

Dynamics of Unstructured Populations
Formal Demography
Stanford Summer Short Course

James Holland Jones
Department of Anthropological Sciences
Stanford University

August 6, 2006

Outline

1. Growth of Unstructured Populations

- (a) Discrete Time
- (b) Continuous Time

2. Mixtures of Populations & Variance in r

3. Logistic Growth

- (a) Solutions
- (b) Stability
- (c) Discrete logistic: nonlinear dynamics and chaos

4. Interacting Populations

Annual Population Growth

Population compounded annually at rate r

We want a relationship for the current year's population size and next year's

Let P_t denote the size of the population at time t

Let P_0 denote the initial population size

How big is the population one year in the future?

$$P_1 = P_0(1 + r)$$

Annual Population Growth

We can continue this process

$$P_1 = P_0(1 + r)$$

$$P_2 = P_1(1 + r) = P_0(1 + r)(1 + r) = P_0(1 + r)^2$$

·

·

·

$$P_t = P_0(1 + r)^t$$

Population Compounded j Times Annually

New members added to the population j times per year

$$P_{t+1} = P_t \left(1 + \frac{r}{j}\right)^j$$

for $P_t = 100$, $r = 0.02$, and $j = 1$:

$$P_{t+1} = 100(1 + 0.02)^1 = 102$$

Now, say that the population is compounded twice annually

$$P_{t+1} = P_t \left(1 + \frac{r}{2}\right)^2$$

for $P_t = 100$, $r = 0.02$, and $j = 2$:

$$P_{t+1} = 100\left(1 + \frac{0.02}{2}\right)^2 = 102.01$$

```
> compound <- expression(pt*(1+(r/j))^j)
> pt <- 100
> r <- 0.02
> j <- 1
> eval(compound)
[1] 102
> j <- 2
> eval(compound)
[1] 102.01
> j <- 3
> eval(compound)
[1] 102.0134
```

Now for a fun fact of calculus

$$\lim_{j \rightarrow \infty} \left(1 + \frac{1}{j}\right)^j = e$$

This is, in fact, one definition of the number e , the base of the natural logarithm

Replace the 1 with r

$$\lim_{j \rightarrow \infty} \left(1 + \frac{r}{j}\right)^j = e^r$$

Continuously compounded populations grow in one year by

$$P_t = P_{t-1}e^r \tag{1}$$

Just checking . . .

```
> compound <- expression(pt*(1+(r/j))^j)
> pt <- 1
> r <- 1
> j <- 100000
> eval(compound)
[1] 2.718268
> exp(1)
[1] 2.718282
```


Continuous Growth Equation

Assume constant growth rate for t years:

$$P_t = P_0(e^r)^t = P_0e^{rt}$$

Note that this means that the ratio between the population size separated by t years is simply e^{rt}

One offspring born now constitutes $1/P(0)$ of the population

One offspring born t years from now constitutes $1/P(0)e^{rt}$ of the population

In terms of proportional representation an offspring born t years in the future must be discounted by a factor e^{-rt}

Go to equation 4

Per Capita Birth and Death Rates

Some Definitions

N_t The total number of individuals in the population at time t

B Number of births per animal per year

D Probability that an animal dies in a year ($1 - D$ is the probability that an animal survives the interval)

$$N_{t+1} = BN_t + (1 - D)N_t$$

Rearrange

$$N_{t+1} = RN_t$$

$$\text{where } R = (1 + B - D)$$

Geometric Growth

Project the population forward t years

$$N_1 = RN_0$$

$$N_2 = RN_1 = R(RN_0) = R^2 N_0$$

It's not difficult to see that

$$N_t = R^t N_0$$

Take logarithms of both sides of this relationship

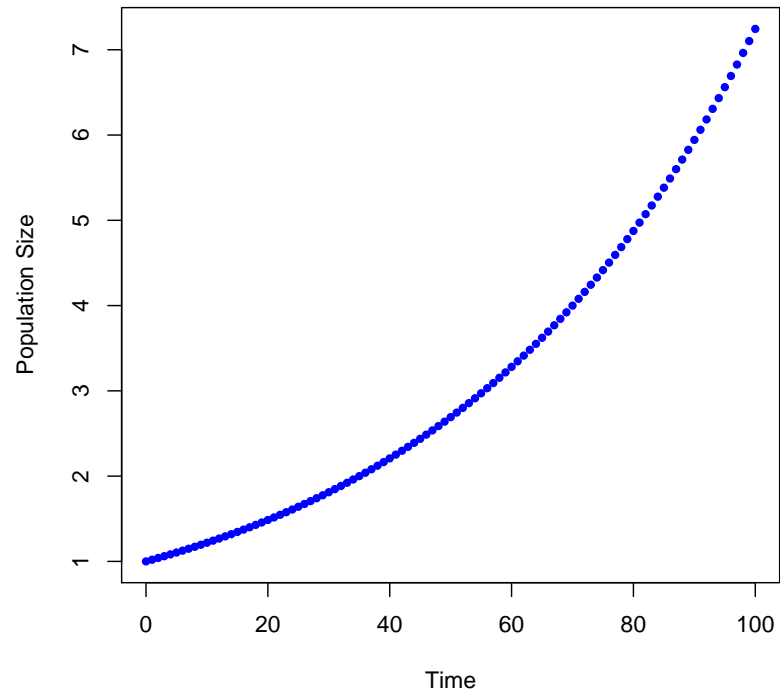
$$\log(N_t) = \log(R)t + \log(N_0)$$

Why should we care about this?

One very good reason is that it provides a simple graphical diagnostic

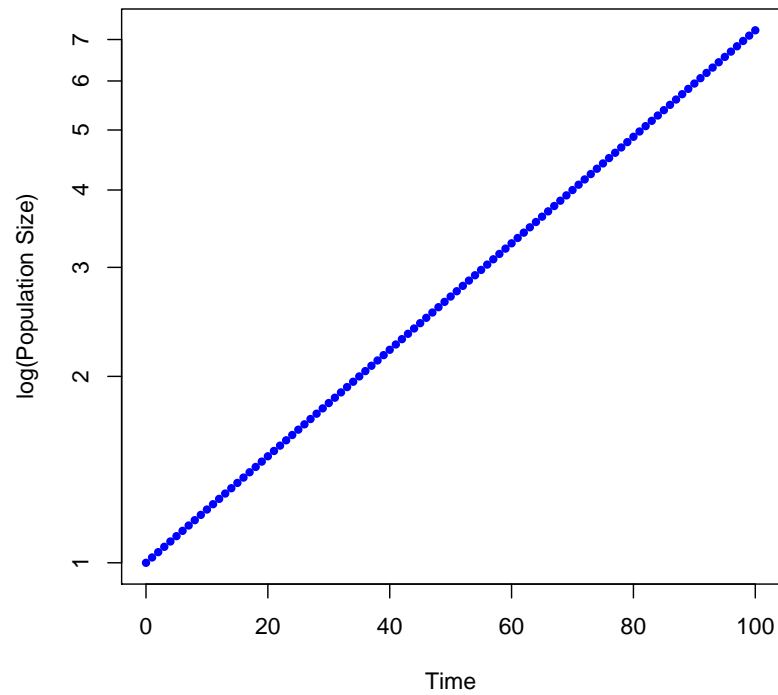
```
R.calc <- expression(1 + (B - D))
B <- 0.05
D <- 0.03
no <- 1
t <- seq(0,100,1)
R <- eval(R.calc)
geo.grow <- expression(R^t*no)
pop.dyn <- eval(geo.grow)
plot(t,pop.dyn,pch=20,col="blue",xlab="Time",ylab="Population Size")
```

