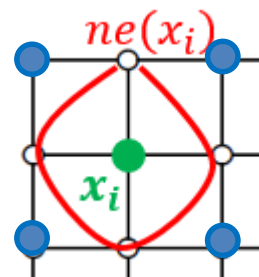


# 第四讲 高斯图模型

2024.3.13

局部马氏性:

$$p(x_i | \mathbf{x}_{-i}) = p(x_i | \mathbf{x}_{ne(x_i)})$$



给定 $x_i$ 的邻域 $ne(x_i)$ ,  $x_i$ 与不相邻的蓝点独立.  
蓝点通过邻域 $ne(x_i)$ 对 $x_i$ 发生影响。

$\Sigma^{-1}$ 称为精度矩阵(precision matrix)

利用分块矩阵的逆，上一讲定理3\*及其推论1将多元正态分布的条件分布用协方差矩阵 $\Sigma$ 的逆矩阵 $\Omega = \Sigma^{-1}$ 表达了出来，但从 $\Sigma$ 到 $\Sigma^{-1}$ 的转换看起来并不自然。

从多元正态概率密度函数的形式来看，这是比较显然的。记 $\Omega = \Sigma^{-1} = (\omega_{ij})$ ，则我们可以展开二次型：

$$\begin{aligned} f(\mathbf{x}) &= C \exp\left(-\frac{\mathbf{x}^\top \Sigma^{-1} \mathbf{x}}{2}\right) = C \exp\left(-\frac{\mathbf{x}^\top \Omega \mathbf{x}}{2}\right) \\ &= C \exp\left(-\frac{1}{2} \sum_{i,j} \omega_{ij} x_i x_j\right), \end{aligned}$$

不妨设 $\boldsymbol{\mu} = \mathbf{0}$

$$C = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} = \frac{|\Omega|^{1/2}}{(2\pi)^{n/2}}$$

例如，如果系数 $\omega_{ij} = 0$ ，上述二次型中没有交叉项 $x_i x_j$ ，则可分拆概率密度，得到独立或条件独立性质。

# 高斯图模型

## 记号约定和准备知识

$\mathbf{x}_A$  与  $\mathbf{x}_{-A}$

记号：随机向量  $\mathbf{x} = (x_1, x_2, \dots, x_p)^\top$ , 假设下标集合  $S \subset \{1, 2, \dots, p\}$ ,  $\mathbf{x}$  中下标属于  $S$  的分量组成向量  $\mathbf{x}_S = (x_i, i \in S)$ , 其它分量组成  $\mathbf{x}_{-S} = \mathbf{x}_{\bar{S}} = (x_i, i \notin S)$ 。

$\Sigma_{AB}$   
 $\Sigma_{AB \cdot C}$

记号：假设  $\Sigma = \text{cov}(\mathbf{x})$  是  $p \times p$  正定矩阵, 设  $A, B \subset \{1, 2, \dots, p\}$ ,

- $\Sigma_{AB} = \text{cov}(\mathbf{x}_A, \mathbf{x}_B)$  为  $\Sigma$  的行标为  $A$ , 列标为  $B$  的子矩阵;
- $\mathbf{x}_A^\perp = \mathbf{x}_A - \Sigma_{AC} \Sigma_{CC}^{-1} \mathbf{x}_C$  为  $\mathbf{x}_A$  关于  $\mathbf{x}_C$  的去相关化,  $\text{cov}(\mathbf{x}_A^\perp, \mathbf{x}_C) = 0$ ;
- $\Sigma_{AB \cdot C} = \text{cov}(\mathbf{x}_A^\perp, \mathbf{x}_B^\perp) = \Sigma_{AB} - \Sigma_{AC} \Sigma_{CC}^{-1} \Sigma_{CB}$ .

条件分布

$p(\mathbf{x}_S | \mathbf{x}_{-S}) = p(\mathbf{x}_S, \mathbf{x}_{-S}) / p(\mathbf{x}_{-S}) \propto p(\mathbf{x})$ :

将联合密度  $p(\mathbf{x}) = p(\mathbf{x}_S, \mathbf{x}_{-S})$  看作是  $\mathbf{x}_S$  的函数 ( $\mathbf{x}_{-S}$  固定), 即得到条件概率密度  $p(\mathbf{x}_S | \mathbf{x}_{-S})$ , 但相差一个仅与  $\mathbf{x}_{-S}$  有关的常数倍数。

我们将会用到独立性、条件独立性的如下简易证明或计算方法：

## 独立与条件独立

引理1.

(1) 假设随机向量 $\mathbf{x}$ 和 $\mathbf{y}$ 的联合概率密度函数 $p(\mathbf{x}, \mathbf{y})$ ，若存在函数 $\varphi_1, \varphi_2$ （未必是概率密度）使得  $p(\mathbf{x}, \mathbf{y}) = \varphi_1(\mathbf{x})\varphi_2(\mathbf{y})$ ，则 $\mathbf{x}$ 和 $\mathbf{y}$ 独立。

(2) 若存在函数 $\psi_1, \psi_2$ （未必是概率密度）使得联合概率密度函数  $p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \psi_1(\mathbf{x}, \mathbf{z})\psi_2(\mathbf{y}, \mathbf{z})$ ，则给定 $\mathbf{z}$ 时， $\mathbf{x}$ 与 $\mathbf{y}$ 条件独立。

证:(1)略。(2) 给定 $\mathbf{z}$ 时的条件概率密度

$$p(\mathbf{x}, \mathbf{y} | \mathbf{z}) = \frac{1}{p(\mathbf{z})} p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \frac{1}{p(\mathbf{z})} \psi_1(\mathbf{x}, \mathbf{z})\psi_2(\mathbf{y}, \mathbf{z}) = \varphi_1(\mathbf{x})\varphi_2(\mathbf{y})$$

