Park	Seonches	, park	
Seoncheol Park			
Seoncheol Park			
, park			
cheol Park			
Seoncheol Park			
1/2			
onche abeol Pa	Computational Satis	stics cheol Paris	
Seonche	Seoncheol Park		
sheol Park			
Seoncheol F	Seoncheol Park		
r park	2024-06-01		
Seoncheol Park	2024-00-01		
Seonche			
sheol Park			
Seonche			
Seon			
Seoncheol Park			
500			
i park			
Seoncheol Park			
ark			
Seoncheol Park			
Seoffer			
l Park			
Seoncheol Park			
ieol Park			
leor .			
Seoncheol Park			
ncheol Park			
Seoncheol Park			
Seoner			
eoncheol Park			
Seoncheol Pa			
_ park			
Seoncheol Park			
Seoncheol	1 La.	2eo	ncheol

Park	Seonche	Loark	
Seoncheol Park			
Seoncheol P			
Seofici			
cheol Park			
Table of	contents		
Seon	Seoncheol Pal		
oncheol Park			
seonch			
, park			
Seoncheol Park			
Se0			
al Park			
Seoncheol Park			
baol Park			
Seoncheol Park			
Seoncheol Park			
Seonche			
Seoncheol Park			
Seon Seon			
Seoncheol Park			
l Park			
(Par.			
Seoncheol Par			
leol Park			
1601,			
Seoncheol ¹			
	Seonche 2		
	oncheol Park		
Se	Our	nork	-ncheol*

Park		Seouches	, park	
1				
1 -				
cheol Park				
	Preface of Park			
oncheol Park	This is a Quarto book.			
	sheol Park	to books visit https://g	Seonoraldoss/books	
. 172	To learn more about Quar	to books visit https://q	uarto.org/docs/books.	
Seoncheol Pa	1 + 1			
,	[1] 2			
Seoncheol				
Seoncheol Park				
Seor				
ark				
l Park				
, park				
ieol Park				
ncheol Park				
ncheoli				
		Seonche al Pa		
	Seoncheol Pa		Sec	-ncheol

Park	Seonches	, park	
Seoncheol F			
Seonches			
olPark			
Seonche			
Seolici			
cheol Park			
conc			
Seon			
oncheol Park			
CP(
, park			
Seoncheol Park			
	Seoncheol Park	oncheol Park	
Park	Se(ncheol Park	
Seoncheol Park	ı park		
	Seoncheol Park In	tro	
Seoncheol Park		Ero Seoncheol Park	
Seoncheo			
Seoncheol Park			
Seonch			
Seoncheol Park			
Seom			
Seoncheol Pa			
l Park			
(Pai			
Seoncheol			
leol Park			
601,			
Seonch			
ncheol Park			
ucheo.			
		4	
Seoncheol Park			
Seolicina	Seoncheol Park		
	Seonchic	nork	-nche ^C

1 Introduction This is a book created from markdown and executable code. See Knuth (1984) for additional discussion of literate programming. [1] 2 Seoncheol Park Seonchegl Park

Park	Seonches	, park	
Seoncheol Par			
Seonches			
olPark			
Seoncheol '			
Seonche			
cheol Park			
Seonch ^e			
Seonch			
oncheol Park			
.0110			
Seon			
Seoncheol Park			
Seom	eoncheol Park Pa	Seonch	
Se	Pa	art II _{Park}	
Seoncheol Park			
Seom	Ontimizati	ion Mothods	
1.	Seopumizau	ion Methods	
Seoncheol Park		Seoncheol Park	
Seons			
Seoncheol Park			
Seom			
l.			
Seoncheol Park			
ark Seo.			
Seoncheol Park			
l Park			
[Pair			
Seoncheol Pa			
, park			
ieol Park			
Seonche ^o			
, park			
ncheol Park			
seonch			
		ne6l Park	
eoncheol Park		Seonche	
	Seoncheol Park		ncheol
		nork.	-ncheu'

2 최적화 방법

2.1 Optimization

-Seoncheol Park 통계에서 최적화를 해야 하는 상황

• Maximum likelihood $\max_{\theta} L(\theta|m{y})$

2.2 Maximum Likelihood Theory Seoncheol Pa

• For independent data: likelihood function park

Seoncheol Park
$$L(\theta) = \prod_{i=1}^n f(\boldsymbol{x}_i|\theta)$$

Seoncheol • Log-likelihood function

$$\ell(\theta) = \sum_{i=1}^n \log(f(\boldsymbol{x}_i; \theta))$$

- Maximum likelihood estimate: $\hat{\theta}_{\mathrm{ML}} = \mathrm{argmax}_{\theta} L(\theta)$
- For smooth likelihoods, necessary requirement:

$$m{s}(m{ heta}) \equiv \ell'(m{ heta}), \quad |m{ heta}|$$
 equations, called score vector $m{J}(m{ heta}) \equiv -\ell''(m{ heta}), \quad$ positive (definite), called observed Fisher information

- Theory:
 - $E[\boldsymbol{s}(\boldsymbol{\theta})] = 0$
 - $-I(\boldsymbol{\theta}) \equiv -E[\ell''(\boldsymbol{\theta})] = E[\boldsymbol{J}(\boldsymbol{\theta})] = Var[\boldsymbol{s}(\boldsymbol{\theta})],$ expected Fisher information Seoncheol Park
 - For large n (and some regularity assumptions)

$$\hat{\pmb{\theta}}_{\rm MI} \approx \mathcal{N}(\pmb{\theta}, \pmb{I}^{-1}(\hat{\pmb{\theta}}_{\rm ML})) \approx \mathcal{N}(\pmb{\theta}, \pmb{J}^{-1}(\hat{\pmb{\theta}}_{\rm ML}))$$

