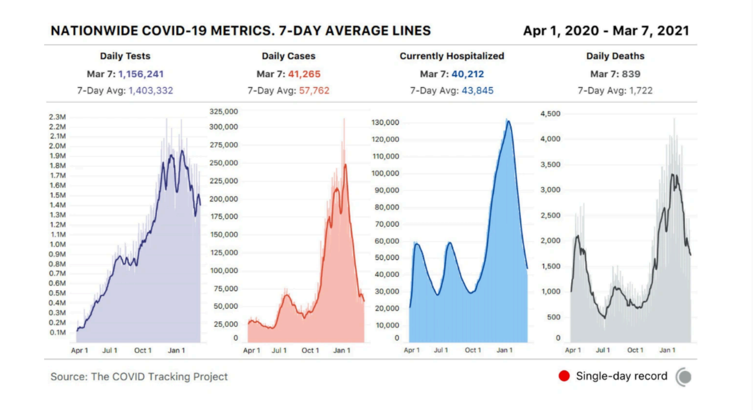
Mathematical Statistics II

STA2212H S LEC9101

Week 10

March2 26 2021

Start recording!



link

Today

Start Recording

1. Calendar, HW, Final HW ← 11 homeworks || → April 9 – (7?) ← 19th
2.
3. Stein's paradox – baseball example
4. multivariate distributions (Ch 14 AoS, SM 6.3)
5. intro to graphical models (SM 6.3)

- Mar 29 3.00 – 4.00 pm EDT
  Data Science ARES
  John Aston, University of Cambridge
  "Functional Data in Constrained Spaces"
- Professor of Statistics in Public Life;
  formerly Chief Scientific Adviser
  to the Home Office, UK Government.

- $X \sim N_k(\mu, \Omega)$ with density

$$f(x;\mu,\Omega) = \frac{1}{\sqrt{2\pi}^k}|\Omega|^{-1/2}\exp\{-\frac{1}{2}(x-\mu)^T\Omega^{-1}(x-\mu)\}$$

single obs :

- $X_1,\ldots,X_n$ i.i.d. $N(\mu,\Omega)$; log-likelihood functon

$$\ell(\mu,\Omega;x) = -\frac{n}{2}\log|\Omega| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Omega^{-1}(x_i-\mu)$$

sample

- let $\Delta = \Omega^{-1}$, then

$$\ell(\mu,\Delta;x) = \frac{n}{2}\log|\Delta| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Delta(x_i-\mu),$$

$\hat{\mu} = \bar{x}$

$$\ell(\widehat{\mu},\Delta;x) = \underbrace{\frac{n}{2}\log|\Delta| - \frac{n-1}{2}\mathrm{tr}(\Delta S)}$$

Matrix cookbook; Waterloo

Matrix-calculus

$$l_{prof}(\Delta; \underline{x}) = l(\hat{\mu}_\Delta, \Delta; \underline{x}) \qquad \hat{\mu}_\Delta \equiv \hat{\mu} = \bar{x}$$

$$\frac{\partial}{\partial \Delta}\left\{ \frac{n}{2}\log|\Delta| - \frac{n-1}{2}\operatorname{tr}(\Delta S)\right\} \qquad S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$$

$$= 0 \quad \text{to def} \quad \hat{\Delta} \qquad \hat{\sigma}^2_{mle} = \underline{\underline{\tfrac{1}{n}}}\sum(x_i - \bar{x})^2$$

$$= \boxed{\frac{n}{2|\Delta|}\cdot\frac{\partial|\Delta|}{\partial\Delta} - \frac{n-1}{2}\frac{\partial}{\partial\Delta}\operatorname{tr}(\Delta S) = \Delta^{-1}\cdot n - (n-1)S}$$

$$= \ldots = \underbrace{\hat{\Delta}^{-1} = \frac{(n-1)S}{n}} = \hat{\Omega} \qquad \frac{\partial f(A)}{\partial A_{ij}} = \ldots$$

Important properties of MV Normal

marginal distribution of a component: $\underset{\sim}{X} \sim N_k(\mu, \Omega)$ $k \times k$ matrix