Geometric Growth: Linear Axes



Another Plot: Semilogarithmic Axes

```
plot(t,pop.dyn,log="y",pch=20,col="blue",xlab="Time",ylab="Population Size")
```



Population Growth in Continuous Time

Some Updated Definitions

N Population Size

b Average per capita birth rate per unit time

d Death rate per unit time

Note the shift to rates that accompanies a continuous model

$$\frac{dN}{dt} = bN - dN$$

Combine the birth and death rates into a summary parameter, r , customarily called the “intrinsic rate of increase” or, more stodgily, “the Malthusian parameter.”

$$r = b - d$$

The equation for population growth in continuous time

$$\frac{dN}{dt} = rN \tag{2}$$

Solving the Continuous-Time Growth Equation

1. Rearrange Equation 2 so like items are on the same side of the equals sign:

$$\frac{dN}{N} = r dt$$

2. Integrate both sides from $t = 0$ to $t = T$

$$\int_{t=0}^{t=T} \frac{dN}{N} = \int_{t=0}^{t=T} r dt$$

3. Compute the integrals

$$\log(N(T)) - \log(N(0)) = rT$$

4. Remember that $\int \frac{1}{N} dN = \log(N)$
5. Take the exponential of both sides

$$e^{\log(N(T))} e^{-\log(N(0))} = e^{rT}$$

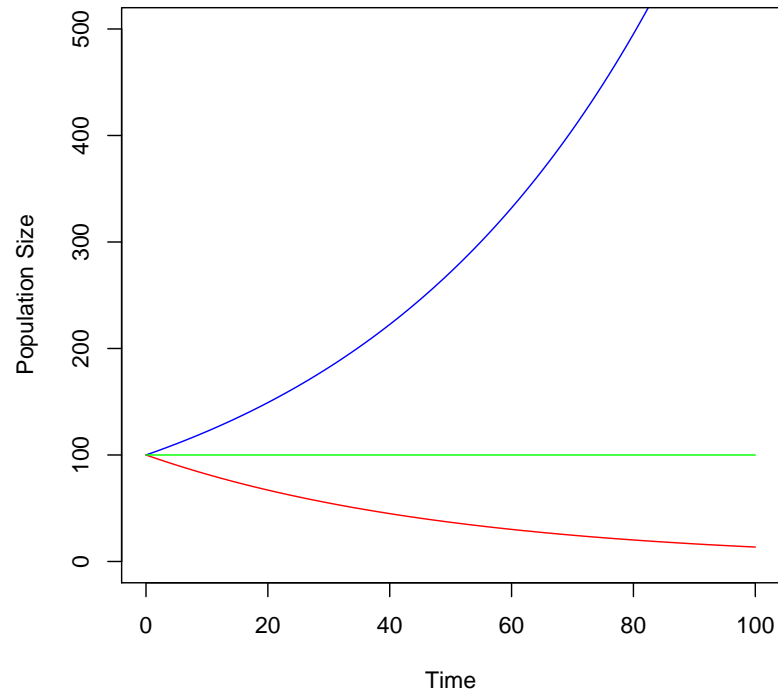
6. Note that $e^{\log a} = a$ and $e^{-\log a} = \frac{1}{a}$.

7. Solve for $N(T)$

$$\frac{N(T)}{N(0)} = e^{rT}$$

$$N(T) = N(0)e^{rT} \tag{3}$$

What Happens for Various Values of r ?



This is very important.

1. Populations can grow exponentially, but “common sense” tells us they can only do so for relatively short periods of time.
2. An equilibrium only occurs when $r = 0$.
3. This result has many applications in applied mathematics.

The Relationship between r and R

If r is small,

$$R \approx 1 + r$$

Why is that?

This derives from a *Taylor Series Approximation*

We use a Taylor polynomial to approximate a function $f(x)$ around some point a

For a continuous function with $n + 1$ derivatives, this polynomial is:

$$P_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \dots + \frac{f^n(a)}{n!}(x - a)^n$$

Most of the time, we only worry about the first couple terms

For the problem of the relationship between r and R we know that

$$R = e^r$$

But the Taylor series about 0 of e^r is:

$$e^r \approx 1 + r + \frac{r^2}{2} + \frac{r^3}{6} \dots$$

If r is small, then the square terms and beyond will be negligible relative to r itself

More on Taylor Polynomials

Taylor series approximations are a very important technique in formal demography

You will see them again and again...

```
> exp(r)
[1] 1.020201
> 1
[1] 1
> 1+r
[1] 1.02
> 1 + r + r^2/2
[1] 1.0202
> 1 + r + r^2/2 + r^3/6
[1] 1.020201
> 1 + r + r^2/2 + r^3/6 + r^4/24
[1] 1.020201
```

Doubling Time of a Population

How long until a population growing at rate r will double?

The relationship for a continuously growing population is

$$N(t) = N(0)e^{rt}$$

We want to know about the doubling from the current population, so

$$N(t) = 2N(0)$$

Substitute and solve for t

$$2 = e^{rt}$$

$$t = \frac{\log(2)}{r} \approx \frac{0.693}{r}$$

How long will it take for a population growing at 2% annually to double?

$$t = \frac{0.693}{0.02} = 34.7 \text{ years}$$

Growth Rate of a Mixture of Populations

A population size Q , with growth rate r , increases in numbers by Qe^{rt} over the interval t

The intrinsic rate of increase, r , is the *per capita rate of increase of the population*

Thus, by definition, we can write it as:

$$r = \frac{1}{N(t)} \cdot \frac{dN(t)}{dt} \quad (4)$$

Aside: it is worth noting the following:

$$\frac{1}{N(t)} \cdot \frac{dN(t)}{dt} \equiv \frac{d \log N(t)}{dt}$$

For a mixture of n subpopulations, each with its own rate of increase r_i , the increase in interval t will simply be

$$N(t) = \sum_i^n Q_i e^{r_i t}$$

The derivative of $N(t)$ is

$$\frac{dN(t)}{dt} = \sum_i^n Q_i r_i e^{r_i t}$$

Substituting these, the overall rate of increase, \tilde{r} is thus

$$\tilde{r} = \frac{1}{N(t)} \cdot \frac{dN(t)}{dt} = \frac{\sum_i^n Q_i r_i e^{r_i t}}{\sum_i^n Q_i e^{r_i t}} \quad (5)$$

This is just a weighted mean of the subpopulation growth rates, with weights the initial population size of the subpopulations

Change in the Mean Rate of Change

Differentiate \tilde{r} with respect to t

$$\frac{d\tilde{r}}{dt} = \frac{\sum_i^n Q_i r_i^2 e^{r_i t}}{\sum_i^n Q_i e^{r_i t}} - \left(\frac{\sum_i^n Q_i r_i e^{r_i t}}{\sum_i^n Q_i e^{r_i t}} \right)^2.$$

This messy looking equation, has the form of

$$\mathbb{E}(X^2) - \mathbb{E}(X)^2$$

which is the definition of variance of x

Thus,

$$\frac{d\tilde{r}}{dt} = \sigma^2(t)$$

What does this mean?

