
Likelihood-Based Approaches to the Analysis of Data with Missing Values (1)

ESC 2024 Summer Session 4차
3조 신재우 양희원 이창현



Contents

1. Theory of Inference Based on Likelihood Function

1.1 Review: Likelihood

1.2 Likelihood-Based Inference

1.3 A Generally Flawed Alternative to Maximum Likelihood

1.4 Likelihood Theory for Coarsened Data

2. Factored Likelihood Methods when Missingness Mechanism is Ignorable

2.1 Introduction

2.2 Bivariate Normal Data with One Variable Subject to Missingness: ML Estimation

2.3 Bivariate Normal Monotone Data: Small-Sample Inference

2.4 Monotone Missingness with More Than Two Variables

2.5 Factored Likelihood for Special Non-Monotone Patterns

1

Theory of Inference Based on Likelihood Function

1.1 Review: Likelihood

1.2 Likelihood-Based Inference

1.3 A Generally Flawed Alternative to Maximum Likelihood

1.4 Likelihood Theory for Coarsened Data

1.1 Review

Probability는 어떠한 분포 내에서 데이터가 가질 확률

Likelihood(우도)는 데이터가 어떠한 분포를 따르는 지에 대한 측도

$$L(\theta; \mathbf{X}) = \prod_{i=1}^n f(X_i; \theta), \theta \in \Omega$$

우도함수

The Likelihood function $L_Y(\theta | y)$ is any function of $\theta \in \Omega_\theta$ proportional to $f_Y(y | \theta)$

; for fixed observed y by definition, $L_Y(\theta | y) = 0$ for any $\theta \notin \Omega_\theta$

로그우도함수

The loglikelihood function $l_Y(\theta | y)$ is the natural logarithm of $L_Y(\theta | y)$

MLE

A maximum likelihood (ML) estimate of θ is a value of $\theta \in \Omega_\theta$ that maximizes the likelihood $L_Y(\theta | y)$, or equivalently, the loglikelihood $l_Y(\theta | y)$

우도함수가 **1) 미분가능하고 2) 위로 유계할 경우** θ 에 대해 미분하여 MLE 구함

$$\text{Score Function : } D_l(\theta) = 0, \text{ where } D_l(\theta) = \frac{\partial l_Y(\theta | y)}{\partial \theta}$$

1.1 Review

Example

Univariate Normal Sample with complete data

$$\mathbf{Y} \sim N(\mu, \sigma^2)$$

$$f_Y(y|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2}\right\}$$

$$l_Y(\mu, \sigma^2 | y) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2} \quad \text{고정된 관찰 데이터 } y \text{에 대한 } \theta = (\mu, \sigma^2) \text{의 함수}$$

$$D_l(\sigma^2) = -\frac{n}{2\hat{\sigma}^2} + \frac{n(\bar{y} - \hat{\mu})^2}{2\hat{\sigma}^4} + \frac{(n-1)s^2}{2\hat{\sigma}^4} = 0$$

$$\hat{\mu} = \bar{y}, \hat{\sigma}^2 = (n-1)s^2/n$$

KEYWORDS..

Direct Likelihood

Bayesian Likelihood

Frequentist Likelihood

P-MAR

IGN

CAR

L_{full}

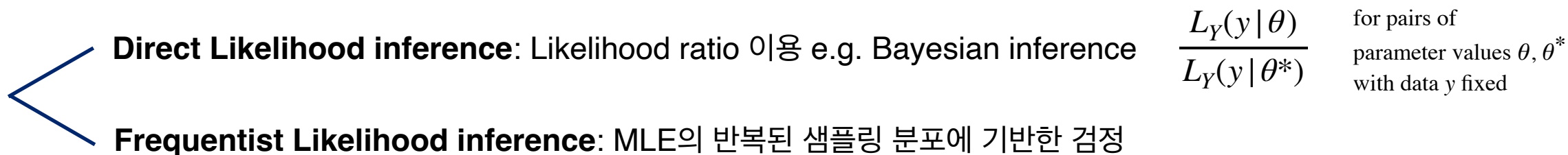
L_{ign}

L_{mispar}

1.2 Likelihood Based Inference

Background

Maximum Likelihood(ML) Inference란 MLE를 구하고 이를 신뢰구간, 베이زي안 사후분포 등을 통해 평가하는 추론 방식



* Bayesian inference에서 사전 분포가 constant할 때 $\hat{\theta}$ 를 사후 분포 $p(\theta|y)$ 의 MLE라고 함

* 충분히 큰 샘플 크기인 경우 두 방법에 큰 차이 없음

Incomplete data에서의 ML inference

Likelihood function 설정 → MLE 구하기 → Posterior Distribution

1) 관찰된 데이터가 일반적으로 iid를 만족하지 않고 2) 결측치가 있는 경우 우도함수가 달라질 수 있다.

1.2 Likelihood Based Inference

Mealli and Rubin (2015) introduced likelihood-based inference

$$Y = (y_{ij}), M = (m_{ij}) \quad i = 1, \dots, n, \quad j = 1, \dots, K, \quad y_{ij} \in \Omega_{ij}$$

Notations

Y : 결측치가 없을 때 $(n \times K)$ 데이터 행렬

M : Binary missingness indicator로 구성된 $(n \times K)$ 행렬

- Fully observed
- y_{ij} 가 결측되었으면 $m_{ij} = 1$, 결측이 아니면 $m_{ij} = 0$
- $m_{ij} = 1$ 의 경우 $y_{ij} = *$ 로 표현, 즉 Ω_{ij} 에서 아무 값이 될 수 있음

Factorization 이용해 Y 와 M 의 **joint pdf**를 다음과 같이 모델링

$$p(Y = y, M = m | \theta, \psi) = f_Y(y | \theta) f_{M|Y}(m | y, \psi)$$

* θ 는 데이터 모델에 대한 모수, ψ 는 missingness mechanism 모델에 대한 모수

Full Likelihood:

$$L_{full}(\theta, \psi | y_{(0)}, m) = \int f_Y(y_{(0)}, y_{(1)} | \theta) f_{M|Y}(m | y_{(0)}, y_{(1)}, \psi) dy_{(1)}$$

Likelihood of ignoring the missingness mechanism:

$$L_{ign}(\theta | y_{(0)}) = \int f_Y(y_{(0)}, y_{(1)} | \theta) dy_{(1)}$$

→ 똑같이 결측값을 marginalize하지만 missingness mechanism을 무시하기 때문에 θ 에 대한 간단한 함수로 나타낼 수 있음

→ 그러나 M 과 Y 의 joint pdf를 모델링은 어렵기에 주로 L_{ign} 사용

1.2 Likelihood Based Inference

Definition

\tilde{m} 과 $\tilde{y}_{(0)}$ 를 $(m, y_{(0)})$ 에서 샘플링한 하나의 관찰 값이라 하자. Missingness mechanism이 *ignorable*하다는 것은 $m = \tilde{m}, y_{(0)} = \tilde{y}_{(0)}$ 일 때 L_{ign} 에서의 모수 추정값이 $m = \tilde{m}, y_{(0)} = \tilde{y}_{(0)}$ 일 때 L_{full} 에서의 추정값과 같다는 것이다.

Methods

1) Direct likelihood 2) Bayes 3) Frequentist

$$\frac{L_{full}(\theta, \psi | \tilde{y}_{(0)}, \tilde{m})}{L_{full}(\theta^*, \psi | \tilde{y}_{(0)}, \tilde{m})} = \frac{L_{ign}(\theta | \tilde{y}_{(0)})}{L_{ign}(\theta^* | \hat{y}_{(0)})} \text{ for all } \theta, \theta^*, \psi$$

- (a) **Parameter distinctness:** 모수 θ 와 ψ 가 서로의 값에 의존하지 않아야(distinct) 하며 joint space가 각 모수의 공간의 곱으로 구성되어야 한다. $\Omega_{\theta, \psi} = \Omega_{\theta} \times \Omega_{\psi}$
- (b) **Factorization of the full likelihood:** 전체 likelihood는 아래와 같이 분해될 수 있어야 한다.

$$L_{full}(\theta, \psi | \tilde{y}_{(0)}, \tilde{m}) = L_{ign}(\theta | \tilde{y}_{(0)}) \times L_{rest}(\psi | \tilde{y}_{(0)}, \tilde{m}) \text{ for all } \theta, \psi \in \Omega_{\theta, \psi}$$

* (b)가 가능하기 위한 충분조건은 MAR, 즉 결측치의 분포가 관측값과 모수 ψ 에만 의존하고, 결측치 자체에는 의존하지 않는 패턴

$$f(\tilde{y}_{(0)}, \tilde{m} | \theta, \psi) = f_{M|Y}(\tilde{m} | \tilde{y}_{(0)}, \psi) \times \int \underline{f_Y(\tilde{y}_{(0)}, y_{(1)} | \theta) dy_{(1)}} = f_{M|Y}(\tilde{m} | \tilde{y}_{(0)}, \psi) \times \underline{f_Y(\tilde{y}_{(0)} | \theta)}$$

Corollary

데이터가 MAR이고 모수들이 서로 distinct하면 Missing mechanism이 ignorable하다.

1.2 Likelihood Based Inference

Methods	1) Direct likelihood 2) Bayes 3) Frequentist
(A) Bayesian inference under full model	(B) Bayesian inference under ignoring MM
$p(\theta, \psi \mid \tilde{y}_{(0)}, \tilde{m}) \propto p(\theta, \psi) \times L_{full}(\theta, \psi \mid \tilde{y}_{(0)}, \tilde{m})$	$p(\theta \mid \tilde{y}_{(0)}) \propto p(\theta) \times L_{ign}(\theta \mid \tilde{y}_{(0)})$

Definition Bayesian inference에서 missingness mechanism이 ignorable함은 모수 θ 의 (A)에서의 사후 분포 추정값이 (B)에서의 사후 분포 추정값과 같다는 것이다.

(a) **Priori independent parameter**: 사전 분포에서 모수 θ 와 ψ 가 독립이어야 한다.

(b) **Factorization of the full likelihood**

* (b)가 가능하기 위한 충분조건은 MAR

Corollary 데이터가 MAR이고 모수들이 서로 priori independent하면 Missing mechanism이 ignorable하다.

* **Note**: Priori independence 조건은 Direct inference에서의 Distinctness 조건보다 더 엄격한 조건이라 할 수 있다. Distinct한 모수들이 사전 분포에서 서로 독립적이지 않을 수 있다.

1.2 Likelihood Based Inference

Methods

1) Direct likelihood 2) Bayes 3) Frequentist

See $y_{(0)}$

(a) **Parameter distinctness** (b) **Factorization of the full likelihood**이 반복적 샘플링을 통해 관찰된 데이터에 대해 만족해야 한다.

$$L_{full}(\theta, \psi | y_{(0)}, m) = L_{ign}(\theta | y_{(0)}) \times L_{rest}(\psi | y_{(0)}, m) \text{ for all } y_{(0)}, m \text{ and } \theta, \psi \in \Omega_{\theta, \psi}$$

$$L_{full}(\theta, \psi | \tilde{y}_{(0)}, \tilde{m}) = L_{ign}(\theta | \tilde{y}_{(0)}) \times L_{rest}(\psi | \tilde{y}_{(0)}, \tilde{m}) \text{ for all } \theta, \psi \in \Omega_{\theta, \psi} \text{ (Direct Likelihood 추론에서 full likelihood)}$$

* (b)가 가능하기 위한 충분조건은 MAAR (Missing Always At Random)

$$f_{M|Y}(m | y_{(0)}, y_{(1)}, \psi) = f_{M|Y}(m | y_{(0)}, y_{(1)}^*, \psi) \text{ for all } m, y_{(0)}, y_{(1)}, y_{(1)}^*, \psi$$

Corollary

데이터가 MAAR이고 변수들이 서로 priori independent하면 Missing mechanism이 ignorable하다.

