Computing_HW1

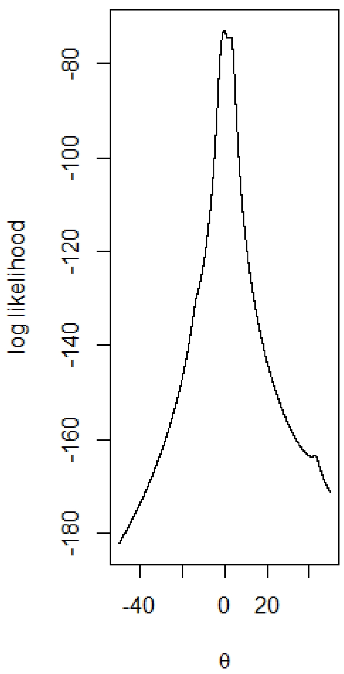
2020321163

엄상준

1.

(a)

- Graph



- MLE using the Newton-Raphson Method

Starting Point	-11	-1	0	1.5	4	4.7	7	8	38
MLE	NaN	-0.1923	-0.1923	1.714	2.817	-0.1923	41.041	NaN	42.795

우선 위의 그래프에서 max값을 갖는 theta의 위치는

> theta[which.max(ll_result)]

[1] -0.19

이다.

따라서 -1, 0, 4.7은 좋은 starting point들이라고 볼 수 있다. 그러나 starting point에 따라서 Na값이 나오기도 하고 -0.19와는 동떨어진 값이 나오기도 하는 것을 알 수 있다. 특히 7과 38은 매우 다른 값을 보인다. 즉, newton-raphson 방법은 starting point를 무엇으로 설정하는지가 굉장히 중요하다는 것을 알 수 있다.

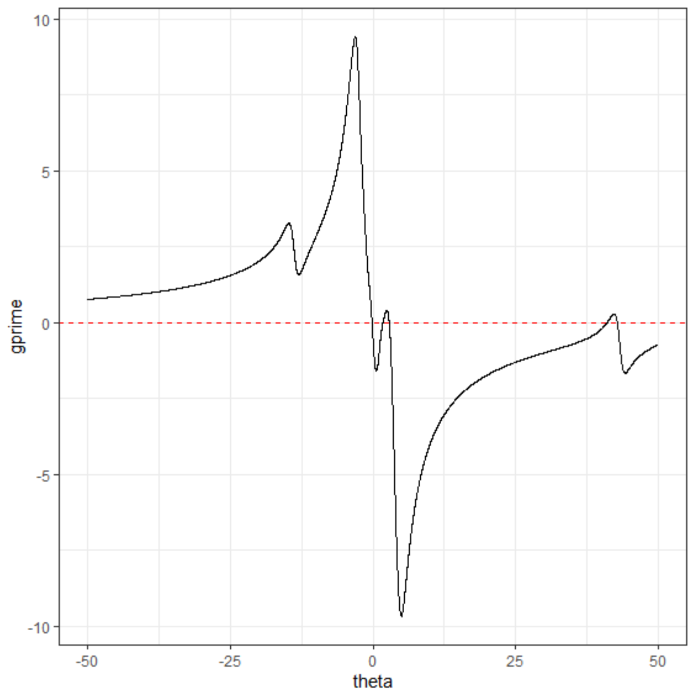
Data의 평균을 starting point로 잡으면,

> newton(mean(data), 0.0001, data, 10000)

[1] 54.87662

값이 나오는데, 이는 -0.19에서 동떨어진 값으로 별로 좋지 않다.

위와 같은 결과가 나온 이유를 log-likelihood의 미분 함수의 그래프를 그려서 살펴보자.



→ 그래프의 모양이 굉장히 복잡하며 0값에 해당하는 theta값이 5개 나온다는 것을 알 수 있다. 이는 local optimum이 5개라는 뜻이며 starting point의 설정에 따라 global maximum이 아니라 local optimum으로 값이 수렴할 수 있음을 의미한다.

(b)

- starting point가 -1과 1인 경우

> bisec(-1,1,0.0001,data, 10000)

[1] -0.1922913

→ -0.19에 가까운 좋은 결과.

- starting point가 -10과 -5인 경우

> bisec(-10,-5,0.0001, data, 10000)

[1] -9.999962

→ global maximum을 찾는 데에 실패한 경우.

(c)

		Alpha				
		3	1	0.64	0.25	0.1
Starting Point	1	-18.739	-0.072	-0.192	-0.192	-0.192
	-1	-2.658	0.795	-0.192	-0.192	-0.192
	-5	6.770	0.156	-0.192	-0.192	-0.192

→ Starting point에 관계없이 alpha가 0.64미만이면 global maximum으로 대체적으로 수렴.