比如取 $\varphi_1(\mathbf{x}) = \psi_1(\mathbf{x}, \mathbf{z})/p(\mathbf{z}), \varphi_2(\mathbf{y}) = \psi_2(\mathbf{y}, \mathbf{z})$ ,

注意： $\mathbf{z}$ 是常数

由(1)知 $\mathbf{x}$ 和 $\mathbf{y}$ 在给定 $\mathbf{z}$ 时条件独立。

这里我们通过展开多元正态密度函数中的二次型，再次得到第4讲定理3\*的推论1，并得到更多结论。

独立和条件独立

定理4. 假设  $\mathbf{x} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1})$ ，即协方差矩阵  $\boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}$ ， $\boldsymbol{\Omega} = (\omega_{ij})$ 。

(1)  $\omega_{ij} = 0 \Leftrightarrow x_i \perp x_j \mid \mathbf{x}_{-(ij)}$  (给定其它变量时， $x_i$ 与 $x_j$ 条件独立)。

(2)  $x_i \mid \mathbf{x}_{-i} \sim N_1\left(\mu_i - \sum_{j \neq i} \omega_{ii}^{-1} \omega_{ij} (x_j - \mu_j), \omega_{ii}^{-1}\right)$ 。

(3) 条件相关系数 (给定 $\mathbf{x}_{-(i,j)}$ 条件下， $x_i$ 与 $x_j$ 的相关系数)：

$$\rho_{ij \cdot -(ij)} = -\omega_{ij} / \sqrt{\omega_{ii} \omega_{jj}}, \quad i \neq j.$$

这也是一般的 (非正态下的) 偏相关系数。

注：(1),(3)说明，对于多元正态，条件独立 $\Leftrightarrow$ 条件不相关

证明:(1) (虽然推论1已证, 但我们还是给出一个富有启发性的证明)

不妨假设均值 $\boldsymbol{\mu} = \mathbf{0}$ 。给定 $\mathbf{x}_{-(ij)} = (x_t, t \neq i, j)$ 时,

$$p(x_i, x_j | \mathbf{x}_{-(ij)}) \propto p(\mathbf{x}) \propto \exp\left(-\frac{1}{2} \omega_{ii} x_i^2 - \frac{1}{2} \omega_{jj} x_j^2 - \omega_{ij} x_i x_j + c_1 x_i + c_2 x_j + c_3\right),$$

其中 $c_1, c_2, c_3$ 只与 $\mathbf{x}_{-(ij)}$ 有关,  $\mathbf{x}_{-(ij)}$ 给定时, 它们都是常数。

当 $\omega_{ij} = 0$ 时, 二次型无交叉项 $x_i x_j$ ,

所以 $p(x_i, x_j | \mathbf{x}_{-(ij)})$ 可分解为 $\varphi_1(x_i)\varphi_2(x_j) \Leftrightarrow x_i, x_j$ 条件独立。

注意 $N(\mu, \sigma^2)$ 的密度函数 $\propto \exp(-\frac{1}{2\sigma^2} x^2 + \frac{\mu}{\sigma^2} x)$ ,  
指数项中平方项 $x^2$ 的系数为 $-1/2\sigma^2$ ,  $x$ 系数为 $\mu/\sigma^2$ .

(2)  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1})$ :

$$p(x_i | \mathbf{x}_{-i}) \propto \exp(-\mathbf{x}^T \boldsymbol{\Omega} \mathbf{x} / 2) \propto \exp(-\omega_{ii} x_i^2 / 2 - (\sum_{j \neq i} \omega_{ij} x_j) x_i),$$

具有一元正态密度的形式。观察其平方项 $x_i^2$ 知, 该条件分布的方差为 $\omega_{ii}^{-1}$ ,

由一次项 $x_i$ 的系数知均值为 $-\omega_{ii}^{-1}(\sum_{j \neq i} \omega_{ij} x_j)$ , 所以 $x_i | \mathbf{x}_{-i} \sim N(-\omega_{ii}^{-1} \sum_{j \neq i} \omega_{ij} x_j, \omega_{ii}^{-1})$ .

(3)给定 $\mathbf{x}_{-(ij)}$ 条件下, $(x_i, x_j)$ 的条件概率密度为

$$p(x_i, x_j | \mathbf{x}_{-(ij)}) \propto \exp\left(-\frac{1}{2} \mathbf{x}^\top \boldsymbol{\Omega} \mathbf{x}\right) \\ \propto \exp\left(-\frac{1}{2} (x_i, x_j) \begin{pmatrix} \omega_{ii} & \omega_{ij} \\ \omega_{ji} & \omega_{jj} \end{pmatrix} \begin{pmatrix} x_i \\ x_j \end{pmatrix} + x_i \text{或} x_j \text{的一次项}\right)$$

指数上是 $x_i, x_j$ 的(正定)二次型, 因此上述条件分布是二元正态分布。

下面只需确定该分布的协方差矩阵, 假设该协方差阵为 $\begin{pmatrix} \gamma_{ii} & \gamma_{ij} \\ \gamma_{ji} & \gamma_{jj} \end{pmatrix}$ , 即

$$\begin{pmatrix} \gamma_{ii} & \gamma_{ij} \\ \gamma_{ji} & \gamma_{jj} \end{pmatrix} = \begin{pmatrix} \omega_{ii} & \omega_{ij} \\ \omega_{ji} & \omega_{jj} \end{pmatrix}^{-1} = \frac{1}{\omega_{ii}\omega_{jj} - \omega_{ij}^2} \begin{pmatrix} \omega_{jj} & -\omega_{ij} \\ -\omega_{ji} & \omega_{ii} \end{pmatrix}$$