1.2 Likelihood Based Inference

Example 1

Bivariate Normal Sample with One Variable Subject to Missingness

Bivariate Normal Sample에서 하나의 변수에 결측치가 발생했을 때를 살펴본다. y_{i1} , y_{i2} 변수가 있을 때 y_{i2} 에서 $i = (r + 1), \dots, n$ 에 결측치가 발생했다고 하자.

$$\begin{bmatrix} y_{11} & y_{12} \\ \vdots & \vdots \\ y_{r1} & y_{r2} \\ y_{r+1,1} & \square \\ \cdot & \square \\ \cdot & \square \\ y_{n1} & \square \end{bmatrix}$$

$$\begin{aligned} l_{ign}(\mu, \Sigma | y_{(0)}) &= \log(L_{ign}(\mu, \Sigma | y_{(0)})) \\ &= -\frac{1}{2}r \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^r (y_i - \mu) \Sigma^{-1} (y_i - \mu)^T - \frac{1}{2}(n-r) \ln \sigma_{11} \\ &\quad - \frac{1}{2} \sum_{i=r+1}^n \frac{(y_{i1} - \mu_1)^2}{\sigma_{11}} \end{aligned}$$

Direct inference가 적절한 추론이 되려면 다음 조건이 필요하다.

1) M 의 조건부 분포(=probability that y_{i2} is missing)가 y_{i2} 값에 의존하지 않는다. 그러나 y_{i1} 값에는 의존할 수 있음.

$$Pr(M = m | y_{i1}, y_{i2}) = Pr(M = m | y_{i1})$$

2) 모수 $\theta = (\mu, \Sigma)$ 가 ψ 와 distinct함

이 조건들이 충족되면 θ 의 MLE $\hat{\theta}$ 를 찾을 수 있다.

Bayesian inference에서는 위의 조건을 만족하고 (μ, Σ, ψ) 의 사전분포가 $p(\mu, \Sigma)p(\psi)$ 형태일 경우 μ 와 Σ 의 joint 사후분포는 $p(\mu, \Sigma) \times \log L_{ign}(\mu, \Sigma | y_{(0)})$ 에 비례한다.

Complete data case

$$\begin{aligned} l(\mu, \Sigma | y) &= \log(L(\mu, \Sigma | y)) \\ &= -\frac{1}{2}n \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^T \Sigma^{-1} (y_i - \mu) \end{aligned}$$

1.2 Likelihood Based Inference

부분 벡터: θ 를 (θ_1, θ_2) 로 작성할 때, 여기서 θ_1 과 θ_2 는 데이터 X 에 대한 모델 구성 요소의 부분 벡터라고 한다.

P-MAR

Full likelihood Function이 아래와 같이 분해될 수 있으면, θ_1 에 대한 Direct likelihood 추론에서 데이터가 부분적으로 MAR, 즉 P-MAR(θ_1)이라고 한다.

$$L_{full}(\theta_1, \theta_2, \psi | \tilde{y}_{(0)}, \tilde{m}) = \underbrace{L_1(\theta_1 | \tilde{y}_{(0)})}_{\text{결측 메커니즘 모델을 포함하지 않음}} \times \underbrace{L_{rest}(\theta_2, \psi | \tilde{y}_{(0)}, \tilde{m})}_{\text{모수 } \theta_1 \text{을 포함하지 않음}} \text{ for all } \theta_1, \theta_2, \psi$$

IGN

데이터가 θ_1 에 대한 Direct likelihood 추론에서 ignorable한 경우 $IGN(\theta_1)$ 는 다음 조건을 만족할 때: 1) 결측 메커니즘이 θ_1 에 대해 부분적으로 무작위 결측(P-MAR(θ_1))이고 2) θ_1 와 (θ_2, ψ) 가 distinct함

$$IGN(\theta_1) \text{인 경우} \quad \frac{L_{full}(\theta_1^*, \theta_2, \psi | \tilde{y}_{(0)}, \tilde{m})}{L_{full}(\theta_1^{**}, \theta_2, \psi | \tilde{y}_{(0)}, \tilde{m})} = \frac{L_1(\theta_1^* | \tilde{y}_{(0)})}{L_1(\theta_1^{**} | \tilde{y}_{(0)})} \times \frac{\cancel{L_{rest}(\theta_2, \psi | \tilde{y}_{(0)}, \tilde{m})}}{\cancel{L_{rest}(\theta_2, \psi | \tilde{y}_{(0)}, \tilde{m})}} = \frac{L_1(\theta_1^* | \tilde{y}_{(0)})}{L_1(\theta_1^{**} | \tilde{y}_{(0)})} \text{ for all } \theta_1^*, \theta_1^{**}, \theta_2, \psi$$

* 베이지 추론에서도 비슷하게 θ_1 과 (θ_2, ψ) 가 priori independent하면 $p(\theta_1 | \tilde{y}_{(0)}) \propto \pi_1(\theta_1) \times L_1(\theta_1 | \tilde{y}_{(0)})$ 성립

* P-MAR 이지만 만약 사전적으로 독립적이지 않다고 해도 결측 메커니즘을 모델링하는 추가 가정을 피하기 위한 방법으로써 어느 정도 유효할 수 있다(Sinha and Ibrahim 2003).

1.2 Likelihood Based Inference

Example 2

Regression where missingness depends on the covariates

공변량에 의존하는 결측치에 대한 회귀 분석을 살펴본다.

Senario:

- (a) (y_i, z_i) , $i = 1, \dots, n$ 은 각각 Y 와 Z 의 랜덤 표본
- (b) 관심사는 Y 를 Z 에 회귀시킨 회귀 모형의 모수 θ_1 , $Y = Z\theta_1 + e$
- (c) $y_{i(1)}, y_{i(0)}, z_{i(1)}, z_{i(0)}$ 는 각각 y_i 와 z_i 의 관측값과 결측값
- (d) $m_i = (m(Y_i), m(Z_i))$ 는 각각 y_i 와 z_i 의 구성 요소 결측 여부

Assumption:

- (a) $i = 1, \dots, r$ 에 대해서는 z_i 가 완전히 관측, y_i 의 적어도 하나의 구성 요소가 관측됨
- (b) $i = r + 1, \dots, n$ 에 대해서는 y_i 가 전부 결측, z_i 의 결측 데이터 패턴은 임의적
- (c) (y_i, z_i, m_i) 는 각 단위에 대해 iid, $(\theta_1, \theta_2, \psi)$ 는 각각 회귀모델, Z 의 분포, 결측 메커니즘의 모수
- (d) 결측 메커니즘은 공변량 Z 에 의존하지만 결과 Y 에는 의존 않음
→ MNAR

$$L(\theta_1, \theta_2, \psi) = L_1(\theta_1 | \tilde{Y}_{(1)}, \tilde{Z}_{(1)}) \times L_{rest}(\theta_2, \psi)$$

$$L_1(\theta_1 | \tilde{Y}_{(1)}, \tilde{Z}_{(1)}) = \prod_{i=1}^r \int f_{Y|Z}(\tilde{y}_{i(0)}, y_{i(1)} | \tilde{z}_i, \theta_1) dy_{i(1)}$$

$$L_{rest}(\theta_2, \psi) = \prod_{i=1}^r f_Z(\tilde{z}_i | \theta_2) f_{M|Z}(\tilde{m}_i | \tilde{z}_i, \psi) \times \prod_{i=r+1}^n f_Z(\tilde{z}_{i(1)}, \tilde{z}_{i(0)} | \theta_2) \int f_{M|Z}(\tilde{m}_i | \tilde{z}_{i(0)}, z_{i(1)}, \psi) dz_{i(1)}$$

→ θ_1 과 (θ_2, ψ) 가 독립적이면 데이터는 θ_1 에 대해 P-MAR하고 IGN

→ 결측 메커니즘 없이 처음 r 개 데이터만으로 유효한 추론 가능

- * (θ_2, ψ) 에 관한 일부 정보 손실 가능성
- * Y 에서 처음 r 개의 데이터에 결측값이 없다면 complete data analysis

1.3 A Generally Flawed Alternative to Maximum Likelihood

Method

결측치를 매개변수로 설정하고 complete data 우도를 최대화하기 위한 결측치와 모수를 찾는 방법

$$L_{mispar}(\theta, y_{(1)} | y_{(0)}) = f_Y(\tilde{y}_{(0)}, y_{(1)} | \theta) \quad \text{결측값을 } y_{(1)}, \text{ 관측값을 } y_{(0)} \text{이라 함}$$

결측 데이터가 MAR이 아니거나 θ 가 ψ 와 distinct하지 않은 경우

$$L_{mispar}(\theta, \psi, y_{(1)} | \tilde{y}_{(0)}, \tilde{m}) = L_{full}(\theta, \psi | \tilde{y}_{(0)}, y_{(1)}, \tilde{m}) = f_Y(\tilde{y}_{(0)}, y_{(1)} | \theta) f_{M|Y}(\tilde{m} | \tilde{y}_{(0)}, y_{(1)}, \psi) \text{ over } (\theta, \psi, y_{(1)})$$

- * Incomplete data를 분석하는 일반적인 방법으로는 적합하지 않음
- * 예외: 샘플 크기가 증가함에 따라 결측 데이터의 비율이 0으로 수렴하는 비대칭적인 상황

Background

고전적 예는 ANOVA에서 결측 플롯을 다루는 방법, 효율적인 계산

$\left\{ \begin{array}{l} L_{full} \text{이나 } L_{ign} \text{는 관측된 데이터를 토대로 } \theta \text{에 대한 알맞은 likelihood를 정의해주는 반면,} \\ L_{mispar} \text{은 } y_{(1)} \text{이 random variable임에도 불구하고 고정된 매개변수로 취급하기 때문에 올바른 likelihood라고 할 수 없다.} \end{array} \right.$

1.3 A Generally Flawed Alternative to Maximum Likelihood

Example

Univariate Normal Samples with Missing Data

Senario:

(a) $y = (y_{(0)}, y_{(1)})$: 평균 μ , 분산 σ^2 인 정규분포에서 추출된 n 개 표본

(b) $y_{(0)}$ 은 r 개의 관측값, $y_{(1)}$ 은 $(n-r)$ 개의 결측값

(c) 결측치들은 MAR 가정

(d) $\theta = (\mu, \sigma^2)$ 은 결측 메커니즘의 모수와 독립 가정

$L_{ign}(\theta | y_{(0)})$ 는 r 개의 관측된 표본(**Complete data**)에 대한 우도

$$\hat{\mu} = \sum_{i=1}^r \frac{y_i}{r} \text{ and } \hat{\sigma}^2 = \sum_{i=1}^r \frac{(y_i - \hat{\mu})^2}{r}$$

