Mathematical Statistics II

STA2212H S LEC9101

Week 11

March 31 2021

Start recording!

link

Today

Start Recording

1. HW 10 (current) not due until Monday April 5
   HW 11 posted April 2 due April 9
   Take-home posted April 9 due April 19 — April 7 if I can
   No class Friday April 2; no office hour Monday April 5

2. Course evaluations available until April 12
3. Bayesian methods for text classification; discriminant analysis
4. Intro to graphical models and causality

Recap

- Multivariate normal distribution
- $X \sim N_k(\mu, \Omega)$
- correlation ρ_{ij}
- partial correlation $\rho_{ij|-(i,j)}$

Multinomial distribution

AoS §14.4; SM Ex.2.36

- $X \sim Mult_k(n; p)$ — $X_j =$ number of obs in category j

- $\mathrm{pr}(X_1 = x_1, \dots, X_k = x_k; p) =$

- $\mathrm{E}(X) =$

- $\mathrm{cov}(X) =$ — AoS Thm 14.4

- $\hat{p} =$

- $\mathrm{cov}(\hat{p}) =$

Multinomial distribution

AoS §14.4; SM Ex.2.36

- $X \sim Mult_k(n; p)$ $\quad$ $X_j =$ number of obs in category j
- $\mathrm{pr}(X_1 = x_1, \ldots, X_k = x_k; p) =$
- $\mathrm{E}(X) =$
- $\mathrm{cov}(X) =$ $\quad$ AoS Thm 14.4
- $\hat{p} =$
- $\mathrm{cov}(\hat{p}) =$

Overview

- multivariate and multinomial distributions used to study the joint distribution
- analogous to unsupervised learning
- AoS §15.1,2: 2 binary variables; 2 discrete variables multinomial
- AoS §14.2: pairs of normal variables
- AoS §15.4 one discrete, one continuous variable
- AoS Ch.15 Inference about independence

Overview

supervised vs unsupervised

categorical vs continuous

text analysis

Models for dependent data

- Markov chains — SM §6.1
- continuous time Markov models — finite state space $\mathcal{S}$ — SM §6.2
- Markov random fields; directed acyclic graphs — SM §6.2
- Multivariate normal — SM §6.3
- Time series — SM §6.4
- Point processes — SM §6.5

- $X_1, \ldots, X_n$ a random sequence
- Markov property:

$$\mathrm{pr}(X_{j+1} = x_{j+1} \mid \qquad\qquad\qquad\qquad) =$$

- set of sites $\mathcal{J} = \{1, \ldots, n\}$
- neighbourhood system $\mathcal{N} = \{\mathcal{N}_j, j \in \mathcal{J}\}$

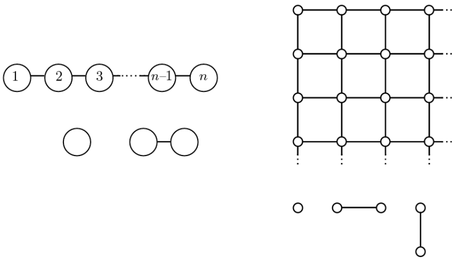

Figure 6.4 Markov random fields. Left: neighbourhood structure for first-order Markov chain and its cliques and their subsets. Right: first-order neighbourhood structure, cliques and their subsets for rectangular grid of sites.

Directed acyclic graphs

- can be convenient for studying relationships between variables
- through a probability distribution on the graph
- that is specified by a factorization of the joint density

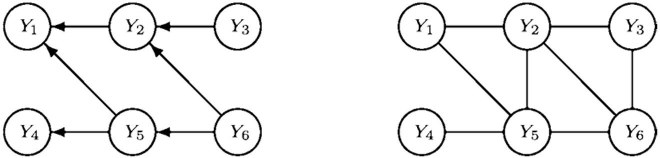

Figure 6.6 Directed acyclic and moral graphs. Left: directed acyclic graph representing (6.17). Right: moral graph, formed by moralizing the directed acyclic graph, that is, 'marrying' parents and dropping arrowheads.

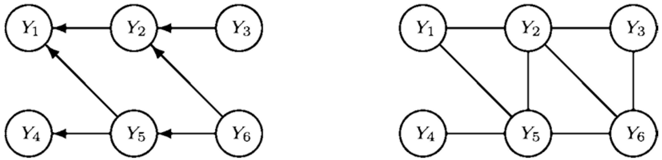

Figure 6.6 Directed acyclic and moral graphs. Left: directed acyclic graph representing (6.17). Right: moral graph, formed by moralizing the directed acyclic graph, that is, 'marrying' parents and dropping arrowheads.

