For comparison, the true MLE can be calculated for the linkage example. As seen last time

$$l(\lambda) = \frac{125}{2+\lambda} - \frac{38}{1-\lambda} + \frac{34}{\lambda} = 0$$

which is equivalent to solving

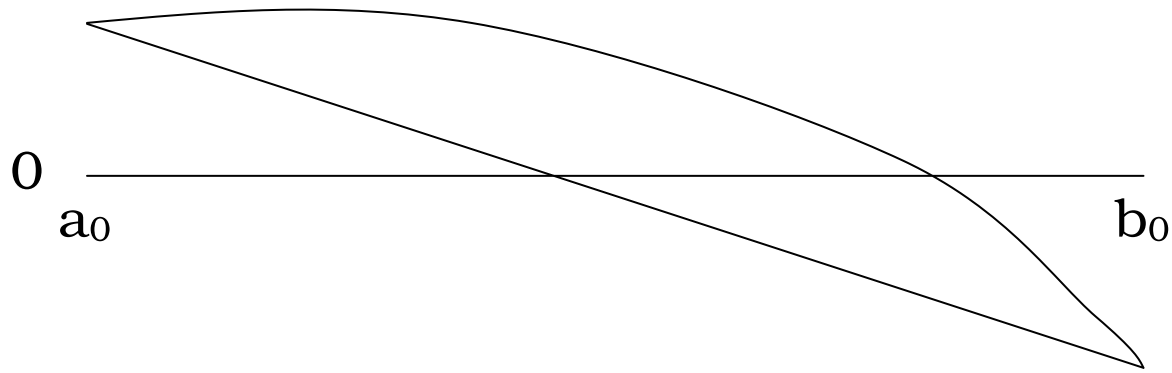
$$125(1-\lambda)\lambda - 38(2+\lambda)\lambda + 34(2+\lambda)(1-\lambda) = 0$$

$$= -197\lambda^2 + 15\lambda + 68$$

The two roots of this equation are 0.6268215 and -0.5506794. Only the first one is valid since λ must be in the range [0.25, 1].

There are other approaches similar to bisection. One useful one is the method of False Position (Regula Falsi).

A motivation behind this method is that the function is approximately linear in the region of interest.



Join points (a_i, $g(a_i)$) and (b_i, $g(b_i)$) with a straight line and find the point where the straight line intersects with the line y = 0 (call point p_i.

$$\text{Line: } l(x) = g(a_i) + \frac{g(b_1) - g(a_i)}{b_i - a_i}(x - a_i)$$

$$l(x) = 0 \Rightarrow p_i = a_i - \frac{(b_i - a_i)g(a_i)}{g(b_1) - g(a_i)}$$

If $g(p_i)g(a_i) > 0$ set $a_{i+1} = p_i$ and $b_{i+1} = b_i$

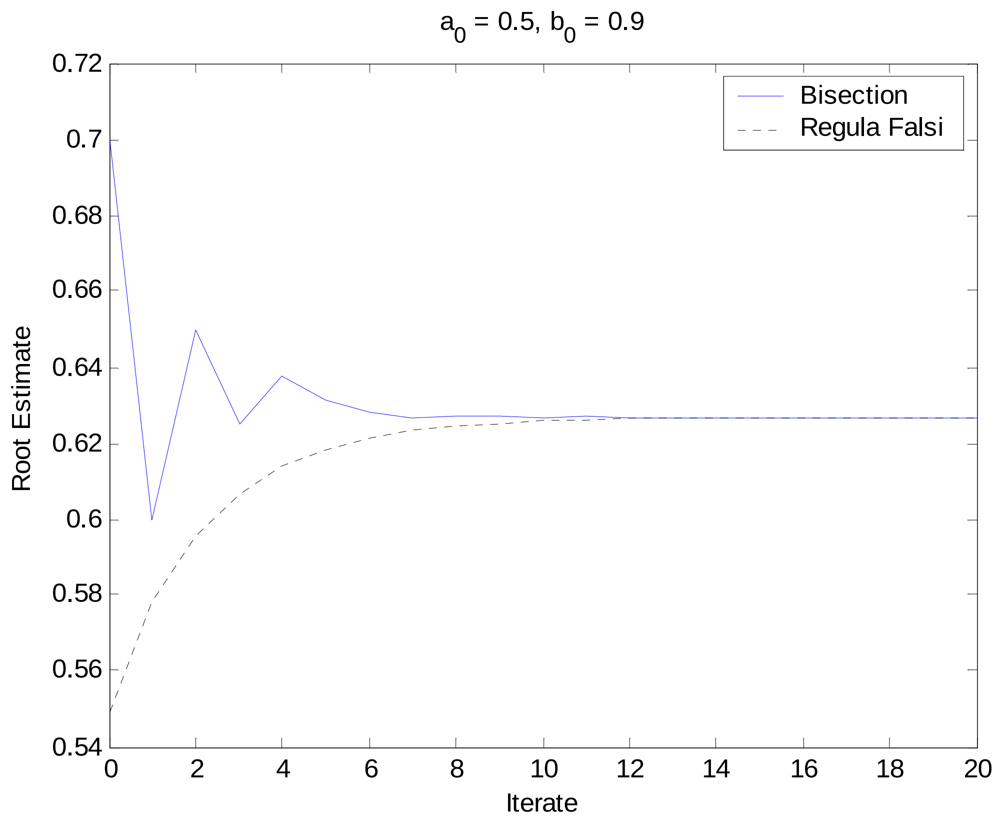
If $g(p_i)g(a_i) < 0$ set $a_{i+1} = a_i$ and $b_{i+1} = p_i$