1. \tilde{r} increases. Can it increase without bound?
2. Since \tilde{r} is the average of an ensemble of constituent r_i , it can never be greater than the largest of its constituents
3. This represents one derivation of Fisher's Fundamental Theorem of Natural Selection
4. It also means that the sum of a mixture of population projections with different growth rates will grow *faster than the the population projected by the mean growth rate*

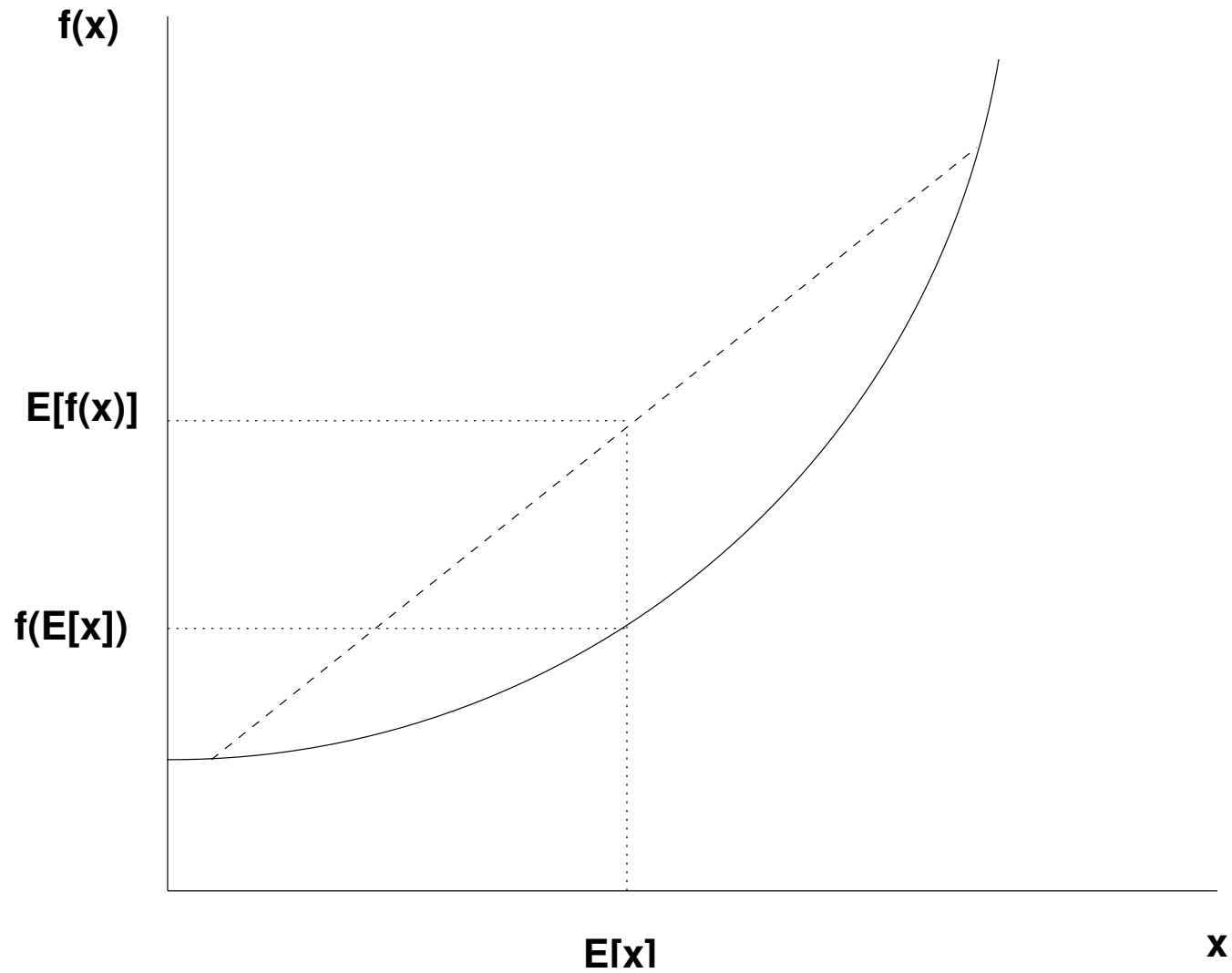
This last fun fact is a demonstration of *Jensen's Inequality*, which states for a convex function:

$$\mathbb{E}[f(x)] \geq f(\mathbb{E}[x])$$

(the inequality is reversed for concave function)

Go to equation [6](#)

Graphical Interpretation of Jensen's Inequality



Stalled Demographic Transition

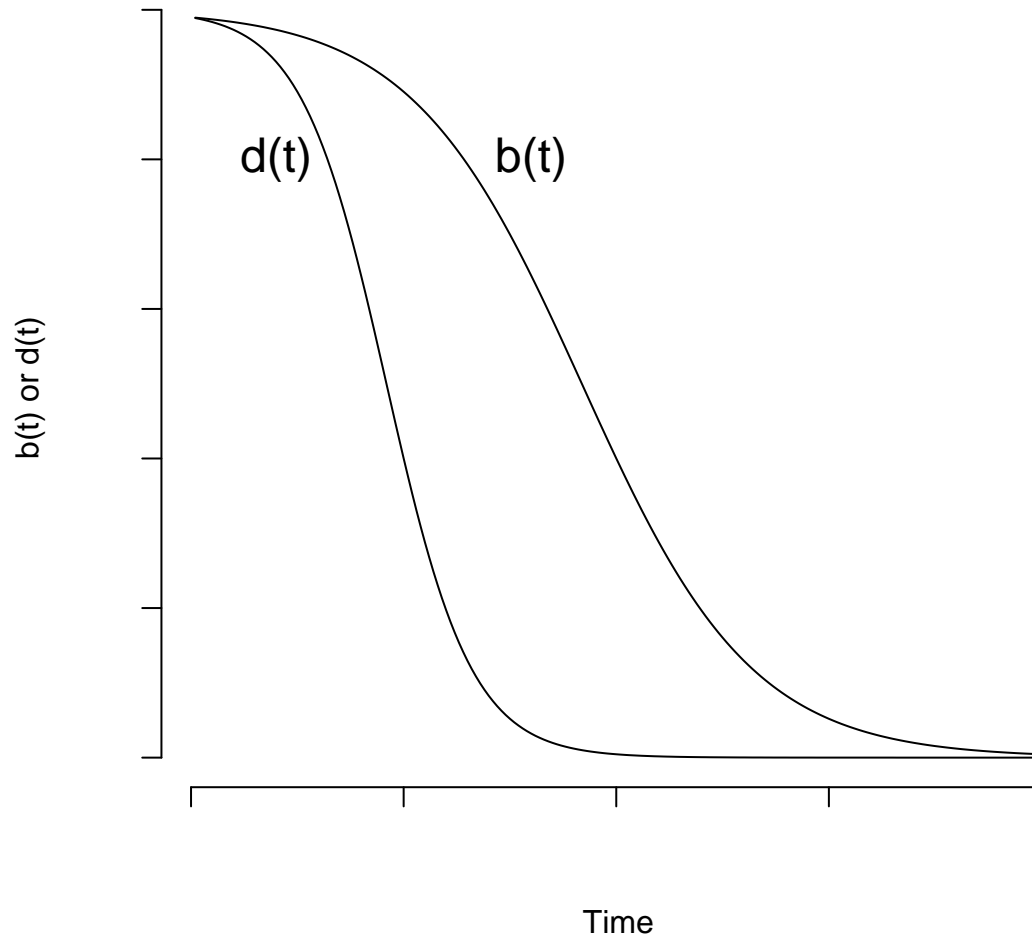
Notestein, the father of demographic transition theory, famously wrote in 1945 of a phenomenon in which “the stage of transitional growth...in which the decline of both fertility and mortality is well established but in which the decline of mortality precedes that of fertility and produces rapid growth.”