所以条件相关系数为

$$\rho_{ij \cdot -(ij)} = \frac{\gamma_{ij}}{\sqrt{\gamma_{ii}\gamma_{jj}}} = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}$$

## 偏相关系数矩阵

协方差矩阵 $\Sigma_{p \times p} = (\sigma_{ij})$ , 精度矩阵 $\Omega_{p \times p} = \Sigma^{-1} = (\omega_{ij})$

- 相关系数矩阵 $R = (\rho_{ij}) = \left( \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \right) = D^{-1/2}\Sigma D^{-1/2}$ ,  $D = \text{diag}(\Sigma)$
- 偏相关系数矩阵 $R_p = (\rho_{ij \cdot -(ij)}) = \left( -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} \right) = 2I_p - C^{-1/2}\Omega C^{-1/2}$ ,

$C = \text{diag}(\Omega)$ . 注:  $-C^{-1/2}\Omega C^{-1/2}$ 的对角元为-1, 所以对角元需加2.

高斯图模型基于精度矩阵或偏相关系数矩阵的0元素, 将变量之间的关系以无向图(undirected graph/network)表达出来。

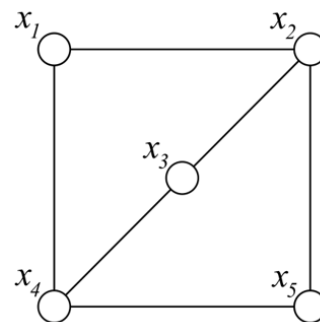
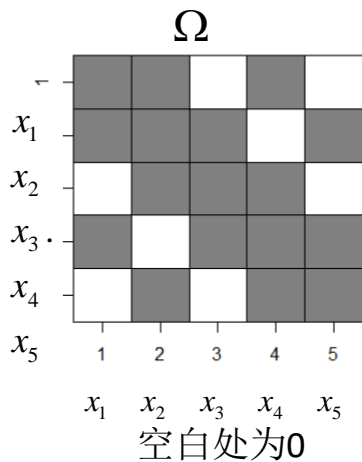


# 高斯图模型

定义3: (高斯图模型或高斯马氏随机场GMRF)

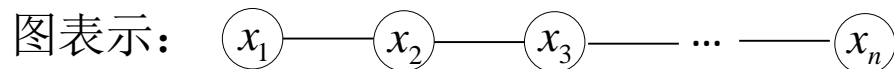
假设随机向量  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Omega^{-1})$ , 其中  $\Omega = (\omega_{ij})$ , 以节点表示图  $\mathbf{x}$  的分量:

$$i \sim j \text{ (即 } x_i \text{ 和 } x_j \text{ 以边相连)} \Leftrightarrow \omega_{ij} \neq 0.$$



例1. 假设  $\mathbf{x} = (x_1, \dots, x_n)^\top \sim N_n(\mathbf{0}, \Sigma), -1 < \rho < 1$

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}, \quad \Sigma^{-1} = (1-\rho^2) \begin{pmatrix} 1 & -\rho & & & \\ -\rho & 1+\rho^2 & -\rho & & \\ & -\rho & 1+\rho^2 & \ddots & \\ & & \ddots & \ddots & \\ & & & & 1+\rho^2 & -\rho \\ & & & & -\rho & 1 \end{pmatrix}$$

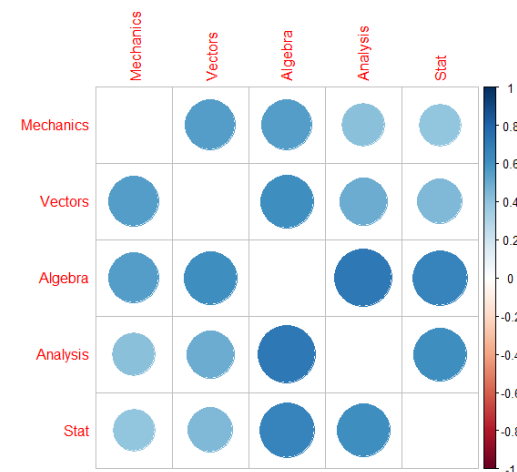


这称为马氏链或时间序列AR1模型

注: 马氏链的常规定义:  $P(x_i | x_{i-1}, x_{i-2}, \dots, x_1) = P(x_i | x_{i-1})$ ,  
 $\Leftrightarrow$  马氏随机场的定义:  $P(x_i | \mathbf{x}_{-i}) = P(x_i | x_{i-1}, x_{i+1})$   
 等价性不平凡。

例2. 88个学生5门数学成绩相关系数矩阵如下，

	Mechanics	Vectors	Algebra	Analysis	Stat
Mechanics	1.00	0.55	0.55	0.41	0.39
Vectors	0.55	1.00	0.61	0.49	0.44
Algebra	0.55	0.61	1.00	0.71	0.66
Analysis	0.41	0.49	0.71	1.00	0.61
Stat	0.39	0.44	0.66	0.61	1.00