Method 적용:

$$L_{mispar}(\mu, \sigma^2, y_{r+1}, \dots, y_n | y_1, \dots, y_r)$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^r \frac{(y_i - \mu)^2}{\sigma^2}\right\} \exp\left\{-\frac{1}{2} \sum_{i=r+1}^n \frac{(y_i - \mu)^2}{\sigma^2}\right\}$$

(c), (d) 시나리오를 만족하지 못하는 경우에는 다음 식에 적용해야 한다.

$$\begin{aligned} L_{mispar}(\theta, \psi, y_{(1)} | \tilde{y}_{(0)}, \tilde{m}) &= L_{full}(\theta, \psi | \tilde{y}_{(0)}, y_{(1)}, \tilde{m}) \\ &= f_Y(\tilde{y}_{(0)}, y_{(1)} | \theta) f_{M|Y}(\tilde{m} | \tilde{y}_{(0)}, y_{(1)}, \psi) \text{ over } (\theta, \psi, y_{(1)}) \end{aligned}$$

모수 $\mu, \sigma^2, y_{r+1}, \dots, y_n$ 추정값:

$$\hat{y}_i = \hat{\mu}, i = r + 1, \dots, n, \quad \underline{\hat{\mu}^* = \hat{\mu}}, \quad \underline{\hat{\sigma}^{*2} = r\hat{\sigma}^2/n}$$

- * 분산의 MLE는 L_{ign} 에서의 MLE에 r/n 을 곱한 값
- * 결측 데이터 비율이 상당히 큰 경우 추정된 분산은 크게 편향되고, $n \rightarrow \infty$ 일 때 편향 r/n 이 1로 가지 않으면 사라지지 않는다.

1.4 Likelihood Theory for Coarsened Data

Coarsened Data란 정확한 데이터 전체가 보이지 않고 정확한 데이터가 포함된 어떤 set만 관찰된 데이터

MAR (Missing At Random)	결측 데이터 패턴이 관측 데이터에만 의존, 누락된 데이터 자체에는 의존하지 않음	
CAR (Coarsened At Random)	코어싱 데이터 패턴이 관측 데이터에만 의존, <u>코어싱되지 않은 데이터 자체에는 의존하지 않음</u> 관측되지 않은 원래 데이터 값	누락(missingness), 그룹화(grouping) 힙핑(heapng), 검열(censoring)

Notations

Specific Case

$Y = \{y_{ij}\}$: coarsing 전의 완전한 데이터 행렬
 $f_Y(y | \theta)$: 완전한 데이터 모델에서의 pdf
 $y_{ij(0)}$: y_{ij} 에서 관측된 데이터이고 y_{ij}, c_{ij} 에 의해 정의됨: $y_{ij(0)} = (y_{ij}, c_{ij})$

 이때 coarsened 부분집합이 unobserved true value를 포함해야 함
 $\rightarrow y_{ij} \in y_{ij(0)}(y_{ij}, c_{ij})$

$C = \{c_{ij}\}$ is the matrix of binary missingness indicators

 $y_{ij(0)} = \begin{cases} \{y_{ij}\}, \text{the set consisting of the single true value,} & \text{if } c_{ij} = 0 \\ \Psi_{ij}, \text{the sample space of } y_{ij}, & \text{if } c_{ij} = 1 \end{cases}$

1.4 Likelihood Theory for Coarsened Data

Example

Censoring with stochastic censoring time

확률적 검열 시간을 가진 검열 문제를 살펴본다.

Y 는 이벤트 발생 시간이고, 일부 값은 관찰되며 일부는 검열된다.

y_i 는 이벤트 발생 시간, c_i 는 검열 시간이다.

완전한 데이터는 $(y_i, c_i), i = 1, \dots, n$ 이며 코어싱 데이터 $y_{(0)i}$ 는

$$y_{(0)i} = y_{(0)i}(y_i, c_i) = \begin{cases} \{y_{ij}\}, & \text{if } y_i \leq c_i \\ (c_i, \infty) & \text{if } y_i > c_i \end{cases}$$

Full likelihood function:

$$\begin{aligned} L_{full}(\theta, \psi | y_{(0)}, c_{(0)}) \\ = \int \int f_{C|Y}(c_{(0)}, c_{(1)} | y_{(0)}, y_{(1)}, \phi) f_Y(y_{(0)}, y_{(1)} | \theta) dy_{(1)} dc_{(1)} \end{aligned}$$

Likelihood function ignoring coarsening mechanism:

$$L_{ign}(\theta | y_{(0)}) = \int f_Y(y_{(0)}, y_{(1)} | \theta) dy_{(1)}$$

데이터가 CAR하다는 것은

$$\begin{aligned} f_{C|Y}(\tilde{c}_{(0)}, c_{(1)} | \tilde{y}_{(0)}, y_{(1)}, \phi) &= f_{C|Y}(\tilde{c}_{(0)}, c_{(1)}^* | \tilde{y}_{(0)}, y_{(1)}^*, \phi) \\ &\text{for all } c_{(1)}, c_{(1)}^*, y_{(0)}, y_{(1)}^*, \phi \end{aligned}$$

Coarsening mechanism이 ignorable할 조건은
Missingness mechanism과 유사하게

- 1) 데이터가 CAR이고
- 2) 모수 θ 와 ϕ 가 독립

2

Factored Likelihood Methods when Missingness Mechanism is Ignorable

2.1 Introduction

2.2 Bivariate Normal Data with One Variable Subject to Missingness : ML Estimation

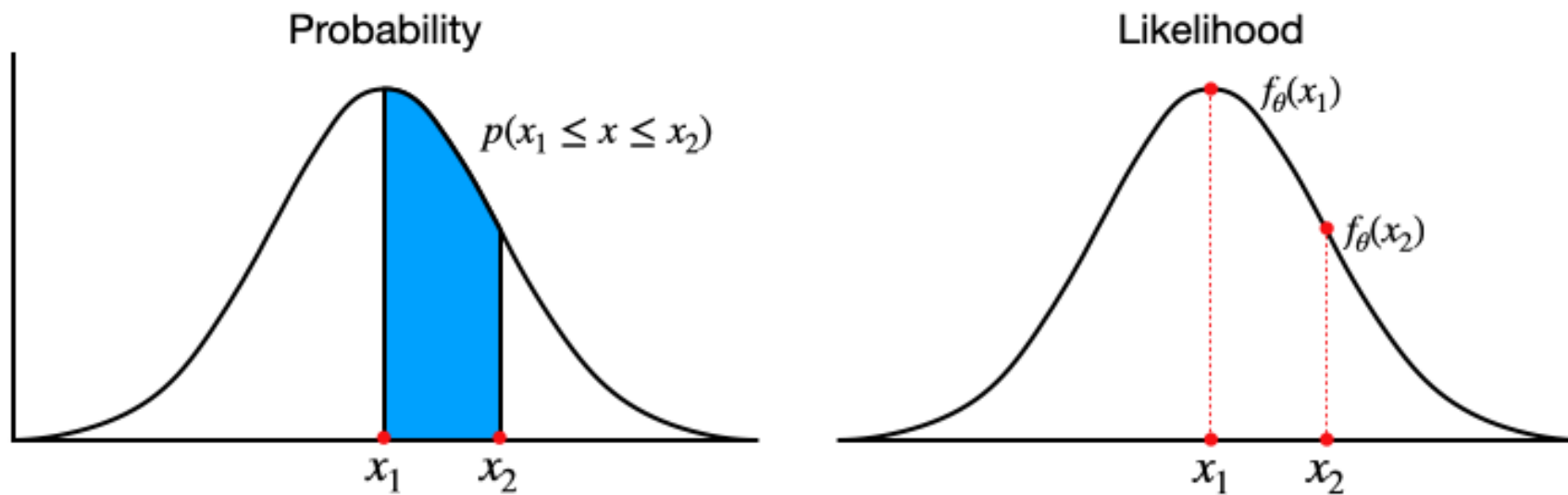
2.3 Bivariate Normal Monotone Data : Small-Sample Inference

2.4 Monotone Missingness with More Than Two Variables

2.5 Factored Likelihoods for Special Non-Monotone Patterns

2.1 Introduction

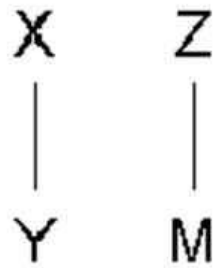
6장 : 우도함수 관련, 어떻게 우리가 미싱 데이터를 우도함수로 표현할 수 있는지



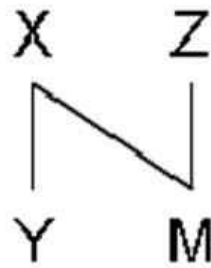
7장 : 만약에 Missingness Mechanism 이 Ignorable 하면 우도함수를 이용해서 어떻게 풀까?

2.1 Introduction

Review on Missingness Mechanism is Ignorable



(a) MCAR



(b) MAR



(c) MNAR

MCAR & MAR : 미싱 여부를 알려주는 변수인 M 이 종속변수인 Y 와 연관성이 없다

MNAR : 미싱 여부를 알려주는 변수인 M 이 종속변수인 Y 와 연관성이 있다

즉 몇몇의 X 에 대한 Y 데이터가 없어도 연관성이 없기에 해당 값들은 Prediction 으로 진행해도 괜찮다는 것!

2.1 Introduction

7장 제목 : Factored Likelihood Methods When the Missingness Mechanism is Ignorable

Factored Likelihood Methods 에 대해서 살펴겠습니다.

우도함수를 구축했는데 굉장히 복잡하게 나올 수도 있지만, 모수들끼리 **독립**이라면 **Factor** 를 시키는 것이 가능합니다.

$$\ell(\phi | y_{(0)}) = \ell_1(\phi_1 | y_{(0)}) + \ell_2(\phi_2 | y_{(0)}) + \dots + \ell_J(\phi_J | y_{(0)})$$

$$\hat{\phi}_1 = \arg \max_{\phi_1} \ell(\phi | y_{(0)}) = \arg \max_{\phi_1} \ell(\phi_1 | y_{(0)})$$

7장은 두 내용만 알면 완벽하게 이해할 수 있습니다.

2.2 Bivariate Normal Case (7.2.1 ML Estimates)

Example 7.1 : Bivariate Normal Sample with One Variable Subject to Non-response

Bivariate Normal 의 경우 데이터가 y_{i1}, y_{i2} 이라면,

$y_{i1}, y_{i2} \sim N_2(\mu, \Sigma)$: 저희의 모수는 μ 와 Σ 입니다.

한개의 Variable 만 Missing 이기 때문에 Set 을 두개로 나누는 것이 가능합니다.

$\{(y_{i1}, y_{i2}), i = 1, \dots, r\}$: Bivariate Unit 중에서 r 개는 둘다 관찰이 된 Set

$\{(y_{i1}), i = r + 1, \dots, n\}$: Bivariate Unit 중에서 $n - r$ 개는 y_{i1} 만 관찰이 된 Set

2.2 Bivariate Normal Case (7.2.1 ML Estimates)

확률밀도함수를 세우면 다음과 같이 세우는 것이 가능합니다.

$$f_Y(y_{i1}, y_{i2} \mid \mu, \Sigma) = f_1(y_{i1} \mid \mu_1, \sigma_{11}) \cdot f_2(y_{i2} \mid y_{i1}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1})$$

근데 이것은 왜 이렇게 유도가 될까요?