If mortality decline precedes fertility growth, how much bigger will the population be when fertility finally drops to replacement?

Define two functions of time $b(t)$ and $d(t)$

These describe the change in birth and death rates, respectively, with time

More Stalling...



If $b(t)$ and $d(t)$ start and end at the same points (i.e., equal, implying a stationary population), then the ratio of the population size at the end of the transition period T and the beginning time will be

$$\int_0^T [b(t) - d(t)] dt = \int_0^T r(t) dt = A$$

where A is just the area (hence “A”) between the two curves

The ratio of the population size at the end of T to that at the beginning of the period (i.e., $t = 0$) is then simply

$$\exp \left[\int_0^T r(t) dt \right] = e^A$$

Now, if $d(t)$ and $b(t)$ also have the same shape, we can simplify even further

Say that the lag between $d(t)$ and $b(t)$ is L years and define K as the absolute drop in $d(t)$ and $b(t)$

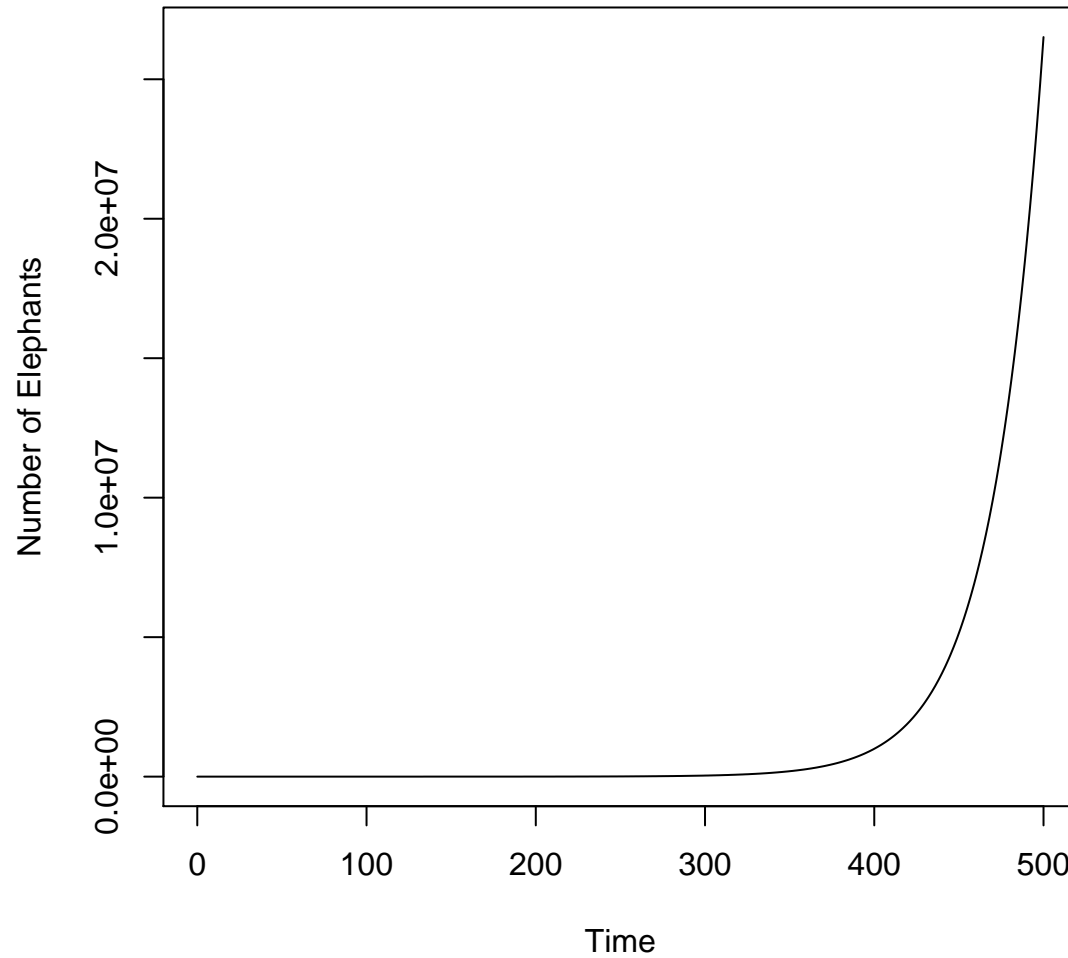
In this special case $A = KL$

Darwin's Elephants

There is no exception to the rule that every organic being naturally increases at so high a rate, that if not destroyed, the earth would soon be covered by the progeny of a single pair. Even slow-breeding man has doubled in twenty-five years, and at this rate, in a few thousand years, there would literally not be standing room for his progeny. Linnaeus has calculated that if an annual plant produced only two seeds and there is no plant so unproductive as this and their seedlings next year produced two, and so on, then in twenty years there would be a million plants. The elephant is reckoned to be the slowest breeder of all known animals, and I have taken some pains to estimate its probable minimum rate of natural increase: it will be under the mark to assume that it breeds when thirty years old, and goes on breeding till ninety years old, bringing forth three pairs of young in this interval; if this be so, at the end of the fifth century there would be alive fifteen million elephants, descended from the first pair.

Darwin, The Origin of Species

It's Usually a Good Idea to Respect Darwin's Observations



Darwin's ciphering amounts to 3.28% annual growth

Feedback

We want a model such that as the population size gets large, the growth rate decreases

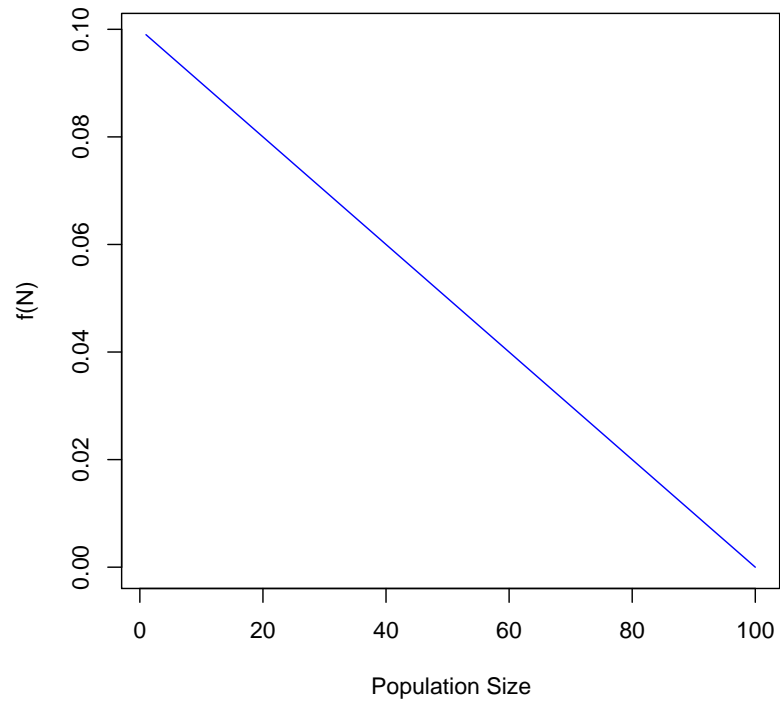
$$\frac{dN}{dt} = N f(N) \quad (6)$$

The simplest form is linear

$$f(N) = r(1 - N/K)$$

where K is known as the carrying capacity of the population

```
> logistic.fn <- expression(r*(1-X/K))
> X <- 1:100
> plot(X,eval(logistic.fn),type="l",col="blue",xlab="Population Size",ylab="f(N)")
```



