Math tools 1

Note: Statistical Inference

Oct 2024

Lecturer: *Typed by: Zhuohua Shen*

Contents

1 Preliminary **5**

2 Statistical inference fundamentals **7**
- 2.1 Statistical Models . . . 7
- 2.2 Principles of Data Reduction . . . 8
  - 2.2.1 Sufficiency Principle . . . 8
  - 2.2.2 Likelihood principle . . . 8

3 Multivariate Inference Fundamentals **9**
- 3.1 Random vectors and distributions . . . 9
  - 3.1.1 Multivariate normal distribution . . . 10
  - 3.1.2 The noncentral χ^2 and F distribution . . . 11
- 3.2 Asymptotic properties . . . 12
  - 3.2.1 Asymptotic distributions of sample means and covariance matrices . . . 12

4 Linear Models **13**
- 4.1 Linear regression for the full-rank model . . . 13

5 Generalized Linear Models **15**

6 Bayesian Inference **17**

7 Structural Equation Model (SEM) **19**
- 7.1 SEM models . . . 19
- 7.2 Bayesian methods for estimating SEM . . . 21
  - 7.2.1 Bayesian estimation using MCMC . . . 21
- 7.3 Hierarchical and Multisample Data . . . 22

Chapter 1

Preliminary

Chapter 2

Statistical inference fundamentals

References: most of the contents are from the undergraduate course STA3020 (by Prof. Jianfeng Mao in 2022-2023 T1, and Prof. Jiasheng Shi in 2023-2024 T2) and postgraduate course STAT5010 (by Kin Wai Keith Chan in 2024-2025 T1), with main textbook Casella and Berger [1]

2.1 Statistical Models

See Chapter 3 of [1]. Suppose $X_i \sim_{\text{iid}} \mathbb{P}_*$, where $\mathbb{P}_*$ refers to the unknown data generating process (DGPg), we find $\widehat{\mathbb{P}} \approx \mathbb{P}_*$. A statistical model is a set of distributions $\mathscr{F} = \{\mathbb{P}_\theta : \theta \in \Theta\}$, where Θ is the parameter space. A parametric model is the model with $\dim(\Theta) < \infty$, while a nonparametric model satisfies $\dim(\Theta) = \infty$.

Definition 2.1.1 (Exponential family). A k-dimensional exponential family (EF) $\mathscr{F} = \{f_\theta : \theta \in \Theta\}$ is a model consisting of pdfs of the form

$$f_\theta(x) = c(\theta)h(x)\exp\left\{\sum_{j=1}^{k}\eta_j(\theta)T_j(x)\right\} \tag{2.1}$$

where $c(\theta), h(x) \geq 0$, $\Theta = \{\theta : c(\theta) \geq 0, \eta_j(\theta) \text{ being well defined for } 1 \leq j \leq k\}$. Let $\eta_j = \eta_j(\theta)$, the canonical form is

$$f_\eta(x) = b(\eta)h(x)\exp\left\{\sum_{j=1}^{k}\eta_j T_j(x)\right\}, \tag{2.2}$$

- k-dim natural exponential family (NEF): $\mathscr{F}' = \{f_\eta : \eta \in \Xi\}$;
- natural parameter $\eta = (\eta_1, \dots, \eta_k)^\top$;
- natural parameter space: $\Xi = \{\eta \in \mathbb{R}^k : 0 < b(\eta) < \infty\}$;
- the NEF $\mathscr{F}'$ is of full rank if Ξ contains an open set in $\mathbb{R}^k$;
- the EF is a curved exponential family if $p = \dim(\Theta) < k$.

Properties of EF:

- Let $X \sim f_\eta$, where $\eta \in \Xi$ such that (i) f_η is of the form (2.2) with $B(\eta) = -\log b(\eta)$, and (ii) Ξ contains an open set in $\mathbb{R}^k$. Then, for $j, j' = 1, \dots, k$, $\mathbb{E}\{T_j(X)\} = \partial B(\eta)/\partial \eta_j$ and $\mathbb{C}\text{ov}\{T_j(X), T_{j'}(X)\} = \partial^2 B(\eta)/(\partial \eta_j \partial \eta_{j'})$.
- Stein's identity:

Definition 2.1.2 (Location-scale family). Let f be a density.

- A location-scale family is given by $\mathscr{F} = \{f_{\mu,\sigma} : \mu \in \mathbb{R}, \sigma \in \mathbb{R}^{++}\}$, where $f_{\mu,\sigma}(x) = f((x-\mu)/\sigma)/\sigma$.
- location parameter: μ; scale parameter: σ; standard density: f;
- A location family is $\mathscr{F} = \{f_{\mu,1} : \mu \in \mathbb{R}\}$.
- A scale family is $\mathscr{F} = \{f_{0,\sigma} : \sigma \in \mathbb{R}^{++}\}$

Representation: $X = \mu + \sigma Z$, $Z \sim f_{0,1}(\cdot)$.

- See some examples in Example 3.9, Keith's note 3, and Table 1 in Shi's note L1.
- Transform between location parameter and scale parameter by taking log.

Definition 2.1.3 (Identifiable family). If $\forall \theta_1, \theta_2 \in \Theta$ that

$$\theta_1 \neq \theta_2 \quad \Rightarrow \quad f_{\theta_1}(\cdot) \neq f_{\theta_2}(\cdot),$$

then $\mathscr{F}$ is said to be an identifiable family, or equivalently $\theta \in \Theta$ is identifiable.

A typical feature of non-identifiable EF is that $p = \dim(\Theta) \geqslant k$. Typically,

- $p < k$, curved (must).
- $p = k$, of full rank.
- $p > k$, non-identifiable.

2.2 Principles of Data Reduction

Statistics: $T = T(X_{1:n})$, a function of $X_{1:n}$ and free of any unknown parameter.

2.2.1 Sufficiency Principle

Sufficiency principle: If $T = T(X_{1:n})$ is a "sufficient statistics" for θ, then any inference on θ will depend on $X_{1:n}$ only through T.

Definition 2.2.1 (Sufficient, minimal sufficient, ancillary, and complete statistics). Suppose $X_{1:n} \sim_{\text{iid}} \mathbb{P}_\theta$, where $\theta \in \Theta$. Let $T = T(X_{1:n})$ be a statistic. Then T is sufficient (SS) for θ

- $\Leftrightarrow$ (def) $[X_{1:n} \mid T = t]$ is free of θ for each t.
- $\Leftrightarrow$ (technical lemma) $T(x_{1:n}) = T(x'_{1:n})$ implies that $f_\theta(x_{1:n})/f_\theta(x'_{1:n})$ is free of θ.
- $\Leftrightarrow$ (Neyman-Fisher factorization theorem) $\forall \theta \in \Theta$, $x_{1:n} \in \mathscr{X}^n$, $f_\theta(x_{1:n}) = A(t, \theta) B(x_{1:n})$.
- $\Leftrightarrow$ Define $\Lambda(\theta', \theta'' \mid x_{1:n}) := f_{\theta'}(x_{1:n})/f_{\theta''}(x_{1:n})$. $\forall \theta', \theta'' \in \Theta$, $\exists$ function $C_{\theta', \theta''}$ such that $\Lambda(\theta', \theta'' \mid x_{1:n}) = C_{\theta', \theta''}(t)$, for all $x_{1:n} \in \mathscr{X}^n$ where $t = T(x_{1:n})$.

T is minimal sufficient (MSS) for θ

- $\Leftrightarrow$ (def) (1) T is a SS for θ; (2) $T = g(S)$ for any other SS S.
- $\Leftrightarrow$ (1) T is a SS for θ; (2) $S(x_{1:n}) = S(x'_{1:n})$ implies $T(x_{1:n}) = T(x'_{1:n})$ for any SS S.
- $\Leftrightarrow$ (Lehmann-Scheffé theorem) $\forall x_{1:n}, x'_{1:n} \in \mathscr{X}^n$, $f_\theta(x_{1:n})/f_\theta(x'_{1:n})$ is free of $\theta \Leftrightarrow T(x_{1:n}) = T(x'_{1:n})$.

$A = A(X_{1:n})$ is ancillary (ANS) if the distribution of A does not depend on θ.
T is complete (CS) if $\forall \theta \in \Theta$, $\mathbb{E}_\theta g(T) = 0$ implies $\forall \theta \in \Theta$, $\mathbb{P}_\theta\{g(T) = 0\} = 1$.