Conditional Distribution 정의에 의하면,

$$f(y_{i2} \mid y_{i1}) = \frac{f(y_{i1}, y_{i2})}{f(y_{i1})}, \quad f(y_{i1}, y_{i2}) = f(y_{i1}) \cdot f(y_{i2} \mid y_{i1})$$

이것을 모수까지 포함해서 풀면 위와 같은 식이 나오게 됩니다.

위의 식에 의하면,

- 확률변수 y_{i1} 와 관련된 모수들은 μ_1, σ_{11}
- 확률변수 y_{i2} 와 관련된 모수들은 $y_{i1}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1}$

2.2 Bivariate Normal Case (7.2.1 ML Estimates)

(y_{i1}, y_{i2}) 는 Bivariate Normal Distribution 을 따르기에 다음과 같은 분포로 표현이 가능합니다.

$$y_{i1}, y_{i2} \sim N_2(\mu, \Sigma) \leftrightarrow N_2 \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right)$$

y_{i1} 와 $y_{i2} | y_{i1}$ 의 분포는 다음과 같이 표현할 수 있습니다.

$$y_{i1} \sim N(\mu_1, \sigma_{11}), \quad y_{i2} | y_{i1} \sim N \left(\mu_2 - \frac{\sigma_{21}}{\sigma_{11}} \mu_1 + \frac{\sigma_{21}}{\sigma_{11}} \cdot y_{i1}, \sigma_{22} - \frac{\sigma_{21}^2}{\sigma_{11}} \right)$$

두번째 분포의 모수들을 다음과 같이 표현하겠습니다.

$$\beta_{21.1} = \frac{\sigma_{12}}{\sigma_{11}}, \quad \beta_{20.1} = \mu_2 - \beta_{21.1} \mu_1, \quad \sigma_{22.1} = \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}$$

이전 슬라이드에서의 모수들이 여기에서 나온 겁니다.

$$f_Y(y_{i1}, y_{i2} | \mu, \Sigma) = f_1(y_{i1} | \mu_1, \sigma_{11}) \cdot f_2(y_{i2} | y_{i1}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1})$$

2.2 Bivariate Normal Case (7.2.1 ML Estimates)

기존 모수들을 $\theta = (\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})^\top$ 으로 표현하면,
새로운 모수들을 $\phi = (\mu_1, \sigma_{11}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1})^\top$ 으로 표현할 수 있습니다.

이때 ϕ 를 θ 에 관한 **One-to-one (일대일 대응) Function** 이라고 할 수 있습니다.
즉 새로운 모수들만 구해도 기존 모수들을 구할 수가 있습니다.

$$\beta_{21.1} = \frac{\sigma_{12}}{\sigma_{11}}, \quad \beta_{20.1} = \mu_2 - \beta_{21.1}\mu_1, \quad \sigma_{22.1} = \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}$$

정리하자면, 다음과 같이 표현이 가능합니다.

$$y_{i1} \sim N(\mu_1, \sigma_{11}), \quad y_{i2} \mid y_{i1} \sim N(\beta_{20.1} + \beta_{21.1} \cdot y_{i1}, \sigma_{22.1})$$

나중에 SWEEP Operator 를 배울 때 중요하지만, 회귀식으로 모수들을 표현할 수 있습니다.

Note : y_{i1} 에서의 모수들이 $y_{i2} \mid y_{i1}$ 에서는 사용이 안됩니다.

2.2 Bivariate Normal Case (7.2.1 ML Estimates)

우도함수를 구축하면, 다음과 같이 Factored Likelihood 으로 표현이 가능합니다.

$$\begin{aligned} f(y_{(0)} \mid \theta) &= \prod_{i=1}^r f_Y(y_{i1}, y_{i2} \mid \theta) \prod_{i=r+1}^n f_1(y_{i1} \mid \theta) \\ &= \left[\prod_{i=1}^r f_1(y_{i1} \mid \theta) f_2(y_{i2} \mid y_{i1}, \theta) \right] \left[\prod_{i=r+1}^n f_1(y_{i1} \mid \theta) \right] \\ &= \left[\prod_{i=1}^r f_1(y_{i1} \mid \theta) \right] \left[\prod_{i=1}^r f_2(y_{i2} \mid y_{i1}, \theta) \right] \left[\prod_{i=r+1}^n f_1(y_{i1} \mid \theta) \right] \\ &= \left[\prod_{i=1}^n f_1(y_{i1} \mid \mu_1, \sigma_{11}) \right] \left[\prod_{i=1}^r f_2(y_{i2} \mid y_{i1}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1}) \right] \end{aligned}$$

2.2 Bivariate Normal Case (7.2.1 ML Estimates)

여기에서 첫번째 부분을 최대화 시키면 (μ_1, σ_{11}) 에 대한 MLE 를 구하는 것이 가능합니다.

$$f(y_{(0)} \mid \theta) = \left[\prod_{i=1}^n f_1(y_{i1} \mid \mu_1, \sigma_{11}) \right] \left[\prod_{i=1}^r f_2(y_{i2} \mid y_{i1}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1}) \right]$$

정규분포를 따라가기에 MLE 는 다음과 같이 나옵니다.

$$\hat{\mu}_1 = \frac{1}{n} \cdot \sum_{i=1}^n y_{i1}, \quad \hat{\sigma}_{11} = \frac{1}{n} \sum_{i=1}^n (y_{i1} - \hat{\mu}_1)^2$$

두번째 부분 역시 정규분포를 따라가기에, 각각을 구하면 다음과 같이 나옵니다.

$$\hat{\beta}_{21.1} = \frac{\tilde{\sigma}_{12}}{\tilde{\sigma}_{11}}, \quad \hat{\beta}_{20.1} = \frac{1}{r} \sum_{i=1}^r y_{i2} - \hat{\beta}_{21.1} \cdot \frac{1}{r} \sum_{i=1}^r y_{i1}, \quad \hat{\sigma}_{22.1} = \tilde{\sigma}_{22} - \frac{\tilde{\sigma}_{12}^2}{\tilde{\sigma}_{11}}$$

2.2 Bivariate Normal Case (7.2.1 ML Estimates)

저희가 궁금한 것은 $\phi = (\mu_1, \sigma_{11}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1})^\top$ 가 아니라 $\theta = (\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})^\top$ 입니다.
하지만 One-to-one Function 이기에 $\hat{\phi}$ 를 이용해서 $\hat{\mu}_2, \hat{\sigma}_{22}, \hat{\rho}$ 를 구할 수 있습니다.

$$\hat{\mu}_2 = \bar{y}_2 + \hat{\beta}_{21.1}(\hat{\mu}_1 - \bar{y}_1), \quad \hat{\sigma}_{22} = \tilde{\sigma}_{22} + \hat{\beta}_{21.1}^2(\hat{\sigma}_{11} - \tilde{\sigma}_{11}),$$

$$\hat{\rho} = \tilde{\rho} \left(\frac{\hat{\sigma}_{11}}{\tilde{\sigma}_{11}} \right)^{1/2} \cdot \left(\frac{\tilde{\sigma}_{22}}{\hat{\sigma}_{22}} \right)^{1/2} = \tilde{\sigma}_{12}(\tilde{\sigma}_{11} \cdot \tilde{\sigma}_{22})^{-1/2} \cdot \left(\frac{\hat{\sigma}_{11}}{\tilde{\sigma}_{11}} \right)^{1/2} \cdot \left(\frac{\tilde{\sigma}_{22}}{\hat{\sigma}_{22}} \right)^{1/2}$$

Notation 정리 :

$\hat{\mu}_1$: Mean of y_{i1} for n units, \bar{y}_1 : Mean of y_{i1} for r units, \bar{y}_2 : Mean of y_{i2} for r units

$\hat{\sigma}_{11}$: Sample Variance of y_{i1} for n units, $\tilde{\sigma}_{11}$: Sample Variance of y_{i1} for r units

$\tilde{\sigma}_{22}$: Sample Variance of y_{i2} for r units

2.2 Bivariate Normal Case (7.2.1 ML Estimates)

$\hat{\mu}_2$ 를 정확하게 살피면, 다음과 같이 표현이 가능합니다.

$$\hat{\mu}_2 = \bar{y}_2 + \hat{\beta}_{21.1}(\hat{\mu}_1 - \bar{y}_1) = \frac{1}{n} \left(\sum_{i=1}^r y_{i2} + \sum_{i=r+1}^n \hat{y}_{i2} \right) \quad \hat{y}_{i2} = \bar{y}_2 + \hat{\beta}_{21.1}(y_{i1} - \bar{y}_1)$$

추정량에 대한 유도가 Non-trivial 했기에, $\hat{\mu}_2$ 에 대한 내용을 살피겠습니다.

$$\begin{aligned} \hat{\mu}_2 &= \bar{y}_2 + \hat{\beta}_{21.1}(\hat{\mu}_1 - \bar{y}_1) = \frac{1}{n}(n \cdot \bar{y}_2) + \hat{\beta}_{21.1} \left(\frac{1}{n} \sum_{i=1}^n y_{i1} - \frac{1}{r} \sum_{i=1}^r y_{i1} \right) \\ &= \frac{1}{n}(r \cdot \bar{y}_2 + (n - r) \cdot \bar{y}_2) + \hat{\beta}_{21.1} \left(\frac{1}{n} \left[\sum_{i=1}^n y_{i1} - n \cdot \bar{y}_1 \right] \right) \\ &= \frac{1}{n}(r \cdot \bar{y}_2 + (n - r) \cdot \bar{y}_2) + \hat{\beta}_{21.1} \left(\frac{1}{n} \left[\sum_{i=1}^r y_{i1} + \sum_{i=r+1}^n y_{i1} - (r + (n - r)) \cdot \bar{y}_1 \right] \right) \end{aligned}$$

2.2 Bivariate Normal Case (7.2.1 ML Estimates)

$$\begin{aligned}\hat{\mu}_2 &= \frac{1}{n}(r \cdot \bar{y}_2 + (n - r) \cdot \bar{y}_2) + \hat{\beta}_{21.1} \left(\frac{1}{n} \left[\sum_{i=1}^r y_{i1} - r \cdot \bar{y}_1 + \sum_{i=r+1}^n y_{i1} - (n - r) \cdot \bar{y}_1 \right] \right) \\&= \frac{1}{n}(r \cdot \bar{y}_2 + (n - r) \cdot \bar{y}_2) + \hat{\beta}_{21.1} \left(\frac{1}{n} \left[\sum_{i=r+1}^n y_{i1} - \sum_{i=r+1}^n \bar{y}_1 \right] \right) \\&= \frac{1}{n}(r \cdot \bar{y}_2 + (n - r) \cdot \bar{y}_2) + \hat{\beta}_{21.1} \left(\frac{1}{n} \left[\sum_{i=r+1}^n (y_{i1} - \bar{y}_1) \right] \right) \\&= \frac{1}{n} \left(\sum_{i=1}^r y_{i2} + \sum_{i=r+1}^n \left[\bar{y}_2 + \hat{\beta}_{21.1}(y_{i1} - \bar{y}_1) \right] \right)\end{aligned}$$

2.2 Bivariate Normal Case (7.2.2 Large-Sample Covariance Matrix)

7.2.1 에서 한 것은 추정량을 구하는 것입니다. 하지만 이 추정량이 정확한지에 대해서도 알아야 되며, 그러기 위해서는 분산 값들을 알아야합니다.