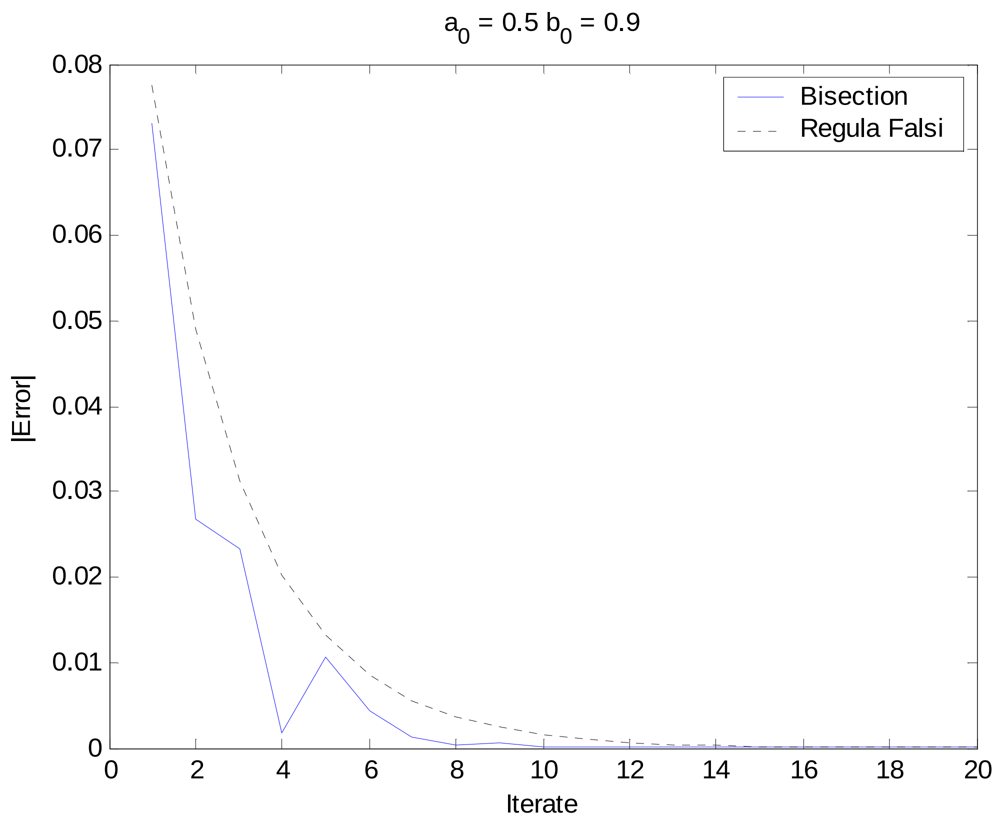
For some problems, this approach can be faster than bisection, but it depends on the shape of the function and the starting endpoints.

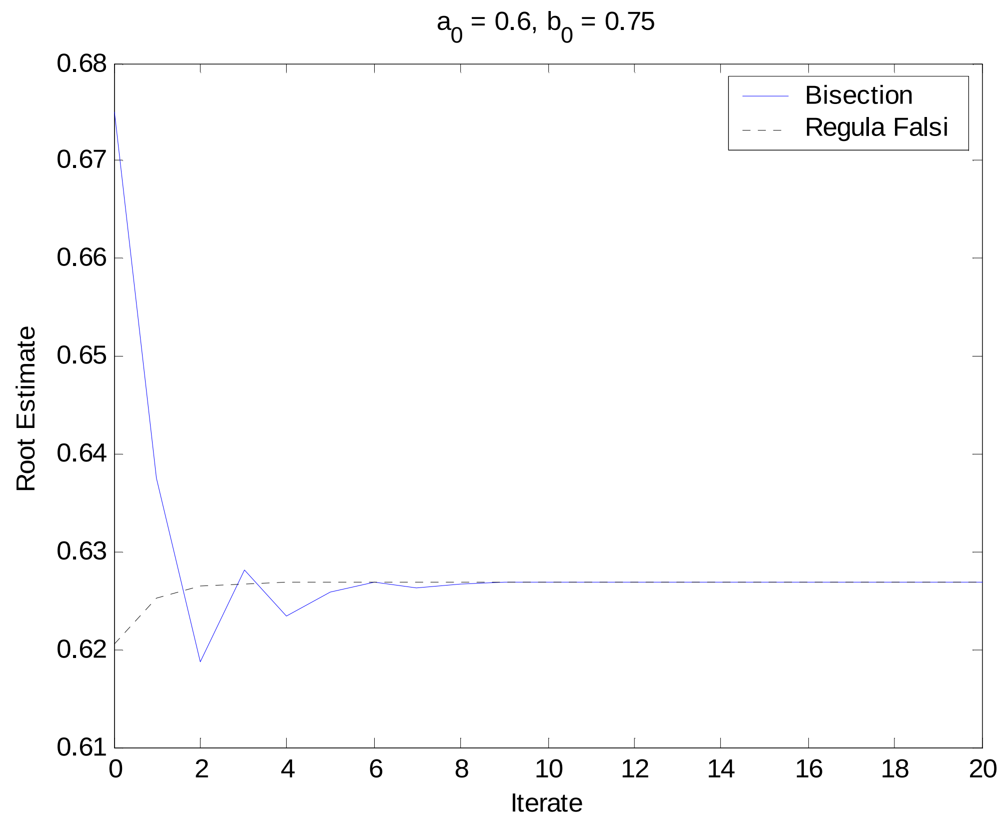
Also its harder to show convergence to the root since the interval size doesn't have to go to zero like with bisection. The can happen with a convex or concave function.

However, this routine will eventually converge to a root. This can be shown since $\{a_i\}$ is an non-decreasing sequence bounded above and $\{b_i\}$ is an non-increasing sequence bounded below.