Properties

- (Transformation) If $T = r(T')$, then (i) T is SS $\Rightarrow$ T' is SS; (ii) T' is CS $\Rightarrow$ T is CS; (iii) r is one-to-one, then if one is SS/MSS/CS, then the another is.
- (Basu's Lemma) $X_i \sim_{\text{iid}} \mathbb{P}_\theta$, A is ANS and T s CSS, then $A \perp\!\!\!\perp T$.
- (Bahadur's theorem) $X_i \sim_{\text{iid}} \mathbb{P}_\theta$, if an MSS exists, then any CSS is also an MSS.
  - Then if a CSS exists, then any MSS is also a CSS $\Rightarrow$ CSS=MSS.
  - All or nothing: start with MSS T, check whether T is CS. (i) Yes, it is both CSS and MSS, then the set of MSS=CSS; (ii) No, there is no CSS at all.
- (Exp-family) If $X_i \sim_{\text{iid}} f_\eta$ in (2.2), then $T = (\sum_{i=1}^n T_1(X_i), \ldots, \sum_{i=1}^n T_k(X_i))$ is a SS, called natural sufficient statistic. If Ξ contains an open set in $\mathbb{R}^k$ (i.e., $\mathscr{F}'$ is of full rank), then T is MSS and CSS.

Proof techniques

- Prove T is not sufficient for θ: show if $\exists x_{1_n}, x'_{1:n} \in \mathcal{X}^n$ and $\theta', \theta'' \in \Theta$, such that $T(x_{1:n}) = T(x'_{1:n})$ and $\Lambda(\theta', \theta'' \mid x_{1:n}) \neq \Lambda(\theta', \theta'' \mid x'_{1:n})$.
- Prove A is an ANS: consider location-scale representation.
- Prove T is a CS: use definition or take $\mathrm{d}\mathbb{E}_\theta g(T)/\mathrm{d}\theta = 0$.
- Disprove T is CS:
  - Construct an ANS $S(T)$ based on T, then $\mathbb{E}S(T)$ is free of θ, then $g(T) = S(T) - \mathbb{E}S(T)$ is free of θ but $g(T) \neq 0$ w.p.1.
  - (Cancel the 1st moment) Find two unbiased estiamtors for θ as a function of T. E.g., $X_1, X_2 \sim_{\text{iid}} \mathrm{N}(\theta, \theta^2)$, $T = (X_1, X_2)$, $g(T) = X_1 - X_2 \sim \mathrm{N}(0, 2\theta^2)$.

Remark 2.2.2.

- ANS A is useless on its own, but useful together with other information.
- $\mathbb{P}(A(\boldsymbol{X}) \mid \theta)$ is free of θ, but for non-SS T, $\mathbb{P}(A(\boldsymbol{X}) \mid T(\boldsymbol{X}))$ is not necessarily free of θ.

2.2.2 Likelihood principle

Chapter 3

Multivariate Inference Fundamentals

Reference:

- Robb J. Muirhead - Aspects of multivariate statistical theory [5].
- CUHK STAT4002 - Applied Multivariate Analysis (2023 Spring), by Zhixiang Lin.
- CUHK STAT5030 - Linear Models (2025 Spring), by Yuanyuan Lin.
- Peng DING - Linear Model and Extensions.
- Ronald Christensen - Plane answers to complex questions: the theory of linear models [2].

3.1 Random vectors and distributions

Definition 3.1.1. Let $\boldsymbol{x} = (x_1, \dots, x_p)^\top \in \mathbb{R}^p$ be a random vector,

- Mean $\mathbb{E}\boldsymbol{x} = \boldsymbol{\mu} = (\mathbb{E}x_1, \dots, \mathbb{E}x_p)^\top = (\mu_j)$.
- Covariance matrix $\mathbb{V}\mathrm{ar}(\boldsymbol{x}) = \mathbb{C}\mathrm{ov}(\boldsymbol{x}) = \Sigma = \mathbb{E}[(\boldsymbol{x} - \mathbb{E}\boldsymbol{x})(\boldsymbol{x} - \mathbb{E}\boldsymbol{x})^\top] = \mathbb{E}\boldsymbol{x}\boldsymbol{x}^\top - \mathbb{E}\boldsymbol{x}\mathbb{E}\boldsymbol{x}^\top = (\sigma_{ij})$, $\Sigma \succeq \mathbf{0}$.
- Correlation matrix $R = D^{-1/2}\Sigma D^{-1/2}$, where $D = \mathrm{diag}(\sigma_{11}, \dots, \sigma_{pp})$. We have $R_{ij} = \rho_{ij} = \sigma_{ij}/(\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}})$.
- If $\boldsymbol{y} \in \mathbb{R}^q$ random vector, then $\mathbb{C}\mathrm{ov}(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{E}[(\boldsymbol{x} - \mathbb{E}\boldsymbol{x})(\boldsymbol{y} - \mathbb{E}\boldsymbol{y})^\top] = \mathbb{E}\boldsymbol{x}\boldsymbol{y}^\top - \mathbb{E}\boldsymbol{x}\mathbb{E}\boldsymbol{y}^\top \in \mathbb{R}^{p\times q}$.

If $\boldsymbol{Z} = (z_{ij}) \in \mathbb{R}^{p\times q}$ is a random matrix,

- $\mathbb{E}\boldsymbol{Z} = (\mathbb{E}z_{ij})$.

Proposition 3.1.2. *Let $\boldsymbol{x} \in \mathbb{R}^p$ be a random vector, $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^p$ be vectors, $A \in \mathbb{R}^{r_1\times p}, B \in \mathbb{R}^{r_2\times p}$ be matrices,*

- $\mathbb{E}\boldsymbol{a}^\top\boldsymbol{x} = \boldsymbol{a}^\top\mathbb{E}\boldsymbol{x}$, $\mathbb{V}\mathrm{ar}(\boldsymbol{a}^\top\boldsymbol{x}) = \boldsymbol{a}^\top\Sigma\boldsymbol{a}$, *and* $\mathbb{C}\mathrm{ov}(\boldsymbol{a}^\top\boldsymbol{x}, \boldsymbol{b}^\top\boldsymbol{x}) = \boldsymbol{a}^\top\Sigma\boldsymbol{b}$.
- $\mathbb{E}A\boldsymbol{x} = A\mathbb{E}\boldsymbol{x}$, $\mathbb{V}\mathrm{ar}(A\boldsymbol{x}) = A\Sigma A^\top$, *and* $\mathbb{C}\mathrm{ov}(A\boldsymbol{x}, B\boldsymbol{x}) = A\Sigma B^\top$.
- *If* $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b}$, *where* $A \in \mathbb{R}^{q\times p}$, $\boldsymbol{b} \in \mathbb{R}^q$, *then* $\boldsymbol{\mu}_{\boldsymbol{y}} = A\boldsymbol{\mu}_{\boldsymbol{x}} + \boldsymbol{b}$ *and* $\Sigma_{\boldsymbol{y}} = A\Sigma_{\boldsymbol{x}}A^\top$.
- $\mathbb{E}(\boldsymbol{x}^\top A\boldsymbol{x}) = \mathrm{tr}(A\Sigma) + \boldsymbol{\mu}^\top A\boldsymbol{\mu}$.

Let $\boldsymbol{Z} \in \mathbb{R}^{p\times q}$ be a random matrix, $B \in \mathbb{R}^{m\times p}$, $C \in \mathbb{R}^{q\times n}$, and $D \in \mathbb{R}^{m\times n}$ constants, then

- $\mathbb{E}(B\boldsymbol{Z}C + D) = B\mathbb{E}(\boldsymbol{Z})C + D$.

- The $\Sigma \in \mathbb{R}^{p\times p}$ is a covariance matrix (i.e., $\Sigma = \mathbb{C}\mathrm{ov}(\boldsymbol{x})$ for some random vector $\boldsymbol{x} \in \mathbb{R}^p$) iff $\Sigma \succeq \mathbf{0}$.
  - ($\Leftarrow$): suppose $\mathrm{r}(\Sigma) = r \le p$, write full rank decomposition $\Sigma = CC^\top$, $C \in \mathbb{R}^{p\times r}$. Let $\boldsymbol{y} \sim [\mathbf{0}_r, I_r]$, then $\mathbb{C}\mathrm{ov}(C\boldsymbol{y}) = \Sigma$.
- If Σ is not PD, then $\exists \boldsymbol{a} \ne \mathbf{0}_p$ s.t. $\mathbb{V}\mathrm{ar}(\boldsymbol{a}^\top\boldsymbol{x}) = 0$ so w.p.1., $\boldsymbol{a}^\top\boldsymbol{x} = k$, i.e., $\boldsymbol{x}$ lies in a hyperplane.

Theorem 3.1.3. *If $\boldsymbol{x} \in \mathbb{R}^p$ random, then its distribution is uniquely determined by the distributions of $\boldsymbol{a}^\top\boldsymbol{x}$, $\forall\boldsymbol{a} \in \mathbb{R}^p$.*

The proof uses the fact that a distribution in $\mathbb{R}^p$ is uniquely determined by its ch.f., see Theorem 1.2.2. [5].

Definition 3.1.4. Dataset contains p variables and n observations are represented by $\boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)^\top$, where the ith row $\boldsymbol{x}_i^\top = (x_{i1}, \dots, x_{ip})$ is the ith observation vector, $i = 1, \dots, n$.