즉, 저희는 Covariance Matrix 인 $\text{Cov}(\hat{\theta}) = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^\top]$ 를 구하는 것이 목표입니다.

이 값을 구하기 위해서는 를 구하면 편하게 구할 수가 있습니다.

그러기 위해서는 Log-likelihood 함수를 구해야 하며, 이것은 우도함수에 로그를 취한 것입니다.

$$f(y_{(0)} | \theta) = \left[\prod_{i=1}^n f_1(y_{i1} | \mu_1, \sigma_{11}) \right] \left[\prod_{i=1}^r f_2(y_{i2} | y_{i1}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1}) \right]$$

$$\log(f(y_{(0)} | \theta)) = \left[\sum_{i=1}^n \log(f_1(y_{i1} | \mu_1, \sigma_{11})) \right] \left[\sum_{i=1}^r \log(f_2(y_{i2} | y_{i1}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1})) \right]$$

2.2 Bivariate Normal Case (7.2.2 Large-Sample Covariance Matrix)

Bivariate Normal 한 케이스이기에 각각 정규분포이기에 전체 로그 우도함수를 구축하는 것이 가능합니다.

$$l(\phi \mid y_{(0)}) = -\frac{1}{2\sigma_{11}} \cdot \sum_{i=1}^n (y_{i1} - \mu_1)^2 - \frac{1}{2}n \cdot \log(\sigma_{11}) \\ - \frac{1}{2\sigma_{22 \cdot 1}} \sum_{i=1}^r (y_{i2} - \beta_{20 \cdot 1} - \beta_{21 \cdot 1} y_{i1})^2 - \frac{1}{2}r \cdot \log(\sigma_{22 \cdot 1})$$

이것을 이용해서 정보행렬을 구하면 됩니다. $I(\theta) = -E \left[\frac{\partial^2 l(\theta)}{\partial \theta^2} \right]$

저희는 ϕ 에 대한 추정량들이 있기 때문에 θ 자리에 ϕ 를 대입시키면 됩니다.

이때 **Likelihood** 를 **Factor** 시켜서 모수들을 따로따로 표현하는 것이 가능했기에 정보행렬 역시 다음과 같이 생기게 됩니다.

$$I(\hat{\phi} \mid y_{(0)}) = \begin{bmatrix} I_{11}(\hat{\mu}_1, \hat{\sigma}_{11} \mid y_{(0)}) & 0 \\ 0 & I_{22}(\hat{\beta}_{20 \cdot 1}, \hat{\beta}_{21 \cdot 1}, \hat{\sigma}_{22 \cdot 1} \mid y_{(0)}) \end{bmatrix}$$

2.2 Bivariate Normal Case (7.2.2 Large-Sample Covariance Matrix)

각각의 부분들을 살펴보면, I_{11} 은 다음과 같이 나오게 됩니다.

$$I_{11} (\hat{\mu}_1, \hat{\sigma}_{11} \mid \mathbf{y}_{(0)}) = \begin{bmatrix} n/\hat{\sigma}_{11} & 0 \\ 0 & n/(2\hat{\sigma}_{11}^2) \end{bmatrix}$$

I_{22} 의 경우 모수가 3개이기 때문에 정보행렬도 3×3 행렬입니다.

$$I_{22} (\hat{\beta}_{20.1}, \hat{\beta}_{21.1}, \hat{\sigma}_{22.1} \mid \mathbf{y}_{(0)}) = \begin{bmatrix} r/\hat{\sigma}_{22.1} & r\bar{y}_1/\hat{\sigma}_{22.1} & 0 \\ r\bar{y}_1/\hat{\sigma}_{22.1} & \sum_{i=1}^r y_{i1}^2/\hat{\sigma}_{22.1} & 0 \\ 0 & 0 & r/(2\hat{\sigma}_{22.1}^2) \end{bmatrix}$$

Fisher Information 을 이용해서 다음과 같이 공분산 행렬을 구할 수 있게 됩니다.

$$\text{Cov}(\hat{\phi}) = \begin{bmatrix} I_{11}^{-1} (\hat{\mu}_1, \hat{\sigma}_{11} \mid \mathbf{y}_{(0)}) & 0 \\ 0 & I_{22}^{-1} (\hat{\beta}_{20.1}, \hat{\beta}_{21.1}, \hat{\sigma}_{22.1} \mid \mathbf{y}_{(0)}) \end{bmatrix}$$

2.2 Bivariate Normal Case (7.2.2 Large-Sample Covariance Matrix)

이것을 대입 시키면, 각각의 Element 는 다음과 같이 나옵니다.

$$I_{11}^{-1} (\hat{\mu}_1, \hat{\sigma}_{11} \mid y_{(0)}) = \begin{bmatrix} \hat{\sigma}_{11}/n & 0 \\ 0 & 2\hat{\sigma}_{11}^2/n \end{bmatrix}$$

$$I_{22}^{-1} (\hat{\beta}_{20.1}, \hat{\beta}_{21.1}, \hat{\sigma}_{22.1} \mid y_{(0)}) = \begin{bmatrix} \hat{\sigma}_{22.1} (1 + \bar{y}_1^2/\tilde{\sigma}_{11})/r & -\bar{y}_1 \hat{\sigma}_{22.1}/(r\tilde{\sigma}_{11}) & 0 \\ -\bar{y}_1 \hat{\sigma}_{22.1}/(r\tilde{\sigma}_{11}) & \hat{\sigma}_{22.1}/(rr\tilde{\sigma}_{11}) & 0 \\ 0 & 0 & 2\hat{\sigma}_{22.1}^2/r \end{bmatrix}$$

전체 공분산 행렬은 다음과 같이 나오게 됩니다.

$$\text{Cov}(\hat{\phi}) = \begin{bmatrix} \hat{\sigma}_{11}/n & 0 & 0 & 0 & 0 \\ 0 & 2\hat{\sigma}_{11}^2/n & 0 & 0 & 0 \\ 0 & 0 & \hat{\sigma}_{22.1} (1 + \bar{y}_1^2/\tilde{\sigma}_{11})/r & -\bar{y}_1 \hat{\sigma}_{22.1}/(r\tilde{\sigma}_{11}) & 0 \\ 0 & 0 & -\bar{y}_1 \hat{\sigma}_{22.1}/(r\tilde{\sigma}_{11}) & \hat{\sigma}_{22.1}/(rr\tilde{\sigma}_{11}) & 0 \\ 0 & 0 & 0 & 0 & 2\hat{\sigma}_{22.1}^2/r \end{bmatrix}$$

2.2 Bivariate Normal Case (7.2.2 Large-Sample Covariance Matrix)

하나 저희가 궁금한 것은 ϕ 에 대한 추정량이 아닌 θ 에 대한 추정량이기 때문에 공분산행렬도 이에 관한 값들이 궁금합니다. → Use Delta Method!

$$\sqrt{n} (X_n - \theta) \xrightarrow{D} N(0, \sigma^2)$$

$$\sqrt{n} (g(X_n) - g(\theta)) \xrightarrow{D} N\left(0, \sigma^2 (g'(\theta))^2\right)$$

즉, 만약에 X_n 의 분산이 σ^2 이면, $g(X_n)$ 의 분산은 $\sigma^2 \cdot (g'(\theta))^2$ 입니다.

저희의 현재 모수는 ϕ 인 벡터이며 궁금한 모수는 μ_2 이기에 μ_2 를 ϕ 에 대해 미분해야 됩니다.

$$\mu_2 = \beta_{20.1} + \beta_{21.1}\mu_1$$

$$D(\mu_2) = \left(\frac{\partial \mu_2}{\partial \mu_1}, \frac{\partial \mu_2}{\partial \sigma_{11}}, \frac{\partial \mu_2}{\partial \beta_{20.1}}, \frac{\partial \mu_2}{\partial \beta_{21.1}}, \frac{\partial \mu_2}{\partial \sigma_{22.1}} \right) = (\hat{\beta}_{21.1}, 0, 1, \hat{\mu}_1, 0)$$

2.2 Bivariate Normal Case (7.2.2 Large-Sample Covariance Matrix)

이것을 이용해서 μ_2 에 대한 분산을 다음과 같이 구할 수 있습니다.

$$\text{Var}(\hat{\mu}_2) = D(\hat{\mu}_2) C(\phi - \hat{\phi}) D(\hat{\mu}_2)^T = \hat{\sigma}_{22.1} \left[\frac{1}{r} + \frac{\hat{\rho}^2}{n(1 - \hat{\rho}^2)} + \frac{(\bar{y}_1 - \hat{\mu}_1)^2}{r\tilde{\sigma}_{11}} \right]$$

MCAR 내에서는 Missing 과 데이터 자체가 전혀 관계가 없습니다. 즉 다음과 같은 추정량이 Unbiased 하다고 표현이 가능합니다.

$$\bar{y}_1 = \frac{1}{r} \sum_{i=1}^r y_{i1}, \quad E[\bar{y}_1] = \mu_1$$

즉 $(\bar{y}_1 - \hat{\mu}_1)^2$ 의 차수는 r 입니다. 그렇기에 Large Sample 에 대해서 3번째 Term 은 0 으로 수렴하면서 무시도 할 수 있게 됩니다.

$$\text{Var}(\hat{\mu}_2) = \text{Var}(\mu_2 - \hat{\mu}_2) \approx \hat{\sigma}_{22.1} \left[\frac{1}{r} + \frac{\hat{\rho}^2}{n(1 - \hat{\rho}^2)} \right] = \frac{\hat{\sigma}_{22}}{r} \left(1 - \hat{\rho}^2 \frac{n - r}{n} \right)$$

2.3 Bivariate Normal, Small-Sample Inference

Inference 단계라고 하면 저희는 보통 신뢰구간을 살핍니다.