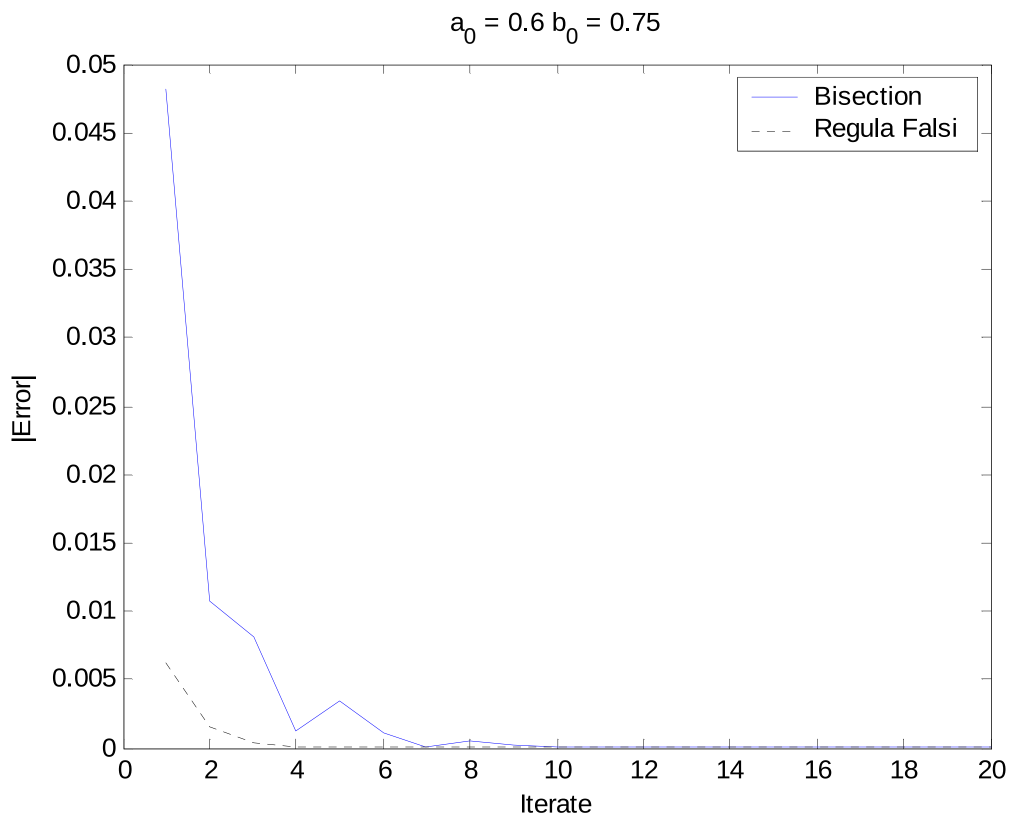
	Bisection		Regula Falsi	
Iterate	Lower	Upper	Lower	Upper
0	0.5000	0.9000	0.5000	0.9000
1	0.5000	0.7000	0.5492	0.9000
2	0.6000	0.7000	0.5779	0.9000
3	0.6000	0.6500	0.5955	0.9000
4	0.6250	0.6500	0.6066	0.9000
5	0.6250	0.6375	0.6137	0.9000
6	0.6250	0.6312	0.6183	0.9000
7	0.6250	0.6281	0.6212	0.9000
8	0.6266	0.6281	0.6232	0.9000
9	0.6266	0.6273	0.6244	0.9000
10	0.6266	0.6270	0.6253	0.9000
11	0.6268	0.6270	0.6258	0.9000
12	0.6268	0.6269	0.6262	0.9000
13	0.6268	0.6269	0.6264	0.9000
14	0.6268	0.6268	0.6265	0.9000
15	0.6268	0.6268	0.6266	0.9000
16	0.6268	0.6268	0.6267	0.9000
17	0.6268	0.6268	0.6267	0.9000
18	0.6268	0.6268	0.6268	0.9000
19	0.6268	0.6268	0.6268	0.9000
20	0.6268	0.6268	0.6268	0.9000


$a_0 = 0.5, b_0 = 0.9$
Root Estimate
Iterate
Bisection
Regula Falsi
0.72
0.7
0.68
0.66
0.64
0.62
0.6
0.58
0.56
0.54
0 2 4 6 8 10 12 14 16 18 20

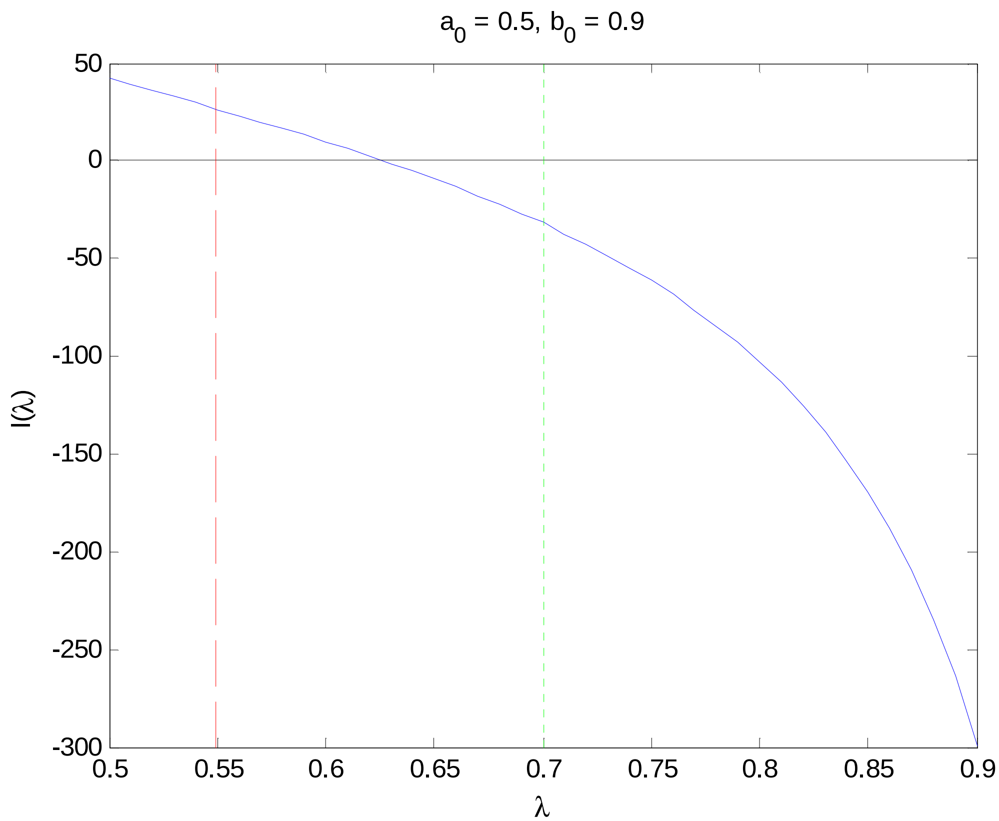

$a_0 = 0.5$ $b_0 = 0.9$
|Error|
Iterate
Bisection
Regula Falsi
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0 2 4 6 8 10 12 14 16 18 20