- (Sample mean vector) $\bar{\boldsymbol{x}} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i = (\bar{x}_1, \dots, \bar{x}_p)^\top$, where $\bar{x}_j = n^{-1}\sum_{i=1}^n x_{ij}$.
- (Sum of squares and cross product (SSCP) matrix) $\boldsymbol{A} = \sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top$.
- (Sample covariance matrix) $\boldsymbol{S} = (n-1)^{-1}\boldsymbol{A}$.
- (Sample correlation matrix) $\boldsymbol{R} = D^{-1/2}\boldsymbol{S}D^{-1/2}$, where $D^{-1/2} = \mathrm{diag}(1/\sqrt{s_{11}}, \dots, 1/\sqrt{s_{pp}})$.

- $\bar{\boldsymbol{x}} = n^{-1}\boldsymbol{X}^\top \mathbf{1}_n$, and

$$\begin{aligned}\boldsymbol{A} &= \sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top - n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})(\bar{\boldsymbol{x}} - \boldsymbol{\mu})^\top \\ &= (\boldsymbol{X} - \mathbf{1}_n\bar{\boldsymbol{x}}^\top)^\top(\boldsymbol{X} - \mathbf{1}_n\bar{\boldsymbol{x}}^\top) \succeq \mathbf{0}.\end{aligned}$$

- $\mathbb{E}\bar{\boldsymbol{x}} = \boldsymbol{\mu}$, $\mathbb{V}\text{ar}(\bar{\boldsymbol{x}}) = n^{-1}\Sigma$, $\mathbb{E}\boldsymbol{A} = (n-1)\Sigma$, and $\mathbb{E}\boldsymbol{S} = \Sigma$.

3.1.1 Multivariate normal distribution

Definition 3.1.5 (Original definition of multivariate normal)**.** The random vector $\boldsymbol{x} \in \mathbb{R}^p$ is said to have an p-variate normal distribution ($\boldsymbol{x} \sim \mathrm{N}_p$) if $\forall \boldsymbol{a} \in \mathbb{R}^p$, the distribution of $\boldsymbol{a}^\top\boldsymbol{x}$ is univariate normal.

Theorem 3.1.6 (Fundamental properties)**.** *Let $\boldsymbol{x} \sim \mathrm{N}_p$, we have*

1. *Both $\boldsymbol{\mu} = \mathbb{E}\boldsymbol{x}$ and $\Sigma = \mathbb{C}\text{ov}(\boldsymbol{x})$ exist and the distribution of $\boldsymbol{x}$ is determined by $\boldsymbol{\mu}$ and Σ. Write $\boldsymbol{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$.*
2. *(Representation) Let $\Sigma \succeq \mathbf{0}_{p\times p}$, $\mathrm{r}(\Sigma) = r \le p$, and $u_{1:r} \sim_{\text{iid}} \mathrm{N}(0,1)$, i.e., $\boldsymbol{u} \sim \mathrm{N}_r(\mathbf{0}_r, I_r)$, then if C is the full rank decomposition of Σ and $\boldsymbol{\mu} \in \mathbb{R}^p$, then $\boldsymbol{x} = C\boldsymbol{u} + \boldsymbol{\mu} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$.*
   - *Let $\Sigma = HDH^\top$ be the spectral decomposition, then $\boldsymbol{x} = HD^{1/2}\boldsymbol{z} + \boldsymbol{\mu}$, where $\boldsymbol{z} \sim \mathrm{N}_p(\mathbf{0}_p, I_p)$.*
3. *If $\boldsymbol{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, then its ch.f. $\phi_{\boldsymbol{x}}(\boldsymbol{t}) = \exp(i\boldsymbol{\mu}^\top\boldsymbol{t} - \boldsymbol{t}^\top\Sigma\boldsymbol{t}/2)$.*
4. *(Density) $\boldsymbol{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ with $\Sigma \succ \mathbf{0}$, then $\boldsymbol{x}$ has pdf*

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}. \tag{3.1}$$

Note that we guarantee the existence of $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ by means of the representation in point 2. By its density, we have MVN kernel: If

$$f(\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}^\top A\boldsymbol{x} - 2\boldsymbol{x}^\top B)\right\} = \exp\left\{-\frac{1}{2}(\boldsymbol{x} - A^{-1}B)^\top A(\boldsymbol{x} - A^{-1}B) - B^\top A^{-1}B\right\},$$

then $\boldsymbol{x} \sim \mathrm{N}_p(A^{-1}B, A^{-1})$.

Theorem 3.1.7 (Properties of multivariate normal)**.** *If $\boldsymbol{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, then we have*

1. *(Linearity) Let $B \in \mathbb{R}^{q\times p}, \boldsymbol{b} \in \mathbb{R}^q$ nonrandom, and $B\Sigma B^\top \succ \mathbf{0}$, then $B\boldsymbol{x} + \boldsymbol{b} \sim \mathrm{N}_q(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B^\top)$.*
2. *(Linear combinations) If $\boldsymbol{x}_k \sim \mathrm{N}_p(\boldsymbol{\mu}_k, \Sigma_k) \perp\!\!\!\perp$ for $k = 1, \dots, N$, then for any fixed constants $\alpha_1, \dots, \alpha_N$, $\sum_{k=1}^N \alpha_k\boldsymbol{x}_k \sim \mathrm{N}_p(\sum_{k=1}^N \alpha_k\boldsymbol{\mu}_k, \sum_{k=1}^N \alpha_k^2\Sigma_k)$.*
   - *The sample mean $\bar{\boldsymbol{x}} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma/N)$.*
3. *(Subset) The marginal distribution of any subset of $k(< p)$ components of $\boldsymbol{x}$ is k-variate normal.*
4. *(Marginal distribution) Partition*

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \boldsymbol{x}_1 \in \mathbb{R}^q, \boldsymbol{x}_2 \in \mathbb{R}^{p-q}, \Sigma_{12} \in \mathbb{R}^{q\times(p-q)}.$$

   Then $\boldsymbol{x}_1 \sim \mathrm{N}_q(\boldsymbol{\mu}_1, \Sigma_{11})$, $\boldsymbol{x}_1 \perp\!\!\!\perp \boldsymbol{x}_2$ iff $\Sigma_{12} = \mathbf{0}$.
5. *(Conditional distribution) Let Σ_{22}^- be a generalized inverse of Σ_{22} (i.e., $\Sigma_{22}\Sigma_{22}^-\Sigma_{22} = \Sigma_{22}$), then*
   (a) $\boldsymbol{x}_1 - \Sigma_{12}\Sigma_{22}^-\boldsymbol{x}_2 \sim \mathrm{N}_q(\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^-\boldsymbol{\mu}_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{21})$, and $\perp\!\!\!\perp \boldsymbol{x}_2$.
   (b) $[\boldsymbol{x}_1 \mid \boldsymbol{x}_2] \sim \mathrm{N}_q(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^-(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{21})$.
6. *(Cramér) If $p \times 1$ random vectors $\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{y}$ and $\boldsymbol{x} + \boldsymbol{y} \sim \mathrm{N}_p$, then both $\boldsymbol{x}, \boldsymbol{y} \sim \mathrm{N}_p$.*
7. *(MLE) of (μ, Σ) is $(\bar{\boldsymbol{x}}, A/n)$.*
8. *(Inverse of Σ and conditional independence) Denote $\Sigma^{-1} = (v^{jk})_{1\le j,k\le p}$. Then $\forall j \ne k$, $v^{jk} = 0 \Leftrightarrow x_j \perp\!\!\!\perp x_k \mid \boldsymbol{x} \setminus \{x_j, x_k\}$.*

For point 3, each component of a random vector is (marginally) normal does not imply that the vector has a multivariate normal distribution. Counterexample: let $U_1, U_2, U_3 \sim_{\text{iid}} \mathrm{N}(0,1)$, $Z \perp\!\!\!\perp U_{1:3}$. Define

$$X_1 = \frac{U_1 + ZU_3}{\sqrt{1+Z^2}}, \quad X_2 = \frac{U_2 + ZU_3}{\sqrt{1+Z^2}}.$$

Then $[X_1|Z] \sim \mathrm{N}(0,1)$, free of Z, so $X_1 \sim \mathrm{N}(0,1)$, and $X_2 \sim \mathrm{N}(0,1)$. But (X_1, X_2) not normal. The converse is true if the components of $\boldsymbol{x}$ are all independent and normal, or if $\boldsymbol{x}$ consists of independent subvectors, each of which is normally distributed.

For the proof of point 5, we use the lemma: if $\Sigma \succeq \mathbf{0}$, then $\ker(\Sigma_{22}) \subset \ker(\Sigma_{12})$, and $\text{range}(\Sigma_{21}) \subset \text{range}(\Sigma_{22})$. So $\exists B \in \mathbb{R}^{q\times(p-q)}$ satisfying $\Sigma_{12} = B\Sigma_{22}$.