분산을 구했기 때문에 표준오차 역시 구할 수 있으며, 신뢰구간까지 구할 수 있게 됩니다.

$$\hat{\mu}_2 \pm Z_{0.975} \cdot \sqrt{\text{Var}(\hat{\mu}_2)} = \hat{\mu}_2 \pm 1.96 \cdot \sqrt{\text{Var}(\hat{\mu}_2)}$$

분산과 관련된 값들의 경우 보통 Scaling-Issue 로 인해서 신뢰구간을 설정하는 것이 어려울 수도 있어서 Delta Method 을 활용해서 다음과 같이 또 표현이 가능합니다.

$$\log(\hat{\sigma}_{22}) \pm 1.96 \cdot \sqrt{\text{Var}(\log(\hat{\sigma}_{22}))}$$

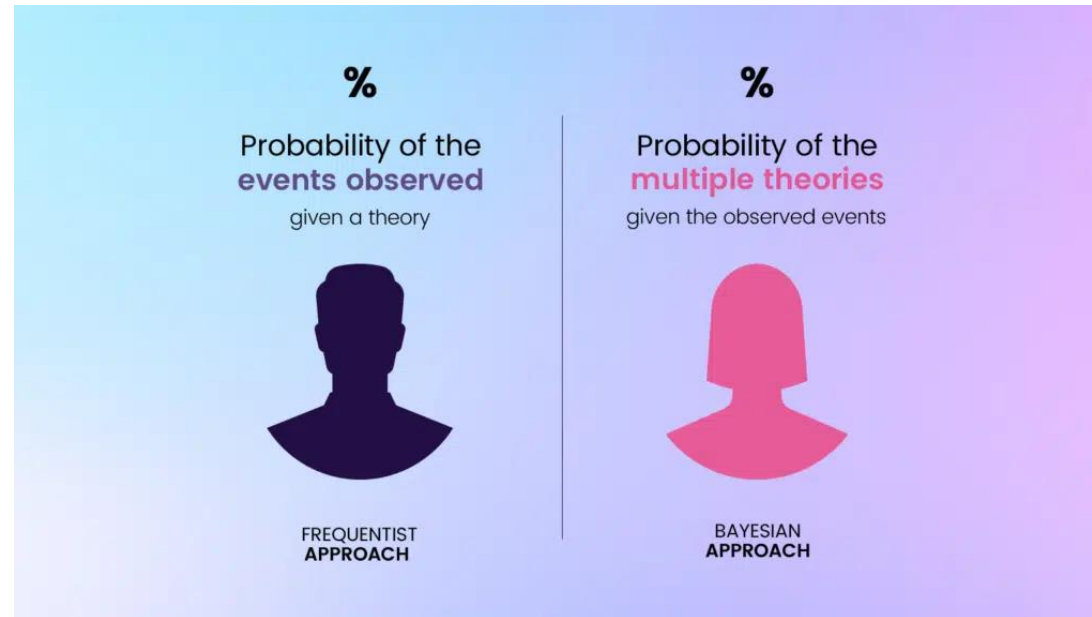
허나 이 모든 것들은 Frequentist 관점에서는 Small-Sample 일 때는 문제가 될 수 있습니다.

비유 → CLT satisfies at $n \geq 30$

베이지즈 관점으로 이 문제를 바라보면 해결이 됩니다.

Detour to 6.1.4 & 6.1.5 (Bayesian)

베이지와 빈도론자 관점의 차이



즉, 모수를 베이지에서는 상수 값이 아닌 확률 변수로 본다는 것입니다.

Detour to 6.1.4 & 6.1.5 (Bayesian)

예시 문제 (교재에는 없습니다) :

$X \sim N(\mu, \sigma^2)$, σ^2 is known, Prior $\mu \sim N(a, b)$, 그럼 $\mu | X$ 의 분포는?

$$f(\mu | X) \propto f(X | \mu) \cdot f(\mu),$$

$$f(X | \mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(\frac{-(x_i - \mu)^2}{2\sigma^2}\right)$$

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

$$f(\mu) = \frac{1}{\sqrt{2\pi b}} \cdot \exp\left(\frac{-(\mu - a)^2}{2b}\right)$$

$$\propto \exp\left(\frac{-(\mu - a)^2}{2b}\right)$$

$$f(\mu | X) = f(X | \mu) \cdot f(\mu) \propto \exp\left(\sum_{i=1}^n \frac{-(x_i - \mu)^2}{2\sigma^2}\right) \cdot \exp\left(\frac{-(\mu - a)^2}{2b}\right)$$

$$= \exp\left(\frac{-1}{2} \cdot \left[\frac{\sum_{i=1}^n (\mu - x_i)^2}{\sigma^2} + \frac{(\mu - a)^2}{b} \right]\right)$$

$$= \exp\left(\frac{-1}{2} \cdot \left[\frac{b \cdot \sum_{i=1}^n (\mu^2 - 2x_i\mu + x_i^2) + \sigma^2 (\mu^2 - 2a\mu + a^2)}{\sigma^2 \cdot b} \right]\right)$$

$$\propto \exp\left(\frac{-1}{2} \cdot \left[\frac{(bn + \sigma^2) \cdot \mu^2 - 2 \cdot (b \cdot \sum_{i=1}^n x_i + \sigma^2 a) \cdot \mu}{\sigma^2 \cdot b} \right]\right)$$

$$= \exp\left(\frac{-1}{2} \cdot \left[\frac{\mu^2 - 2 \cdot \frac{b \cdot \sum_{i=1}^n x_i + \sigma^2 a}{bn + \sigma^2} \cdot \mu}{\frac{\sigma^2 \cdot b}{bn + \sigma^2}} \right]\right)$$

$$\propto \frac{1}{\sqrt{2\pi \cdot \frac{\sigma^2 \cdot b}{bn + \sigma^2}}} \cdot \exp\left(\frac{-1}{2} \cdot \left[\frac{\left(\mu - \frac{b \cdot \sum_{i=1}^n x_i + \sigma^2 a}{bn + \sigma^2}\right)^2}{\left(\frac{1}{b} + \frac{n}{\sigma^2}\right)^{-1}} \right]\right)$$

Detour to 6.1.4 & 6.1.5 (Bayesian)

정리를 하자면, 마지막 식이 다음과 같이 나옵니다.

$$f(\mu | X) \propto \frac{1}{\sqrt{2\pi \cdot \frac{\sigma^2 \cdot b}{bn + \sigma^2}}} \cdot \exp \left(-\frac{1}{2} \cdot \left[\frac{\left(\mu - \frac{b \cdot \sum_{i=1}^n x_i + \sigma^2 \cdot a}{bn + \sigma^2} \right)^2}{\left(\frac{1}{b} + \frac{n}{\sigma^2} \right)^{-1}} \right] \right)$$

이 pdf 를 가진 분포는 정규분포이며, 모수는 다음과 같이 나옵니다.

$$\mu | X \sim N \left(\frac{b \cdot \sum_{i=1}^n x_i + \sigma^2 \cdot a}{bn + \sigma^2}, \left(\frac{1}{b} + \frac{n}{\sigma^2} \right)^{-1} \right)$$

이런 식으로 모수의 Posterior Distribution 을 구하는 것이 가능합니다.

Detour to 6.1.4 & 6.1.5 (Bayesian)

저희가 궁금한 모수의 Posterior 에 대한 분포를 알았기 때문에 이 분포에서 뽑을 수도 있게 됩니다.

$$\mu \mid X \sim N \left(\frac{b \cdot \sum_{i=1}^n x_i + \sigma^2 \cdot a}{bn + \sigma^2}, \left(\frac{1}{b} + \frac{n}{\sigma^2} \right)^{-1} \right)$$

이 분포에서 D 개 만큼 뽑은 뒤에 관련 모수의 표본 Mean 과 Variance 를 구할 수 있게 됩니다.

위의 예시에 대입하면, $\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(D)}$ 개만큼 가지게 되는 것이며, 이 값들의 평균 값과 분산 값을 구하면 **Posterior Mean & Interval** 를 구할 수 있게 됩니다.

Why? → Monte-Carlo Approximation

2.3 Bivariate Normal, Small-Sample Inference

Bivariate Normal 케이스의 경우 다음과 같이 Prior 를 Assume 한다고 합니다.

$$f(\mu_1, \sigma_{11}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1}) \propto \sigma_{11}^{-a} \sigma_{22.1}^{-c}$$

이것을 이전과 같이 구하면, 다음과 같은 Property 들을 가져올 수 있습니다.

1. μ_1, σ_{11} 의 Posterior Distribution 인 $\frac{n \cdot \hat{\sigma}_{11}}{\sigma_{11}}$ 의 분포는 $\chi^2(n + 2a - 3)$ 입니다.
2. $\mu_1 \mid \sigma_{11}$ 의 Posterior Distribution 은 $N(\hat{\mu}_1, \frac{\sigma_{11}}{n})$ 입니다.
3. $\beta_{20.1}, \beta_{21.1}, \sigma_{22.1}$ 의 Posterior Distribution 인 $\frac{r \cdot \hat{\sigma}_{22.1}}{\sigma_{22.1}}$ 의 분포는 $\chi^2(r + 2c - 4)$ 입니다.
4. $\beta_{21.1} \mid \sigma_{22.1}$ 의 Posterior Distribution 은 $N(\hat{\beta}_{22.1}, \frac{\sigma_{22.1}}{r \cdot \hat{\sigma}_{11}})$ 입니다.
5. $\beta_{20.1} \mid \beta_{21.1}, \sigma_{22.1}$ 의 Posterior Distribution 은 $N(\bar{y}_2 - \beta_{21.1} \cdot \bar{y}_1, \frac{\sigma_{22.1}}{r})$ 입니다.
6. μ_1, σ_{11} 와 $\beta_{20.1}, \beta_{21.1}, \sigma_{22.1}$ 의 Posterior 들은 서로 독립입니다.

2.3 Bivariate Normal, Small-Sample Inference

전에 얘기한 Random Draws 를 여기에도 적용할 수 있습니다.

1. Draw from $x_{1t}^2 \stackrel{i.i.d}{\sim} \chi^2(n + 2a - 3)$, $x_{2t}^2 \stackrel{i.i.d}{\sim} \chi^2(r + 2c - 4)$
and also Draw from $z_{1t}, z_{2t}, z_{3t} \stackrel{i.i.d}{\sim} N(0, 1)$
2. 이 값들을 이용해서 ϕ 를 구축할 것이며, d-번째 Sample 을 $\phi^{(d)}$ 이라고 부르겠습니다.

즉 저희는 $\phi^{(d)} = \left(\sigma_{11}^{(d)}, \mu_1^{(d)}, \sigma_{22 \cdot 1}^{(d)}, \beta_{20 \cdot 1}^{(d)}, \beta_{21 \cdot 1}^{(d)} \right)^T$ 를 구하는 것이 목표가 됩니다.

$$\sigma_{11}^{(d)} = \frac{n \hat{\sigma}_{11}}{x_{1t}^2} \quad \mu_1^{(d)} = \hat{\mu}_1 + z_{1t} \left(\frac{\sigma_{11}^{(d)}}{n} \right)^{1/2} \quad \sigma_{22 \cdot 1}^{(d)} = \frac{r \hat{\sigma}_{22 \cdot 1}}{x_{2t}^2}$$

$$\beta_{21 \cdot 1}^{(d)} = \hat{\beta}_{21 \cdot 1} + z_{2t} \left(\frac{\sigma_{22 \cdot 1}^{(d)}}{r \tilde{\sigma}_{11}} \right)^{1/2} \quad \beta_{20 \cdot 1}^{(d)} = \bar{y}_2 - \beta_{21 \cdot 1}^{(d)} \bar{y}_1 + z_{3t} \left(\frac{\sigma_{22 \cdot 1}^{(d)}}{r} \right)^{1/2}$$

3. $\phi^{(d)}$ 를 구했기 때문에 이것을 토대로 $\theta^{(d)}$ 를 구합니다.

2.3 Bivariate Normal, Small-Sample Inference

이렇게 Posterior 를 구하기 위해서 Posterior Distribution 에서 직접 뽑아서 진행되는 것을 MCMC (Markov Chain Monte Carlo) 기법이라고 하며, 저희가 한 것은 그 중에서도 Gibbs Sampling 부분에 해당합니다.