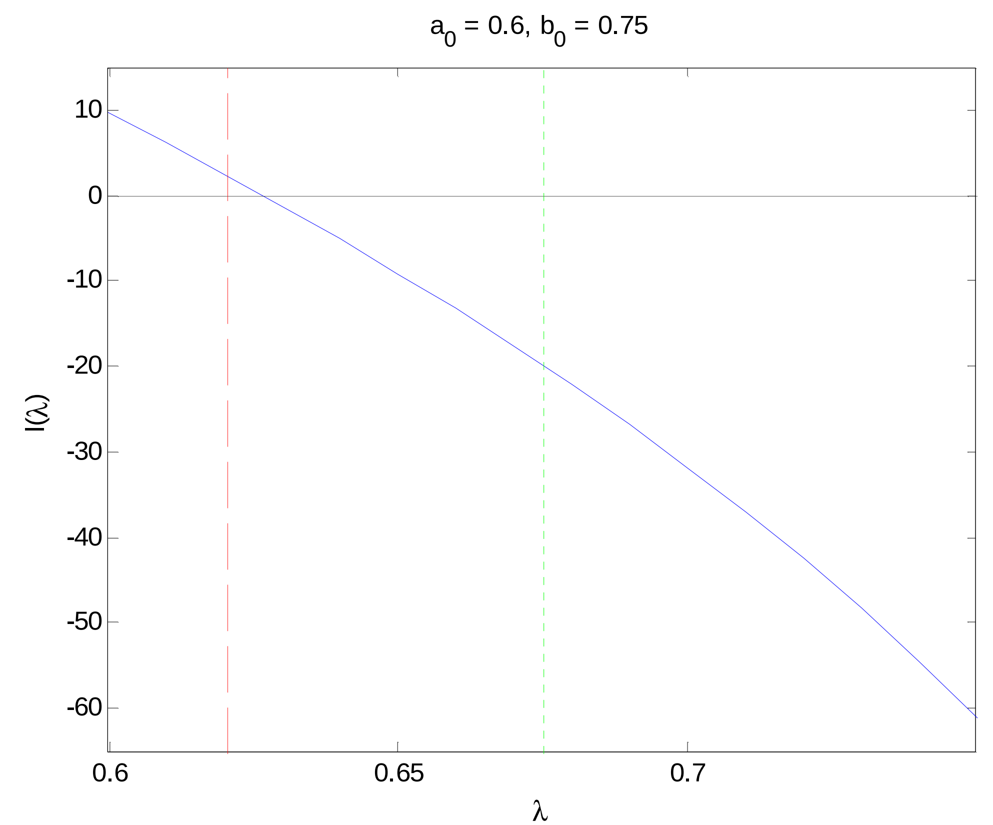


a0 = 0.6, b0 = 0.75
Bisection
Regula Falsi
Root Estimate
Iterate
0.68
0.67
0.66
0.65
0.64
0.63
0.62
0.61
0
2
4
6
8
10
12
14
16
18
20



a0 = 0.6 b0 = 0.75
Bisection
Regula Falsi
|Error|
Iterate
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
2
4
6
8
10
12
14
16
18
20


a0 = 0.5, b0 = 0.9
50
0
-50
-100
-150
-200
-250
-300
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
I(λ)
λ


a0 = 0.6, b0 = 0.75
10
0
-10
-20
-30
-40
-50
-60
0.6
0.65
0.7
I(λ)
λ

Functional Iteration (Fixed Point Approaches)

