Solution of non-linear equations

Finding MLEs, posterior modes (MAP estimates), minimizing loss functions, etc.

In many cases, this problem reduces to solving a nonlinear equation as

> arg min f(x) or arg max f(x) usually satisfies $f'(x) = 0$.

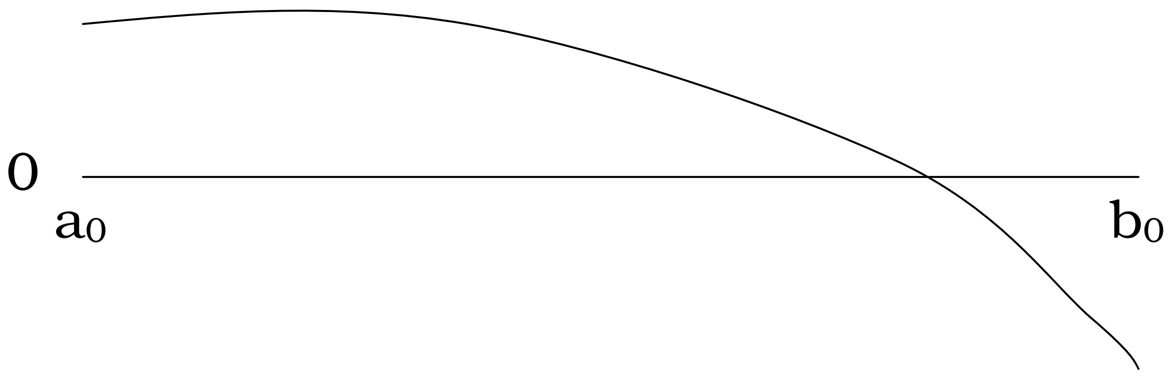
Usually easier to solve $f'(x) = 0$ than to deal with f(x) directly.

There are lots of ways to do this. Three popular approaches are bisection, functional iteration, and Newton-Raphson.

Bisection (for 1 dimension problems)

Have continuous function g(x) and two values a_0 and b_0 such that

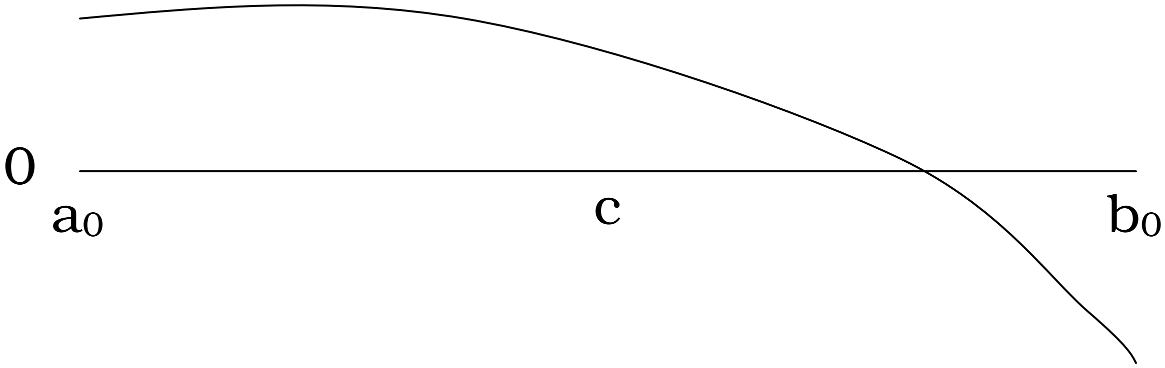
$$g(a_0) > 0 \text{ and } g(b_0) < 0 \quad \text{(or vice versa)}$$



We know that there exists at least one point, x^* in (a_0, b_0) such that $g(x^*) = 0$ by the intermediate value theorem.

