

A, D, G Optimal

Fisher Information

Score statistics = $\log L(\theta|X)'$

$E(\log L(\theta|X)') = 0$

“Score의 표본평균이 0으로부터 심각하게 벗어나면 우리가 생각하는 값은 θ 가 아니다” 라고 할 수 있다.

Score statistics의 분산 : $\text{var}(Y) = E(Y^2) - [E(Y)]^2$

Score statistics의 기대값은 0이므로 $\text{var}(Y) = E(Y^2)$ 가 된다.

-> $\log L(\theta|X)'$ 의 분산 = $E(\log L(\theta|X)'^2) = -E(\log L(\theta|X)'')$

-> 이것을 Fisher Information이라고 한다.

Fisher Information

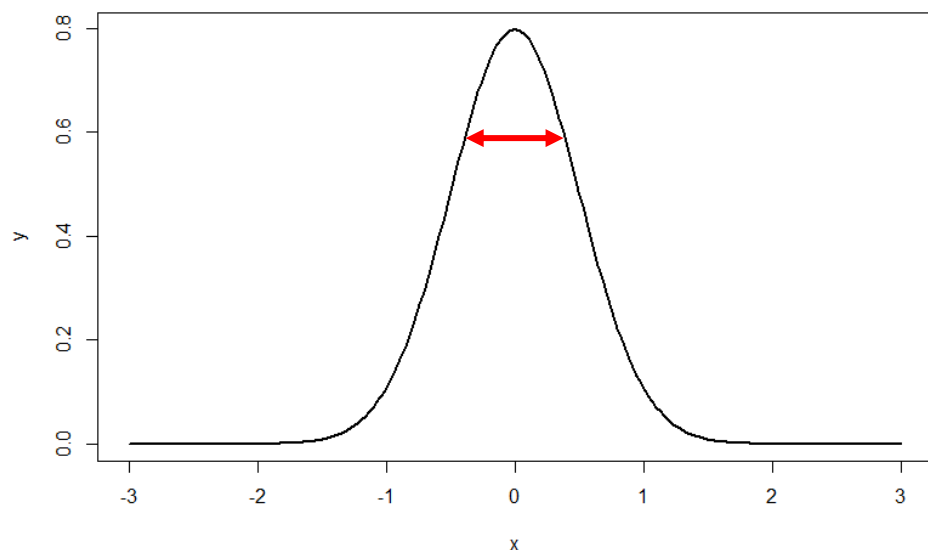
Fisher Information : $-\mathbf{E}(\log L(\theta|X)'')$

기울기 2와 -2의 크기는 같다 방향만 다를 뿐...

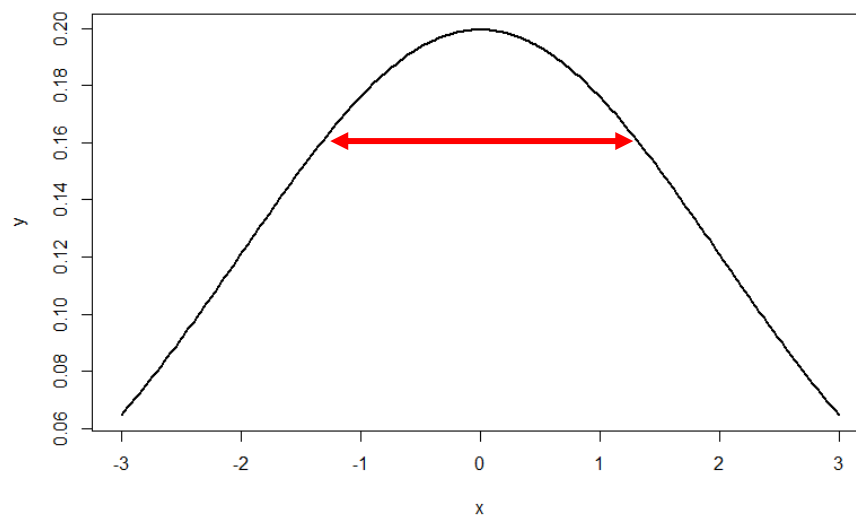
Fisher Information이 크다는 것은 Likelihood함수의 기울기에 대한 변화속도가 커진다는 것을 의미
즉 Fisher Information이 커진다는 것은 Likelihood 함수의 기울기가 가파르다는 것을 의미한다.