Instead of solving $g(x) = 0$, we can investigate the function

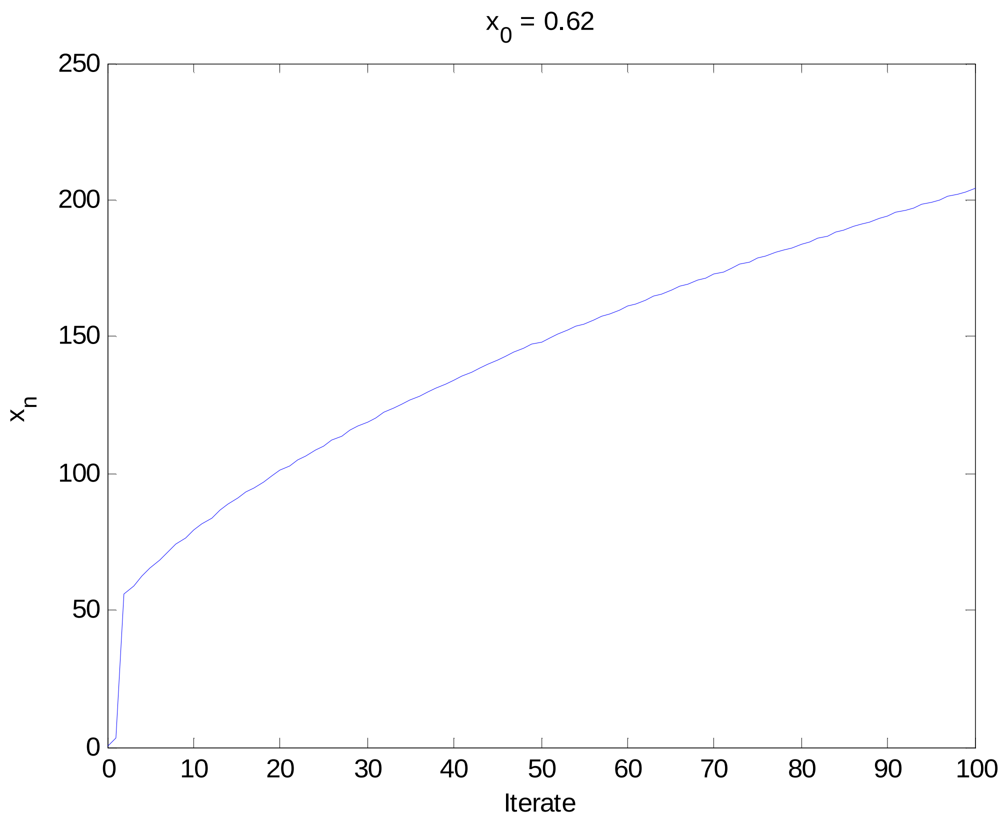
$$f(x) = g(x) + x$$

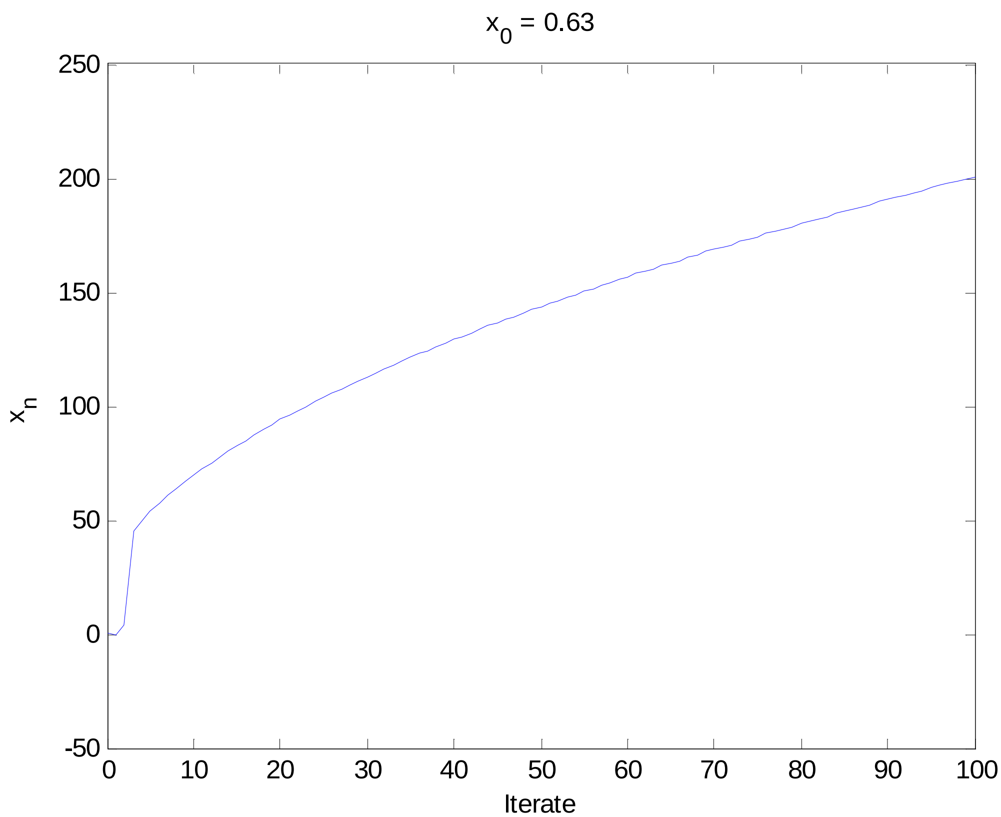
Solving $g(x) = 0$ is the same as solving $f(x) = x$.

In many situations, iterates of the sequence $x_n = f(x_{n-1})$ converge to a root of $g(x)$ starting from any point x_0 nearby.

But it doesn't have to!

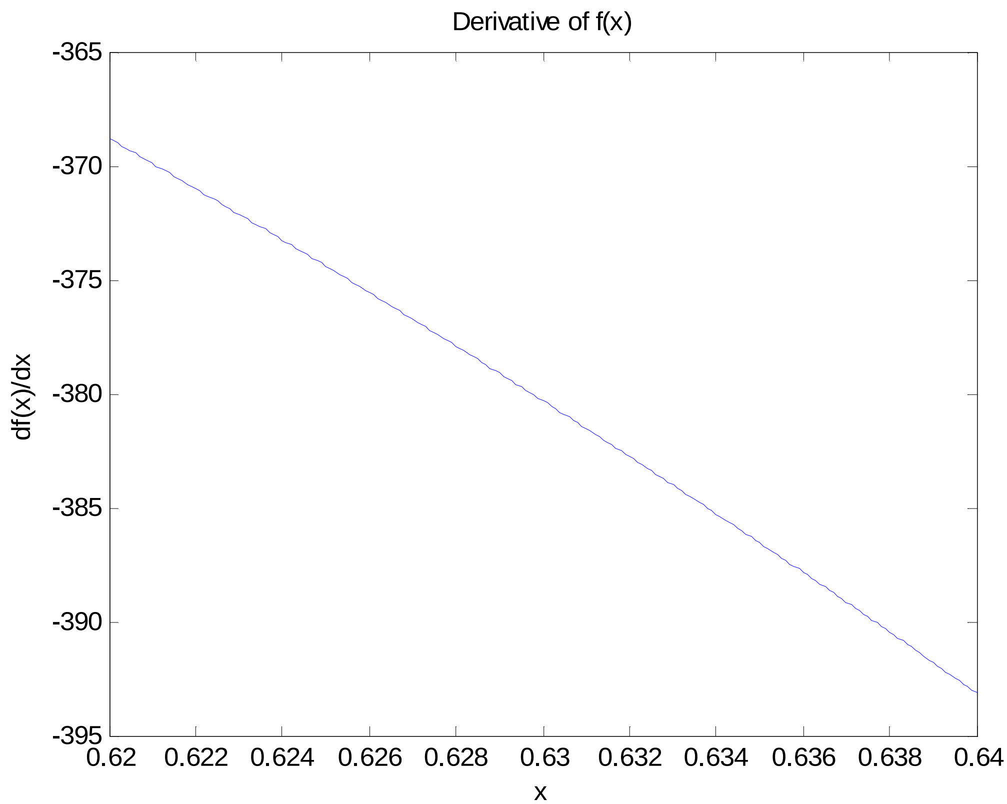
Lets run this algorithm starting at $x_0 = 0.62$ and $x_0 = 0.63$, which are both close to the true root of 0.6268215.


x0 = 0.62
250
200
150
100
50
0
xn
0
10
20
30
40
50
60
70
80
90
100
Iterate


x0 = 0.63
250
200
150
100
50
0
-50
xn
0
10
20
30
40
50
60
70
80
90
100
Iterate

In both cases, the iterates seem to diverge, or at least don't seem to converge in the right region.

Lets look at the function f(x), in particular its derivative.



So small changes in x lead to large changes in f(x), even very close to the fixed point.

So we need conditions on when fixed point methods can work

Proposition 5.3.1: Suppose f(x) defined on a closed interval I satisfies the conditions

1) $f(x) \in I$ whenever $x \in I$

2) $|f(y) - f(x)| \le \lambda |y - x|$ for any two points x & y in I.