Idea: try midpoint of interval

$$c = \frac{a_0 + b_0}{2} \text{ \& evaluate } g(c)$$



If $g(c)g(a_0) > 0$ set $a_1 = c$, $b_1 = b_0$ and continue

> {g(c) and g(a) are both > 0 or both < 0 so must be a root between c and b}

If $g(c)g(a_0) < 0$ set $a_1 = a_0$, $b_1 = c$ and continue

> {g(c) and g(a) are on opposite side so there must be a root between a and c}

If $g(c) = 0$ stop

Continue until the interval width gets small enough.

After n steps, the interval width = $b_n - a_n$
= $\frac{b_0 - a_0}{2^n}$.

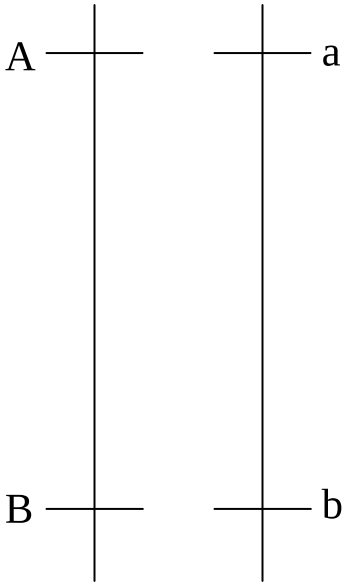
Set $\hat{x}^* = \frac{a_n + b_n}{2}$, the midpoint of the last interval as the estimate of the root $\hat{x}$.

Since it's the midpoint of the last interval, the maximum error satisfies

$$|\hat{x}^* - \hat{x}| \le \frac{b_0 - a_0}{2^{n+1}}$$

Example: Linkage Analysis (Rao, 1973, pp 268-269)

2 gene on a chromosome are separated by a recombination fraction θ (θ ≤ ½).



This organism can pass 4 possible haplotypes to its offspring

Haplotype	Probability
AB	(1 – θ)/2
Ab	θ/2
aB	θ/2
ab	(1 – θ)/2

An experiment was performed to estimate θ. The breeding experiment crossed AB|ab x AB|ab and recorded the observed phenotypes.

In this experiment, 2 dominate traits were observed (A dominant to a, B dominate to b).

While there are 16 possible joint haplotypes in the offspring (4 from father times 4 from mother), there are only 4 possible phenotypes

Phenotype	Probability	Counts
AB	$(3 - 2\theta + \theta^2)/4$	125
Ab	$(2\theta - \theta^2)/4$	18
aB	$(2\theta - \theta^2)/4$	20
ab	$(1 - 2\theta + \theta^2)/4$	34

Note that this problem is easier to solve with the transformation (Lange page 126, problem 7)

$$\lambda = 1 - 2\theta + \theta^2 = (1 - \theta)^2$$

$$\theta = 1 - \sqrt{\lambda}$$

Under this transformation, the probabilities are

Phenotype	Probability	Counts
AB	$(2 + \lambda)/4$	125
Ab	$(1 - \lambda)/4$	18
aB	$(1 - \lambda)/4$	20
ab	$\lambda/4$	34

