

Branching Processes, Part 2
Dem260 Math Demog
Spring 2020
Lecture 8

Joshua R. Goldstein

March 19, 2020

Agenda for today

1. Review
2. Means and variances
3. Closed form solutions with the geometric offspring distribution
4. Covid-19 tracking

Review

- ▶ What a branching process is and potential applications
- ▶ How probability generating functions can be used to analyze
- ▶ Sub-critical, critical, and super-critical extinction probs.

Means and variances

What happened in your simulation?

Means of offspring in generation n

- ▶ Is it “meaningful”? (If we have a lot of zeros)
- ▶ We'll show that unconditional mean

$$\mathbb{E}Z_n = m^n$$

- ▶ What if $m = 1$?

Mean size of surviving lines?

Total probability is sum of conditional probabilities, times the chance of each condition:

$$\mathbb{E}Z_n = \mathbb{E}(Z_n|Z_n > 0)P(Z_n > 0) + \mathbb{E}(Z_n|Z_n = 0)P(Z_n = 0)$$

What is mean size of surviving lines?

- ▶ Hint 1: $P(Z_n = 0) = d_n$
- ▶ Hint 2: $\mathbb{E}Z_n = m^n$

Let's check our result using simulation

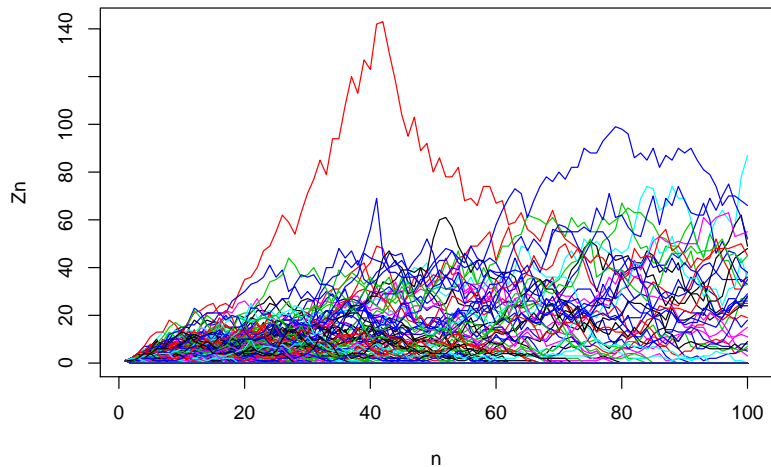
Our set-up (note: code is revised slightly)

```
branch <- function(n_max = 30, p_k = c(p0, p1, p2), Z1 = 1)
{
  ## note: this returns 0s when extinct
  k <- 0:(length(p_k)-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)
}
p0 = .3; p1 = .4; p2 = .3 ## what is m?
```


1000 trials, code

```
n_trials = 1000; n_gen = 100
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")
```

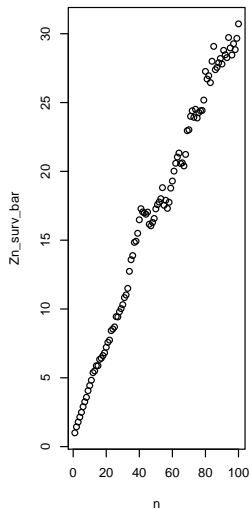
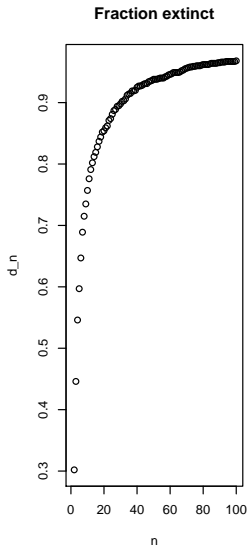
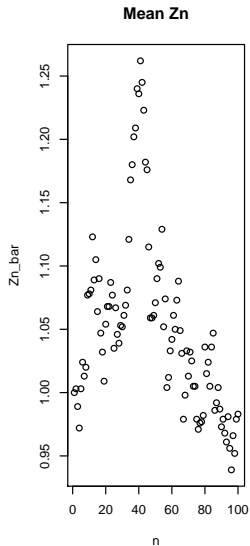
1000 trials, picture