which is clearly an equation for a straight line

Another Way to Look at the Logistic Function

For a given population size, how many recruits are there?

To visualize this, we plot the *Recruitment Curve*

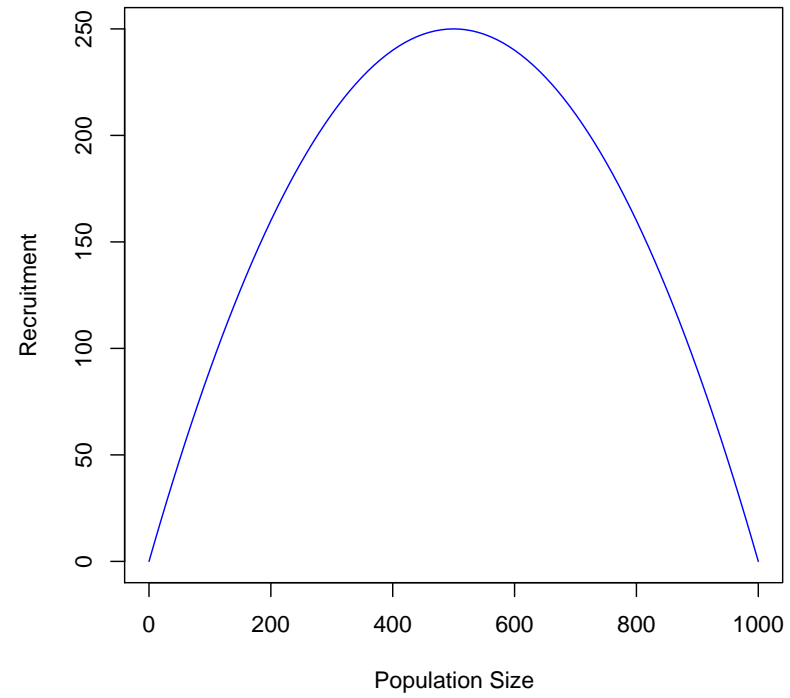
This is just the recruitment function ($r(1 - N/K)$) multiplied by the population size N (this is the number of recruits), plotted against N

```
> Dn <- expression(r * (1 - N/K) * N)
> r <- 1
> K <- 1000
> N <- seq(0,1000,by=10)
> plot(X,eval(Dn),type="l",col="blue",xlab="Population Size",
+ ylab="Recruitment")
```

If you were a hunter/forester, etc., at what population size would you want to harvest your population?

The *Maximum Sustainable Yield* is the peak of the recruitment curve

MSY



Go to equation 8

Solving the Logistic Equation

The logistic model is one that can be solved analytically

$$\frac{dN}{dt} = rN(1 - N/K)$$

Separate variables and integrate

$$\int_{N(0)}^{N(T)} \frac{dN}{N(1 - N/K)} = \int_0^T r dt$$

To do the integration on the lefthand side, we need to do integration by partial fractions

$$\frac{1}{N(1 - N/K)} = \frac{1}{N} + \frac{1/K}{1 - N/K}$$

$$\int_{N(0)}^{N(T)} \frac{1}{N} + \frac{1/K}{1 - N/K} dN = [\log(N) - \log(1 - N/K)]_{N(0)}^{N(T)}$$

$$= \log(N(T)) - \log(1 - N(T)/K) - \log(N(0)) + \log(1 - N(0)/K)$$

The solution of the righthand side is

$$\int_0^T r dt = rT$$

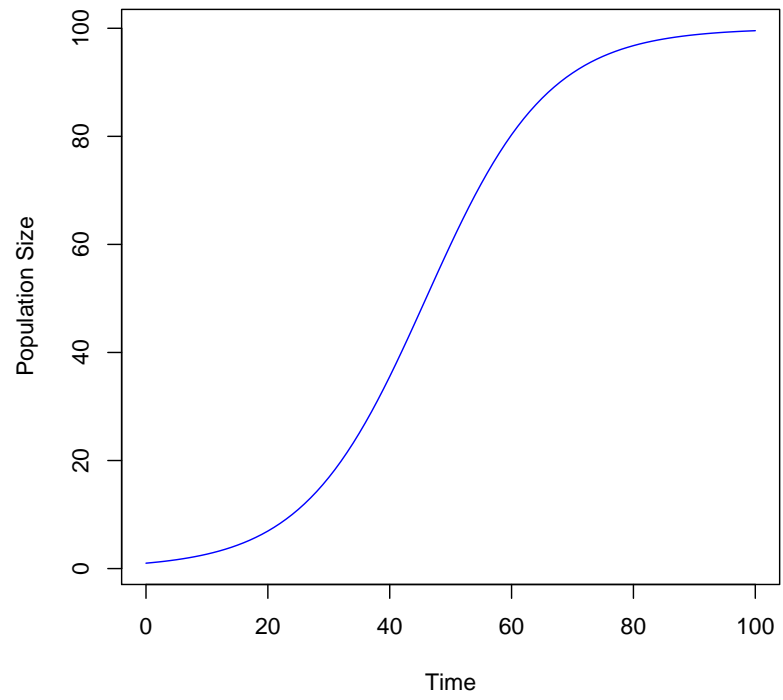
Put these together, and take exponentials of both sides

$$\frac{N(T)(1 - N(0)/K)}{(1 - N(T)/K)N(0)} = e^{rt}$$

Now, we solve for $N(T)$

$$N(T) = \frac{N(0)e^{rT}}{1 + N(0)(e^{rT} - 1)/K}$$

```
> logistic.int <- expression(n0*exp(r*t)/((1+n0*(exp(r*t)-1)/K)))
> n0 <- 1
> r <- 0.1
> K <- 100
> t <- 0:100
> plot(t,eval(logistic.int),type="l",col="blue",xlab="Time",ylab="Population Size")
```

Equilibria and Equilibrium Analysis

Solve for equilibria

$$\frac{dN}{dt} = rN(1 - N/K) = 0$$

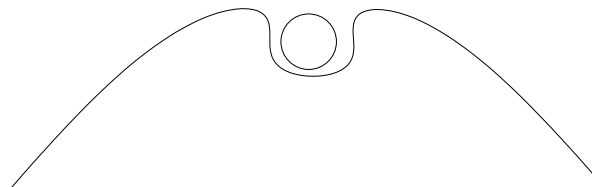
There are two equilibria:

- $N = 0$
- $N = K$

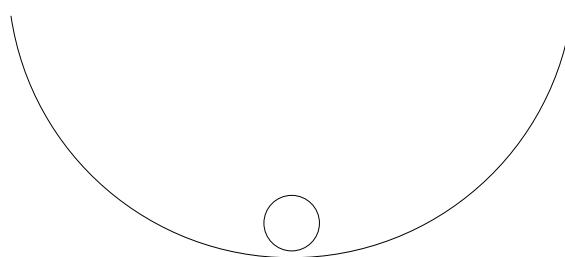
Types of Equilibria

- stable
- unstable
- neutral

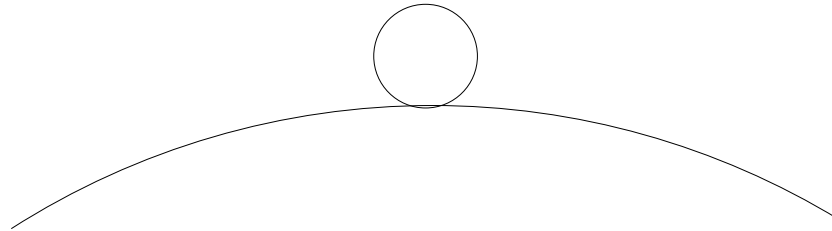
Locally Stable



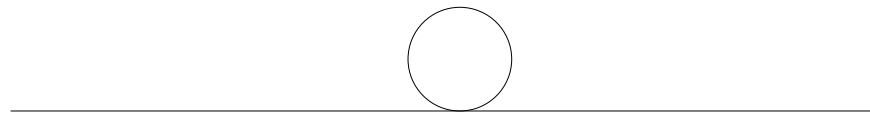
Globally Stable



Unstable



Neutrally Stable



Stability of Continuous-Time Models

The idea: Tweak a population that is at equilibrium. What happens?

- Continue to move in the direction of the tweak?
- Move back to the equilibrium value?

Stability of Continuous-Time Models II: A Recipe

Write the production function in generic form

$$\frac{dN}{dt} = F(N) \quad (7)$$

Assuming the logistic model $F(N)$ is

$$F(N) = rN(1 - N/K)$$

Determine equilibria: Solve for N in

$$F(N) = 0$$

Again, for the logistic model

$$\hat{N} = 0, \quad \hat{N} = K$$

Define a deviation from an equilibrium point \hat{N}

$$n = N - \hat{N}$$

Rearrange, putting N on the left-hand side $N = \hat{N} + n$, and substitute back into the generic equation.

$$\frac{d(\hat{N} + n)}{dt} = F(\hat{N} + n)$$

Now, \hat{N} is a fixed number (i.e., it's the equilibrium), so it won't change. All the change in this differential will therefore come from n which is free to vary (its change is, in fact, what we care about here)

$$\frac{dn}{dt} = F(\hat{N} + n)$$

We don't (necessarily) know what $F(\hat{N} + n)$ will be. For many interesting models, there is no closed-form solution to this differential equation. Our strategy is therefore to approximate it with a **Taylor Series**

By using a Taylor series approximation, we assume that n is small and we content ourselves to investigate the behavior of our model near the equilibrium

$$F(\hat{N} + n) \approx F(\hat{N}) + F'(\hat{N})n$$

$F'(\hat{N})$ is the derivative of F with respect to N evaluated at \hat{N}

The equation for the dynamics of our perturbation becomes

$$\frac{dn}{dt} = F(\hat{N}) + F'(\hat{N})n$$

But, we are evaluating at an equilibrium where, by definition, $F(\hat{N}) = 0$, so we are left with

$$\frac{dn}{dt} = F'(\hat{N})n$$

For grins, rename as follows

$$F'(\hat{N}) = \lambda$$

This gives us our final step

$$\frac{dn}{dt} = \lambda n$$

Which we recognize as the exponential growth model, the solution of which is

$$n(t) = n(0)e^{\lambda t}$$

- ▷ $\lambda > 0$, the equilibrium is unstable
- ▷ $\lambda < 0$, the equilibrium is stable

This process is known as *linearizing around the equilibrium* or *local linearization*

Stability Analysis of the Logistic Model

The generic model for the production function

$$\frac{dN}{dt} = F(N)$$

The logistic production function

$$F(N) = rN(1 - N/K) \quad (8)$$

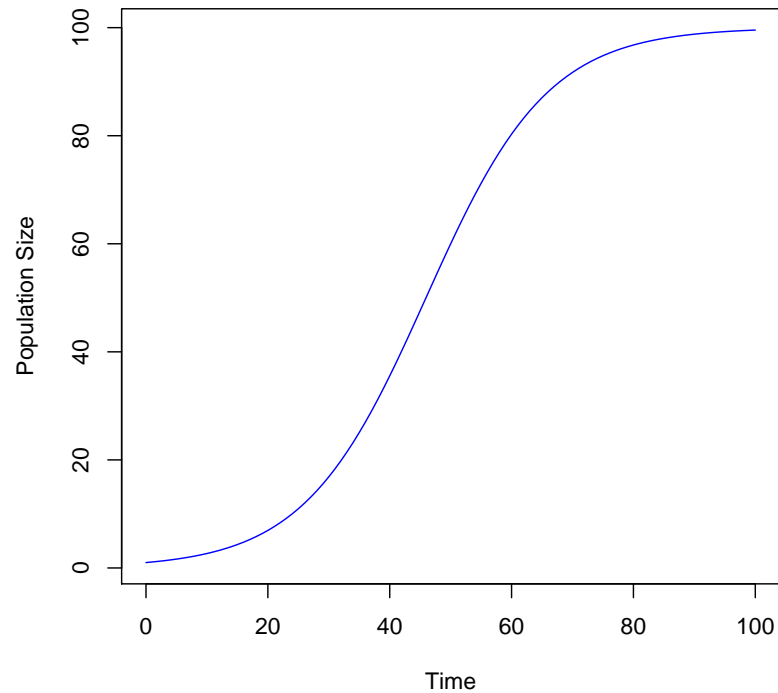
The logistic model has two equilibria: $N = 0$ and $N = K$

Calculate $\lambda = F'(\hat{N})$

For $\hat{N} = 0$, $\lambda = r$ and

For $\hat{N} = K$, $\lambda = -r$

- For small deviations near the equilibrium of $N = 0$, the population will increase exponentially at rate r
- For small deviations near the equilibrium of $N = K$, the population will decay back to the equilibrium exponentially at rate $-r$



Deriving $F'(N)$ for the Logistic

Where do the values $F'(N) = r$ and $F'(N) = -r$ for the two equilibria of the logistic model come from?

Use the Product Rule for Differentiation

Define

$$h(x) = f(x)g(x)$$

The Product Rule Specifies:

$$h'(x) = f'(x)g(x) + f(x)g'(x)$$

For the logistic model

$$f(x) = rN$$

$$g(x) = (1 - N/K)$$

$$f'(x) = r$$

$$g'(x) = -\frac{1}{K}$$

$$h'(x) = r(1 - N/K) - \frac{rN}{K}$$

Substitute back in the values for the equilibria ($N = 0$ and $N = K$)

$$N \rightarrow 0, \quad h'(x) = r$$

$$N \rightarrow K, \quad h'(x) = -r$$

Fitting the Logistic Model to Data

Is the Logistic model a good one for human populations?