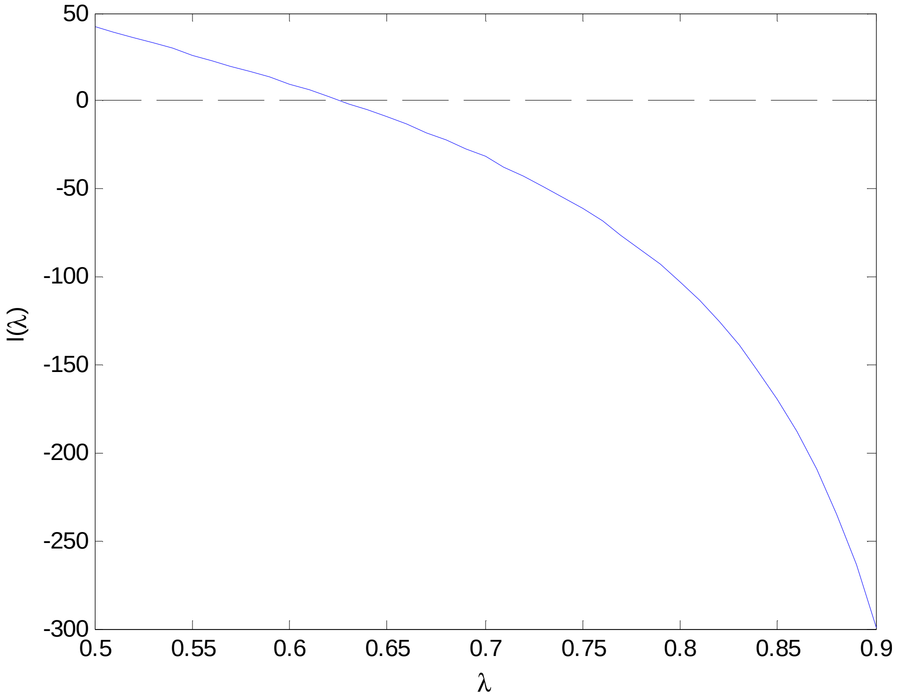
The likelihood and log likelihood functions are

$$L(\lambda) \propto (2+\lambda)^{125}(1-\lambda)^{18+20}\lambda^{34}$$

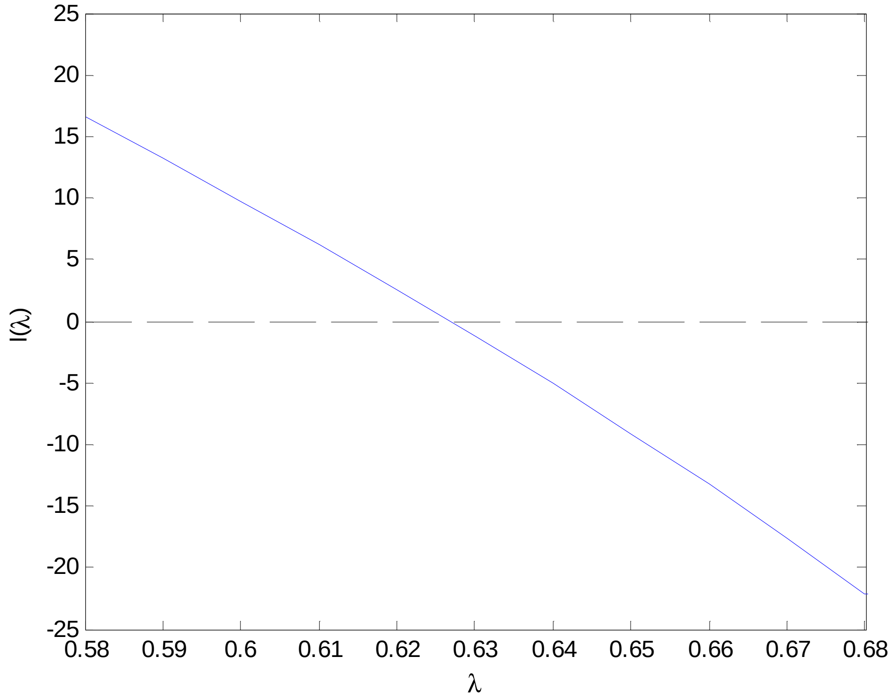
$$\log L(\lambda) = 125\log(2+\lambda) + 38\log(1-\lambda) + 34\log\lambda$$

which gives the score function

$$l(\lambda) = \frac{d}{d\lambda}\log L(\lambda) = \frac{125}{\lambda} - \frac{38}{1-\lambda} + \frac{34}{\lambda}$$

