

3. Inter-variable Statistics

樋口さぶろお

龍谷大学大学院理工学研究科数理情報学専攻

理論物理学特論 L02(2020-10-05 Mon)

最終更新: Time-stamp: "2020-10-05 Mon 18:55 JST hig"

今日の目標

- 共分散行列, 相関行列が行列記法で書ける



解答例

KA MDA §3.2 共分散 Covariance

共分散 (行列) covariance matrix

$$\begin{aligned}
 v_{jk} &= \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \\
 &= \frac{1}{n} (\mathbf{J}\mathbf{x}_j)' \mathbf{J}\mathbf{x}_k \\
 &= \frac{1}{n} \mathbf{x}_j' \mathbf{J}\mathbf{x}_k \\
 &= \frac{1}{n} \mathbf{y}_j' \mathbf{y}_k
 \end{aligned}$$

性質 中心化スコア $\mathbf{y} = \mathbf{J}\mathbf{x}$ の共分散も同じ. KA MDA (3.6)

相関係数 Correlation Coefficient

$$\begin{aligned}
 r_{jk} &= \frac{v_{jk}}{\sqrt{v_{jj}}\sqrt{v_{kk}}} \\
 &= \frac{(\mathbf{J}\mathbf{x}_j)'(\mathbf{J}\mathbf{x}_k)}{\|\mathbf{J}\mathbf{x}_j\|\|\mathbf{J}\mathbf{x}_k\|} \\
 &= \frac{\mathbf{y}'_j\mathbf{y}_k}{\|\mathbf{y}_j\|\|\mathbf{y}_k\|} =: \cos\theta_{jk}
 \end{aligned}$$

どこかで見た式?

$$-1 \leq r_{jk} \leq +1$$

標準化変数の共分散, 相関係数は, もとの変数の相関係数に等しい.

データ行列

変数 x_{ij} , $j = 1, \dots, 8$ それぞれの平均 $\bar{x}_j = \frac{1}{n} \mathbf{1}' \mathbf{x}_j$ を, データ行列 \mathbf{X} から一斉に計算するには,

$$\frac{1}{n} \mathbf{1}' \mathbf{X} \quad \text{KA MDA (2.28)}$$

データ行列 \mathbf{X} を一斉に中心化するには, $\mathbf{Y} = \mathbf{JX}$. KA MDA (2.29)

データ行列 \mathbf{X} を一斉に標準化するには, $\mathbf{Z} = \mathbf{JXD}^{-1}$. KA MDA (2.32)

$p \times p$ 対角行列 $\mathbf{D} = \text{diag}(\sqrt{v_{11}}, \dots, \sqrt{v_{pp}})$.

共分散行列と相関係数の行列

$p \times p$ 共分散行列 $V = \frac{1}{n} X' J X = \frac{1}{n} Y' Y$.

$p \times p$ 相関係数の行列 $R = \frac{1}{n} D^{-1} X' J X D^{-1}$.

KA MDA §3.7 不偏標本分散

不偏 (unbiased) 標本分散を求めるには $\frac{n}{n-1}$ 倍しとけ.
相関係数は, 元の (biased), 不偏 (unbiased) どちらから計算しても同じ.

データ行列 X が中心化されている $\stackrel{\text{def}}{\Leftrightarrow}$

$$\mathbf{1}'X = \mathbf{0}'.$$

KA MDA Note 3.1 $\mathbf{1}'F = \mathbf{0}' \Leftrightarrow JF = F.$

individual の数 n が等しい2つのデータセット $X, F (n \times p, n \times m)$ の間の
相関行列

$$V_{XF} = \frac{1}{n}X'JF. \quad \text{KA MDA (3.24)}$$

KA MDA §3.9 階数

略

問

Exercises 3.4, 3.5, 3.6, and 3.7.