

行列分解因子分析の正体について

大阪大学大学院基礎工学研究科，理化学研究所革新知能統合研究センター 寺田 吉壺

1 はじめに

因子分析では，最尤法をはじめとして，様々な推定方法が提案されており，それらの統計的性質は詳細に議論されている (e.g., Anderson and Rubin, 1956). 因子分析は， p 次元の観測 x に対して，以下のような統計モデルを考える．

$$x = \mu + \Lambda f + \epsilon$$

ここで， $\mu \in \mathbb{R}^p$ は母平均ベクトル， $\Lambda \in \mathbb{R}^{p \times m}$ は因子負荷行列， m ($< p$) は因子数， f は共分散行列が単位行列の中心化された m 次元確率ベクトル， ϵ は f とは無相関な中心化された p 次元確率ベクトルである． f と ϵ は，それぞれ共通因子，独自因子と呼ばれ，独自因子の各要素は無相関であると仮定する．観測を中心化することで，以下では $\mu = 0_p$ とする．独自因子の共分散行列を $\text{Var}(\epsilon) = \Psi^2$ と記すと，因子分析モデル（直交モデル）の下で x の共分散行列は，

$$\text{Var}(x) = \Lambda \Lambda^\top + \Psi^2$$

と表される．多くの因子分析法は，標本共分散行列 $\hat{S}_n$ とモデル化した共分散行列 $\Sigma(\Phi) := \Lambda \Lambda^\top + \Psi^2$ の差を最小化することでパラメータ $\Phi := [\Lambda, \Psi]$ を推定する方法として定式化される．このような方法で得られる推定量は，MDF 推定量 (Minimum Discrepancy Function estimator) と呼ばれる．

2 行列分解因子分析 (MDFA)

2000 年代はじめに，従来の因子分析法とは大きく異なる行列分解に基づく因子分析が提案された．この方法は，行列分解因子分析 (**Matrix Decomposition Factor Analysis, MDFA**) と呼ばれる． n 個の観測 $x_1, \dots, x_n$ から構成されるデータ行列 $X_n = (x_1, \dots, x_n)^\top$ に対して，MDFA は以下の損失関数の最小化として定式化される．

$$\mathcal{L}_n(\Lambda, \Psi, F, E) := \frac{1}{n} \|X_n - (F \Lambda^\top + E \Psi)\|_F^2 = \frac{1}{n} \sum_{i=1}^n \|x_i - (\Lambda f_i + \Psi e_i)\|^2$$

ここで， $F = (f_1, \dots, f_n)^\top \in \mathbb{R}^{n \times m}$ と $E = (e_1, \dots, e_n)^\top \in \mathbb{R}^{n \times p}$ は以下の制約を満たす行列である．

$$1_n^\top F = 0_m^\top, \quad 1_n^\top E = 0_p^\top, \quad \frac{1}{n} F^\top F = I_m, \quad \frac{1}{n} E^\top E = I_p, \quad \text{and} \quad F^\top E = O_{m \times p}$$

以下では，上述の制約を満たす Z の集合を Θ_Z と記す．MDFA は，シンプルな交互最適化アルゴリズムによって，容易かつ安定的に解くことができる．MDFA の詳細に関しては，Adachi and Trendafilov (2018) を参照されたい．

3 行列分解因子分析の正体

MDFA は, サンプルサイズ n に依存する F と E をパラメータに含んでおり, 以下のセミパラメトリック・モデルに対する最尤推定量として捉えることができる.

$$x_i = \Lambda f_i + \Psi e_i + \xi_i \quad (i = 1, \dots, n)$$

ここで, $f_1, \dots, f_n, e_1, \dots, e_n$ は上述の制約を満たす定数ベクトルであり, $\xi_1, \dots, \xi_n$ は独立に分散既知の p 次元の正規分布 $N_p(0_p, \tau_0^2 I_p)$ に従う確率ベクトルである. このセミパラメトリック最尤推定において $Z = [F, E]$ を局外母数と考え, 有限次元パラメータ $\Phi = [\Lambda, \Psi]$ に対する新たなプロファイル尤度を考えることで, 行列分解因子分析の正体が明らかとなる.

Proposition. 任意の $\Phi \in \Theta_\Phi$ に対して,

$$\min_{Z \in \Theta_Z} \mathcal{L}_n(\Phi, Z) = \text{tr}(\hat{S}_n) + \text{tr}\{\Sigma(\Phi)\} - 2\text{tr}\left\{\left(\hat{S}_n^{1/2}\Sigma(\Phi)\hat{S}_n^{1/2}\right)^{1/2}\right\} =: d_{\text{BW}}^2(\hat{S}_n, \Sigma(\Phi))$$

が成り立つ. ここで, $\Sigma(\Phi) := \Phi\Phi^T$ であり, $d_{\text{BW}}(A, B)$ は正定値行列 A と B の *Bures-Wasserstein* 距離 (e.g., *Bhatia et al. (2019)*) を表す.

この命題は, MDFA 推定量が Bures-Wasserstein 距離に基づく MDF 推定量として定式化できることを示している. 2つの正定値行列 Σ_1 と Σ_2 に対して, $d_{\text{BW}}(\Sigma_1, \Sigma_2)$ は Σ_1 と Σ_2 を共分散行列にもつ中心化された2つ楕円分布間の L_2 -Wasserstein 距離を表している (Gelbrich, 1990). より具体的には, 以下の密度関数で定義される楕円分布を考える.

$$f(x \mid \mu, \Sigma) := \frac{1}{|\Sigma|^{1/2}} g\left((x - \mu)^\top \Sigma^{-1}(x - \mu)\right).$$

ここで, μ は平均ベクトル, Σ は共分散行列である. このとき, 2つの楕円分布 $f_1(\cdot) = f(\cdot \mid \mu_1, \Sigma_1)$ と $f_2(\cdot) = f(\cdot \mid \mu_2, \Sigma_2)$ の間の L_2 -Wasserstein 距離は以下で与えられる.

$$W_2^2(f_1, f_2) := \inf_{X_1 \sim f_1, X_2 \sim f_2} \mathbb{E}[\|X_1 - X_2\|_2^2] = \|\mu_1 - \mu_2\|_2^2 + \text{tr}\left(\Sigma_1 + \Sigma_2 - 2\{\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}\}^{1/2}\right).$$

興味深い事実として, L_2 -Wasserstein 距離は関数 g に依存しない. つまり, MDFA 推定量は一般の楕円対称分布族を想定した頑健な推定量と捉えることができる.

参考文献

- [1] Anderson, T. and Rubin, H. (1956). Statistical Inference in Factor Analysis. *In Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability 111*, University of California Press.
- [2] Adachi, K. and Trendafilov, N. T. (2018). Some mathematical properties of the matrix decomposition solution in factor analysis. *Psychometrika*, **83**, 407–424.
- [3] Bhatia, R., Jain, T. and Lim, Y. (2019). On the Bures-Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, **37**, 165–191.
- [4] Gelbrich, M. (1990). On a formula for the L^2 -Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, **147**, 186–203.