기울기가 가파를 수록 분산은 작아진다.

<기울기가 경사진 분포>



<기울기가 완만한 분포>



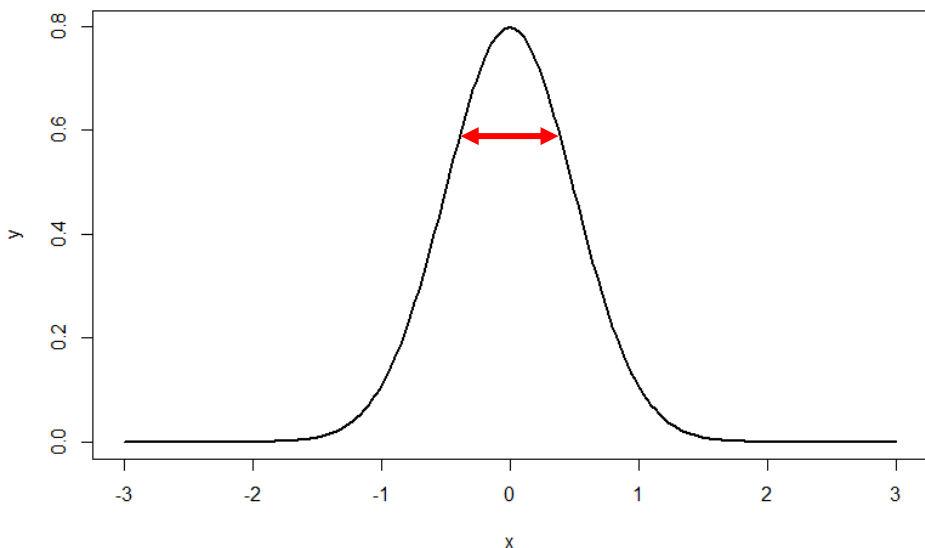
Fisher Information

Fisher Information : $-\mathbf{E}(\log L(\theta|X)'')$

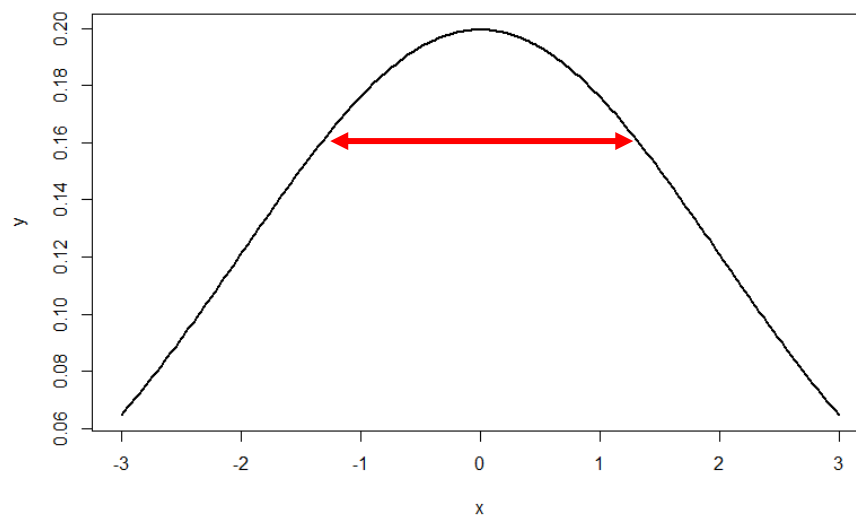
Fisher Information은 특정 관측치 X 가 모수일 확률의 평균을 의미하며,
분산이 작을수록 모수에 대해 더 정확한 정보를 얻을 수 있다는 것이다.

즉 **Fisher Information이 큰 방향** 혹은 **분산이 작은 방향**으로 모수를 추정할 경우 더 정확하게 할 수 있다는 것을 의미한다.