Pearl et al. (1940) thought so

The Data: Total Population Size of the United States, as estimated in the decennial census

The Method: Least-squares minimization

The Tool in R: `optim()`, R's tool for minimizing a function

```
> logistic.int <- expression(n0 * exp(p[1] * t)/(1 + n0 * (exp(p[1] * t) - 1)/p[2]))
> fit.logistic <- function(p,y){
  n0 <- y[1]
  t <- seq(0,140,10)
  sumsq <- sum((y - eval(logistic.int))^2)
}

> year <- seq(1790,1990,10) # decennial census
> r.guess <- (log(usa[15])-log(usa[1]))/140
> k.guess <- usa[15] #1930 US population
```

```
> par <- c(r.guess,k.guess)
> usa1930 <- usa[1:15] # Just want the data up to when Pearl had them
> usa1930.fit <- optim(par,fit.logistic,y=usa1930)
> usa1930.fit
$par
[1] 0.03126604 198.55566623

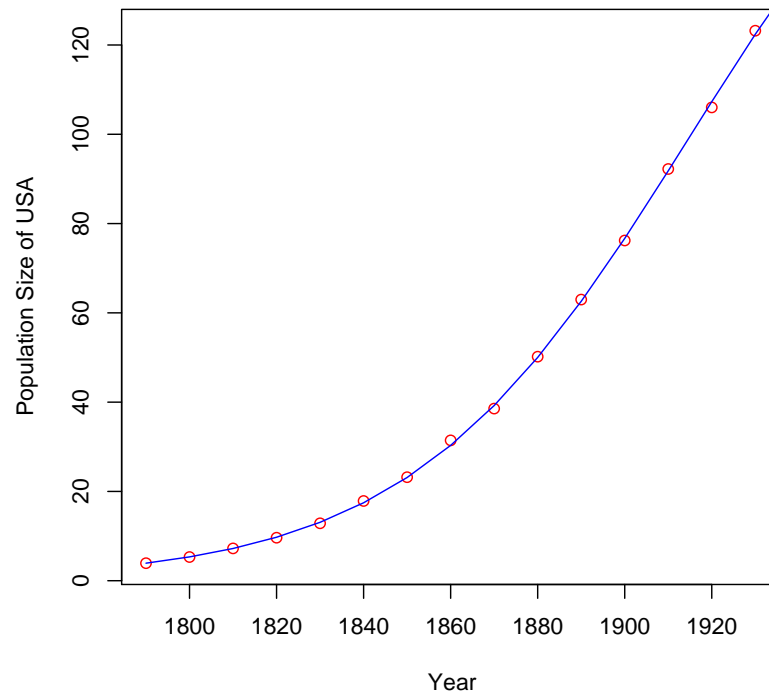
$value
[1] 4.830206

$counts
function gradient
      115      NA

$convergence
[1] 0

$message
NULL

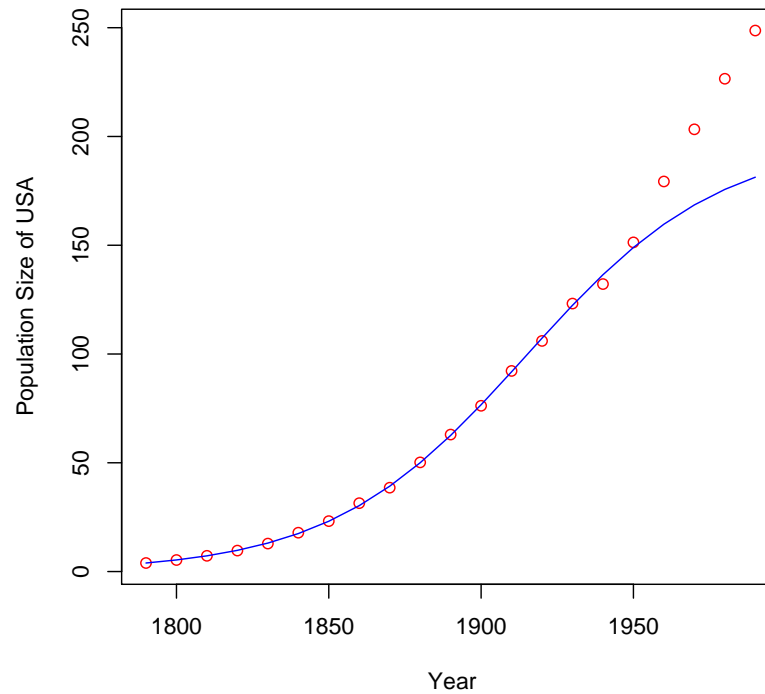
> p <- usa1930.fit$par
> plot(year[1:15],usa1930,col="red",type="p",xlab="Year",ylab="Population Size of USA")
> t <- year - 1790 # convert calendar year to 0:200
> lines(year,eval(logistic.int),col="blue")
```

Perfect!

Doh! (Famously)

Not so well-behaved after 1930 . . .



Discrete-Time Logistic Model

Take the logistic model in continuous time:

$$\frac{dN}{dt} = r_0 N (1 - N/k)$$

Now, we want a discrete-time equivalent of this. One possibility is simply to write down the continuous-time logistic model as a discrete-time model, assuming time increments, $\Delta t = 1$.

$$\frac{\Delta N}{\Delta t} \approx \frac{dN}{dt} = r_0 N (1 - N/K)$$

Now $\Delta N = N_{t+1} - N_t$. We can do a little algebra:

$$N_{t+1} - N_t = r_0 N_t (1 - N_t/K)$$

Rearrange a bit, solving for N_{t+1}

$$N_{t+1} = N_t r_0 + N_t^2 r_0 / K + N_t \quad (9)$$

$$= N_t(r_0 + 1) - N_t^2 r_0 / K \quad (10)$$

We can write $a = 1 + r_0$ and $b = r_0 / K$, giving us the following form for the discrete-time logistic model:

$$N_{t+1} = aN_t - bN_t^2 \quad (11)$$

What is wrong with this procedure?

Write down the inequality

$$0 > aN_t - bN_t^2$$

and solve

$$N_t > \frac{K(1 + r_0)}{r_0}$$

In other words, if the population size at time t exceeds $K(1 + r_0)/r_0$, the population size at time $t + 1$ is negative! That's no good.

An alternative discretization of the logistic model suggested by Turchin (2003) is as follows. Take the continuous-time logistic model, write $r(t) = r_0(1 - N/K)$, and solve for a one year interval, assuming that $r(t)$ remains constant over that interval.

$$N(t + 1) = N(t) \exp[r(t)] \quad (12)$$

Now substitute the expression for $r(t)$ and re-write using subscripts to emphasize that the model is now in discrete time.

$$N_{t+1} = N_t \exp[r_0(1 - N_t/K)] \quad (13)$$

Notice that the feedback term $r_0(1 - N/K)$ is now safely inside the exponential, meaning that as N_t exceeds K , the population will be multiplied by a term that will lie between zero and one.

This model is known as the Ricker model (Ricker 1954)

Note that similar problems beset discretizing the *SIR* epidemic model.