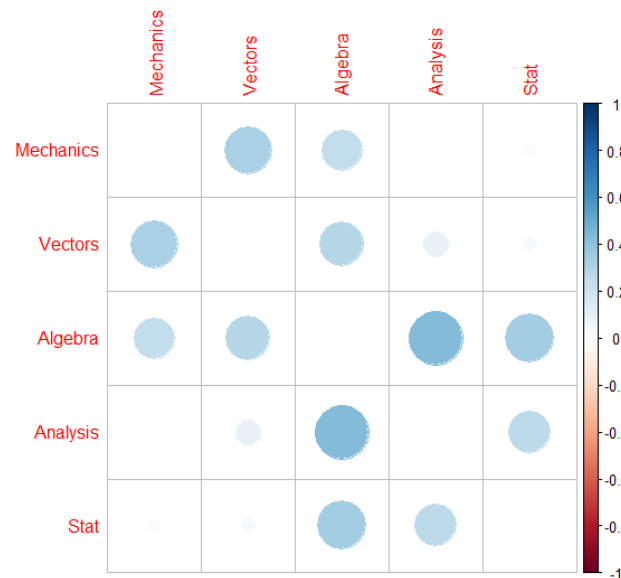


我们关心的问题比如有：统计课成绩与力学成绩的相关系数是0.39，学好力学对学好统计有帮助吗？我们假设5门课程成绩联合服从多元正态分布。

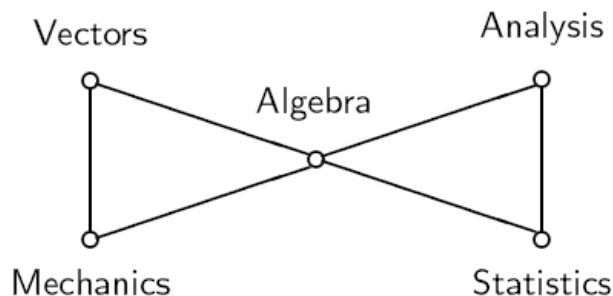
条件相关性系数矩阵如下：

	Mechanics	Vectors	Algebra	Analysis	Stat
Mechanics	1.00	0.32	0.24	-0.01	0.02
Vectors	0.32	1.00	0.28	0.09	0.03
Algebra	0.24	0.28	1.00	0.43	0.34
Analysis	-0.01	0.09	0.43	1.00	0.26
Stat	0.02	0.03	0.34	0.26	1.00

高亮部分认为是0



图表示：



给定Algebra时, (Analysis, Stat) 与 (Vectors, Mechanics) 独立

# 附录1: GLASSO - Graphical lasso

例2中我们将接近于0的偏相关系数当成0, 推断除了图模型的结构。  
Friedman et al. (2007) 提出graphical lasso 方法, 通过在多元正态似然函数基础上惩罚精度矩阵的L1模, 将较小的样本偏相关系数自动估计为0. 极大化惩罚似然:

$$\log \det(\Omega) - \text{tr}(S\Omega) - \lambda \|\Omega\|_1$$

其中 $S$ 为样本协方差矩阵,  $\|\Omega\|_1 = \sum |\omega_{ij}|$ .

细节:

基于数据 $\mathbf{x}_1, \dots, \mathbf{x}_n$  iid  $\sim N_n(\mu, \Sigma)$ , 极大似然法求解 $\Sigma$ 归结于极大化

$$\log \det(\Omega) - \text{tr}(S\Omega), \Omega = \Sigma^{-1}$$

其中 $S$ 为样本协方差矩阵。极大似然估计为 $\hat{\Sigma} = S$ 。GLASSO在此基础上增加了一个惩罚。

参见: Friedman, J., Hastie, T. and Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso, *Biostatistics* 9: 432-441.

# 附录2: 一般的图模型/马氏随机场

## 图graph

(无向)图由节点(node,vertex)和它们之间的连线(边,edge)组成。  
图 $G = (V, E)$ , 节点集合 $V$ , 边集合 $E$ 。  
节点 $i, j \in V$ 之间有边, 记作 $(i, j) \in E$ , 或 $i \sim j$ .

## 图模型/ 马氏随 机场

图模型(graphical model)或马氏随机场(Markov random fields, MRF)  
以无向图(graph)表示多个变量之间的条件相依或条件独立关系。  
图的节点是随机变量, 条件不独立的节点/变量之间连线, 构成图  
的边, 条件独立变量之间不连线。

对任一节点 $i$  (或 $x_i$ ), 定义其领域为所有与 $i$ 有连结的节点集合

$$ne(i) = \{j: j \sim i\},$$

即所有与 $x_i$ 条件不独立的随机变量的集合。换言之, 图模型用  
不连线的方式表示条件独立 (马氏性质):

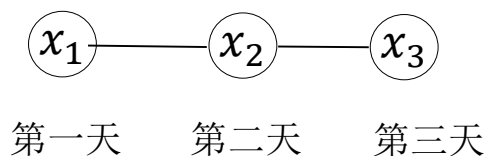
$$x_i \perp\!\!\!\perp \mathbf{x}_{-ne(i)} | \mathbf{x}_{ne(i)} \Leftrightarrow p(x_i | \mathbf{x}_{-i}) = p(x_i | \mathbf{x}_{ne(i)})$$

这里以及后面,  $p(\bullet)$ 表示概率密度(连续)或概率函数(离散).