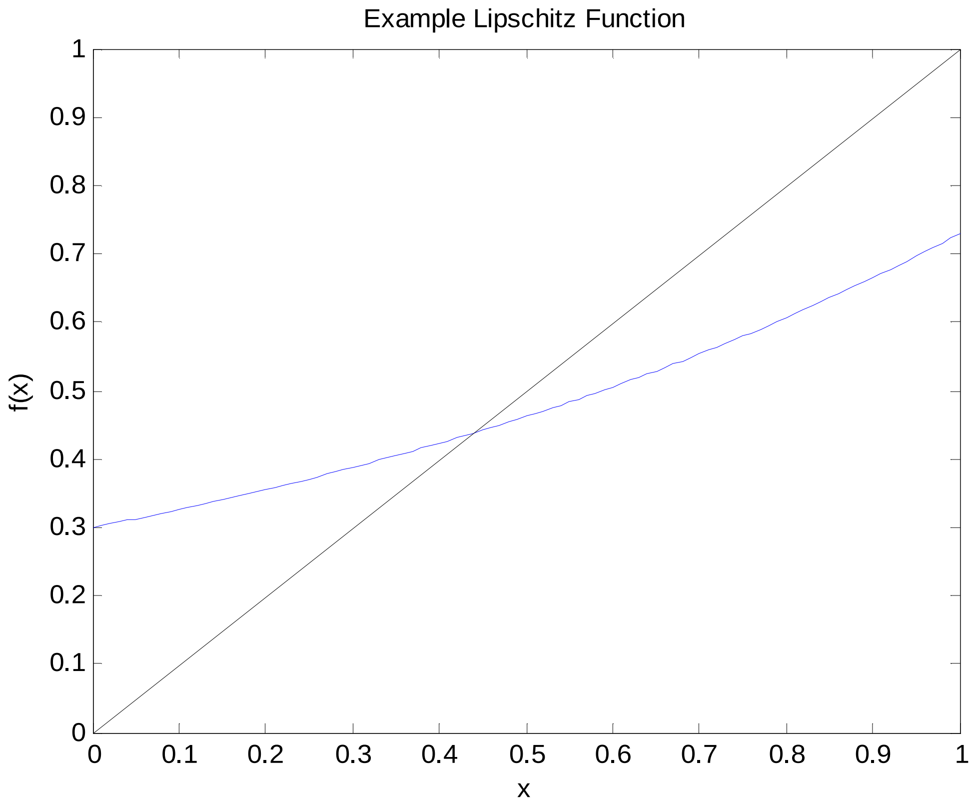
Then provided the Lipschitz constant λ is in [0, 1), f(x) has a unique fixed point $x_\infty \in I$, and the functional iterates $x_n = f(x_{n-1})$ converge to x_∞ regardless of the starting point $x_0 \in I$. Furthermore, we have the precise error estimate

$$|x_n - x_\infty| \le \frac{\lambda^n}{1-\lambda} |x_1 - x_0|$$

(For proof, see Lange)

So if f(x) (and g(x)) is nice enough, we can be guaranteed to find the desired root.

Need to get a handle on the Lipschitz constant λ. Usually you can use an upper bound on $|f'(x)|$.



What happened in the linkage example. Well $\lambda > 360$. So for an x near (but not equal to x_∞)

$|f(x) - x_\infty| = |f'(z)(x - x_\infty)|$ (Mean value theorem)

$$\geq 360\ |x - x_\infty| \geq |x - x_\infty|$$

So there was no way that this approach could work. This situation is known as repulsive. You have to end up further from the fixed point than where you started.

If $|f'(x_\infty)| < 1$, the situation is known as attractive.

If $|f'(x_\infty)| = 1$, the situation is indeterminant and investigation of the function is required.

Example: Extinction Probabilities of Branching Processes (Section 5.3.2)

Stochastic process that describes a model of population growth.

Start with 1 particle. This particle has k offspring with probability p_k. Each of these k particles generates offspring by the same mechanism. And so on for these offspring.

One question of interest is whether the population will completely die out.

This question can be answered by investigating the generating function of the process

$$P(s) = \sum_{k=0}^{\infty} p_k s^k$$

If $p_0 = 0$, the population can never die out, so we will only consider the case $p_0 > 0$.

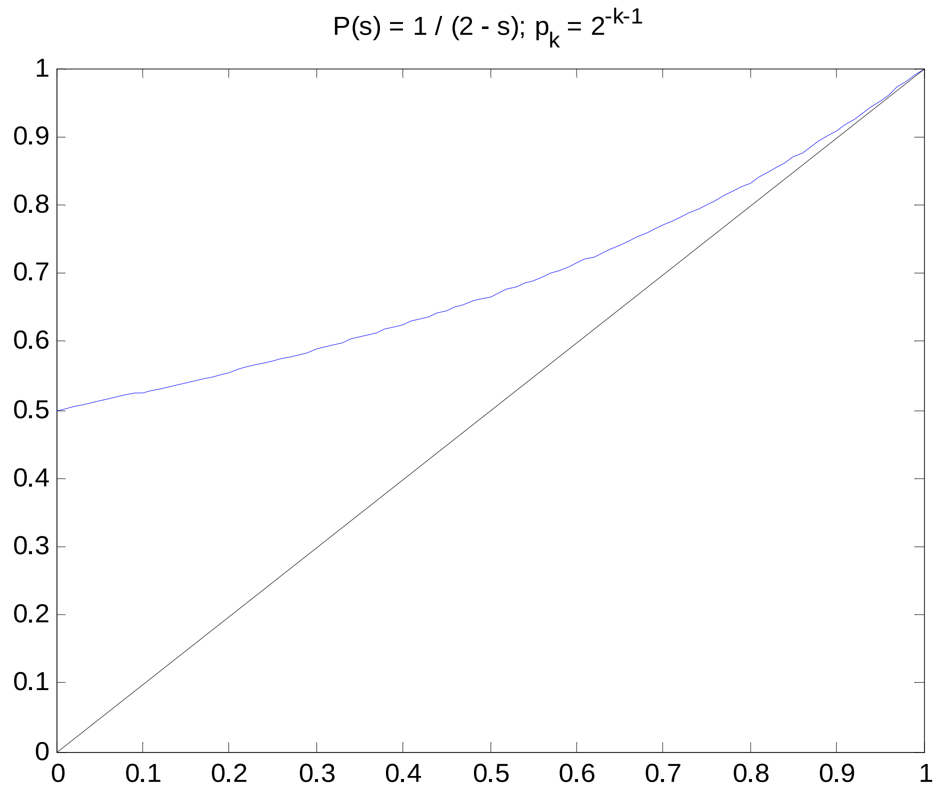
It ends up that the probability that the population will eventually die out satisfies the fixed point equation