Quadratic form of MVN

Proposition 3.1.8 (MGF)**.** *If $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$, A is symmetric and σ IS non-singular, then*

$$M_{\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}(t) = |I - 2tA\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\mu}^\top [I - (I - 2t\boldsymbol{A}\Sigma)^{-1}]\Sigma^{-1}\boldsymbol{\mu}\right\}.$$

Proposition 3.1.9 (Variance)**.** *If $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$, then*

$$\mathbb{V}\mathrm{ar}(\boldsymbol{x}^\top A\boldsymbol{x}) = 2\mathrm{tr}(A\Sigma A\Sigma) + 4\boldsymbol{\mu}^\top A\Sigma A\boldsymbol{\mu},$$
$$\mathbb{C}\mathrm{ov}(\boldsymbol{x}^\top A_1\boldsymbol{x}, \boldsymbol{x}^\top A_2\boldsymbol{x}) = 2\mathrm{tr}(A_1\Sigma A_2\Sigma) + 4\boldsymbol{\mu}^\top A_1\Sigma A_2\boldsymbol{\mu}.$$

See Problem B 2.14 in Peng DING.

Theorem 3.1.10 (Independence of quadratic form)**.** *Assume $\boldsymbol{x} \sim N_p(\mu, \Sigma)$.*

1. *For two symmetric matrix $A, B \in \mathbb{R}^{p\times p}$, $\boldsymbol{x}^\top A\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{x}^\top B\boldsymbol{x}$ iff*

$$\Sigma A\Sigma B\Sigma = \mathbf{0}, \quad \Sigma A\Sigma B\boldsymbol{\mu} = \Sigma B\Sigma A\boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\mu}^\top A\Sigma B\boldsymbol{\mu} = 0.$$

   If $\Sigma \succ \mathbf{0}$, then iff $A\Sigma B = \mathbf{0}$.

2. *If $\Sigma \succ \mathbf{0}$, $A \in \mathbb{R}^{p\times p}$ symmetric, and $B \in \mathbb{R}^{r\times p}$, then $\boldsymbol{x}^\top A\boldsymbol{x} \perp\!\!\!\perp B\boldsymbol{x}$ iff $B\Sigma A = \mathbf{0}$.*

See Theorem B.11 in Peng DING, and Problem 1.22–1.23 in [5].

Theorem 3.1.11 (Quadratic form of Σ)**.** *If $\boldsymbol{x}, \boldsymbol{x}_{1:N} \sim_{\mathrm{iid}} \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, where Σ is nonsingular, then*

- $(\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \sim \chi^2_p$,
- $\boldsymbol{x}^\top \Sigma^{-1}\boldsymbol{x} \sim \chi^2_p(\boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{\mu})$,
- *partition*

$$\boldsymbol{x} = \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix},\ \boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix},\ \Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}, \quad \boldsymbol{x}_1, \boldsymbol{\mu}_1 \in \mathbb{R}^k,\ \Sigma_{11} \in \mathbb{R}^{k\times k},\ \textit{then}$$

$$Q = (\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) - (\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^\top \Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1) \sim \chi^2_{p-k}.$$

- $N(\bar{\boldsymbol{x}}_N - \boldsymbol{\mu})^\top \Sigma^{-1}(\bar{\boldsymbol{x}}_N - \boldsymbol{\mu}) \sim \chi^2_p$,
- *the Mahalanobis distance $d_i^2 = (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_N)^\top \boldsymbol{S}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_N) \xrightarrow{\mathrm{d}} \chi^2_p$.*

If $\mathrm{r}(\Sigma) = k \le p$, then

1. $$(\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{-}(\boldsymbol{x} - \boldsymbol{\mu}) \sim \chi^2_k,$$
   $$(\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{+}(\boldsymbol{x} - \boldsymbol{\mu}) \sim \chi^2_k.$$
2. *If Σ is idempotent with $\mathrm{r}(\Sigma) = k$, then $(\boldsymbol{x} - \boldsymbol{\mu})^\top(\boldsymbol{x} - \boldsymbol{\mu}) \sim \chi^2_k$.*
   - *If $\boldsymbol{\mu} \in \mathrm{Col}(\Sigma)$, then $\boldsymbol{x}^\top\boldsymbol{x} \sim \chi^2_k(\boldsymbol{\mu}^\top\boldsymbol{\mu})$.*

Theorem 3.1.12 (Quadratic form of any matrices)**.** *If $\boldsymbol{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$,*

1. *if Σ is nonsingular, $B \in \mathbb{R}^{p\times p}$ is symmetric, then $\boldsymbol{x}^\top B\boldsymbol{x} \sim \chi^2_k(\delta)$ iff $B\Sigma$ is idempotent (equiv., $B\Sigma B = B$), in which case $k = \mathrm{r}(B)$ and $\delta = \boldsymbol{\mu}^\top B\boldsymbol{\mu}$;*
2. *for $A \in \mathbb{R}^{p\times p}$, $\boldsymbol{x}^\top A\boldsymbol{x} \sim \chi^2_{\mathrm{r}(A\Sigma)}(\boldsymbol{\mu}^\top A\boldsymbol{\mu})$ if*

$$(1)\Sigma A\Sigma A\Sigma = \Sigma A\Sigma, \quad (2)\boldsymbol{\mu}^\top A\Sigma A\boldsymbol{\mu} = \boldsymbol{\mu}^\top A\boldsymbol{\mu}, \quad (3)\Sigma A\Sigma A\boldsymbol{\mu} = \Sigma A\boldsymbol{\mu};$$

3. *let $A_i \in \mathbb{R}^{p\times p}$ be symmetric with rank k_i for $i = 1, \ldots, m$. Denote $A = \sum_{i=1}^m A_i$, which is symmetric with rank k. Then $\boldsymbol{x}^\top A_i\boldsymbol{x} \sim \chi^2_{k_i}(\boldsymbol{\mu}^\top A_i\boldsymbol{\mu})$, $\boldsymbol{x}^\top A_i\boldsymbol{x}$ are pairwise independent and $\boldsymbol{x}^\top A\boldsymbol{x} \sim \chi^2_k(\boldsymbol{\mu}^\top A\boldsymbol{\mu})$, iff*
   (I) *any two of the following are true: (a) $A_i\Sigma$ idempotent, $\forall i$; (b) $A_i\Sigma A_j = 0$, $\forall i < j$; and (c) $A\Sigma$ idempotent;* ***OR***
   (II) *(c) is true and (d) $k = \sum_{i=1}^m k_i$;* ***OR***
   (III) *(c) is true and (e) $A_1\Sigma, \ldots, A_{m-1}\Sigma$ are idempotent and $A_m\Sigma \succeq \mathbf{0}$.*
4. *(Cochran's theorem) $\boldsymbol{x} \sim \mathrm{N}_p(\mathbf{0}_p, I_p)$ and A_i is symmetric of rank k_i, for $i = 1, \ldots, m$ with $\sum_{i=1}^m A_i = I_p$, then $\boldsymbol{x}^\top A_i\boldsymbol{x} \sim \chi^2_{k_i}$ independently iff $\sum_{i=1}^m k_i = p$.*

3.1.2 The noncentral χ^2 and F distribution

3.2 Asymptotic properties

3.2.1 Asymptotic distributions of sample means and covariance matrices

Refer to section 1.2.2, [5].

Theorem 3.2.1 (CLT for sample means). *Let* $\boldsymbol{x}_{1:n} \sim_{\text{iid}} [\boldsymbol{\mu}, \Sigma]$*, then*

$$\sqrt{n}(\bar{\boldsymbol{x}}_n - \boldsymbol{\mu}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu}) \xrightarrow{\mathrm{d}} \mathrm{N}_p(\mathbf{0}_p, \Sigma).$$

Theorem 3.2.2 (CLT for sample covariance matrices). *Let* $\boldsymbol{x}_{1:n} \sim_{\text{iid}} [\boldsymbol{\mu}, \Sigma]$ *with finite fourth moments, SSCP matrix* $\boldsymbol{A} = \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top$*, and* $\boldsymbol{S} = (n-1)^{-1}\boldsymbol{A}$*. Let* $V = \mathbb{C}\mathrm{ov}[\mathrm{vec}((\boldsymbol{x}_1 - \boldsymbol{\mu})(\boldsymbol{x}_1 - \boldsymbol{\mu})^\top)]$*, then*

$$\frac{1}{\sqrt{n}}(\mathrm{vec}(\boldsymbol{A}) - n \cdot \mathrm{vec}(\Sigma)) \xrightarrow{\mathrm{d}} \mathrm{N}_{p^2}(\mathbf{0}, V),$$

$$\sqrt{n-1}(\mathrm{vec}(\boldsymbol{S}) - \mathrm{vec}(\Sigma)) \xrightarrow{\mathrm{d}} \mathrm{N}_{p^2}(\mathbf{0}, V).$$

Note that $V \in \mathbb{R}^{p^2 \times p^2}$ is singular as the LHS vectors above have repeated elements.