$X_a \sim N(\mu_a, \sigma_{aa})$ or $N(\mu_a, \sigma_a^2)$ ω_{aa}

$\begin{pmatrix} X_a \\ X_b \end{pmatrix} \sim N_2\left[\begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \underset{2\times 2}{\Omega_{ab}}\right]$ $\begin{pmatrix} \sigma_{aa} & \sigma_{ab} \\ \sigma_{ab} & \sigma_{bb} \end{pmatrix} = \Omega_{ab}$

$\binom{k}{2}$ prs. of r.v. $\sim N_2$ $\Rightarrow X \sim N_k$

marginal distribution of a subvector:

$\begin{pmatrix} X_{(1)} \quad k_1 \\ X_{(2)} \end{pmatrix}$ $X_{(1)} \sim N_{k_1}(\mu_{(1)}, \Omega_{(1,1)})$

Important properties of MV Normal

conditional distribution of a component: $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ & \sigma_{22} \end{pmatrix}\right]$

$$\Rightarrow \overset{y}{X_1} \mid X_2 = \overset{x}{\underset{}{x_2}} \sim N\Big(\underbrace{\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2)}_{\text{"regression of } X_1 \text{ on } X_2\text{" (slope} < 1)} ; \underbrace{\sigma_{11} - \sigma_{12}^2/\sigma_{22}}_{\text{"variance about the regression"}}\Big)$$

$y_i = \cancel{x\beta + \varepsilon}\ \beta_0 + \beta_1 x_i + \varepsilon_i$

$\hat\beta_1 = S_{xy}/S_{xx}$

$\rho_{12} = \dfrac{\sigma_{12}}{(\sigma_{11}\sigma_{22})^{1/2}} \quad \left(\neq \dfrac{\sigma_{12}}{\sigma_{22}}\right)$

conditional distribution of a subvector:

$$(\underset{k_1}{X_{(1)}} \mid X_{(2)} = x_{(2)}) \sim N_{k_1}\big(\mu_{(1)} + \Omega_{12}\Omega_{22}^{-1}(x_2 - \mu_2),\ (\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21})\big)$$

$Z \sim N(0, I)$ then

$$X = \underline{\mu} + \underline{\underline{\Omega}}^{1/2} Z \sim N(\mu, \Omega)$$

$X \sim N(\mu, \Omega)$ then $AX \sim N(A\mu, A\Omega A^T)$

goodness-of-fit $\longrightarrow$ $(X-\mu)^T \underline{\underline{\Omega}}^{-1}(X-\mu) \sim \chi^2_{\underline{\underline{k}}}$ if Ω is of full rank.

Partial corr$^{\text{n}}$ $\uparrow$ from Ω^{-1}:

Marginal Corr$^{\text{n}}$

$$\rho_{rs} = \frac{Cor(X_r, X_s)}{\{Var(X_r)\, Var(X_s)\}^{1/2}} = \frac{\sigma_{rs}}{(\sigma_{rr}\sigma_{ss})^{1/2}}$$

$$X_s \mid X_{-s} \sim N\left(\mu_s - \Omega_{s,-s}\Omega^{-1}_{-s,-s}(x_{-s} - \mu_{-s}),\ \underbrace{\Omega_{s,s} - \Omega_{s,-s}\Omega^{-1}_{-s,-s}\Omega_{-s,s}}\right)$$

$\begin{pmatrix} X_a \\ X_b \end{pmatrix}$

$S = (a, b)$ 2 elements

$$\Omega_{ab|-s} = (2\times 2) = \begin{pmatrix} \omega_{aa|-s} & \omega_{ab|-s} \\ \omega_{ba|-s} & \omega_{bb|-s} \end{pmatrix}$$

$\in \mathbb{R}$

$\rho_{ab|s} = 0 \Rightarrow \omega_{ab|-s} = 0$

$$\rho_{ab|-s} = \frac{\omega_{ab|-s}}{(\omega_{aa|-s}\,\omega_{bb|-s})^{1/2}}$$

$\in (-1, 1)$

partial corr$^{\text{n}}$ from $\Omega_{ab|s}$

corr$^{\text{n}}$ of (X_a, X_b) cond'l on $X_{(-s)}$ i.e. other X's

If $\rho_{ab|-s} = 0$ $\quad X_a \perp\!\!\!\perp X_b \mid X_{(-s = \{a,b\})}$ $\leftarrow$ causal inf. ("confounders")

