MIT 18.6501: 통계학의 수학적 기초

Fundamentals of Statistics Notes

Contents

Course Structure Current Focus

- Unit 0: Probability Review (Prerequisites)
- **Unit 1: Modeling, Identifiability, Likelihood (현재 단원)**
  - 1.1 Statistical Model
  - 1.2 Identifiability
  - 1.3 Likelihood Function
- Unit 2: Maximum Likelihood Estimation (MLE)
- Unit 3: Method of Moments
- Unit 4: Hypothesis Testing

1 Unit 1. 통계적 모델링과 우도 (Modeling, Identifiability, Likelihood)

지난 시간(*Unit 0*)까지 우리는 확률론의 기초와 큰 수의 법칙(*LLN*), 중심극한 정리(*CLT*)라는 무기를 손에 넣었습니다. 이제 이 무기를 가지고 현실 세계의 데이터를 분석하기 위해, 가장 먼저 해야 할 일은 ***'* 데이터를 담을 수학적 그릇(모델)'**을 만드는 것입니다.

□ 개요 (Overview)

통계학을 제대로 하기 위해서는 현실의 불확실한 데이터를 수학적으로 엄밀하게 정의해야 합니다. 이 단원에서는 데이터를 수학적 집합으로 변환하는 **통계적 모델(Statistical Model)**의 정의, 모델의 파라미터가 유일한지 확인하는 **식별 가능성(Identifiability)**, 그리고 데이터를 통해 파라미터를 추정하기 위한 도구인 **우도(Likelihood)**의 개념을 배웁니다. 이는 추후 배울 '최대우도추정(MLE)'을 위한 필수적인 기초 작업입니다.

□ 핵심 용어 사전

용어 (Term)	직관적 의미 (Meaning)
Sample Space (E)	데이터가 나올 수 있는 모든 후보들의 집합 (무대)
Parameter Space (Θ)	우리가 찾고 싶은 정답(θ)이 숨어있는 범위
Statistical Model	데이터가 발생할 수 있는 확률 규칙들의 모음집
Identifiability	범인의 지문이 유일한가? (파라미터 구별 가능 여부)
Likelihood (L_n)	이 데이터가 관측되었을 때, 범인이 θ일 가능성

1.1 1. 통계적 모델 (Statistical Model)

□ 개념 1: 통계적 모델이란 무엇인가?

한 줄 요약: 현실의 막연한 데이터를 엄밀한 수학적 집합(Set)과 확률(Probability)로 번역하는 과정입니다.

1) 직관적 비유: 맞춤 정장 만들기

현실의 데이터는 '사람의 몸'과 같습니다. 통계적 모델은 이 몸에 딱 맞는 '정장(수트)'을 만드는 과정입니다.

- **표본 공간 (E):** 옷감이 될 수 있는 재료들입니다. (면인가? 실크인가?)
- **파라미터 공간 (Θ):** 옷의 치수 조절 범위입니다. (허리 20 40인치 사이)
- **확률 분포 족 ($\{P_\theta\}$):** 치수(θ)에 따라 만들어지는 다양한 옷 디자인들입니다.

우리는 이 중에서 '내 몸(데이터)'에 가장 잘 맞는 옷(θ)을 찾고 싶은 것입니다.

2) 기술적 정의 및 수학적 구조

통계적 모델은 쌍(Pair) $(E, \{P_\theta\}_{\theta\in\Theta})$으로 정의됩니다.

1. **Sample Space (E):** 관측 데이터 X가 취할 수 있는 값들의 집합.
2. **Parameter Space (Θ):** 미지수 θ가 속한 집합. $\Theta \subseteq \mathbb{R}^k$.
3. **Probability Family (P_θ):** θ값 하나가 정해지면, 그에 대응하는 확률 분포 하나가 결정됩니다.

3) 구체적 예시: 동전 던지기 (베르누이 모델)

우리가 앞면(1), 뒷면(0)이 나오는 동전을 던진다고 합시다.