Chapter 4

Linear Models

4.1 Linear regression for the full-rank model

Consider the linear model

$$\boldsymbol{Y}_n = \boldsymbol{X}_{n\times p}\boldsymbol{\beta}_p + \boldsymbol{\varepsilon}_n, \quad \mathrm{r}(\boldsymbol{X}) = p. \tag{4.1}$$

Chapter 5

Generalized Linear Models

Chapter 6

Bayesian Inference

Chapter 7

Structural Equation Model (SEM)

Reference:

- CUHK STAT5020 - Topics in multivariate analysis (2025 Spring), by Xin Yuan SONG.
- Sik-Yum Lee and Xin-Yuan Song - Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences [4].

7.1 SEM models

Goal: to examine the relationships among the variables of interest.
Approach: group observed variables to form laten variables. Advantages:

- Reduce the number of variables compared to direct regression.
- As highly correlated observed variables are grouped into latent variables, the problem induced by multicollinearity is alleviated.
- It gives better assessments on the interrelationships of latent constructs.

Definition 7.1.1 (Linear SEMs)**.** Assume that the observed variables $\boldsymbol{y}_p \sim_{\text{iid}} \mathrm{N}_p$ with mean $\boldsymbol{\mu}_p$. Let $\boldsymbol{\omega}_q$ be latent variables. $\boldsymbol{\omega}_q = (\eta_{q_1}^\top, \boldsymbol{\xi}_{q_2}^\top)^\top$, where η_{q_1} is the key outcome latent variables, and $\boldsymbol{\xi}_{q_2}$ is the explanatory latent variables. Define

$$\begin{aligned} &(\textbf{Measurement equation }) & \boldsymbol{y}_p &= \boldsymbol{\mu}_p + \boldsymbol{\Lambda}_{p\times q}\boldsymbol{\omega}_q + \boldsymbol{\epsilon}_p, \\ &(\textbf{Structural equation}) & \eta_{q_1} &= \boldsymbol{\Gamma}_{q_1\times q_2}\boldsymbol{\xi}_{q_2} + \boldsymbol{\delta}_{q_1}, \end{aligned} \tag{7.1}$$

where $\boldsymbol{\Lambda}$ is the unknown factor loading matrix, $\boldsymbol{\Gamma}$ is the unknown matrix of regression coeffcients, and $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are measurement (residual) errors. An extension:

$$\eta_{q_1} = \boldsymbol{\Pi}_{q_1\times q_1}\eta_{q_1} + \boldsymbol{\Gamma}_{q_1\times q_2}\boldsymbol{\xi}_{q_2} + \boldsymbol{\delta}_{q_1},$$

where $\boldsymbol{\Pi}$ is a matrix of unknown coeffcients such that $I_{q_1} - \boldsymbol{\Pi}$ is nonsingular and the diagonal elements of $\boldsymbol{\Pi}$ are zero.

Assumption 7.1.2 (Standard linear SEMs)**.** For $i = 1, \ldots, n$,
(A1) $\boldsymbol{\epsilon}_i \sim_{\text{iid}} \mathrm{N}[\mathbf{0}_p, \boldsymbol{\Psi}_{\boldsymbol{\epsilon}}]$, where $\boldsymbol{\Psi}_{\boldsymbol{\epsilon}} \in \mathbb{R}^{p\times p}$ is diagonal.
(A2) $\boldsymbol{\xi}_i \sim_{\text{iid}} \mathrm{N}[\mathbf{0}_{q_2}, \boldsymbol{\Phi}]$, where $\boldsymbol{\Phi}$ is a general.
(A3) $\boldsymbol{\delta}_i \sim_{\text{iid}} \mathrm{N}[\mathbf{0}_{q_1}, \boldsymbol{\Psi}_{\boldsymbol{\delta}}]$, where $\boldsymbol{\Psi}_{\boldsymbol{\delta}}$ is diagonal.
(A4) $\boldsymbol{\delta}_i \perp\!\!\!\perp \boldsymbol{\xi}_i$, and $\boldsymbol{\epsilon}_i \perp\!\!\!\perp \boldsymbol{\omega}_i, \boldsymbol{\delta}_i$.

These assumptions imply that

$$\eta_i \sim_{\text{iid}} \mathrm{N}_{q_1}(\mathbf{0}_{q_1}, \boldsymbol{\Gamma\Phi\Gamma}^\top + \boldsymbol{\Psi}_{\boldsymbol{\delta}}), \quad \boldsymbol{\omega}_i \sim_{\text{iid}} \mathrm{N}_q\left(\mathbf{0}_q, \boldsymbol{\Sigma}_{\boldsymbol{\omega}} = \begin{bmatrix} \boldsymbol{\Gamma\Phi\Gamma}^\top + \boldsymbol{\Psi}_{\delta} & \boldsymbol{\Gamma\Phi} \\ \boldsymbol{\Phi\Gamma}^\top & \boldsymbol{\Phi} \end{bmatrix}\right), \quad \boldsymbol{y}_i \sim_{\text{iid}} \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda\Sigma}_{\boldsymbol{\omega}}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}_{\boldsymbol{\epsilon}}).$$

There is an identifiability issue relevant to all SEMs: The measurement equation is identfied if $\forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, $\mathrm{MeaEq}(\boldsymbol{\theta}_1) = \mathrm{MeaEq}(\boldsymbol{\theta}_2)$ implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. The structural equation is identfied if $\forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, $\mathrm{StEq}(\boldsymbol{\theta}_1) = \mathrm{StEq}(\boldsymbol{\theta}_2)$ implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. The SEM is identified if both of its MeaEq and StEq are identfied.

- A simple and common method is using a $\boldsymbol{\Lambda}$ with the non-overlapping structure, e.g.,

$$\boldsymbol{\Lambda}^\top = \begin{bmatrix} 1 & \lambda_{21} & \lambda_{31} & \lambda_{41} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_{62} & \lambda_{72} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{93} & \lambda_{10,3} \end{bmatrix}$$

  where 1's are fixed to introduce a scale to latent variables.

Example 7.1.3. Study the kidney disease of type 2 diabetic patients. We observe: plasma creatine (PCr), urinary albumin creatinine ratio (ACR), systolic blood pressure (SBP), diastolic blood pressure (DBP), body mass index (BMI), waist hip ratio (WHR), glycated hemoglobin (HbAlc), fasting plasma glucose (FPG). Group

- {PCr, ACR}: 'kidney disease (KD)'
- {SBP, DBP}: 'blood pressure (BP)'
- {BMI, WHR}: 'obesity (OB)'
- {HbA1c, FPG}: 'glycemic control (GC)'

$\boldsymbol{y} = (\mathrm{PCr}, \mathrm{ACR}, \mathrm{SBP}, \mathrm{DBP}, \mathrm{BMI}, \mathrm{WHR})^\top$, $\boldsymbol{\omega} = (\mathrm{KD}, \mathrm{BP}, \mathrm{OB})^\top$, $\boldsymbol{\eta} = \mathrm{KD}$, $\boldsymbol{\xi} = (\mathrm{BP}, \mathrm{OB})^\top$, $p = 6$, $q = 3, q_1 = 1, q_2 = 2$. Then the measurement equation is

$$\begin{bmatrix} \mathrm{PCr} \\ \mathrm{ACR} \\ \mathrm{SBP} \\ \mathrm{DBP} \\ \mathrm{BMI} \\ \mathrm{WHR} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ 0 & \lambda_{32} & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & 0 & \lambda_{53} \\ 0 & 0 & \lambda_{63} \end{bmatrix} \begin{bmatrix} \mathrm{KD} \\ \mathrm{BP} \\ \mathrm{OB} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}.$$

We know that KD is only linked with PCr and ACR, ... The structural equation can be defined as:

$$\mathrm{KD} = \gamma_1 \mathrm{BP} + \gamma_2 \mathrm{OB} + \delta.$$

Suppose we wish to study the effects of $\boldsymbol{\xi}$ on $\boldsymbol{\eta} = (\mathrm{KD}, \eta_A)^\top$, $q_1 = 2$,

$$\begin{pmatrix} \mathrm{KD} \\ \eta_A \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \pi & 0 \end{pmatrix} \begin{pmatrix} \mathrm{KD} \\ \eta_A \end{pmatrix} + \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4 \end{pmatrix} \begin{pmatrix} \mathrm{BP} \\ \mathrm{OB} \end{pmatrix} + \begin{pmatrix} \delta \\ \delta_A \end{pmatrix}.$$

Definition 7.1.4 (More SEM models)**.**

- SEMs with fixed covariates

$$\begin{aligned} \boldsymbol{y}_p &= \boldsymbol{A}_{p\times r_1}\boldsymbol{c}_{r_1} + \boldsymbol{\Lambda}_{p\times q}\boldsymbol{\omega}_q + \boldsymbol{\epsilon}_p, \\ \boldsymbol{\eta}_{q_1} &= \boldsymbol{B}_{q_1\times r_2}\boldsymbol{d}_{r_2} + \boldsymbol{\Pi}_{q_1\times q_1}\boldsymbol{\eta}_{q_1} + \boldsymbol{\Gamma}_{q_1\times q_2}\boldsymbol{\xi}_{q_2} + \boldsymbol{\delta}_{q_1}, \end{aligned} \tag{7.2}$$

  where $\boldsymbol{A}, \boldsymbol{B}$ are a matrix of unknown coeffcients, $\boldsymbol{c}, \boldsymbol{d}$ are a vector of fixed covariates (known).

