Computational Satistics

Seoncheol Park

2024-06-01

Table of contents

Preface

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

```
1 + 1
```

```
[1] 2
```

Part I

Intro

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

Part II

Optimization Methods

2 최적화 방법

2.1 Optimization

통계에서 최적화를 해야 하는 상황

- Maximum likelihood $\max_\theta L(\theta|\boldsymbol{y})$

2.2 Maximum Likelihood Theory

- For independent data: **likelihood function**

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

- **Log-likelihood function**

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

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

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

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

$$\hat{\boldsymbol{\theta}}_{\text{Ml}} \approx \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}}_{\text{ML}})) \approx \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}}_{\text{ML}}))$$

2.3 Newton의 방법

Q. 뉴턴법은 언제 쓰는가?

- $f(x) = 0$의 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.

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

이것을 θ^* 근처에서 테일러 근사로 전개해보자.

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

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

- Solving the maximum of the approximation

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

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)

2.3.2 Log likelihood의 score와 observed information

- Likelihood fct (σ 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

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

- Score fct

$$s(\mu) = \ell'(\mu) = \sum_{i=1}^{n} -0 - 0 - \frac{x_i - \mu}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$$

- Information

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

2.3.3 Multidimensional extension

- Consider log likelihood:

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

- 그러면 다차원에서 테일러 근사는 다음과 같다.

$$\begin{aligned}\ell(\boldsymbol{\theta}) &\approx \ell(\boldsymbol{\theta}^*) + (\boldsymbol{\theta}-\boldsymbol{\theta}^*)^T\ell'(\boldsymbol{\theta}^*) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)^T\boldsymbol{H}(\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\\ &\approx \ell(\boldsymbol{\theta}^*) + (\boldsymbol{\theta}-\boldsymbol{\theta}^*)^T\boldsymbol{s}(\boldsymbol{\theta}^*) - \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)^T\boldsymbol{J}(\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\end{aligned}$$

- Score function:

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

- Observed information:

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

- 여기서는 반복법을 이용해 다음의 근사를 풀어 최대화한다.

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

2.3.4 $\mathbb{R}^p$ 에서의 log likelihood의 score와 observed information

- Likelihood: Σ가 알려져 있다고 할 때

$$L(\mu) = \prod_{i=1}^{n}\frac{1}{\sqrt{(2\pi)^p|\Sigma|}}\exp\left\{-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right\}$$

- Log-likelihood

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