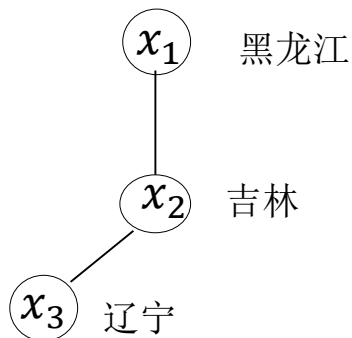
当节点变量服从多元正态分布, 称为高斯图模型或高斯马氏随机场

例如，随机变量 $x_1, x_2, x_3$ ，假设  $x_1 \perp\!\!\!\perp x_3 | x_2 \Leftrightarrow P(x_3 | x_1, x_2) = P(x_3 | x_2)$ ，图表示如下：

□ 当下标有次序时（比如时间）称为马氏链：



□ 当下标无次序时（比如空间位置）称为马氏随机场



与定义3等价的定义方式还有局部马氏性和整体马氏性。

定理2. 假设 $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$ , 对应的图模型为 $G = (V, E)$ , 下述三个性质等价

- 两两马氏性质 (Pairwise Markov property, 定义3) :

$$x_i \perp x_j \mid \mathbf{x}_{-ij}, (i, j) \notin E \text{ (节点 } i, j \text{ 之间没有边/连线)}$$

- 局部马氏性质 (Local Markov property) :

$$x_i \perp \mathbf{x}_{-ne(i)} \mid \mathbf{x}_{ne(i)}, ne(i) = \{k \in V : k \sim i, k \neq i\}$$

$$\text{或 } p(x_i \mid \mathbf{x}_{-i}) = p(x_i \mid \mathbf{x}_{ne(i)})$$

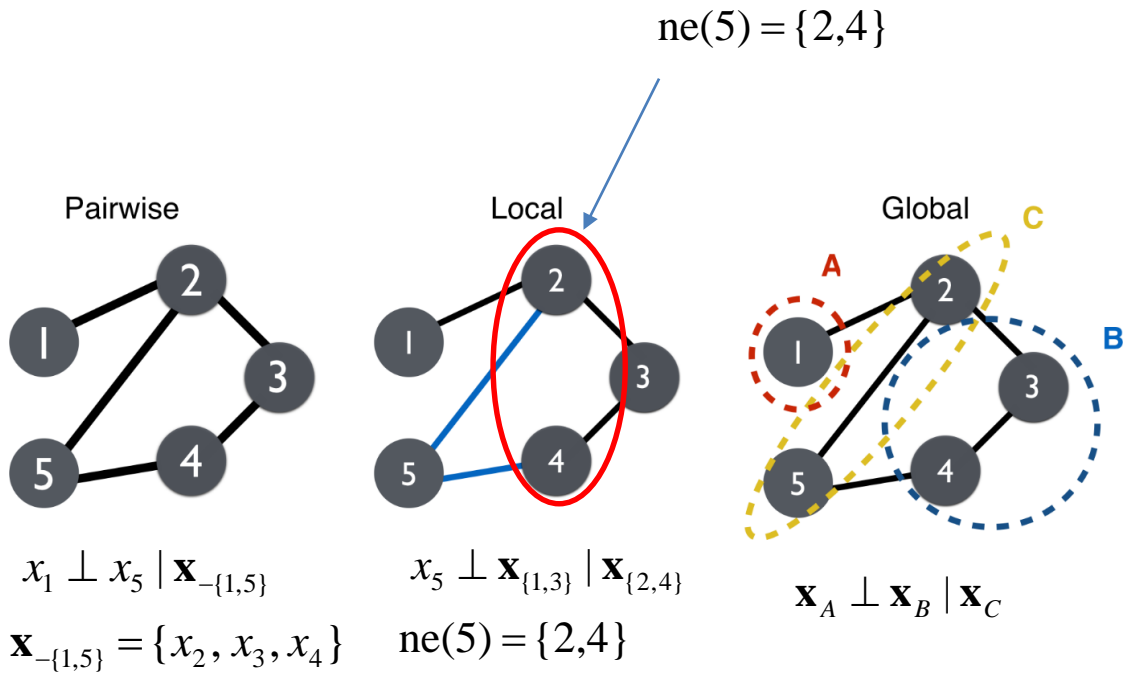
- 整体马氏性质 (Global Markov property) :

$$\mathbf{x}_A \perp \mathbf{x}_B \mid \mathbf{x}_C, \text{ 节点集合 } C \text{ 分离 } A \text{ 和 } B$$

(即 $A, B$ 之间的连结路径都通过 $C$ 中的某些节点,  $A, B, C$ 互斥)。

证明. 参见Koller, Friedman (2009), Probabilistic Graphical Models





# 反问题：Hammersley-Clifford定理

前面的例1-2是从精度矩阵推断高斯图模型结构。更常见的反问题，即根据实际问题建立一系列局部条件分布，最终形成整个系统的联合模型（即联合密度）：

依据实际背景（时间或空间结构）假设局部马氏性

$$p(x_i|\mathbf{x}_{-i}) = p(x_i|\mathbf{x}_{ne(i)}), i = 1, 2, \dots$$

问题是，指定所有局部条件分布，比如假设上述条件分布是一元正态，所有这些一元条件分布是否构成了一个相容、自治的马氏随机场或多元分布？

例3（1）我们考虑最简单情形：假设两个随机变量 $x, y$ ,

$E(x) = E(y) = 0$ , 满足

$$y|x \sim N(ax, \sigma^2), x|y \sim N(by, \tau^2),$$

$\Leftrightarrow$

$$\begin{aligned} y &= ax + \varepsilon_1, \varepsilon_1 \sim N(0, \sigma^2), \varepsilon_1 \perp\!\!\!\perp x \\ x &= by + \varepsilon_2, \varepsilon_2 \sim N(0, \tau^2), \varepsilon_2 \perp\!\!\!\perp y \end{aligned}$$

它们是否矛盾，何时 $(x, y)$ 服从二元正态？

$$ab < 1, a/\sigma^2 = b/\tau^2.$$

例3 (2). 假设三个随机变量 $x_1, x_2, x_3$  (均值都为0) 的关系如下 (东北三省的GDP) :



假设它们满足3个线性模型:

$$x_1 = \beta_{12}x_2 + \varepsilon_1, \quad \varepsilon_1 \sim N(0, \kappa_1^{-1}),$$

$$x_2 = \beta_{21}x_1 + \beta_{23}x_3 + \varepsilon_2, \quad \varepsilon_2 \sim N(0, \kappa_2^{-1}),$$

$$x_3 = \beta_{32}x_2 + \varepsilon_3, \quad \varepsilon_3 \sim N(0, \kappa_3^{-1}),$$

何时这些模型是合理的/相容的 (可唯一确定一个多元正态) ?