- Nonlinear SEMs

$$\begin{aligned} \boldsymbol{y}_p &= \boldsymbol{\mu}_p + \boldsymbol{\Lambda}_{p\times q}\boldsymbol{\omega}_q + \boldsymbol{\epsilon}_p, \\ \boldsymbol{\eta}_{q_1} &= \boldsymbol{\Pi}_{q_1\times q_1}\boldsymbol{\eta}_{q_1} + \boldsymbol{\Gamma}_{q_1\times t}\boldsymbol{F}_t(\boldsymbol{\xi}_{q_2}) + \boldsymbol{\delta}_{q_1}, \end{aligned} \tag{7.3}$$

  where $\boldsymbol{F}(\boldsymbol{\xi}) = (f_1(\boldsymbol{\xi}), \dots, f_t(\boldsymbol{\xi}))^\top$ with nonzero, known, and linearly independent differentiable functions $f_1, \dots, f_t$, $t \geq q_2$.

- Nonlinear SEMs with fixed covariates

$$\begin{aligned} \boldsymbol{y}_p &= \boldsymbol{A}_{p\times r_1}\boldsymbol{c}_{r_1} + \boldsymbol{\Lambda}_{p\times q}\boldsymbol{\omega}_q + \boldsymbol{\epsilon}_p, \\ \boldsymbol{\eta}_{q_1} &= \boldsymbol{B}_{q_1\times r_2}\boldsymbol{d}_{r_2} + \boldsymbol{\Pi}_{q_1\times q_1}\boldsymbol{\eta}_{q_1} + \boldsymbol{\Gamma}_{q_1\times t}\boldsymbol{F}_t(\boldsymbol{\xi}_{q_2}) + \boldsymbol{\delta}_{q_1}. \end{aligned} \tag{7.4}$$

  A simple extension of the StEq is

$$\boldsymbol{\eta}_{q_1} = \boldsymbol{\Pi}_{q_1\times q_1}\boldsymbol{\eta}_{q_1} + \boldsymbol{\Lambda}_{\boldsymbol{\omega}}\boldsymbol{G}_t(\boldsymbol{d}, \boldsymbol{\xi}) + \boldsymbol{\delta}_{q_1}$$

  where $\boldsymbol{G}(\boldsymbol{d}\boldsymbol{\xi}) = (g_1(\boldsymbol{d}, \boldsymbol{\xi}), \dots, g_t(\boldsymbol{d}, \boldsymbol{\xi}))^\top$ is a vector-valued function with nonzero, known, and linearly independent differentiable functions.

Let $\boldsymbol{\Lambda}_k^\top$ be the kth row of $\boldsymbol{\Lambda}$, and $\boldsymbol{\Lambda}_k^\top = (\boldsymbol{\Lambda}_{k\boldsymbol{\eta}}^\top, \boldsymbol{\Lambda}_{k\boldsymbol{\xi}}^\top)$ be a partition correspondings to the partition of $\boldsymbol{\omega} = (\boldsymbol{\eta}^\top, \boldsymbol{\xi}^\top)^\top$. For model (7.3),

$$\begin{aligned} \mathbb{E}(\boldsymbol{\xi}) = \mathbf{0}_{q_2}, \quad \mathbb{E}(\boldsymbol{\eta}) = [(I_{q_1} - \boldsymbol{\Pi})^{-1}\boldsymbol{\Gamma}]\mathbb{E}(\boldsymbol{F}(\boldsymbol{\xi})), \\ \mathbb{E}(y_k) = \mu_k + \boldsymbol{\Lambda}_{k\boldsymbol{\eta}}^\top[(I_{q_1} - \boldsymbol{\Pi})^{-1}\boldsymbol{\Gamma}]\mathbb{E}(\boldsymbol{F}(\boldsymbol{\xi})). \end{aligned}$$

For model (7.4), let $\boldsymbol{A}_k^\top$ be the kth row of $\boldsymbol{A}$,

$$\mathbb{E}(y_k) = \boldsymbol{A}_k^\top\boldsymbol{c} + \boldsymbol{\Lambda}_{k\boldsymbol{\eta}}^\top\mathbb{E}(\boldsymbol{\eta}) = \boldsymbol{A}_k^\top\boldsymbol{c} + \boldsymbol{\Lambda}_{k\boldsymbol{\eta}}^\top[(I_{q_1} - \boldsymbol{\Pi})^{-1}\boldsymbol{\Lambda}_{\boldsymbol{\omega}}]\mathbb{E}(\boldsymbol{G}(\boldsymbol{d}, \boldsymbol{\xi})).$$

7.2 Bayesian methods for estimating SEM

In developing the Bayesian methods for analyzing SEMs, we usually assign fixed known values to the hyperparameters in the conjugate prior distributions. Consider the modified model (7.4):

$$\begin{aligned}\boldsymbol{y}_i &= \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i,\\ \boldsymbol{\eta}_i &= \boldsymbol{B}\boldsymbol{d}_i + \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i = \boldsymbol{\Lambda}_{\boldsymbol{\omega}}\boldsymbol{G}(\boldsymbol{\omega}_i) + \boldsymbol{\delta}_i,\end{aligned}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{\omega}} = (\boldsymbol{B}, \boldsymbol{\Pi}, \boldsymbol{\Gamma}) \in \mathbb{R}^{q_1\times(r_2+q_1+t)}$, and $\boldsymbol{G}(\boldsymbol{\omega}_i) = (\boldsymbol{d}_i^\top, \boldsymbol{\eta}_i^\top, \boldsymbol{F}(\boldsymbol{\xi}_i)^\top)^\top \in \mathbb{R}^{r_2+q_1+t}$. Assumption 7.1.2 is satisfied. $\boldsymbol{\xi}_i \sim_{\text{iid}} \mathrm{N}[\boldsymbol{0}_{q_2}, \boldsymbol{\Phi}]$, $\boldsymbol{\epsilon}_i \sim_{\text{iid}} \mathrm{N}[\boldsymbol{0}_p, \boldsymbol{\Psi}_{\boldsymbol{\epsilon}} = \mathrm{diag}(\psi_{\boldsymbol{\epsilon}k})]$,and $\boldsymbol{\delta}_i \sim_{\text{iid}} \mathrm{N}[\boldsymbol{0}_{q_1}, \boldsymbol{\Psi}_{\boldsymbol{\delta}} = \mathrm{diag}(\psi_{\boldsymbol{\delta}k})]$.

- Prior (conjugate) for $\boldsymbol{\theta}_{\boldsymbol{y}} = (\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}_{\boldsymbol{\epsilon}})$: let $\boldsymbol{\Lambda}_k^\top$ be the kth row of $\boldsymbol{\Lambda}$,

$$\begin{aligned}&\psi_{\boldsymbol{\epsilon}k} \sim \mathrm{IG}(\alpha_{0\boldsymbol{\epsilon}k}, \beta_{0\boldsymbol{\epsilon}k}), \quad [\boldsymbol{\Lambda}_k \mid \psi_{\boldsymbol{\epsilon}k}] \sim \mathrm{N}_q(\boldsymbol{\Lambda}_{0k}, \psi_{\boldsymbol{\epsilon}k}\boldsymbol{H}_{0\boldsymbol{y}k}), \; k = 1, \ldots, p,\\ &\boldsymbol{\mu} \sim \mathrm{N}_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0).\end{aligned}$$

- Prior (conjugate) for $\boldsymbol{\theta}_{\boldsymbol{\omega}} = (\boldsymbol{\Lambda}_{\boldsymbol{\omega}}, \boldsymbol{\Psi}_{\boldsymbol{\delta}}, \boldsymbol{\Phi})$: let $\boldsymbol{\Lambda}_{\boldsymbol{\omega}k}^\top$ be the kth row of $\boldsymbol{\Lambda}_{\boldsymbol{\omega}}$, $\boldsymbol{\Psi}_{\boldsymbol{\delta}}$,

$$\begin{aligned}&\boldsymbol{\Phi} \sim \mathrm{IW}_{q_2}(\boldsymbol{R}_0^{-1}, \rho_0), \text{ or } \boldsymbol{\Phi}^{-1} \sim \mathrm{W}_{q_2}(\boldsymbol{R}_0, \rho_0)\\ &\psi_{\boldsymbol{\delta}k} \sim \mathrm{IG}(\alpha_{0\boldsymbol{\delta}k}, \beta_{0\boldsymbol{\delta}k}), \quad [\boldsymbol{\Lambda}_{\boldsymbol{\omega}k} \mid \psi_{\boldsymbol{\delta}k}] \sim \mathrm{N}_{r_2+q_1+t}(\boldsymbol{\Lambda}_{0\boldsymbol{\omega}k}, \psi_{\boldsymbol{\delta}k}\boldsymbol{H}_{0\boldsymbol{\omega}k}), \; k = 1, \ldots, q_1.\end{aligned}$$

- Assume the prior $\boldsymbol{\theta}_{\boldsymbol{y}} \perp\!\!\!\perp \boldsymbol{\theta}_{\boldsymbol{\omega}}$.