Dawson's not^n $S \subseteq \{1, \dots, k\}$

$$\begin{pmatrix} X_{(S)} \\ X_{(-S)} \end{pmatrix} \sim N_k\left[\begin{pmatrix} \mu_S \\ \mu_{-S} \end{pmatrix}, \begin{pmatrix} \Omega_{SS} & \Omega_{S,-S} \\ \Omega_{-S,S} & \Omega_{-S,-S} \end{pmatrix}\right]$$

partition of MVN

$$X_{(S)} \mid X_{(-S)} = x_{(-S)} \sim N\left(\mu_S + \Omega_{S,-S}\,\Omega_{-S,-S}^{-1}(x_{(-S)} - \mu_{(-S)}),\ \Omega_{S,S} - \Omega_{S,-S}\,\Omega_{-S,-S}^{-1}\,\Omega_{-S,S}\right)$$

$= 0$

$X_{(S)} \perp\!\!\!\perp x_{(-S)}$ "cond'l independence"

Example

SM Ex 6.17

If $X_{a} \perp\!\!\!\perp X_b \mid X_{(-S)}$ then Ω^{-1} has a zero in a,b element

Mechanics (C)	Vectors (C)	Algebra (O)	Analysis (O)	Statistics (O)
77	82	67	67	81
63	78	80	70	81
75	73	71	66	81
55	72	63	70	68
63	63	65	70	63
⋮	⋮	⋮	⋮	⋮
15	38	39	28	17
5	30	44	36	18
12	30	32	35	21
5	26	15	20	20
0	40	21	9	14

Table 6.7 Marks out of 100 in five mathematics examinations for the first and last five of 88 students (Mardia *et al.*, 1979, pp. 3–4). Some of the examinations were closed-book (C), and others were open-book (O).

library("SMPracticals")

data(mathmarks)

Example 6.17 (Maths marks data) Table 6.7 gives marks out of 100 for the first and last five students out of 88 who took five mathematics examinations. As we would

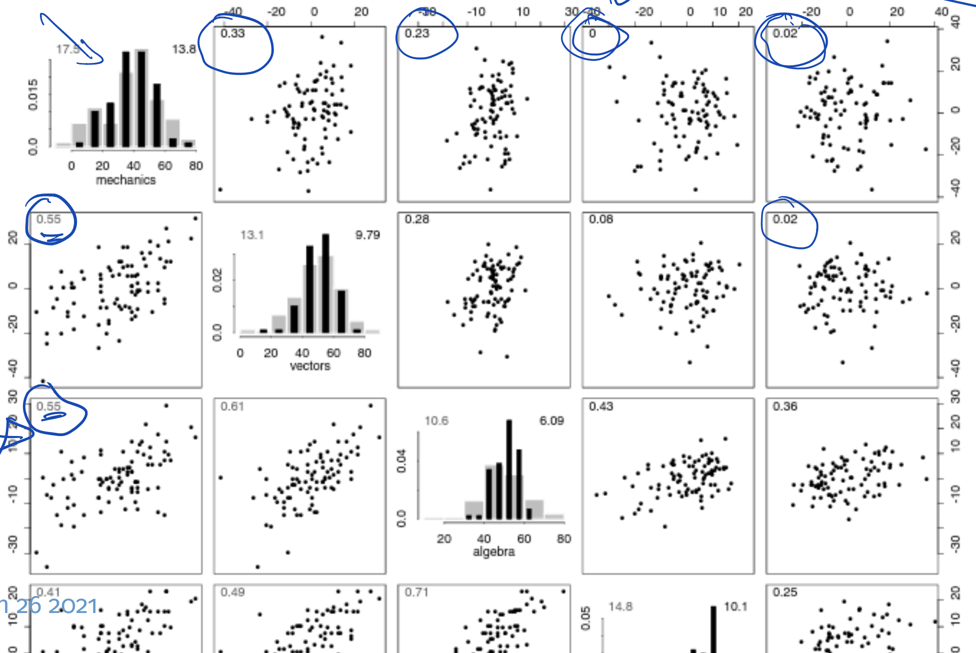
Example

SM Ex 6.19

pairs (mathmarks$)