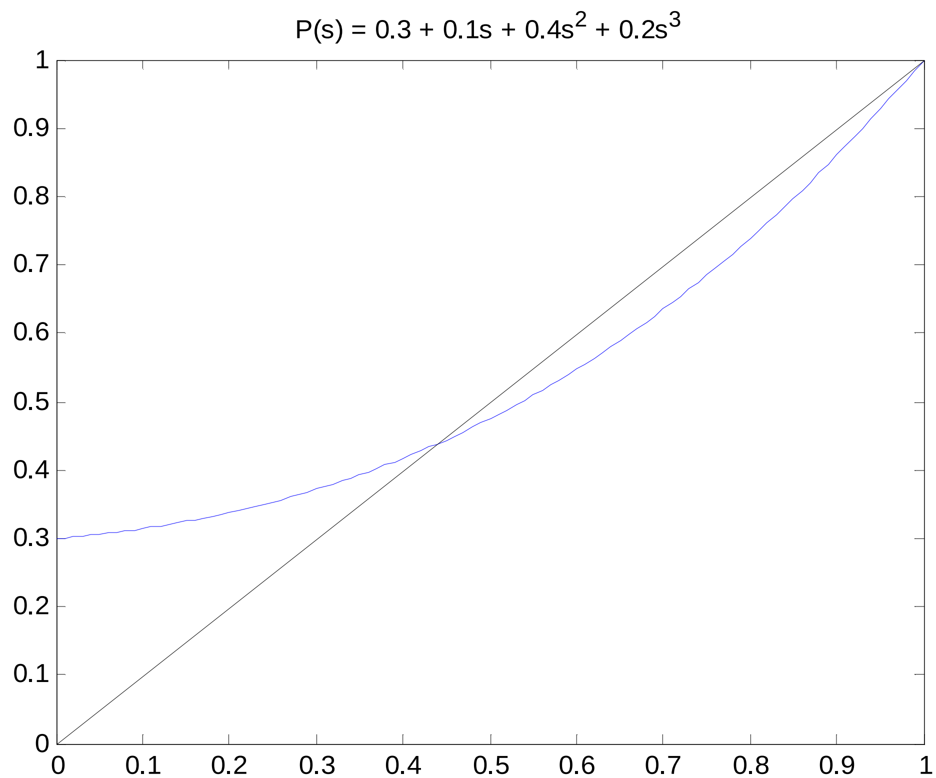
$$s = P(s)$$

This equation can have 1 or 2 fixed points. The point s = 1 must be one as $\sum_{k=0}^{\infty} p_k = 1$. It can be two since P(0) > 0 (we'll also ignore the case where $p_0 = 1$) and P(s) is a convex function in [0,1] since

$$P''(s) = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2} > 0$$

Since its convex, it will intersect a straight line at most twice. However the second point of intersection may not be in [0,1] (if it exists). The function must look like one of the following.


P(s) = 1 / (2 - s); p_k = 2^-k-1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


P(s) = 0.3 + 0.1s + 0.4s^2 + 0.2s^3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

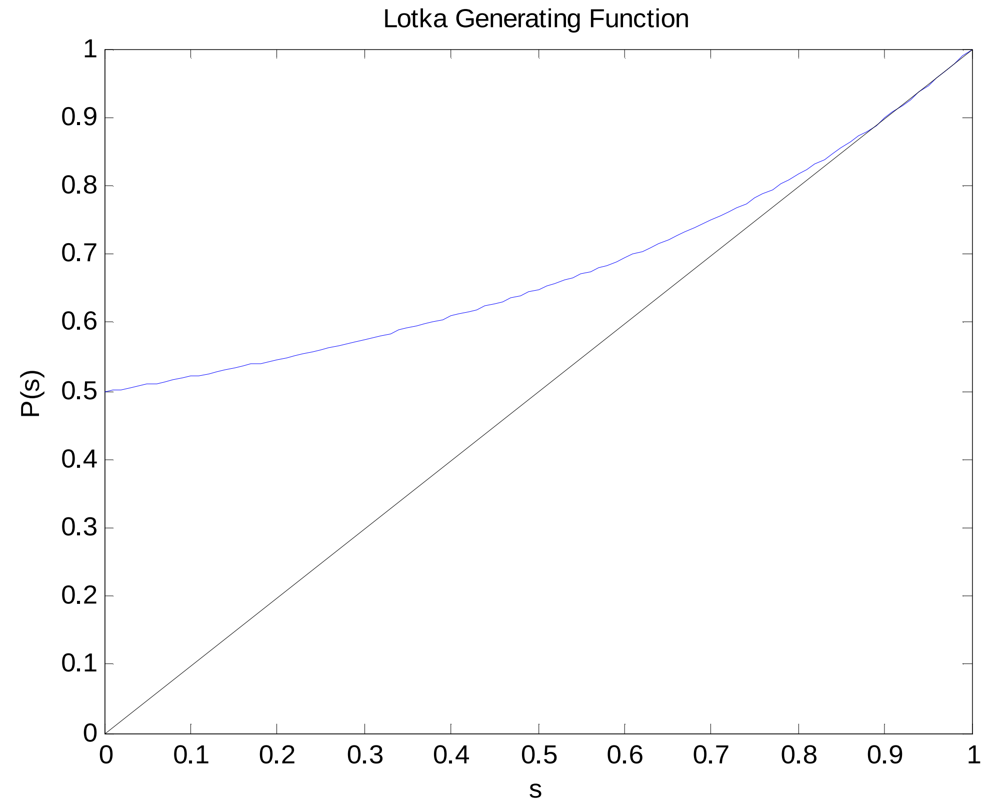
Which occurs depends on P'(1), the mean number of offspring for each particle. If P'(1) ≤ 1, the first situation must happen. The second situation will occur with If P'(1) > 1.