Hyperparameter selection: If we have good prior information about a parameter – select the prior distribution with a small variance. E.g.,

- if $\boldsymbol{\Lambda}_k \approx \boldsymbol{\Lambda}_{0k}$, then $\boldsymbol{H}_{0yk} = 0.5I_q$. If not, select the prior with a larger variance;
- since $\epsilon_{ik} \sim \mathrm{N}(0, \psi_{\boldsymbol{\epsilon}k})$, if the variation is small, $\psi_{\boldsymbol{\epsilon}k}$ is small, then choose small $\mathbb{E}(\psi_{\boldsymbol{\epsilon}k}) = \beta_{0\boldsymbol{\epsilon}k}/(\alpha_{0\boldsymbol{\epsilon}k} - 1)$ and $\mathbb{V}\mathrm{ar}(\psi_{\boldsymbol{\epsilon}k}) = \beta_{0\boldsymbol{\epsilon}k}^2/\{(\alpha_{0\boldsymbol{\epsilon}k} - 1)^2(\alpha_{0\boldsymbol{\epsilon}k} - 2)\}$;
- if $\boldsymbol{\Phi} \approx \boldsymbol{\Phi}_0$, since $\mathbb{E}(\boldsymbol{\Phi}) = \boldsymbol{R}_0^{-1}/(\rho_0 - q_2 - 1)$, choose $\boldsymbol{R}_0^{-1} = (\rho_0 - q_2 - 1)\boldsymbol{\Phi}_0$.

If the sample size is large, can use a portion of the data to estimate $\boldsymbol{\Lambda}_{0k}$, $\boldsymbol{\Lambda}_{0\boldsymbol{\omega}k}$ and $\boldsymbol{\Phi}_0$. If the sample size is moderate, can use the same data twice.

Noninformative prior (Jeffrey): If information is not available and the sample size is small,

$$\begin{aligned}&\mathbb{P}(\boldsymbol{\Lambda}, \boldsymbol{\Psi}_{\boldsymbol{\epsilon}}) \propto \mathbb{P}(\psi_{\boldsymbol{\epsilon}1}, \cdots, \psi_{\boldsymbol{\epsilon}p}) \propto \prod_{k=1}^{p} \psi_{\boldsymbol{\epsilon}k}^{-1}, \quad \mathbb{P}(\boldsymbol{\Lambda}_{\boldsymbol{\omega}}, \boldsymbol{\Psi}_{\boldsymbol{\delta}}) \propto \mathbb{P}(\psi_{\boldsymbol{\delta}1}, \cdots, \psi_{\boldsymbol{\delta}q_1}) \propto \prod_{k=1}^{q_1} \psi_{\boldsymbol{\delta}k}^{-1},\\ &\mathbb{P}(\boldsymbol{\Phi}) \propto |\boldsymbol{\Phi}|^{-(q_2+1)/2}.\end{aligned}$$

7.2.1 Bayesian estimation using MCMC

Model: Linear SEM with fixed covariates (7.2) without intercept:

$$\begin{aligned}\boldsymbol{y}_i &= \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i,\\ \boldsymbol{\eta}_i &= \boldsymbol{B}\boldsymbol{d}_i + \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i = \boldsymbol{\Lambda}_{\boldsymbol{\omega}}\boldsymbol{v}_i + \boldsymbol{\delta}_i,\end{aligned}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{\omega}} = (\boldsymbol{B}, \boldsymbol{\Pi}, \boldsymbol{\Gamma}) \in \mathbb{R}^{q_1\times(r_2+q_1+q_2)}$, and $\boldsymbol{v}_i = (\boldsymbol{d}_i^\top, \boldsymbol{\eta}_i^\top, \boldsymbol{\xi}_i^\top)^\top \in \mathbb{R}^{r_2+q_1+q_2}$. That is, assume $\boldsymbol{\mu} = \boldsymbol{0}_p$ and $\boldsymbol{F}(\boldsymbol{\xi}_i) - \boldsymbol{\xi}_i$.

Denote data $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_p)^\top \in \mathbb{R}^{p\times n}$, $\boldsymbol{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) = (\boldsymbol{V}_1, \ldots, \boldsymbol{V}_{r_2+q_1+q_2})^\top$, $\boldsymbol{\Xi}_k = (\eta_{1k}, \cdots, \eta_{nk})^\top$ for $k = 1, \ldots, q_1$, matrix of latent variables $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n) \in \mathbb{R}^{q\times n}$, $\boldsymbol{\Omega}_1 = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n)$, $\boldsymbol{\Omega}_2 = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n)$, and

$$\boldsymbol{\theta} = (\boldsymbol{\Lambda}, \boldsymbol{B}, \boldsymbol{\Pi}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_{\boldsymbol{\epsilon}}, \boldsymbol{\Psi}_{\boldsymbol{\delta}}) = (\underbrace{\boldsymbol{\Lambda}, \boldsymbol{\Psi}_{\boldsymbol{\epsilon}}}_{\boldsymbol{\theta}_{\boldsymbol{y}}}, \underbrace{\boldsymbol{\Lambda}_{\boldsymbol{\omega}}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_{\boldsymbol{\delta}}}_{\boldsymbol{\theta}_{\boldsymbol{\omega}}}).$$

Proposition 7.2.1. *The above model has the following posterior distributions:*

1. *Conditional distribution* $\mathbb{P}(\boldsymbol{\Omega} \mid \boldsymbol{Y}, \boldsymbol{\theta}) = \prod_{i=1}^n \mathbb{P}(\boldsymbol{\omega}_i \mid \boldsymbol{y}_i, \boldsymbol{\theta}) \propto \prod_{i=1}^n \mathbb{P}(\boldsymbol{\omega}_i \mid \boldsymbol{\theta})\mathbb{P}(\boldsymbol{y}_i \mid \boldsymbol{\omega}_i, \boldsymbol{\theta})$, *where*

$$\begin{aligned}&[\boldsymbol{\omega}_i \mid \boldsymbol{\theta}] \sim \mathrm{N}_q(\boldsymbol{\mu}_{\boldsymbol{\omega}_i}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}}), \quad [\boldsymbol{y}_i \mid \boldsymbol{\omega}_i, \boldsymbol{\theta}] \sim \mathrm{N}_p(\boldsymbol{\Lambda}\boldsymbol{\omega}_i, \boldsymbol{\Psi}_{\boldsymbol{\epsilon}}),\\ &[\boldsymbol{\omega}_i \mid \boldsymbol{y}_i, \boldsymbol{\theta}] \sim \mathrm{N}_q({\boldsymbol{\Sigma}^*}^{-1}(\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\omega}_i} + \boldsymbol{\Lambda}^\top\boldsymbol{\Psi}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{y}_i), {\boldsymbol{\Sigma}^*}^{-1})\end{aligned}$$

where

$$\begin{aligned}&\boldsymbol{\Pi}_0 = I_{q_1} - \boldsymbol{\Pi}, \; \boldsymbol{\mu}_{\boldsymbol{\omega}_i} = \begin{pmatrix}\boldsymbol{\Pi}_0^{-1}\boldsymbol{B}\boldsymbol{d}_i\\ \boldsymbol{0}_{q_2}\end{pmatrix}, \; \boldsymbol{\Sigma}_{\boldsymbol{\omega}} = \begin{bmatrix}\boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^\top + \boldsymbol{\Psi}_{\boldsymbol{\delta}})\boldsymbol{\Pi}_0^{-\top} & \boldsymbol{\Pi}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\\ \boldsymbol{\Phi}\boldsymbol{\Gamma}^\top\boldsymbol{\Pi}_0^{-\top} & \boldsymbol{\Phi}\end{bmatrix},\\ &\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1} + \boldsymbol{\Lambda}^\top\boldsymbol{\Psi}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\Lambda}.\end{aligned}$$

2. *Assume all elements of* $\boldsymbol{\Lambda}_k$ *and* $\boldsymbol{\Lambda}_{\boldsymbol{\omega}}$ *are unknown, conditional distribution* $\mathbb{P}(\boldsymbol{\theta} \mid \boldsymbol{Y}, \boldsymbol{\Omega}) = \mathbb{P}(\boldsymbol{\theta}_{\boldsymbol{y}} \mid \boldsymbol{Y}, \boldsymbol{\Omega})\mathbb{P}(\boldsymbol{\theta}_{\boldsymbol{\omega}} \mid$