- $E = \{0, 1\}$ (동전은 0 아니면 1만 나옵니다.)
- $\Theta = [0, 1]$ (앞면이 나올 확률 p는 0과 1 사이입니다.)
- $\{P_\theta\}$ = 베르누이 분포 $\mathrm{Ber}(p)$ 들의 집합.

오해 방지: 잘 정의된 모델(Well-specified)이란?

"모델을 세운다"는 것은 가정을 한다는 뜻입니다. 만약 실제 데이터는 '주사위 (1 6)'인데, 모델을 '동전(0,1)'으로 세우면 어떻게 될까요?

- 실제 데이터 생성 원리 P가 우리 모델 집합 $\{P_\theta\}$ 안에 없을 때, 이를 **Mis-specified Model**이라고 합니다.
- 반대로, 실제 자연의 법칙 P가 우리 모델 안에 포함되어 있다면 **Well-specified Model**이라고 합니다.

—

1.2 2. 식별 가능성 (Identifiability)

□ 개념 2: 이 모델로 정답을 찾을 수 있는가?

한 줄 요약: 서로 다른 원인(파라미터)이 서로 다른 결과(분포)를 만들어야만, 결과를 보고 원인을 역추적할 수 있습니다.

1) 직관적 비유: 범인의 지문

범죄 현장에서 지문을 채취했습니다(데이터).

- **식별 가능:** 모든 사람의 지문이 다르다면, 지문을 보고 범인을 특정할 수 있습니다.
- **식별 불가능:** 만약 철수와 영희의 지문이 똑같이 생겼다면? 지문을 확보해도 둘 중 누가 범인인지 알 수 없습니다.

2) 기술적 정의

함수 $\theta \mapsto P_\theta$ 가 **단사함수(Injective)** 여야 합니다. 즉,

$$\theta \neq \theta' \implies P_\theta \neq P_{\theta'}$$

만약 서로 다른 θ가 같은 P_θ를 만든다면, 데이터가 무한히 많아도 θ를 추정할 수 없습니다.

3) 숫자 예시: 식별 불가능한 경우 (Non-identifiable)

어떤 데이터 X가 평균이 $\mu_1 + \mu_2$인 정규분포를 따른다고 가정해 봅시다.

$$X \sim \mathcal{N}(\mu_1 + \mu_2, 1)$$

우리가 데이터에서 평균이 5라는 사실을 알아냈습니다.

- 경우 A: $\mu_1 = 2, \mu_2 = 3 \to$ 합 5
- 경우 B: $\mu_1 = 1, \mu_2 = 4 \to$ 합 5

μ_1과 μ_2의 조합은 무수히 많습니다. 데이터만으로는 절대 진짜 (μ_1, μ_2)를 찾을 수 없습니다. 이것이 **식별 불가능** 입니다.
—

1.3 3. 우도 (Likelihood)

□ 개념 3: 관점의 대전환

한 줄 요약: "데이터가 주어졌을 때, 이 파라미터가 정답일 점수는 몇 점인가?"를 계산하는 함수입니다.

1) 직관적 비유: 명탐정 코난

- **확률(PDF):** 범인(θ)이 정해져 있을 때, 어떤 증거(x)를 남길지 예측하는 것. (미래 예측)
- **우도(Likelihood):** 증거(x)가 이미 확보되었을 때, 누가 범인(θ)일지 추리하는 것. (과거 추론)

2) 기술적 정의: i.i.d와 결합 확률

데이터 $X_1, ..., X_n$이 서로 독립(Independent)이고 같은 분포(Identically Distributed)를 따른다고 가정합니다. 우도 함수 $L_n(\theta)$는 결합 확률 밀도 함수(Joint PDF)와 식은 같지만, **주인공(변수)** 이 다릅니다.

$$L_n(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

여기서 X_i는 이미 관측된 고정값(상수)이고, θ가 변수입니다.

3) 계산 예시: 불량품 찾기

공장에서 부품을 3개 뽑았는데 (양품, 양품, 불량)이 나왔습니다. 양품확률을 p라고 합시다. (1=양품, 0=불량) 데이터: $X_1 = 1, X_2 = 1, X_3 = 0$.