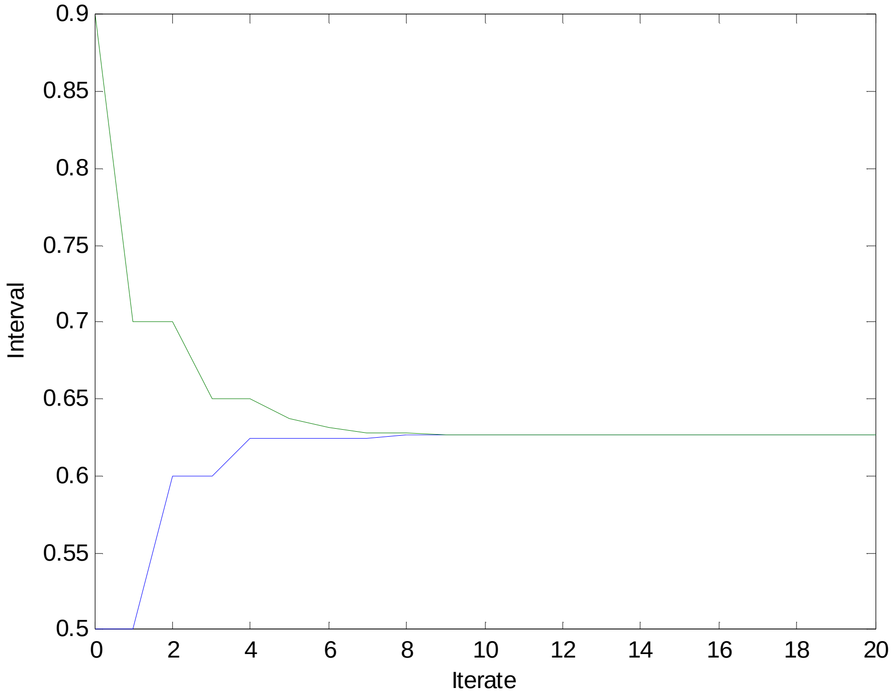


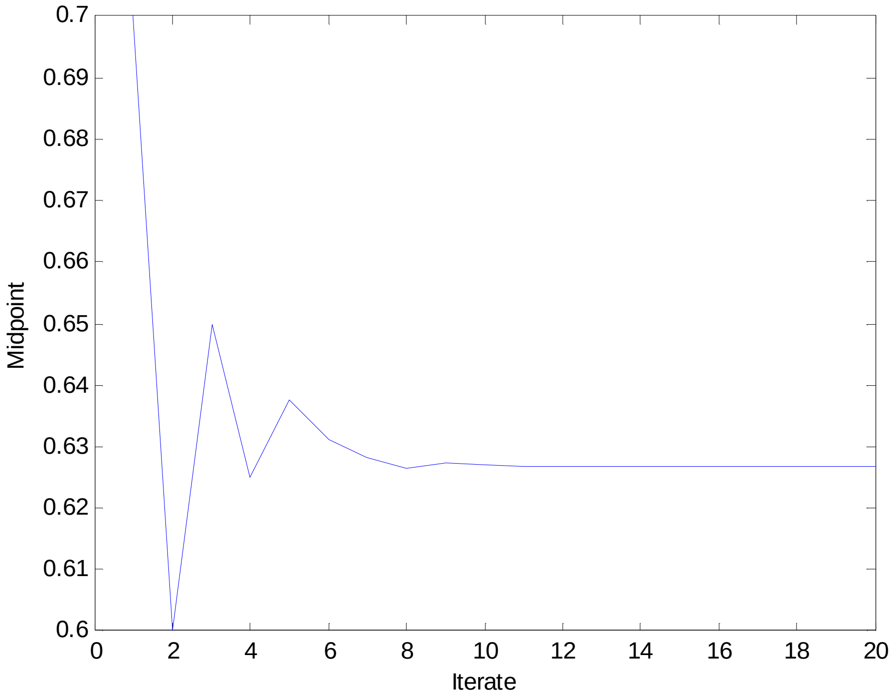
50
0
-50
-100
-150
-200
-250
-300
I(λ)
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
λ