$\boldsymbol{Y}, \boldsymbol{\Omega})$, *where we can show* $[\boldsymbol{\Lambda}_k, \psi_{\boldsymbol{\epsilon} k} \mid \boldsymbol{Y}, \boldsymbol{\Omega}] \perp\!\!\!\perp$ *and* $[\boldsymbol{\Lambda}_{\boldsymbol{\omega} k}, \psi_{\boldsymbol{\delta} k} \mid \boldsymbol{\Omega}] \perp\!\!\!\perp$, *and*

$$\mathbb{P}(\boldsymbol{\theta}_{\boldsymbol{y}} \mid \boldsymbol{Y}, \boldsymbol{\Omega}) \propto \prod_{k=1}^{p} \mathbb{P}(\boldsymbol{\Lambda}_k, \psi_{\boldsymbol{\epsilon} k} \mid \boldsymbol{Y}, \boldsymbol{\Omega}),$$

$$\mathbb{P}(\boldsymbol{\theta}_{\boldsymbol{\omega}} \mid \boldsymbol{Y}, \boldsymbol{\Omega}) \propto \left[\prod_{k=1}^{q_1} \mathbb{P}(\boldsymbol{\Lambda}_{\boldsymbol{\omega} k}, \psi_{\boldsymbol{\delta} k} \mid \boldsymbol{\Omega})\right] \mathbb{P}(\boldsymbol{\Phi} \mid \boldsymbol{\Omega}_2)$$

where

$$\begin{aligned}
\mathbb{P}(\boldsymbol{\Lambda}_k, \psi_{\boldsymbol{\epsilon} k}^{-1} \mid \boldsymbol{Y}, \boldsymbol{\Omega}) &\propto \mathrm{N}_q(\boldsymbol{a}_k, \psi_{\boldsymbol{\epsilon} k} \boldsymbol{A}_k) \cdot \mathrm{Ga}(n/2 + \alpha_{0\boldsymbol{\epsilon} k}, \beta_{\boldsymbol{\epsilon} k}), \\
\mathbb{P}(\boldsymbol{\Lambda}_{\boldsymbol{\omega} k}, \psi_{\boldsymbol{\delta} k}^{-1} \mid \boldsymbol{\Omega}) &\propto \mathrm{N}_{r_2+q_1+q_2}(\boldsymbol{a}_{\boldsymbol{\omega} k}, \psi_{\boldsymbol{\delta} k} \boldsymbol{A}_{\boldsymbol{\omega} k}) \cdot \mathrm{Ga}(n/2 + \alpha_{0\boldsymbol{\delta} k}, \beta_{\boldsymbol{\delta} k}), \\
[\boldsymbol{\Phi} \mid \boldsymbol{\Omega}_2] &\sim \mathrm{IW}_{q_2}[(\boldsymbol{\Omega}_2 \boldsymbol{\Omega}_2^\top + \boldsymbol{R}_0^{-1}), n + \rho_0].
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{A}_k &= (\boldsymbol{H}_{0yk}^{-1} + \boldsymbol{\Omega}\boldsymbol{\Omega}^\top)^{-1}, \ \boldsymbol{a}_k = \boldsymbol{A}_k(\boldsymbol{H}_{0yk}^{-1} \boldsymbol{\Lambda}_{0k} + \boldsymbol{\Omega} \boldsymbol{Y}_k), \\
\beta_{\boldsymbol{\epsilon} k} &= \beta_{0\boldsymbol{\epsilon} k} + \frac{1}{2}(\boldsymbol{Y}_k^\top \boldsymbol{Y}_k - \boldsymbol{a}_k^\top \boldsymbol{A}_k^{-1} \boldsymbol{a}_k + \boldsymbol{\Lambda}_{0k}^\top \boldsymbol{H}_{0yk}^{-1} \boldsymbol{\Lambda}_{0k}), \\
\boldsymbol{A}_{\boldsymbol{\omega} k} &= (\boldsymbol{H}_{0\boldsymbol{\omega} k}^{-1} + \boldsymbol{V}_k \boldsymbol{V}_k^\top)^{-1}, \ \boldsymbol{a}_{\boldsymbol{\omega} k} = \boldsymbol{A}_{\boldsymbol{\omega} k}(\boldsymbol{H}_{0\boldsymbol{\omega} k}^{-1} \boldsymbol{\Lambda}_{0\boldsymbol{\omega} k} + \boldsymbol{V}_k \boldsymbol{\Xi}_k), \\
\beta_{\boldsymbol{\delta} k} &= \beta_{0\boldsymbol{\delta} k} + \frac{1}{2}(\boldsymbol{\Xi}_k^\top \boldsymbol{\Xi}_k - \boldsymbol{a}_{\boldsymbol{\omega} k}^\top \boldsymbol{A}_{\boldsymbol{\omega} k}^{-1} \boldsymbol{a}_{\boldsymbol{\omega} k} + \boldsymbol{\Lambda}_{0\boldsymbol{\omega} k}^\top \boldsymbol{H}_{0\boldsymbol{\omega} k}^{-1} \boldsymbol{\Lambda}_{0\boldsymbol{\omega} k}).
\end{aligned}$$

Remark 7.2.2. For general nonlinear SEMs with fixed covariates (7.4), we can defind $\boldsymbol{u} = [\boldsymbol{c}^\top, \boldsymbol{\omega}^\top]^\top \in \mathbb{R}^{r_1+q}$ and use similar procedure to derive the full conditional distributions. But by the nonlinear structure $\boldsymbol{G}(\boldsymbol{\omega})$, $\mathbb{P}(\boldsymbol{\Omega} \mid \boldsymbol{Y}, \boldsymbol{\theta})$ may not have closed form like normal, while $\mathbb{P}(\boldsymbol{\theta} \mid \boldsymbol{Y}, \boldsymbol{\Omega})$ is not affected and keeps normal-Gamma. To handle fixed parameters, see Appendix 3.3 in [4] and Sec 4.3.1 and Appendix 4.3 in [3]. Also see STAT5020 HW1 Q3.

7.3 Hierarchical and Multisample Data

Definition 7.3.1 (Two-level nonlinear SEM with mixed type variables)**.** Consider a collection of p-variate random vectors $\boldsymbol{u}_{gi}$, $i = 1, \ldots, N_g$, nested within groups $g = 1, \ldots, G$.

$$\begin{aligned}
&\text{(Within-groups) } \boldsymbol{u}_{gi} = \boldsymbol{v}_g + \boldsymbol{\Lambda}_{1g} \boldsymbol{\omega}_{1gi} + \boldsymbol{\epsilon}_{1gi}, \quad g = 1, \cdots, G, \quad i = 1, \cdots, N_g, \\
&\text{(Between-groups) } \boldsymbol{v}_g = \boldsymbol{\mu} + \boldsymbol{\Lambda}_2 \boldsymbol{\omega}_{2g} + \boldsymbol{\epsilon}_{2g}, \quad g = 1, \ldots, G
\end{aligned}$$

where $\boldsymbol{\Lambda}_{1g} \in \mathbb{R}^{p \times q_1}$, $\boldsymbol{\epsilon}_{1gi} \in \mathbb{R}^{q_1}$, $\boldsymbol{\epsilon}_{1gi} \in \mathbb{R}^p \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi}_{1g})$ independent of $\boldsymbol{\omega}_{1gi}$, where $\boldsymbol{\Psi}_{1g}$ is diagonal. Then

$$\mathbf{u}_{\mathrm{gi}} = \mu + \boldsymbol{\Lambda}_2 \omega_{2\mathrm{g}} + \epsilon_{2\mathrm{g}} + \boldsymbol{\Lambda}_{1\mathrm{g}} \omega_{1\mathrm{gi}} + \epsilon_{1\mathrm{gi}}.$$

The SEM is

$$\begin{aligned}
\eta_{1gi} &= \Pi_{1g} \eta_{1gi} + \Gamma_{1g} \mathbf{F}_1(\xi_{1gi}) + \delta_{1gi}, \\
\eta_{2g} &= \Pi_2 \eta_{2g} + \Gamma_2 \mathbf{F}_2(\xi_{2g}) + \delta_{2g},
\end{aligned}$$

Bibliography

[1] G. Casella and R. L. Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002. 2, 2.1

[2] R. Christensen et al. *Plane answers to complex questions*, volume 35. Springer, 2002. 3

[3] S.-Y. Lee. *Structural equation modeling: A Bayesian approach.* John Wiley & Sons, 2007. 7.2.2

[4] S.-Y. Lee and X.-Y. Song. *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences.* John Wiley & Sons, 2012. 7, 7.2.2

[5] R. J. Muirhead. *Aspects of multivariate statistical theory.* John Wiley & Sons, 1982. 3, 3.1, 3.1.1, 3.2.1