1000 trials, means

```
Zn_bar = apply(Z.mat, 2, mean)
n <- 1:ncol(Z.mat)
proportion.zero <- function(x){prop.table(table(x == 0))["T"]}
d_n = apply(Z.mat, 2, proportion.zero)
Z.mat.na <- Z.mat; Z.mat.na[Z.mat == 0] <- NA
Zn_surv_bar = apply(Z.mat.na, 2, mean, na.rm = T)
par(mfrow = c(1,3))
plot(n, Zn_bar, main = "Mean Zn")
plot(n, d_n, main = "Fraction extinct")
plot(n, Zn_surv_bar)
## insert code here for Zn_surv_bar.hat and add a line
```

1000 trials, means, picture



Proving $\mathbb{E}Z_n = m^n$

Ingredients

- ▶ $h'(1) = m = \bar{Z}$
- ▶ $h'_n(1) = \bar{Z}_n$
- ▶ $h_{n+1}(z) = h(h_n(z))$

Derivation

Variance

Variance result. (Too much algebra to do here).

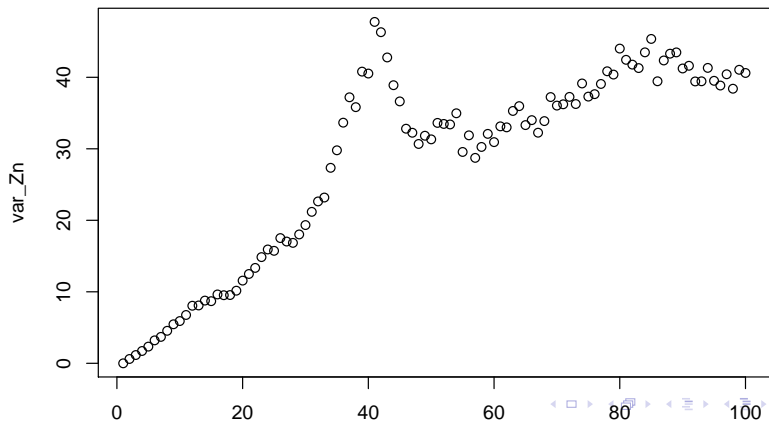
For $m = 1$,

$$\sigma_n^2 = n\sigma^2$$

- ▶ Also a result for $m \neq 1$
- ▶ What does increasing variance mean for critical case? (Does this make sense?)
- ▶ What happens to variance of lines that survive? Is it bigger or smaller than unconditional variance?

Variance in our simulation

```
var_Zn = apply(Z.mat, 2, var)
n <- 1:ncol(Z.mat)
plot(n, var_Zn)
```



Geometric offspring distribution

For $k = 1, 2, \dots$,

$$p_k = bc^{k-1}$$

For $k = 0$,

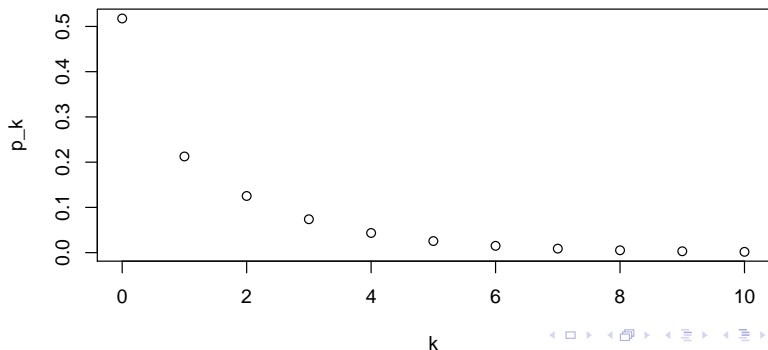
$$p_0 = 1 - p_1 - p_2 - \dots$$

Let's solve for p_0 , using the geometric series sum, for $a < 1$,

$$1 + a + a^2 + \dots = 1/(1 - a)$$

A picture, Lotka's parameters for 1920

```
b = 0.2126 ;    c = 0.5893  
kk = 1:10 ;    p_kk = b * c^(kk-1)  
p0 = b/(1-c)  
k = c(0, kk) ;    p_k = c(p0, p_kk)  
plot(k, p_k)
```



Realism?