→ 다만 global maximum에 수렴하는 것이 아닌 경우 starting point에 따라 MLE값이 차이를 보인다.

(d)

		$\theta^{(1)}$		
		-1	1	3
$\theta^{(0)}$	-1	NaN	-0.192	1.714
	-2	-0.192	1.714	2.817
	-3	-0.192	1.714	2.817

→ global maximum으로 수렴하는 값도 있는 한편, 발산하거나 local optimum 값으로 가는 값도 있다.

(e)

-Speed 비교

일반적인 system.time 방법을 비교했을 때엔 값이 모두 0으로 나와서 microseconds의 값을 잴 수 있는 package를 사용하였다. (convergence criteria는 10e-8로, max iteration은 10000번으로 통일하였고 모두 global maximum으로 수렴한 starting point를 사용)

그 결과

		Microseconds		
		Min	Mean	Max
Method	Newton	3.2	4.384	42.9
	Bisec	6.6	7.461	24.4
	Fixed	48.4	50.384	84.8
	Secant	4.2	4.88	24.5

- ➔ 전체적으로 봤을 때 Fixed Method가 가장 오래 걸린다는 것을 알 수 있다.
- ➔ Newton의 경우 Min과 Max의 차이가 크다.
- ➔ Fixed 외의 방법들은 대체적으로 비슷하였다.
- ➔ 하지만 각 방법들은 초기값이 무엇인지, 그리고 어떤 알고리즘인지에 따라 소요시간의 차이가 많이 날 수 있다.
- ➔ 또한 microseconds는 매우 작은 값으로 실제로는 거의 차이가 나지 않는 것들이다. 따라서 large dataset의 경우에는 결과가 달라질 수 있다.

- Stability

위의 경우는 loglikelihood 미분함수가 여러 해를 갖는 multimodality의 경우이다. 따라서 2차 미분 함수를 이용하는 newton method와 secant method의 경우 초기값에 따라 알고리즘이 발산해버릴 수 있다. Bisection의 경우에도 multimodality에 의해 안정적이지 못할 수 있다. Fixed method의 경우도 마찬가지.

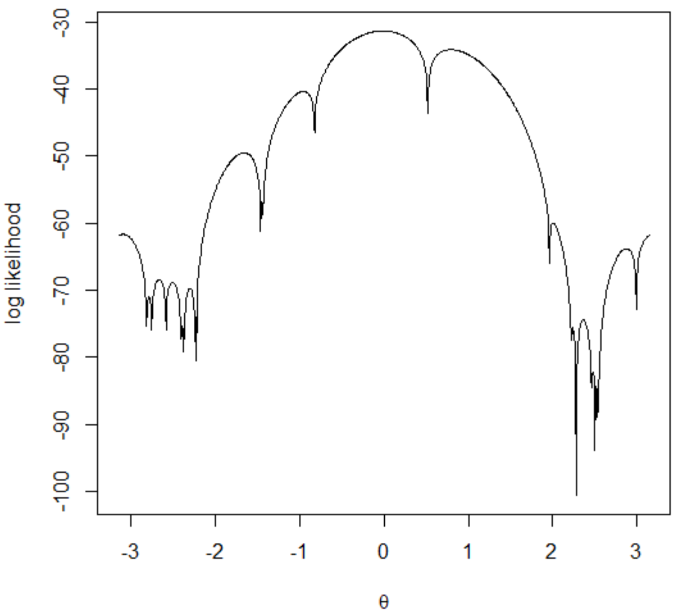
-Normal

Normal의 경우는 loglikelihood의 1차 미분함수가 단조 감소 함수로 1개의 peak만을 가지는 loglikelihood function을 가진다. 따라서 위에서 발생했던 multimodality로 인한 문제들이 발생하지 않는다. 따라서 이론적인 결과에 따라 Newton과 secant method가 가장 빠르게 MLE에 수렴할 것이며, Bisection의 초기값 구간이 표본평균을 포함하면 Global maximum으로 수렴할 것이다.

Fixed의 경우도 scaling factor를 너무 작은 값으로 하지 않는 이상 수렴할 것이다.

2.

(a)



(b)

Method-of-moments Estimator.