- **가설 A ($p = 0.5$):** 공장이 반반 확률로 만듦.

$$L_3(0.5) = 0.5 \times 0.5 \times (1 - 0.5) = 0.125$$

- **가설 B ($p = 0.9$):** 공장이 90% 확률로 잘 만듦.

$$L_3(0.9) = 0.9 \times 0.9 \times (1 - 0.9) = 0.081$$

결과: 이 데이터(양,양,불) 기준으로는 가설 A($p = 0.5$)의 우도(0.125)가 더 높습니다. 즉, $p = 0.5$일 가능성이 더 높다고 추론할 수 있습니다.

2 로그 우도 (Log-Likelihood)와 실전 적용

왜 로그를 취하는가?

우도 L_n은 확률의 곱셈($\prod$) 입니다. n이 커지면 값이 0에 너무 가깝게 작아져서 컴퓨터가 계산하지 못합니다(Underflow). 따라서 로그를 취해 **덧셈($\sum$)**으로 바꿉니다. 로그는 단조증가 함수이므로 최대값의 위치는 변하지 않습니다.

$$\ell_n(\theta) = \sum_{i=1}^{n} \log f(X_i; \theta)$$

실전 시나리오: 사용자 체류 시간 분석

당신이 앱 서비스 기획자라고 가정합시다. 사용자의 체류 시간(T)이 지수 분포 $f(t; \lambda) = \lambda e^{-\lambda t}$를 따른다고 모델링했습니다.

1. **데이터 수집:** 3명의 유저가 각각 2분, 3분, 5분을 머물렀습니다. ($x_1 = 2, x_2 = 3, x_3 = 5$)
2. **우도 함수 구성:**

$$L(\lambda) = (\lambda e^{-2\lambda}) \times (\lambda e^{-3\lambda}) \times (\lambda e^{-5\lambda}) = \lambda^3 e^{-10\lambda}$$

3. **로그 우도:**

$$\ell(\lambda) = \log(\lambda^3 e^{-10\lambda}) = 3 \log \lambda - 10\lambda$$

4. **최적화:** 이것을 미분해서 0이 되는 λ를 찾으면, 그것이 바로 가장 그럴듯한 (Likely) 파라미터입니다.

3 자주 묻는 질문 (FAQ)

Q1. 확률(Probability)과 우도(Likelihood)는 같은 거 아닌가요? A. 아니요, 정반대 입니다!

- 확률: θ가 고정 → 데이터 x가 변수. (적분하면 1)
- 우도: 데이터 x가 고정 → 파라미터 θ가 변수. (적분해도 1이 아님)

Q2. 왜 식별 가능성(Identifiability)을 먼저 확인하나요? A. 식별 불가능한 모델을 가지고 우도 계산을 하는 것은, 답이 없는 문제를 열심히 푸는 것과 같습니다. 열심히 계산해서 우도가 최대인 지점을 찾아도, 그 θ가 유일한 정답이라고 확신할 수 없기 때문입니다.

Next Step: 우리는 이제 '우도 함수'라는 강력한 점수판을 만들었습니다. 다음 단원인 **Unit 2**에서는 미분(Calculus)을 사용하여 이 우도 함수를 **최대화(Maximum)** 시키는 θ, 즉 **MLE(Maximum Likelihood Estimator)** 를 구하는 법을 본격적으로 배웁니다.

Unit 1 핵심 요약

- **통계적 모델:** 데이터를 설명하기 위한 수학적 가정 $(E, \{P_\theta\})$.
- **식별 가능성:** 파라미터가 다르면 분포도 달라야 한다 ($\theta \neq \theta' \implies P_\theta \neq P_{\theta'}$). 이것이 보장되어야 추정이 가능하다.
- **우도(Likelihood):** 관측된 데이터 X를 고정하고, 파라미터 θ를 변화시키며 '가능성'을 측정하는 함수.
- **로그 우도:** 계산의 편의성과 미분을 위해 우도에 로그를 취한 형태 ($\sum \log f$).