2.3 Newton의 방법

- **Q**. 뉴턴법은 언제 쓰는가?
 - f(x)=0의 해를 근사적으로 구할 때
 - g(x) = h(x)인 x를 근사적으로 구할 때
 - f(x)의 최소값 또는 최대값을 구할 때
- Q. 뉴턴법의 한계
 - 해가 여러개 일 경우
 - 수렴속도가 초깃값에 따라 달라짐
- **Q**. 통계에서의 뉴턴법? ML estimation 1차원에서의 ML 추정을 생각해보자.

$$\operatorname{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta} | \boldsymbol{y}) = \operatorname{argmax}_{\boldsymbol{\theta}} \underbrace{\log L(\boldsymbol{\theta} | \boldsymbol{y})}_{\boldsymbol{\ell}(\boldsymbol{\theta} | \boldsymbol{y})}$$

이것을 $heta^*$ 근처에서 테일러 근사로 전개해보자.

$$\begin{split} \ell(\theta) &\approx \ell(\theta^*) + (\theta - \theta^*)\ell'(\theta^*) + \frac{1}{2}(\theta - \theta^*)^2\ell''(\theta^*) \\ &= \ell(\theta^*) + (\theta - \theta^*)s(\theta^*) - \frac{1}{2}(\theta - \theta^*)^2J(\theta^*) \end{split}$$

즉 score 함수는 $s(\theta) = \ell'(\theta)$, observed information은 $J(\theta) = -\ell''(\theta)$ 이다.

Seo • Solving the maximum of the approximation

$$\theta = \theta^* + \frac{s(\theta^*)}{J(\theta^*)} = \theta^* - \frac{\ell'(\theta^*)}{l''(\theta^*)}$$

2.3.1 Stoping criteria

- 반복법에서는 어느 시점에 어떤 기준을 가지고 멈춰야 할지를 정하는 것이 중요하다.
- 1. Absolute convergence: 이는 x가 클 경우에 시간이 오래 걸릴 수 있다.

$$|x^{(t+1)} - x^{(t)}| < \varepsilon \text{ or } \|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\| < \varepsilon$$

2. Relative convergence: 이는 $|x^{(t)}|$ 가 작을 경우에 불안정할 수 있다.

$$\frac{|x^{(t+1)}-x^{(t)}|}{|x^{(t)}|} < \varepsilon \text{ or } \frac{\|x^{(t+1)}-x^{(t)}\|}{\|x^{(t)}\|} < \varepsilon$$

3. After N iterations (additional criteria) seoncheol Park

2.3.2 Log likelihood의 score와 observed information

Seoncheol Park Likelihood fct (σ is known)

Fct (
$$\sigma$$
 is known)
$$L(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right\}$$

• Log-likelihood Seoncheol Park

$$\ell(\mu) = \sum_{i=1}^n -\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma^2 - \frac{1}{2}\Big(\frac{x_i - \mu}{\sigma}\Big)^2$$

• Score fct

• Score fct
$$s(\mu) = \ell'(\mu) = \sum_{i=1}^n -0 - 0 - \frac{x_i - \mu}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)$$
 Seen check Park

Information

$$Seoncheol Park J(\mu) = -\ell''(\mu) = -s'(\mu) = -\frac{1}{\sigma^2} \sum_{i=1}^n \binom{\text{heal Park}}{-1} = \frac{n}{\sigma^2}$$

2.3.3 Multidimensional extension

• Consider log likelihood:

$$\mathrm{argmax}_{\pmb{\theta}} \ell(\pmb{\theta}) = \mathrm{argmax}_{\pmb{\theta}} L(\pmb{\theta}|\pmb{y}) = \mathrm{argmax}_{\pmb{\theta}} \log(L(\pmb{\theta}|\pmb{y}))$$

• 그러면 다차원에서 테일러 근사는 다음과 같다.

$$\ell(\boldsymbol{\theta}) \approx \ell(\boldsymbol{\theta}^*) + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T \ell'(\boldsymbol{\theta}^*) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T \boldsymbol{H}(\boldsymbol{\theta}^*) (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

$$\approx \ell(\boldsymbol{\theta}^*) + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T \boldsymbol{s}(\boldsymbol{\theta}^*) - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T \boldsymbol{J}(\boldsymbol{\theta}^*) (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$
 efunction:

• Score function:

seoncheol Park
$$s(\boldsymbol{\theta}) = \nabla \ell(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\theta})$$

Seoncheol Par Observed information:

$$\mathbf{J}(\boldsymbol{\theta}) = -\nabla^2 \ell(\boldsymbol{\theta}) = \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ell(\boldsymbol{\theta})$$

Seoncheol Park • 여기서는 반복법을 이용해 다음의 근사를 풀어 최대화한다.

$$\boldsymbol{\theta} = \boldsymbol{\theta}^* + \boldsymbol{J}(\boldsymbol{\theta}^*)^{-1}\boldsymbol{s}(\boldsymbol{\theta}^*) = \boldsymbol{\theta}^* - \boldsymbol{H}(\boldsymbol{\theta}^*)^{-1}\nabla\ell(\boldsymbol{\theta}^*)$$

Seoncheol Park 2.3.4 \mathbb{R}^p 에서의 \log likelihood의 score 와 observed information

• Likelihood: Σ 가 알려져 있다고 할 때

• Likelihood:
$$\Sigma$$
가 알려져 있다고 할 때
$$L(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{(2\pi)^p|\Sigma|}} \exp\Big\{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)\Big\}$$
 Seen the of Park

Log-likelihood

$$\ell(\mu) = \sum_{i=1}^n -\frac{p}{2} \log 2\pi - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$$