由Hammersley – Clifford定理, 条件是

$$\Omega = \begin{pmatrix} \kappa_1 & -\kappa_1\beta_{12} & 0 \\ -\kappa_2\beta_{21} & \kappa_2 & -\kappa_2\beta_{23} \\ 0 & -\kappa_3\beta_{32} & \kappa_3 \end{pmatrix} > 0 \quad (\text{对称正定}),$$

对称性要求 $\kappa_1\beta_{12} = \kappa_2\beta_{21}$ 等等。  $\Omega > 0$ 时,  $\mathbf{x} \sim N_3(\mathbf{0}, \Omega^{-1})$ 。

一般地，假设局部正态线性回归模型如下，这些模型是否相容？

$$x_i | \mathbf{x}_{-i} \sim N \left( \mu_i - \sum_{j:j \sim i} \beta_{ij} (x_j - \mu_j), \kappa_i^{-1} \right), \quad i = 1, \dots, p$$

定理3 (Hammersley-Clifford定理的特殊情形). 假设  $\mathbf{x} = (x_1, \dots, x_p)^\top$  服从边缘正态回归模型（条件正态）：

$$x_i | \mathbf{x}_{-i} \sim N \left( \mu_i - \sum_{j:j \sim i} \beta_{ij} (x_j - \mu_j), \kappa_i^{-1} \right), \quad i = 1, \dots, p$$

若  $j \not\sim i$ ，则令  $\beta_{ij} = \beta_{ji} = 0$ . 记  $\Omega = (\omega_{ij})$ ，其中  $\omega_{ii} = \kappa_i$ ， $\omega_{ij} = \kappa_i \beta_{ij}$ ， $i \neq j$ 。

如果  $\Omega > 0$ （正定），则  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Omega^{-1})$ 。

$$\Omega = K^{-1}B$$

回归系数：  $B = \begin{pmatrix} 1 & \beta_{12} & \cdots & \beta_{1p} \\ \beta_{12} & 1 & \cdots & \beta_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ \beta_{p1} & \beta_{p2} & \cdots & 1 \end{pmatrix}$ ， 误差方差  $K = \text{diag}(1/\kappa_1, \dots, 1/\kappa_p)$

HC定理给出了MRF的联合分布形式

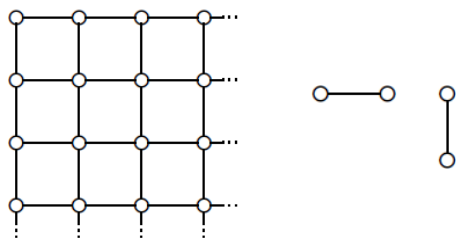
定理4 (Hammersley-Clifford定理). 若马氏随机场的联合分布 $p(\mathbf{x}) > 0$  (未必正态), 则具有如下形式

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left(-\sum_C \phi_C(\mathbf{x}_C)\right)$$

其中 $C$ 是最大的全连通节点子集(Clique),  $Z$ : partition 归一化常数.

难点在于验证条件 $p(\mathbf{x}) > 0$ , 即何时这才是一个正在的分布? 参见例3, 定理3

例如, 如果 $\mathbf{x} = (x_i, i = 1, 2, \dots)$ 的分布是马氏随机场, 下标 $i$ 代表格点(下左图), 则clique如右图



$\Rightarrow$

联合分布具有形式

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left(-\sum_{i \sim j} \phi_{ij}(x_i, x_j)\right)$$

其中 $i \sim j$ 代表水平或竖直方向相邻的两个格点

如果前述格点图中每个 $x_i$ 只取两个值，比如1, -1, 则二元函数 $\phi_{ij}(x_i, x_j)$ 一定具有形式

$$\phi_{ij}(x_i, x_j) = a_{ij} + b_i x_i + c_j x_j + d_{ij} x_i x_j,$$

这是Ising模型, 其中系数参数 $d_{ij}$ 代表了相邻格点交互作用程度。

若假设一致性(homogeneous), 即系数a,b,c,d没有下标, 比如 $d_{ij} \equiv d$ , 代表相邻节点交互作用处处相同(homegeneous)。

对统计学来说, Ising模型 $p(\mathbf{x})$ 是多个二值随机变量的联合分布(统计中也称为loglinear模型)。它描述联合概率

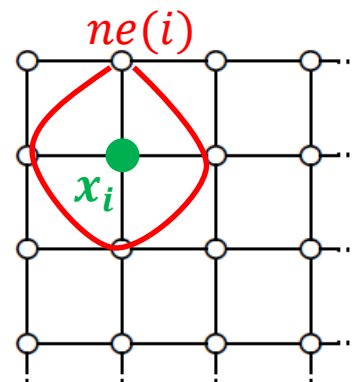
$$P(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{Z} \exp(-a - \sum_i b_i x_i - \sum_{i \sim j} d_{ij} x_i x_j), x_i = \pm 1$$