See Table 10.3, p 386, Grinstead and Snell.

The Geometric Distribution's simple PGF

$$h(z) = p_0 + p_1z + p_2z^2 + \dots$$

With geometric p_k

$$h(z) = p_0 + \sum_{k=1}^{\infty} bc^{k-1}z^k.$$

Substituting for p_0 and rewriting

$$h(z) = (1 - b/1 - c) + bz \sum_{k=1}^{\infty} (cz)^{k-1}.$$

Substituting $j = k - 1$,

$$h(z) = (1 - b/1 - c) + bz \sum_{j=0}^{\infty} (cz)^j = (1 - b/1 - c) + \frac{bz}{1 - cz}$$

The PGF is now “tractable”

m and extinction

$$h(z) = (1 - b/1 - c) + \frac{bz}{1 - cz}$$

Please solve for m . (Hint: $h'(1)$). What is m with Lotka's b and c ?

We solve $z = h(z)$ with a bunch of algebra to get

$$d = \frac{1 - b - c}{c(1 - c)}$$

How does d depend on b and c ?

Big payoff: the full distribution of Z_n

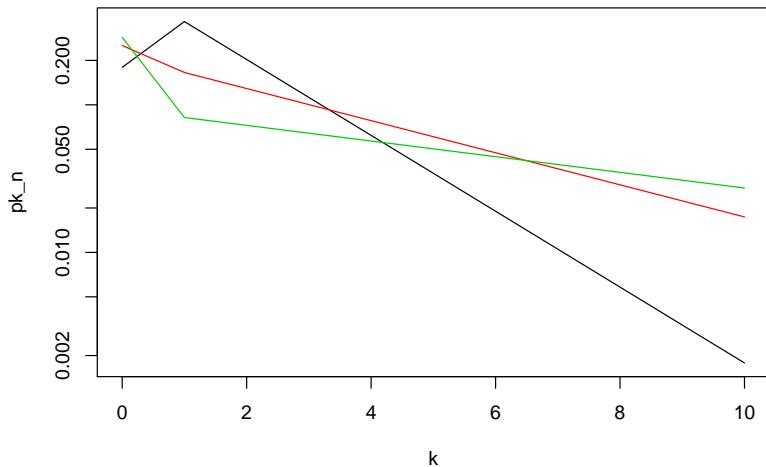
See Grinstead and Snell p. 385

A plot of Keyfitz's numbers for generations 1, 2, and 3. Is it exponential for $k > 0$?

```
## b = 0.2126 ;   c = 0.5893 ## lotka
b = 0.3666; c = .5533 ## Keyfitz (from GS)
m = b / (1-c)^2 ## [1] 1.260416
d = (1 - b - c) / (c * (1-c))      #[1] 0.8185088
par(mfrow = c(1,1))
for (i in 1:3)
{
  n = i
  p0_n = d * (m^n - 1) / (m^n - d)
  j = kk
  pj_n = m^n *
    ((1-d) / (m^n - d))^2 *
    ((m^n - 1) / (m^n - d))^(j-1)
  pk_n <- c(p0_n, pj_n)
  if (i == 1)
    plot(k, pk_n, type = "l", log = "")
  if (i > 1)
    lines(k, pk_n, col = i)
}
```

log scale

```
## [1] 1
```



Applications

We have exponential distribution with a few very large lines, and a lot of small lines.

- ▶ Distribution of neutral alleles
- ▶ Distribution of family lines (Y-chromosome, mtDNA, last names)

Our result

With geometric p_k , we get geometric Z_n , for all n .

Conjecture: geometric is to BP as gamma is to frailty?

Branching Processes and Covid-19

Discussion questions:

- ▶ What is the BP that they are studying? Is it contagion, social contacts, or ?
- ▶ What do they assume about the BP?
- ▶ Do they use any analytical results or just simulation? Why?
- ▶ Best feature of paper?
- ▶ Worst feature of paper?
- ▶ Inspire any other approaches?

Spring Break!