Figure 6.8 Modified scatterplot matrix for the full maths marks data. Below the diagonal are scatterplots (and sample correlation coefficient) for the centred pairs of marks; for example, the lower left panel shows results for statistics plotted against those for mechanics. Above the diagonal are scatterplots of residuals (and sample partial correlation coefficients): for example, the top right panel shows the dependence remaining between mechanics and statistics after adjusting for the other variables. The diagonal shows histograms of the variables (light) and of the residuals from regression on all other variables, centred at the variable

marginal



partial corr's

Sab|-s

Example

SM Ex 6.20

	Mechanics	Vectors	Algebra	Analysis	Statistics
Mechanics	17.5/13.8	0.33	0.23	−0.00	0.03
Vectors	$\boxed{0.55}$	13.2/9.8	0.28	0.08	0.02
Algebra	0.55	0.61	10.6/6.1	0.43	0.36
Analysis	0.41	0.49	0.71	14.8/10.1	0.25
Statistics	0.39	0.44	0.66	0.61	17.3/12.5
Average	39.0	50.6	50.6	46.7	42.3

Table 6.9 Summary statistics for maths marks data. The sample correlations between variables are below the diagonal, and the sample partial correlations are above the diagonal. The diagonal contains sample standard deviation/ sample partial standard deviation.

marg'l corr$^{\underline{n}}$

partial corr$^{\underline{n}}$

Example

Example 6.21 (Maths marks data) The above-diagonal part of Table 6.9 suggests a graphical model in which the upper right 2×2 corner of Δ is set equal to zero. The likelihood ratio statistic for comparison of this model with the full model is 0.90, which is not large relative to the χ_4^2 distribution. This suggests strongly that the simpler model fits as well as the full one, an impression confirmed by comparing the original and fitted partial correlations,

exact

approx

$$
\begin{array}{cccc}
0.33 & 0.23 & -0.00 & 0.03 \\
 & 0.28 & 0.08 & 0.02 \\
 & & 0.43 & 0.36 \\
 & & & 0.25
\end{array}
\qquad
\begin{array}{cccc}
0.33 & 0.24 & 0.00 & 0.00 \\
 & 0.33 & 0.00 & 0.00 \\
 & & 0.45 & 0.37 \\
 & & & 0.26
\end{array}
$$

Figure 6.10 shows the graphs for these two models. In the full model every variable is joined to every other, and there is no simple interpretation. The reduced model has a butterfly-like graph whose interpretion is that given the result for algebra, results for mechanics and vectors are independent of those for analysis and statistics. Thus a result for mechanics can be predicted from those for algebra and vectors alone, while prediction for algebra requires all four other results. ■

Example

SM Ex 6.21

264 6 · Stochastic Models

undirected graph

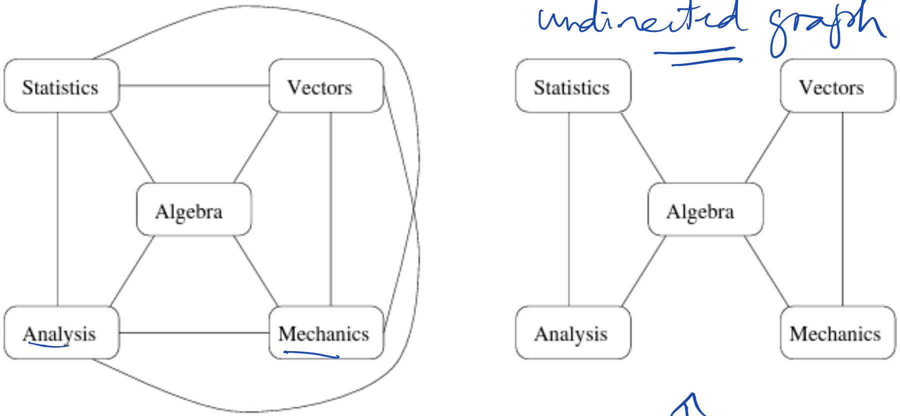


Figure 6.10 Graphs for the full model (left) and a reduced model (right) for the maths marks data. The interpretation of the reduced model is that given the result for algebra, results for vectors and mechanics are independent of those for analysis and statistics.

all components connected

embodies cond'l independencies

Ω^{-1}

Multinomial distribution

AoS §14.4; EH §5.34