显然 $p(\mathbf{x})$ 是一个恰当多元分布(与正态相像)。

# 伊辛模型 Ising model

对物理学家来说，Ising模型是一个物理模型，用于描述磁化等物理现象。以二维空间为例，平面晶格上有 $x_i = -1$  or  $1$ （小磁针,每个磁针有两个自旋状态spin），其中每个格点的邻域 $ne(i)$ 由与其直接连结的4个点组成。假设 $(x_1, \dots, x_n)^T$ 是一个马氏随机场：

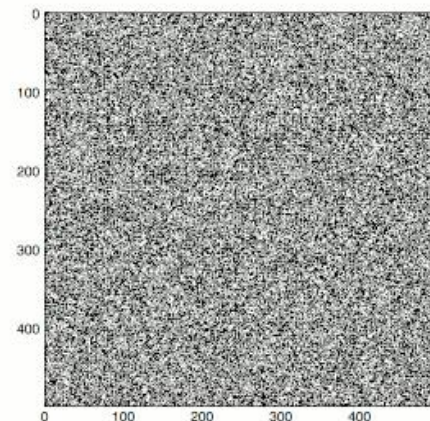
$$x_i \perp\!\!\!\perp \mathbf{x}_{-ne(i)} \mid \mathbf{x}_{ne(i)}$$



所有格点的联合分布（一致性假设下）在物理中通常表述为：

$$p(\mathbf{x}) = \frac{1}{Z} \exp(-H(\mathbf{x})/T),$$

其中 $H(\mathbf{x}) = -J \sum_{i \sim j} x_i x_j - h \sum_i x_i$ 为能量函数 (Hamiltonian),  $J$ 代表交互作用,  $h$ 代表外部磁场,  $T$ 代表温度。当 $J > 0$ , 相邻磁针趋于同号, 能量较低, 若温度较低, 显现磁性。当温度升高到一定程度, 磁性突然消失, 称为相变。



Partition (配分)  $Z$  是归一化常数,

$$Z = Z(J, h) = \sum_{\mathbf{x} \in \{-1, 1\}^n} \exp(-H(\mathbf{x})/T).$$

$Z$  包含了概率分布所有的性质 (参见指数族)。  $Z$  的零点的极限性质与相变性质有关, Lee-Yang circle theorem 证明了零点在复平面单位圆上 (模型不局限于 Ising), 当温度趋于极限状态时, 复零点趋于实数, 在该点发生相变或突变。

1d Ising: 有精确解没有相变

2d Ising: 有精确解有相变 (Onsager, 1944; Yang CN, 1952)

3d Ising: open

指数族分布:

$$f(x, \theta) = h(x) \exp(\eta(\theta)x - A(\theta)) = \frac{1}{Z(\theta)} \exp(\eta(\theta)x - B(x))$$

$$Z(\theta) = \exp(A(\theta))$$

$Z(\theta)$  包含了  $f$  的所有信息:

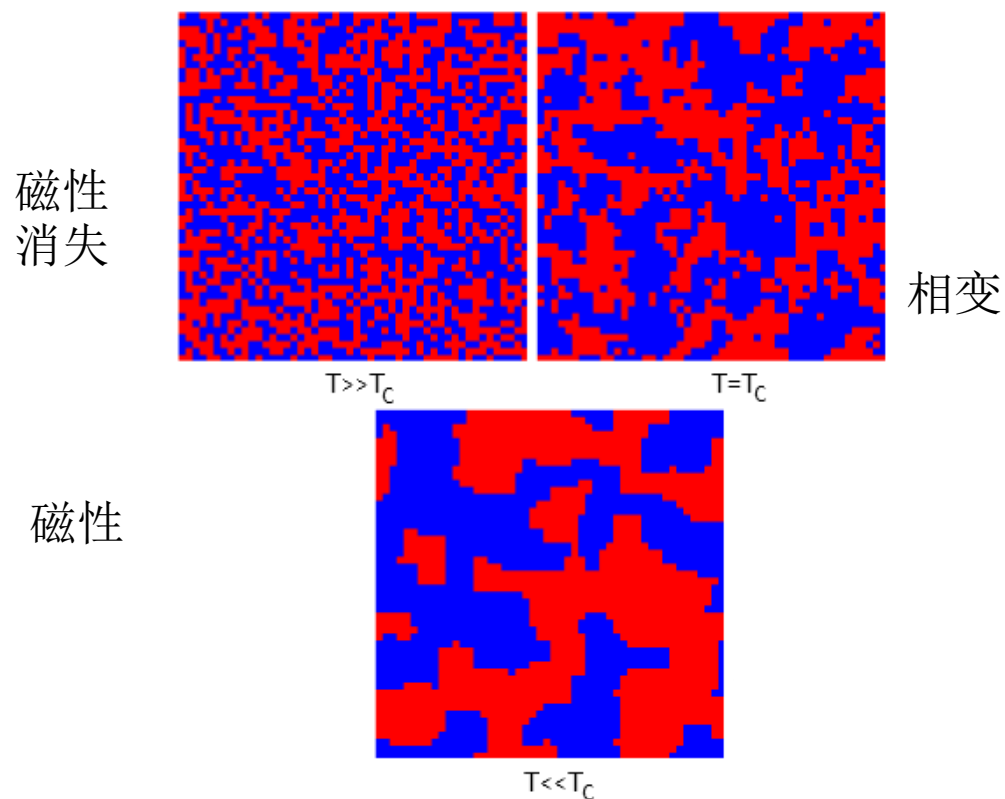
$$\text{矩母函数: } E \exp(tx) = \frac{Z(\theta+t)}{Z(\theta)} = \exp(A(\theta+t) - A(\theta))$$