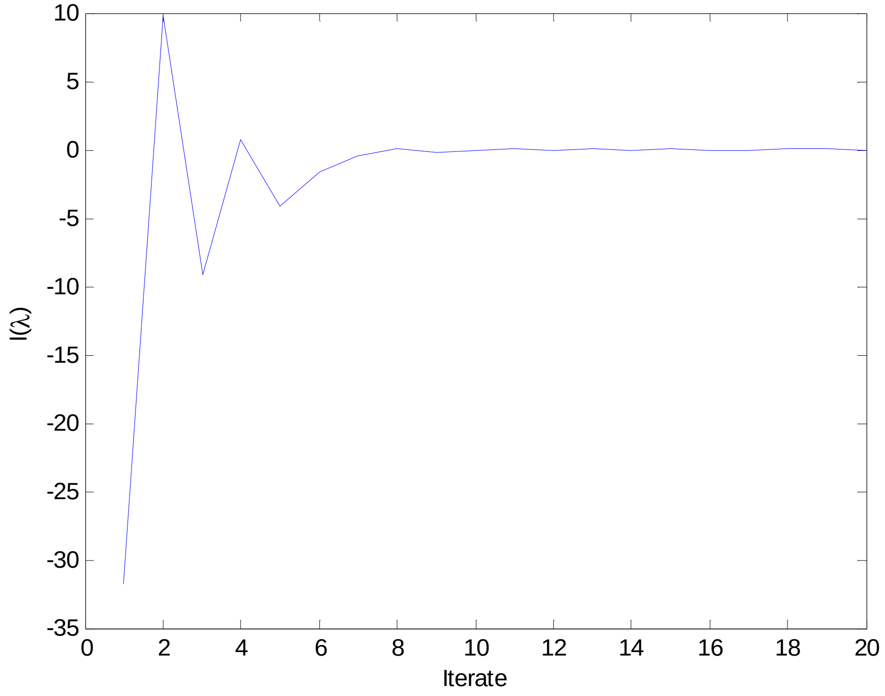


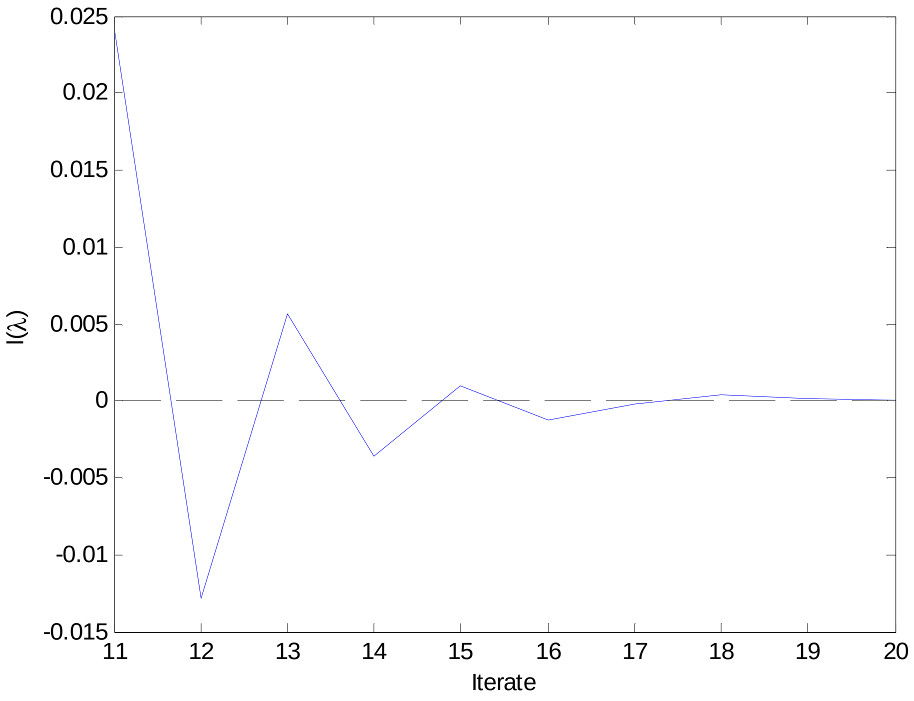
25
20
15
10
5
0
-5
-10
-15
-20
-25
I(λ)
0.58
0.59
0.6
0.61
0.62
0.63
0.64
0.65
0.66
0.67
0.68
λ