31/2 and p.250

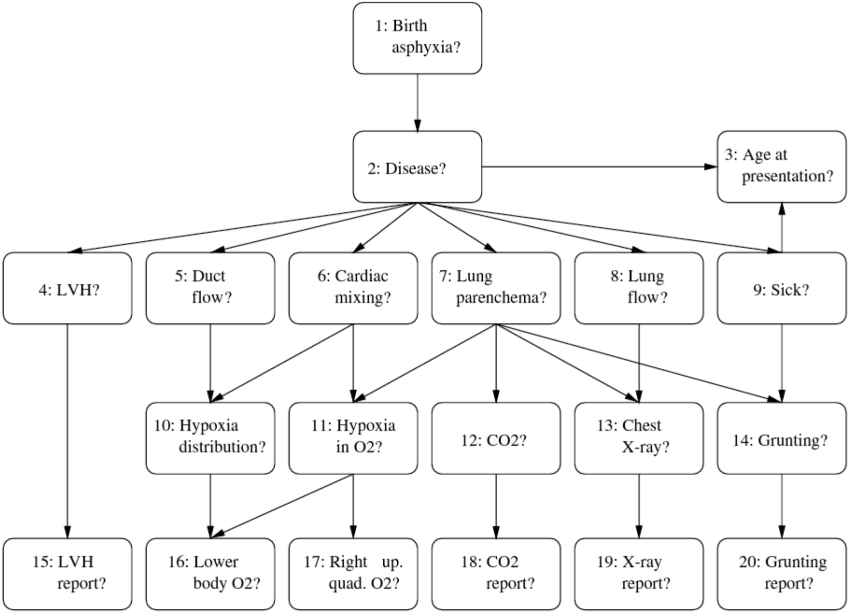

Figure 6.7 Directed acyclic graph representing the incidence and presentation of six possible diseases that would lead to a 'blue' baby (Spiegelhalter *et al.*, 1993). LVH means left ventricular hypertrophy.

Graphical Gaussian models

SM §6.3.3

264 6 · *Stochastic Models*

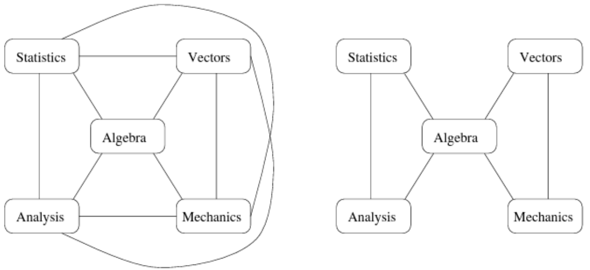

Figure 6.10 Graphs for the full model (left) and a reduced model (right) for the maths marks data. The interpretation of the reduced model is that given the result for algebra, results for vectors and mechanics are independent of those for analysis and statistics.

- Fig 17.2, Example 17.4
- Fig 17.3

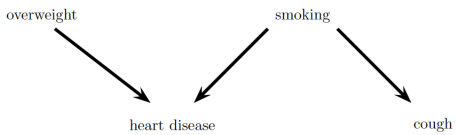

FIGURE 17.2. DAG for Example 17.4.

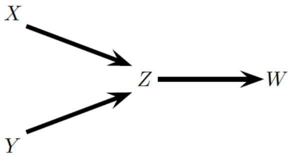

FIGURE 17.3. Another DAG.

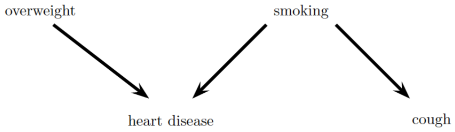

FIGURE 17.2. DAG for Example 17.4.

17.4 Example. Figure 17.2 shows a DAG with four variables. The probability function for this example factors as

$$\begin{aligned} f(\text{overweight}, \text{smoking}, \text{heart disease}, \text{cough}) \\ &= f(\text{overweight}) \times f(\text{smoking}) \\ &\times f(\text{heart disease} \mid \text{overweight}, \text{smoking}) \\ &\times f(\text{cough} \mid \text{smoking}). \quad \blacksquare \end{aligned}$$

17.5 Example. For the DAG in Figure 17.3, $\mathbb{P} \in M(\mathcal{G})$ if and only if its probability function f has the form

$$f(x, y, z, w) = f(x) f(y) f(z \mid x, y) f(w \mid z). \quad \blacksquare$$