$f(x) = \frac{1}{2\pi}(1-\cos(x-\theta))$, $0 \le x \le 2\pi$

$E(X) = \int_0^{2\pi} \frac{x}{2\pi}(1-\cos(x-\theta))\,dx$

$= \frac{1}{2\pi}\left[[x^2 - x\sin(x-\theta)]_0^{2\pi} - \int_0^{2\pi} x - \sin(x-\theta)\,dx\right]$ by integration by part

$= \frac{1}{2\pi}[2\pi^2 + 2\pi\sin\theta]$

$= \pi + \sin\theta$, $-\pi \le \theta \le \pi$

1st method of moments estimator : $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$\bar{x} = \pi + \sin\theta$

$\Leftrightarrow \theta = \sin^{-1}(\bar{x} - \pi)$

이에 따른 method of moments estimator값은

> asin(mean(data2)-pi)

[1] 0.05844061

(c)

- Mom

> newton2(mom, 10e-5, data2, max_iter = 1000)

[1] -0.011972

- -2.7

> newton2(-2.7, 10e-5, data2, max_iter= 1000)

[1] -2.6667

- 2.7

> newton2(2.7, 10e-5, data2, max_iter= 1000)

[1] 2.873095

(d)

200개의 구간을 나눈 뒤 netwon-raphson방법을 통해 200개의 구간으로 나눈 starting point에서 MLE값을 구해서 각 구간별 local optimum을 구해 본 결과

총 19개의 mode가 나왔다.

(e)

위의 그래프의 형태를 보았을 때 local minimum이 뾰족한 형태를 보인다. 즉, local minimum의 우극한과 좌극한을 starting point로 한다면 서로 다른 MLE값을 가질 것이라고 추측하였다.

Local minimum 중 global minimum 값인 2.279935값을 기준으로 0.0001을 빼고 더해준 값을 비교하였다.

> newton2(x=2.279935-0.0001, epsilon = 0.0001, data = data2, max_iter = 1000)

[1] 2.236219

> newton2(x=2.279935+0.0001, epsilon = 0.0001, data = data2, max_iter = 1000)

[1] 2.28007

그 결과 다른 MLE값을 가진다는 것을 알 수 있다.

추가숙제

1. g(x)=log(x)/(1+x) (1<x<5) Find max, using (1) Bisection, (2) Newton, (3) Secant, (4) Fixed point Method.

> newton3(2, 10e-8,10000)

[1] 3.591121

> secant3(2,3,10e-8,10000)

[1] 3.591121

> bisec3(2,4,10e-8,10000)

[1] 3.591121

> fixed3(2,0.7,10e-8,10000)

[1] 3.591113

→ 모두 약 3.59의 MLE값을 가진다.

2. IRLS for the Poisson Regression

Iterative reweighted least square for the poisson regression

Linear predictor $\eta_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$ 라 하자.

그리고 $\eta_i = \log(\theta_i)$, $\theta_i = e^{\eta_i}$. 즉 link function 을 log로 하자.

Likelihood for poisson regression

$$: L(\theta) = \prod_{i=1}^{n} \frac{e^{-\theta_i}\theta_i^{y_i}}{y_i!} = \exp\left(-\sum_i \theta_i + \sum_i y_i \log\theta_i\right)\left(\prod_{i=1}^{n} y_i^{-1}!\right)$$

$$\log L(\theta) = -\sum_{i=1}^{n}\theta_i + \sum_{i=1}^{n} y_i \log\theta_i - \text{constant}$$

$$= -\sum_{i=1}^{n} e^{\eta_i} + \sum_{i=1}^{n} y_i \eta_i - C$$

$$= -\sum_{i=1}^{n} e^{x_i^T\beta} + \sum_{i=1}^{n} y_i x_i^T\beta - C$$

$$\frac{\partial \log L(\theta)}{\partial \beta_i} = -\sum_{i=1}^{n} x_i^T e^{x_i^T\beta} + \sum_{i=1}^{n} x_i^T y$$

$$= \sum_{i=1}^{n} x_i^T (y_i - e^{x_i^T\beta})$$

$$= X^T(Y - \theta)$$

$$\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} = \begin{bmatrix} e^{x_1^T\beta} \\ \vdots \\ e^{x_n^T\beta} \end{bmatrix}$$ ← Corresponding to mean of poisson distribution

$$\frac{\partial^2 \log L(\theta)}{\partial \beta_i^2} = -\sum_{i=1}^{n} x_i^T e^{x_i^T\beta} x_i$$

$$= -X^T W X$$

이 때 $W = \begin{pmatrix} \exp(x_1^T\beta) & & 0 \\ & \ddots & \\ 0 & & \exp(x_n^T\beta) \end{pmatrix}$ ⇒ Corresponding to mean of poisson distribution

$$\text{IRLS} : \beta^{(t+1)} = \beta^{(t)} + \underbrace{(X^T W^{(t)} X)^{-1}}_{\{-\nabla^2 \ell(\beta)\}^{-1}} \underbrace{X^T(Y - \theta^{(t)})}_{\nabla \ell(\beta)}$$