The bisection algorithm for $l(\lambda)$ with $a_0 = 0.5$ and $b_0 = 0.9$ gives $\hat{\lambda}^* = 0.6268$ after 20 steps. The convergence pattern can be seen with




0.7
0.69
0.68
0.67
0.66
0.65
0.64
0.63
0.62
0.61
0.6
Midpoint
0
2
4
6
8
10
12
14
16
18
20
Iterate


10
5
0
-5
-10
-15
-20
-25
-30
-35
I(λ)
0
2
4
6
8
10
12
14
16
18
20
Iterate



Iterate	λ	l(λ)
1.0000	0.7000	-31.7989
2.0000	0.6000	9.7436
3.0000	0.6500	-9.0939
4.0000	0.6250	0.6857
5.0000	0.6375	-4.1009
6.0000	0.6312	-1.6835
7.0000	0.6281	-0.4931
8.0000	0.6266	0.0977
9.0000	0.6273	-0.1973
10.0000	0.6270	-0.0497
11.0000	0.6268	0.0240
12.0000	0.6269	-0.0128
13.0000	0.6268	0.0056
14.0000	0.6268	-0.0036
15.0000	0.6268	0.0010
16.0000	0.6268	-0.0013
17.0000	0.6268	-0.0002
18.0000	0.6268	0.0004
19.0000	0.6268	0.0001
20.0000	0.6268	-0.0000

For this example,

$$\text{max error} \le \frac{0.9-0.5}{2^{21}} = 1.9 \times 10^{-7}$$

To get $\hat{\theta}^*$, use $\theta = 1 - \sqrt{\lambda}$, which gives

$$\hat{\theta}^* = 1 - \sqrt{0.6268} = 0.2083$$

Note that the maximum error in with the estimate $\hat{\theta}^*$ needs to be carefully thought about, since the transformation is non-linear.

How many iterations for the bisection algorithm?

Once a_0 and b_0 are determined its easy. Base on a maximum desired error

Want $\frac{b_0 - a_0}{2^{n+1}} \le M$. Then set n to satisfy

$$n \ge \log_2\left(\frac{b_0 - a_0}{M}\right) - 1$$

$$= \log\left(\frac{b_0 - a_0}{M}\right) / \log 2 - 1$$

For example, for $M = 0.0001$

$$n \geq \log\left(\frac{0.9-0.5}{0.0001}\right)/\log 2 - 1 = 10.9$$

so use at least 11 iterates.

Advantages of the bisection method:

- Must terminate
- Guaranteed to find a zero of the function do desired accuracy

Disadvantages:

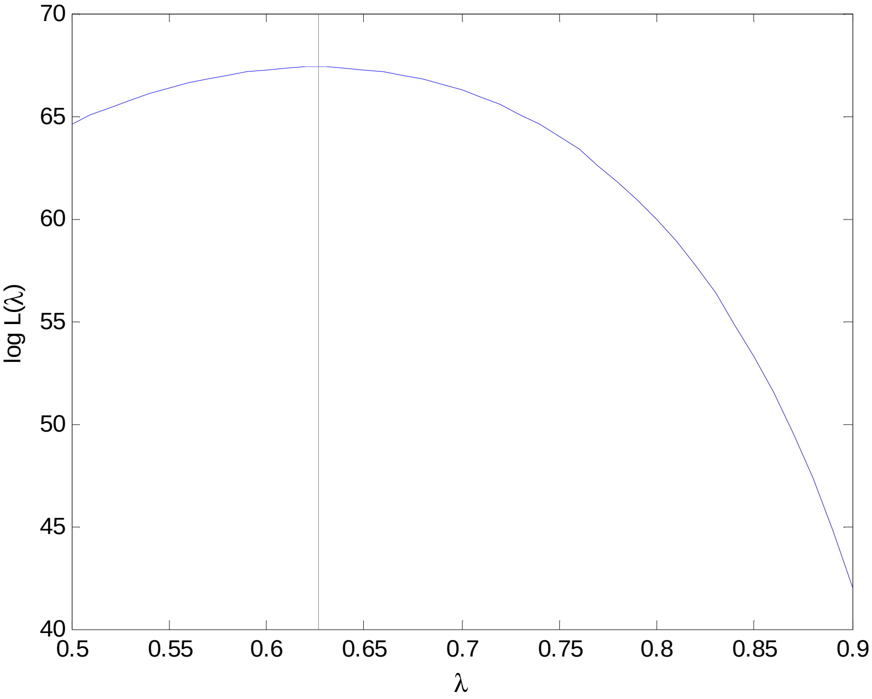
- Can only handle univariate problems
- Linear convergence (Other algorithms, such as Newton-Raphson can be faster)
- From optimization point of view, not guaranteed to find a optima.