<기울기가 경사진 분포>



<기울기가 완만한 분포>



A-optimal

A-optimal

$$\min \sum \text{tr}(|(X'X)^{-1}|) \quad \text{tr: 대각합(trace)}$$

A optimal은 회귀계수의 분산들만 다룬다.

$|(X'X)^{-1}|$ 의 주대각 요소의 합이 최소가 되는 설계다.

-> 회귀계수들의 분산의 합을 최소로 한다.

회귀계수의 분산

$$\text{var}(\hat{\beta}) = (X'X)^{-1} s^2$$

$(X'X)^{-1}$ 이 최소화 된다는 것은 회귀계수의 분산이 최소화 된다는 것을 의미한다.

회귀계수의 분산을 줄여서 조금이라도 더 정확한 모델을 만든다는 사상을 가지고 있다.

A-optimal

$(X'X)^{-1}$ with Airquality data

	cons	Solar.R	wind	Temp
cons	1.1847784925	5.969440e-05	-2.329795e-02	-1.227913e-02
Solar.R	0.0000596944	1.198358e-06	-7.763005e-07	-3.514939e-06
wind	-0.0232979458	-7.763005e-07	9.545821e-04	1.793637e-04
Temp	-0.0122791315	-3.514939e-06	1.793637e-04	1.432766e-04

$tr((X'X)^{-1})$ with Airquality data

cons	Solar.R	wind	Temp
1.184778e+00	1.198358e-06	9.545821e-04	1.432766e-04

A-optimal은 $\hat{\beta}$ 의 분산의 합을 최소화 하겠다는 것을 의미

D-optimal

D-optimal

$$\min |(X'X)^{-1}| \rightarrow \min \det((X'X)^{-1})$$

Determinant는 벡터의 넓이, 차원이 더 높을 경우 부피를 의미

넓이 혹은 부피를 줄인다는 것은 분산을 최소화 해서 불확실성을 최소화 한다는 것을 의미

$(X'X)^{-1}$ with Airquality data

	cons	solar.R	wind	Temp
cons	1.1847784925	5.969440e-05	-2.329795e-02	-1.227913e-02
solar.R	0.0000596944	1.198358e-06	-7.763005e-07	-3.514939e-06
wind	-0.0232979458	-7.763005e-07	9.545821e-04	1.793637e-04
Temp	-0.0122791315	-3.514939e-06	1.793637e-04	1.432766e-04

G-optimal

G-optimal

$$\min \max(\text{tr}(\text{hatmatrix})) \longrightarrow \min \max(\text{tr}(\text{var}(\hat{y})))$$

G-optimal은 \hat{y} 의 최대 분산을 최소화 하는 설계 기법이다.

Hatmatrix

$$\hat{y} = \text{Hatmatrix} * y$$

Hatmatrix는 \hat{y} 을 찾기 위한 개별 관측치(y)의 정보를 나타낸다.

Hatmatrix는 \hat{y} 의 정보라고 할 수 있으며 정보는 곧 분산을 말한다.

즉 Hatmatrix는 \hat{y} 의 분산을 의미한다.

G-optimal

Hatmatrix with Airquality data

	1	2	3	4	7	8
1	0.042135255	0.026094178	0.003408967	0.04001773	0.04596746	0.02616731
2	0.026094178	0.023864961	0.007117039	0.01201403	0.01823166	0.02076589
3	0.003408967	0.007117039	0.014935842	0.00695647	0.00210577	0.01788822
4	0.040017733	0.012014031	0.006956470	0.07184509	0.06581747	0.03133104
7	0.045967461	0.018231662	0.002105770	0.06581747	0.06645355	0.02568741
8	0.026167307	0.020765890	0.017888221	0.03133104	0.02568741	0.04581198

tr(Hatmatrix) with Airquality data

	1	2	3	4	7	8
	0.04213526	0.02386496	0.01493584	0.07184509	0.06645355	0.04581198

G-optimal은 \hat{y} 의 최대 분산을 최소화 하겠다는 것을 의미

Thank You