比如,  $E(x) = A'(\theta) = \partial \log(Z(\theta)) / \partial \theta$ 。

一般,  $Z(\theta)$  作为光滑函数没有实数零点, 也不包含突变信息, 但其虚数根在极限情况下, 却有实际意义 (包含了相变信息)。



再强调一下， $J > 0$ 时，可磁化，临近的磁针趋向于同向，温度低于某个阈值  $T_c$ 时，呈现磁性（不混乱），当温度高于  $T_c$ 时会发生相变 (phase transition), 磁性消失（混乱）。



一个著名的例子：应用Ising模型研究社会意见的趋同和分裂。

## The Spontaneous Magnetization of a Two-Dimensional Ising Model

C. N. YANG

*Institute for Advanced Study, Princeton, New Jersey*

(Received September 18, 1951)

The spontaneous magnetization of a two-dimensional Ising model is calculated exactly. The result also gives the long-range order in the lattice.

IT is the purpose of the present paper to calculate the spontaneous magnetization (i.e., the intensity of magnetization at zero external field) of a two-dimensional Ising model of a ferromagnet. Van der Waerden<sup>1</sup> and Ashkin and Lamb<sup>2</sup> had obtained a series expansion of the spontaneous magnetization that converges very rapidly at low temperatures. Near the critical temperature, however, their series expansion cannot be used. We shall here obtain a closed expression for the spontaneous magnetization by the matrix method which was introduced into the problem of the statistics of a two-dimensional Ising model by Montroll<sup>3</sup> and Kramers and Wannier.<sup>4</sup> Onsager gave in 1944 a complete solution<sup>5</sup> of the matrix problem. His method was subsequently greatly simplified by Kaufman,<sup>6</sup> and the result has been used to calculate the short-range order in the crystal lattice.<sup>7</sup>

The Onsager-Kaufman solution of the matrix problem will be used in the present paper to calculate the spontaneous magnetization. In Sec. I we define the specific magnetization  $I$  and express it as an off diagonal element in the matrix problem. By introducing an artificial limiting process its calculation is reduced to an eigenvalue problem in Sec. II. This is solved in the next three sections and the final result given in Sec. VI. The relation between  $I$  and the usual long-range order is

where

$$V_1 = \exp\{H^* \sum_1^n C_r\}, \quad (2)$$

and

$$V_2 = \exp\{H \sum_1^n \mathbf{s}_r \mathbf{s}_{r+1}\}. \quad (3)$$

$H^*$  and  $H$  are given by

$$e^{-2H} = \tanh H^* = \exp[-(1/kT)\{V_{\uparrow\downarrow} - V_{\uparrow\uparrow}\}]. \quad (4)$$

The following abbreviation will be useful:

$$x = e^{-2H}. \quad (5)$$

If a weak magnetic field is introduced the partition function becomes

$$Z_{\mathcal{J}C} = (2 \sinh 2H)^{n/2} \text{trace}(\mathbf{V}_3 \mathbf{V}_2 \mathbf{V}_1)^m, \quad (6)$$

where

$$\mathbf{V}_3 = \exp\{\mathcal{J}C \sum_1^n \mathbf{s}_r\}. \quad (7)$$

For a large crystal only the eigenvector of  $\mathbf{V} = \mathbf{V}_3 \mathbf{V}_2 \mathbf{V}_1$  with the largest eigenvalue is important. We shall be interested in the limiting form of this eigenvector as  $\mathcal{J}C \rightarrow 0$ .

It has been shown by Onsager<sup>5</sup> that below the critical

## Statistical Theory of Equations of State and Phase Transitions. II. Lattice Gas and Ising Model

T. D. LEE AND C. N. YANG

*Institute for Advanced Study, Princeton, New Jersey*

(Received March 31, 1952)

The problems of an Ising model in a magnetic field and a lattice gas are proved mathematically equivalent. From this equivalence an example of a two-dimensional lattice gas is given for which the phase transition regions in the  $p$ - $v$  diagram is exactly calculated.

A theorem is proved which states that under a class of general conditions the roots of the grand partition function always lie on a circle. Consequences of this theorem and its relation with practical approximation methods are discussed. All the known exact results about the two-dimensional square Ising lattice are summarized, and some new results are quoted.

---

### INTRODUCTION

IN paper I<sup>1</sup> we have seen that the problem of a statistical theory of phase transitions and equations of state is closely connected with the distribution of roots of the grand partition function. It was shown there that the distribution of roots determines completely the equation of state, and in particular its behavior near the positive real axis prescribes the properties of the system in relation to phase transitions. It was also shown there that the equation of state of the condensed phases as well as the gas phase can be correctly obtained from a knowledge of the distribution of roots. While this general and abstract theory clarifies the problems underlying the statistical theory of phase transitions and condensed phases, it is natural to ask whether it also provides us with a means of obtaining

as a theorem in Sec. IV of the present paper and proved in the appendix. Implications of the theorem are discussed in Sec. V.

Also in this paper we shall give a proof (Sec. II) that the problem of an Ising model with a magnetic field is mathematically identical with that of a "lattice gas." From this identification we were able to trace exactly the transition region in the  $p$ - $v$  diagram of a two-dimensional lattice gas in detail. This will be presented in Sec. III and forms a clear illustration of the discussions of paper I and of Sec. V of the present paper.

At the end of Sec. V we give a summary of all the exact knowledge known to us about the two-dimensional Ising model in a magnetic field and its relationship with the distribution of roots of the partition function.

### II. ISING MODEL AND LATTICE GAS