Note that this disadvantage is not really specific to bisection, but to using root finders on the derivative of the function to be optimized.

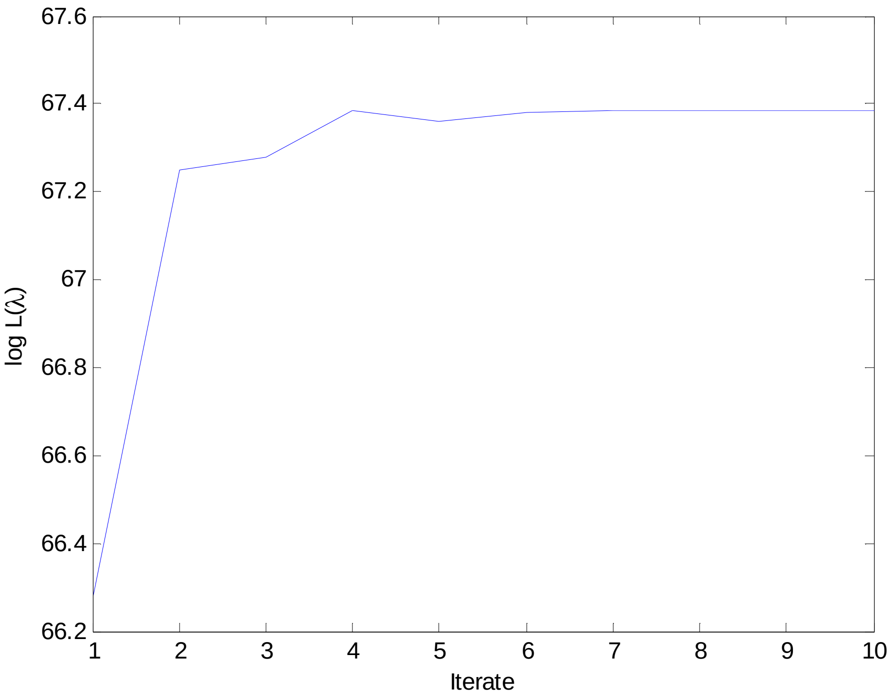
Solving $f'(x) = 0$ may give a minimum or a saddle point when a maximum is desired.

Need to check $f(x)$ or $f''(x)$ to see if $\hat{x}^*$ is a local maximum (e.g. is $f''(\hat{x}^*) < 0$ or is $f'(x)$ a decreasing function around $\hat{x}^*$)

In this case, the log likelihood is definitely concave, so we have found the MLE



One thing to note with the bisection algorithm when used to optimize a likelihood function, the likelihood (or log likelihood) does not have to increase at each step, particularly for the early iterates. However it will tend to do this once you are in the area of the optima, as can be seen in the following figure.


67.6
67.4
67.2
67
66.8
66.6
66.4
66.2
log L(λ)
1
2
3
4
5
6
7
8
9
10
Iterate