If $m > 1$, there exists d bigger than 0 and less than unity.
- This means there's some positive chance of extinction.
- But also some chance of never-extinction. (What form does never-extinction take?)

Return to our simulation



Relevance to Corona virus?

Good and bad set-ups for branching process

Bad

- When offspring of 1 depends on offspring of other (e.g., brothers inheriting a farm)
- When resource constraints slow growth rates (e.g., Malthus: fertility of next gen depends on fertility of last; SIR model in disease spread)
- Analysis. PGF is powerful but still we often have to deal with listing all possibilities.
- Big populations – law of large numbers means randomness doesn't matter.

Good

- Unrestricted growth (frontier, new disease, start of a reaction)
- A "null" model for understanding how apparent structure is just random outcomes. Families that die out didn't have to have low *NRR*. Just because most new viruses don't break out, doesn't mean they aren't potentially dangerous ($R_0 >> 1.0$).
- A model that corresponds our mental model of running a generative process forward. (cf. Fisher-Wright)

Next time

- Do the PS exercises.
- We'll have BP part 2, next week. (means and variances of Z_n, the one tractable offspring distribution, time to extinction, etc.)
- I hope to have a mini-project with Missippi last names.
- We may start Fisher-Wright. (Kennedy and I recommend Gillespie.)

Questions?