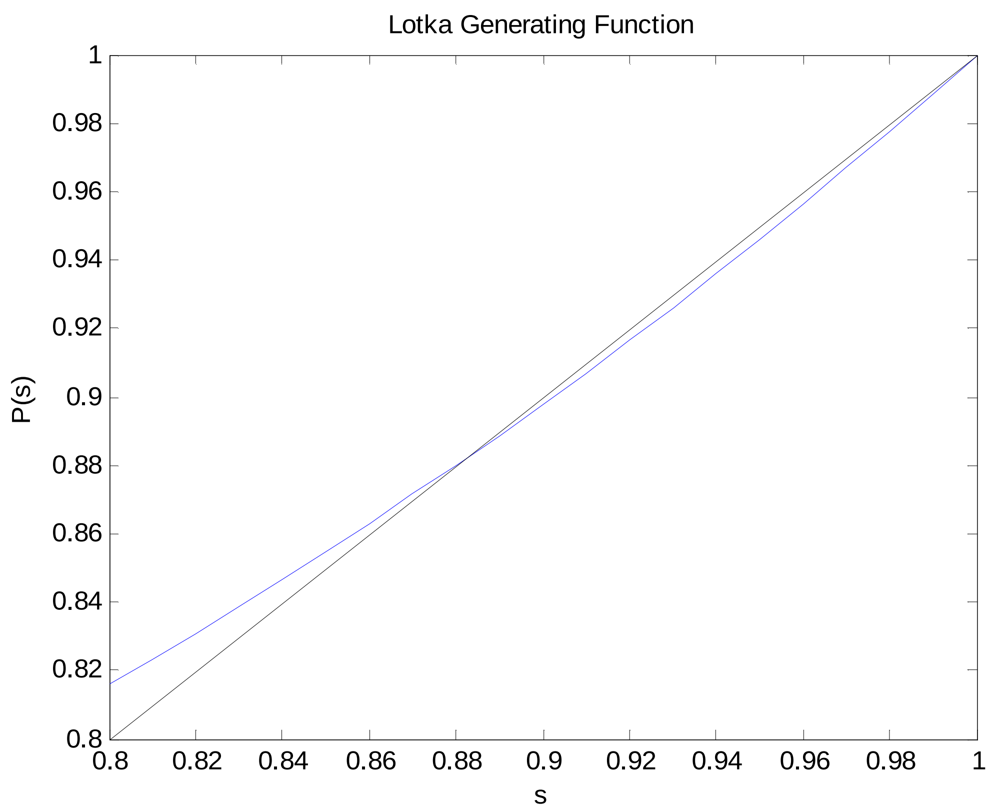
Note that the extinction probability is the smaller fixed point when P'(1) > 1. (s = 1 is a point of repulsion).

When P'(1) > 1, we can find the fixed point by iterating starting at $s_0 = 0$. This works since 0< P'(s) < 1 and P(s) ≤ s for $s \in [0, s_\infty]$. Usually it will be for $s \in [0, s_\infty + \delta]$, where $\delta > 0$.

Lets look at Lotka's example examining the extinction of surnames among white male in the US based on 1920 census data.

$$P(s) = 0.4982 + 0.2103s + 0.1270s^2 + 0.7330s^3 + 0.0418s^4 + 0.0241s^5 + 0.0132s^6 + 0.0069s^7 + 0.0035s^8 + 0.0015s^9 + 0.0005s^{10}$$


Lotka Generating Function
P(s)
s
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

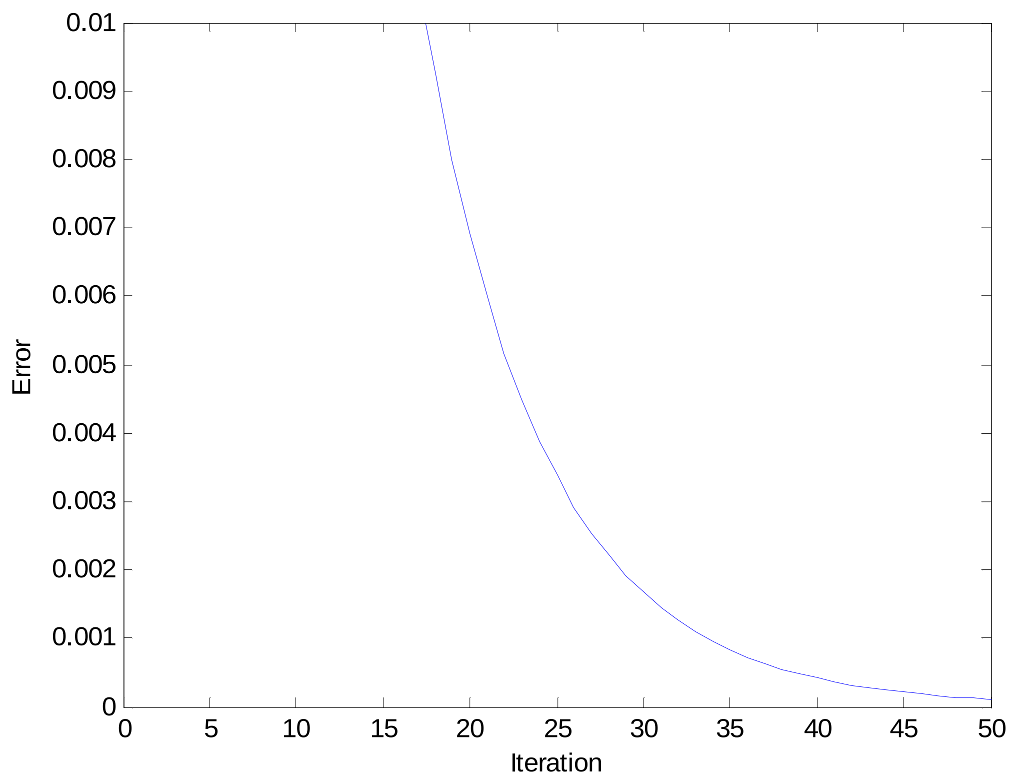

Lotka Generating Function
P(s)
s
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1

The method converges slowly to the extinction probability of 0.879755.

After 50 steps, we can show that

$$\text{Bound: } |s_n - s_\infty| \le \frac{\lambda^n}{1-\lambda}|s_1 - s_0| = 0.0039$$

$$\text{Actual: } |s_n - s_\infty| = 0.000105$$



Why is convergence slow?

Error estimate bound

$$|s_n - s_\infty| \le \frac{\lambda^n}{1-\lambda}|s_1 - s_0|$$

So

$$\frac{\max|s_{n+1} - s_\infty|}{\max|s_n - s_\infty|} = \lambda$$

A reasonable Lipschitz constant for this problem is $P'(s_\infty) = 0.8713$.

So each step is only bringing you about 13% closer to the truth.

If we were to use the bisection method to solve this problem, based on $g(s) = P(s) - s = 0$.

$$\frac{\max|s_{n+1} - s_\infty|}{\max|s_n - s_\infty|} = ½$$

So each step is bringing us about half the way there.

After 50 bisection steps ($a_0 = 0$, $b_0 = 1$)

$$|s_n - s_\infty| \leq \frac{1}{2^{51}} = 4.4409\text{e-}016$$

Functional iteration methods such as these aren't commonly used to directly find roots much in statistics from what I've seen, but other methods, such as Newton-Raphson to have a functional iteration property underlying them.