MCMC : 베이지안 추정을 위해서 Posterior 에서 직접적으로 샘플을 추출하는 기법.

Gibbs Sampling : 그 중에서도 Posterior 분포를 정확히 알고 있을 때 직접 뽑아서 이전 모수들도 점차 업데이트 하는 과정.

예시 : 모수가 3개 있을 때

$$\theta_1^{(t+1)} \sim X \mid \theta_2^{(t)}, \theta_3^{(t)}$$

$$\theta_2^{(t+1)} \sim X \mid \theta_1^{(t+1)}, \theta_3^{(t)}$$

$$\theta_3^{(t+1)} \sim X \mid \theta_1^{(t+1)}, \theta_2^{(t+1)}$$

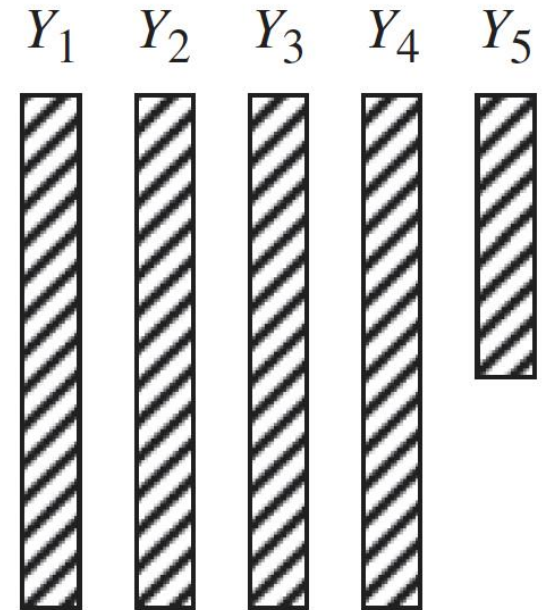
2.4 Monotone Missingness with More Than Two Variables (7.4.1)

여태까지는 변수가 2개, 한개가 Complete, 한개가 Missing 인 경우만 살펴보았으며, 다변량에서의 경우를 살펴보겠습니다.

우선은 다변량이지만 딱 1개의 변수가 Missingness 가 존재한다고 하겠습니다.
이것은 그림 Monotone Missingness Pattern 을 따르게 됩니다.

$K + 1$ 개의 변수들이 있으며, Complete 한 것이 K 개,
Missing 이 존재하는 것이 1개 라고 하겠습니다.

그러면 y_{i1} 를 K 개의 변수로 다변량으로 생각할 수 있습니다.
이들이 모두 정규분포를 따르는 MAR 데이터라고 가정한다면,
 $(y_{i1}, y_{i2}) \sim N_{K+1}(\mu, \Sigma)$ 으로 표현이 가능합니다.



2.4 Monotone Missingness with More Than Two Variables (7.4.1)

정규분포를 따르기에 Bivariate-case 와 같이 MLE 를 구하면,

$$\hat{\mu}_2 = \bar{y}_2 + (\hat{\mu}_1 - \bar{y}_1)^\top \hat{\beta}_{21.1}, \quad \hat{\sigma}_{22} = s_{22} + \hat{\beta}_{21.1}^\top (\hat{\sigma}_{11} - \tilde{\sigma}_{11}) \hat{\beta}_{21.1}$$

Notations :

$\hat{\mu}_1$: $(K \times 1)$ 벡터, n Units 에 대해서 y_{i1} 의 평균

\bar{y}_1 : $(K \times 1)$ 벡터, r Units 에 대해서 y_{i1} 의 평균

$\hat{\beta}_{21.1}$: $(K \times 1)$ 벡터, y_{i2} 를 y_{i1} 에 대해서 회귀식으로 그은 것에 대한 Regression Coefficient

$\hat{\sigma}_{11}$: $(K \times K)$ 메트릭스, n Units 에 대해서 y_{i1} 의 Covariance Matrix

$\tilde{\sigma}_{11}$: $(K \times K)$ 메트릭스, r Units 에 대해서 y_{i1} 의 Covariance Matrix

$\hat{\mu}_2$: Scalar, y_{i2} 이지만 Missing Units 들은 y_{i1} 를 이용한 회귀 식으로 채운 것

$\hat{\sigma}_{22}$: Scalar, y_{i2} 에 대한 Variance

2.4 Monotone Missingness with More Than Two Variables (7.4.1)

여기에서 저희는 다음과 같은 조건 + MAR 이기에 $\hat{\mu}_2$ 에 대한 Estimate 을 MLE 라고 부를 수가 있습니다.

1. $y_{i2} \mid y_{i1} \sim N((\beta_{20.1} + y_{i1} \cdot \beta_{21.1}), \sigma_{22.1})$
2. y_{i1} 의 분포는 다음을 만족하는 분포이다
 - A. y_{i1} 의 평균의 MLE 가 $\hat{\mu}_1$
 - B. μ_1 와 $\beta_{20.1}, \beta_{21.1}, \sigma_{22.1}$ 는 y_{i1} 의 모수들과는 별개인 값들

이것을 또 다르게 진행할 수 있는 방법은 바로 Dummy Variable Regression 입니다.

2.4 Monotone Missingness with More Than Two Variables (7.4.1)

예시로,

- 만약에 y_{i1} 에서 4-th Unit 이 Group 1 에 해당한다면 $y_{41} = (1, 0, 0, \dots, 0)$
- 만약에 y_{i1} 에서 j-th Unit 이 Group K 에 해당한다면 $y_{j1} = (0, 0, 0, \dots, 0, 1)$

그러면 Group $K + 1$ 만 남게 되며, 이것은 $y_{i1} = (0, 0, 0, \dots, 0)$ 으로 두며, 이것을 Reference Group 이라고 부르기도 합니다.

이러한 상황에서 MLE 를 구하면,

- $\hat{\mu}_1$ 은 n Sampled 된 Unit 들이 Group K 에 있을 비율
- μ_1 은 해당 비율들에 대한 기댓값, (두개를 이용하면 이전 슬라이드의 2번째 조건 만족)
- Dummy Variable Regression 도 회귀식이며 y_{i1} 도 정규인 것은 여전하기에 $y_{i2} \mid y_{i1}$ 역시 정규분포를 따라가기에 1번째 조건도 만족합니다.

2.4 Monotone Missingness with More Than Two Variables (7.4.2)

이제는 이전의 확장판을 생각해보겠습니다.

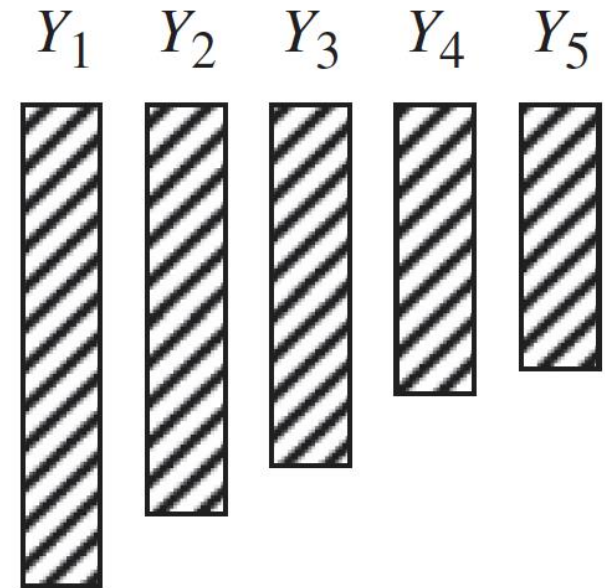
우측 사진과 같은 패턴을 저희는



Pattern 이라고 부릅니다.

즉,

- Y_1 가 n 개 만큼의 Complete Units
- Y_2 가 r_2 개 만큼은 Observed Units
- Y_3 가 r_3 개 만큼은 Observed Units
- 그러면, $n > r_2 \geq r_3 \geq \dots \geq r_j$ 의 관계가 성립이 됩니다.



2.4 Monotone Missingness with More Than Two Variables (7.4.2)

그러면 우도함수는 다음과 같이 나오게 됩니다.

$$\begin{aligned} \prod_{i=1}^n f_Y(y_{i1}, \dots, y_{iJ} \mid \phi) \\ = \prod_{i=1}^n f_1(y_{i1} \mid \phi_1) \prod_{i=1}^{r_2} f_2(y_{i2} \mid y_{i1}, \phi_2) \cdots \prod_{i=1}^{r_J} f_J(y_{iJ} \mid y_{i1}, \dots, y_{i,J-1}, \phi_J) \end{aligned}$$

의미론적으로 해석하자면,

- y_{i1} 의 경우 Complete 하기에 그저 $f_1(y_{i1} \mid \phi_1)$ 으로 표현
- y_{i2} 의 경우 r_2 까지만 Observed 됐기에 Missing 된 부분들을 y_{i1} 으로 표현
- y_{i3} 의 경우 r_3 까지만 Observed 됐기에 Missing 된 부분들을 y_{i1} 과 y_{i2} 으로 표현

2.4 Monotone Missingness with More Than Two Variables (7.4.2)

전체가 다변량 정규분포이기에 우뚝수의 각각의 분포도 정규분포를 따르게 됩니다.

그러면 실제에서는 그저

1. y_{i1} 에 대한 Mean 과 Variance 추정량들을 먼저 구하기
2. y_{i2} 를 y_{i1} 에 대한 Multivariate Linear Regression 을 진행한 다음에 평균과 분산 구하기
3. J 번까지 2번을 반복하기

허나 이것을 회귀식을 씌우면서 하게 된다면 Inverse Matrix 를 구하면서 연산량이 엄청나게 늘어날 수도 있습니다 → Utilize the SWEEP Operator (7.4.3)

2.4 Monotone Missingness with More Than Two Variables (7.4.3)

What is the SWEEP Operator?

G 를 크기가 $(p \times p)$ 크기의 매트릭스라고 생각하며, 이것에 "Swept on row and column k " 를 진행하면 다음의 요소를 가지는 매트릭스 H 가 만들어집니다.

- $h_{kk} = -1/g_{kk}$
- $h_{jk} = h_{kj} = \frac{g_{jk}}{g_{kk}}, \quad j \neq k$
- $h_{jl} = g_{jl} - \frac{g_{jk} \cdot g_{kl}}{g_{kk}}, \quad j \neq k, l \neq k$

$$G = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{12} & g_{22} & g_{23} \\ g_{13} & g_{23} & g_{33} \end{bmatrix}$$

$$H = \text{SWP}[1]G = \begin{bmatrix} -1/g_{11} & g_{12}/g_{11} & g_{13}/g_{11} \\ g_{12}/g_{11} & g_{22} - g_{12}^2/g_{11} & g_{23} - g_{13}g_{12}/g_{11} \\ g_{13}/g_{11} & g_{23} - g_{13}g_{12}/g_{11} & g_{33} - g_{13}^2/g_{11} \end{bmatrix}$$

2.4 Monotone Missingness with More Than Two Variables (7.4.3)

이것을 꾸준히 적용시킬 수도 있습니다.