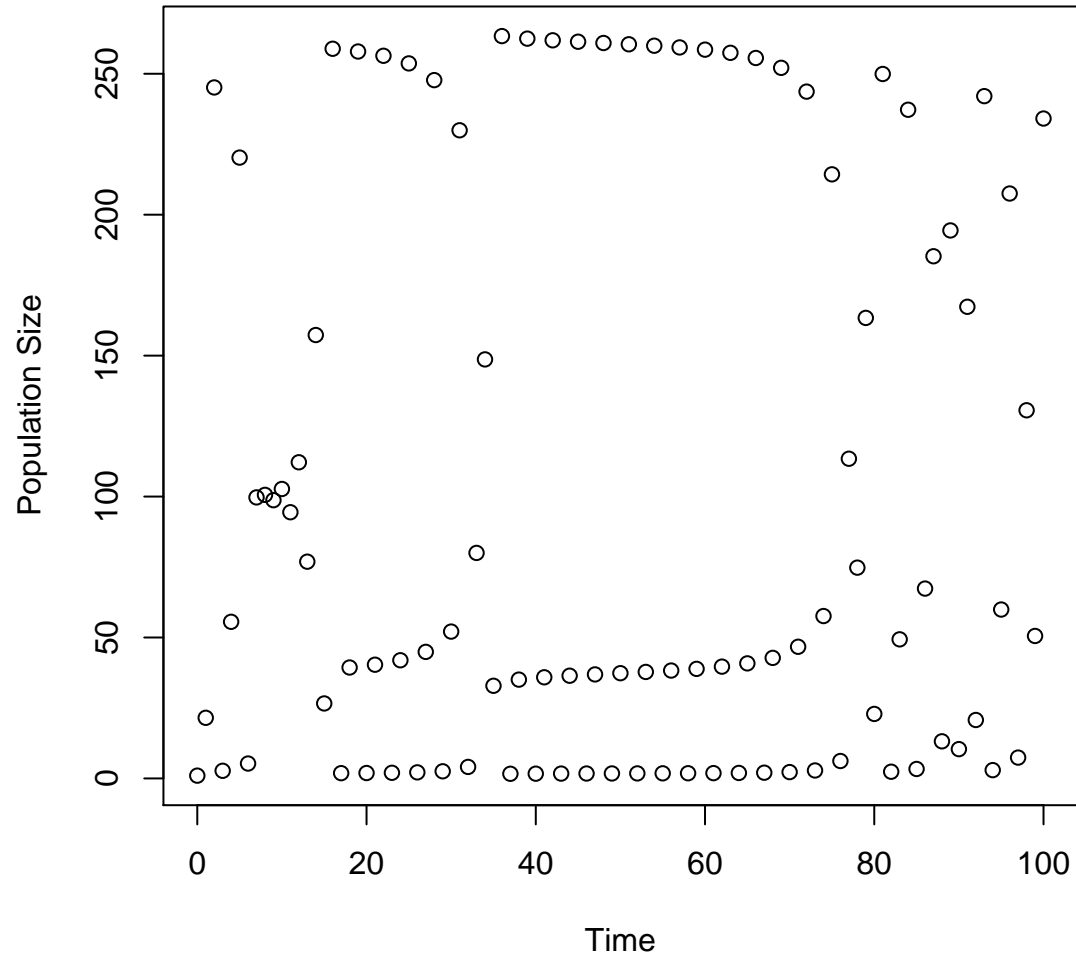
Crazy Dynamics

The Ricker model is subject to some extremely funky dynamics, including chaos

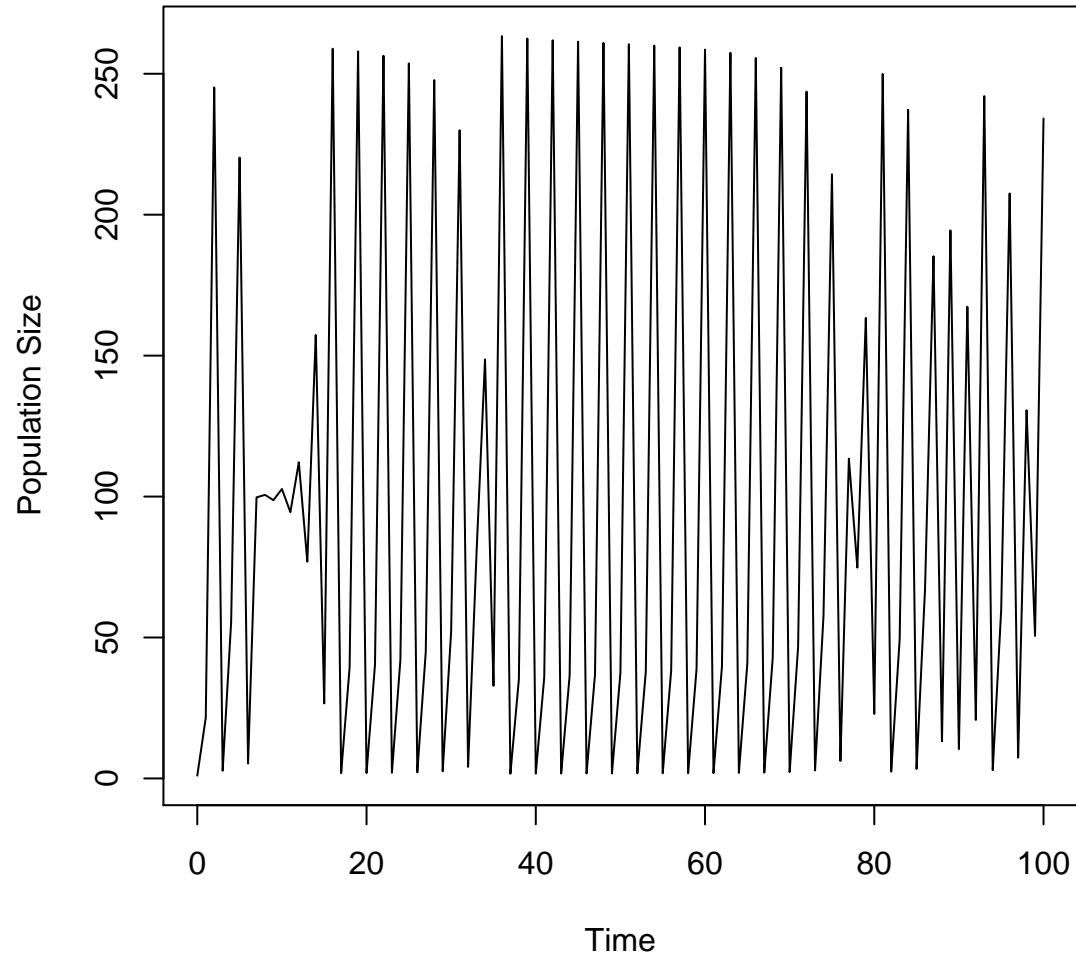
This is a common feature of nonlinear difference equations, particularly those with overcompensatory recruitment function (like Ricker)

```
> NN <- rep(0,101)
> K <- 100
> r <- 3.1
> for(i in 2:101) NN[i] <- NN[i-1]*exp(r*(1-(NN[i-1]/K)))
> plot(0:100,NN,xlab="Time",ylab="Population Size")
> plot(0:100,NN,type="l",xlab="Time",ylab="Population Size")
```

Yikes!



Maybe A Line Plot Will Help



Or not...