$SWP[k_1], SWP[k_2], SWP[k_3]$ 를 G 에 진행한다면, 순서에 상관없이 $SWP[k_1, k_2, k_3]G$ 으로 표현이 가능합니다.

- 실제로 진행할 때는 h_{kk} 를 구하고 나서 h_{jk}, h_{kl} 를 구하기에 효율적입니다.
- Commutative 합니다. 즉 $SWP[1, 2]G = SWP[2, 1]G$

이 Operator 가 Linear Regression 에서 굉장히 유용하다고 하며,

예로 G 가 (2×2) 인 Covariance Matrix 이면, $H = SWP[1]G$ 를 진행하게 된다면,

- h_{12} 는 $Y_2 = \beta \cdot Y_1 + c$ 의 Regression Coefficient 이다.
- h_{22} 는 $Y_2 | Y_1$ 의 Residual Variance 이다.

2.4 Monotone Missingness with More Than Two Variables (7.4.3)

전에 있던 G 를 확장하면,

$$G = \begin{bmatrix} 1 & \bar{y}_1 & \cdots & \bar{y}_j & \cdots & \bar{y}_K \\ \bar{y}_1 & n^{-1} \sum y_1^2 & & & \cdots & n^{-1} \sum y_K y_1 \\ \vdots & \vdots & \ddots & & & \vdots \\ \bar{y}_k & & & n^{-1} \sum y_j y_k & & \\ & & & & \ddots & \\ \bar{y}_K & n^{-1} \sum y_1 y_K & & & \cdots & n^{-1} \sum y_K^2 \end{bmatrix}$$

첫번째 행과 열 값들만 바뀌었으며, $(K + 1) \times (K + 1)$ 행렬입니다.

하나 편의성을 위해서 Index 를 $0 \sim K$ 이라고 하겠습니다.

2.4 Monotone Missingness with More Than Two Variables (7.4.3)

Sweeping 을 진행하면 다음과 같이 나옵니다.

$$\text{SWP}[0]G = \begin{bmatrix} -1 & \bar{y}_1 & \cdots & \bar{y}_j & \cdots & \bar{y}_K \\ \bar{y}_1 & \hat{\sigma}_{11} & & & \cdots & \hat{\sigma}_{K1} \\ \vdots & \vdots & \ddots & & & \vdots \\ \bar{y}_k & & & \hat{\sigma}_{jk} & & \\ & & & & \ddots & \\ \bar{y}_K & \hat{\sigma}_{1K} & & & \cdots & \hat{\sigma}_{KK} \end{bmatrix}$$

첫번째 값을 Y_1, \dots, Y_K 를 Constant 인 $Y = 1$ 에 Regression 을 시킨 Coefficient 값이 됩니다.

$\hat{\sigma}_{jk}$ 값들은 Regression 에서 나오는 Residual Covariance 가 됩니다.

2.4 Monotone Missingness with More Than Two Variables (7.4.3)

Sweeping 을 다시 또 진행하면 다음과 같이 나옵니다.

$$\text{SWP}[0, 1]G = \begin{bmatrix} -(1 + \bar{y}_1/\hat{\sigma}_{11}) & \bar{y}_{11}/\hat{\sigma}_{11} & \bar{y}_2 - (\hat{\sigma}_{12}/\hat{\sigma}_{11}) \bar{y}_1 & \cdots & \bar{y}_K - (\hat{\sigma}_{1K}/\hat{\sigma}_{11}) \bar{y}_1 \\ & -1/\hat{\sigma}_{11} & \hat{\sigma}_{12}/\hat{\sigma}_{11} & \cdots & \hat{\sigma}_{1K}/\hat{\sigma}_{11} \\ & & \hat{\sigma}_{22} - \hat{\sigma}_{12}^2/\hat{\sigma}_{11} & \cdots & \hat{\sigma}_{2K} - \hat{\sigma}_{1K}\hat{\sigma}_{12}/\hat{\sigma}_{11} \\ & & & \ddots & \\ \bar{y}_K - (\hat{\sigma}_{1K}/\hat{\sigma}_{11}) \bar{y}_1 & & & & \hat{\sigma}_{KK} - \hat{\sigma}_{1K}^2/\hat{\sigma}_{11} \end{bmatrix}$$

이 행렬을 다음과 같이 4 부분으로 나눌 수 있습니다.

$$= \begin{bmatrix} -A & B \\ B^\top & C \end{bmatrix}$$

2.4 Monotone Missingness with More Than Two Variables (7.4.3)

값은 다음과 같습니다.

$$A = \begin{bmatrix} (1 + \bar{y}_1/\hat{\sigma}_{11}) & -\bar{y}_{11}/\hat{\sigma}_{11} \\ -\bar{y}_{11}/\hat{\sigma}_{11} & 1/\hat{\sigma}_{11} \end{bmatrix}$$
$$B = \begin{bmatrix} \bar{y}_2 - (\hat{\sigma}_{12}/\hat{\sigma}_{11}) \bar{y}_1 & \cdots & \bar{y}_K - (\hat{\sigma}_{1K}/\hat{\sigma}_{11}) \bar{y}_1 \\ \hat{\sigma}_{12}/\hat{\sigma}_{11} & \cdots & \hat{\sigma}_{1K}/\hat{\sigma}_{11} \end{bmatrix}$$
$$C = \begin{bmatrix} \hat{\sigma}_{22} - \hat{\sigma}_{12}^2/\hat{\sigma}_{11} & \cdots & \hat{\sigma}_{2K} - \hat{\sigma}_{1K}\hat{\sigma}_{12}/\hat{\sigma}_{11} \\ \vdots & & \\ \hat{\sigma}_{KK} - \hat{\sigma}_{1K}^2/\hat{\sigma}_{11} & & \end{bmatrix}$$

이때, 다음과 같은 결론을 도출할 수 있습니다.

- A : 적절한 Residual Covariance 값에 곱하고 n 으로 나누면 B 에 있는 Coefficient 에 대한 분산과 공분산을 구해줍니다.
- B : j -번째 열은 각각 Y_{j+1} 을 Y_1 에 Regression 을 진행한 절편과 기울기를 의미합니다.
- C : Y_2, \dots, Y_K 를 Y_1 에 Regression 을 진행한 것의 Residual Covariance Matrix

2.4 Monotone Missingness with More Than Two Variables (7.4.3)

q 번만큼 이것을 반복하면, 다음과 같이 표현이 가능합니다.

$$\text{SWP}[0, 1, \dots, q]G = \begin{bmatrix} -D & E \\ E^T & F \end{bmatrix} \quad \begin{array}{l} D : (q+1) \times (q+1) \\ E : (q+1) \times (K-q) \\ F : (K-q) \times (K-q) \end{array}$$

Reverse-Sweep 은 SWEEP Operator 가 해주는 것의 정반대입니다.

$$(\text{RSW}[k])(\text{SWP}[k])G = (\text{SWP}[k])(\text{RSW}[k])G = G$$

7.4.4 의 경우 이전의 베이즈로 구했을 때 SWEEP Operator 를 적용시킨 것 뿐입니다.

2.5 Factored Likelihoods for Special Nonmonotone Patterns

특수한 패턴들도 존재할 수가 있으며, 다음이 한 예시입니다.

Units
↓

	Y_3			Y_2			Y_1		
	0	...	0	1	...	1	×	...	×
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	0	...	0	1	...	1	×	...	×
	×	...	×	×	...	×	1	...	1
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	×	...	×	×	...	×	1	...	1

Figure 7.1 Data pattern where Y_3 is more observed than Y_1 , and Y_1 and Y_2 are never jointly observed; 0 = observed, 1 = missing, × = possibly observed.
Source: Adapted from Rubin (1974).

- 0 : Observed
 - 1 : Missing
 - ×
- Possibly Observed

2.5 Factored Likelihoods for Special Nonmonotone Patterns

위쪽부분 :

Y_2 를 무시하고 Y_3 과 Y_1 만 바라본다면 그저

Bivariate Monotone Data 입니다.

아래쪽 부분 :

Y_1 을 무시하고 Y_2, Y_3 만 바라본다면 그저

Marginal Distribution 인 케이스로 바라볼 수 있습니다.

	Y_3			Y_2			Y_1		
Units ↓	0	...	0	1	...	1	×	...	×
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	0	...	0	1	...	1	×	...	×
	×	...	×	×	...	×	1	...	1
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	×	...	×	×	...	×	1	...	1

이것이 만약에 MAR 이라면 위쪽과 아래쪽 부분에
해당하는 파라미터를 다르게 설정하면 2가지로 나눈
Factored Likelihood 으로 표현이 가능합니다.

0 : Observed

1 : Missing

×

Roadmap of Our Progress

Previously :

- **1장** : 완전한 Introduction
- **2장** : Missingness 가 존재할 때 실험계획을 하는 방법과 Regression Imputation 에 대한 내용을 배웠다.
- **3장** : 일반적인 상황에서 적용이 가능한 Weighting 방법을 적용했을 때 Missing Pattern 관련된 처리를 어떻게 하는지에 대해서 배웠다.
- **4장** : Imputation 을 토대로 어떻게 하면 결측값들을 위험을 감수하면서 채울 수 있는지에 대해서 배웠다.
- **5장** : 샘플링으로 Imputation 을 진행하면 생기는 Uncertainty 과 이것을 더 좋게 접근하는 방법에 대해서 배웠다.

Today :

- **6장** : 우도함수와 MLE 에 대해서 배웠으며 General 한 Missing 데이터에 이것을 어떻게 적용하는지에 대해서 배웠다.
- **7장** : 만약에 Missingness Mechanism 을 무시하면 이것에 관한 우도함수를 구축하는 방법과 SWEEP Operator 에 대해서 배웠다.

What Awaits :

- **8장** : MLE 를 이용한 기법인 EM-Algorithm 에 대한 내용을 배운다.
- **9장** : Large Sample 인 경우에서 MLE 에 관한 정확한 Inference 를 구하는 방법
- **10장** : 베이지스 기법들을 활용해서 다양한 Imputation 을 진행하는 방법

Assignment

$Y \sim \exp(\theta)$ 의 값을 갖는 데이터들이 있다고 하자. 이때 observed data는 $y_{(0)} = (y_1, \dots, y_r)^T$, missing data는 $y_{(1)} = (y_{r+1}, \dots, y_n)^T$ 로 나뉘고, $m = (m_1, \dots, m_n)^T$, where $m_i = 0$ ($i = 1, \dots, r$), $m_i = 1$ ($i = r + 1, \dots, n$)인 상황이다. 다음의 ML estimate를 구하고 이에 대해 생각해보자.

i) missing data가 MAR이며 θ 와 ϕ 가 distinct한 경우, θ 에 대한 ML Estimate을 구하시오.

이때 ϕ 는 각 unit이 관측될 확률을 의미한다.

(ii) censoring point c 가 존재하여, $y_i \geq c$ 의 경우 결측이 발생하고 $y_i < c$ 의 경우에는 결측이 발생하지 않았다고 하자.

이와 같은 incomplete case에서 θ 에 대한 ML Estimate을 구하시오.

(iii) 이번에는 missing data를 매개변수로 잡고 complete data log-likelihood를 최대화하는 missing data와 θ 에 대한 ML Estimate을 구하여라. 해당 접근법의 결함도 생각해보자.

감사합니다