课程主页：http://staff.ustc.edu.cn/~ynyang/vector

第五讲　高斯图模型

2025.3.12

$$\Sigma = \mathrm{var}\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix} = \begin{pmatrix}\mathrm{var}(\mathbf{x}_1) & *\\ * & *\end{pmatrix},$$

$$\Sigma^{-1} = \begin{pmatrix}(\mathrm{var}(\mathbf{x}_1|\mathbf{x}_2))^{-1} & *\\ * & *\end{pmatrix}$$

内容

- 多元生成分布：球对称正态、Boltzmann分布
- 椭球分布
- 多元正态分布
- 高斯图模型
- 概率图模型：马氏随机场
- Hammersley-Clifford定理

Recap

高斯积分方法计算球表面积（参见第一讲P27定理1的证明2）及其与Mawell-Hershell定理的关系

注意不能应用任何与球面积或体积有关的先验知识，比如不能使用球内均匀分布等

取球对称多元正态密度（略去常数）：

$$p(\mathbf{x}) = e^{-\|\mathbf{x}\|^2} = \prod_{i=1}^{p} e^{-x_i^2}$$

下面用两种方式计算其积分$\int p(\mathbf{x})d\mathbf{x}$，注意$p(\mathbf{x})$有两个特点：

(1) 球对称：$p(\mathbf{x})$仅依赖于模长， 利用球坐标可把径向与球面分离开来：

$$\int p(\mathbf{x})d\mathbf{x} = \int e^{-\|\mathbf{x}\|^2} d\mathbf{x} = \int \int_{(0,\infty)\times S^{n-1}} e^{-r^2} r^{n-1} dr d\sigma$$

$$= \int e^{-r^2} r^{n-1} dr \times \int_{S^{n-1}} d\sigma = \frac{1}{2}\Gamma(n/2)\ |S^{n-1}|$$

(2) 分量独立：$p(\mathbf{x})$可分解，多重积分转化为n个一重积分

$$\int p(\mathbf{x})d\mathbf{x} = \int \cdots \int e^{-x_1^2 - \cdots - x_n^2} dx_1 \ldots dx_n = \left(\int e^{-x_1^2} dx_1\right)^n = \pi^{n/2}$$

$$\Rightarrow \Gamma(n/2)|S^{n-1}|/2 = \pi^{n/2} \Rightarrow |S^{n-1}| = 2\pi^{n/2}/\Gamma(n/2)$$

同时具有球对称和分量独立两个性质的分布只有球对称多元正态分布(Mawell-Hershell定理), 故$p(\mathbf{x})$取球对称正态是唯一的选择。

正态变量相依性与协方差矩阵有关，条件相依性与协方差矩阵的逆有关

第4讲性质6，6*：假设 $\mathbf{x}\sim N_p(\mu,\Sigma),\Sigma>0,\Omega=\Sigma^{-1}$ ，划分

$$\mathbf{x}=\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix},\ \boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix},\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix},\ \Omega=\begin{pmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{pmatrix}$$

其中 $\mathbf{x}_1,\boldsymbol{\mu}_1$ 是 $q\times 1$ 向量，Σ_{11},Ω_{11} 是 $q\times q$ 方阵，则

- $\mathbf{x}_1\sim N_q(\boldsymbol{\mu}_1,\Sigma_{11})$；
- $\mathbf{x}_1|\mathbf{x}_2\sim N_q(\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2),\Sigma_{11\bullet 2})$
  $=N_q(\boldsymbol{\mu}_1-\Omega_{11}^{-1}\Omega_{12}(\mathbf{x}_2-\boldsymbol{\mu}_2),\Omega_{11}^{-1})$.

← $\mathrm{var}(\mathbf{x}_1)$ 为Σ的子矩阵Σ_{11}

← $\mathrm{var}(\mathbf{x}_1|\mathbf{x}_2)$为$\Omega$子矩阵$\Omega_{11}$的逆

$$\Sigma=\begin{pmatrix}\Sigma_{11} & *\\ * & *\end{pmatrix},\quad \Sigma^{-1}=\Omega=\begin{pmatrix}\Omega_{11} & *\\ * & *\end{pmatrix}$$

$\mathrm{var}(\mathbf{x}_1)$

$(\mathrm{var}(\mathbf{x}_1|\mathbf{x}_2))^{-1}$

$\mathrm{var}(\mathbf{x}_1|\mathbf{x}_2)=\Omega_{11}^{-1}$

高斯图模型

当关注条件相依和条件独立时，多元正态分布
称为高斯图模型（并以无向图表示）

记号约定

$\mathbf{x}_A$与$\mathbf{x}_{-A}$

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

Σ_{AB}
$\Sigma_{AB\bullet C}$

假设$\Sigma=\mathrm{cov}(\mathbf{x})$是$p\times p$正定矩阵，设下标集合$A,B,C\subset\{1,2,...,p\}$，互斥。

- $\Sigma_{AB}=\mathrm{cov}(\mathbf{x}_A,\mathbf{x}_B)$为$\Sigma$的行标为$A$，列标为$B$的子矩阵；
- $\Sigma_{AB\bullet C}=\Sigma_{AB}-\Sigma_{AC}\Sigma_{CC}^{-1}\Sigma_{CB}$.

条件分布
∝联合分布

常用技巧：

$p(\mathbf{x}_S \mid \mathbf{x}_{-S}) = p(\mathbf{x}_S, \mathbf{x}_{-S}) / p(\mathbf{x}_{-S}) \propto p(\mathbf{x})$，$p(\mathbf{x})$中的$\mathbf{x}_{-S}$视作常数

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$p(\mathbf{x}_S \mid \mathbf{x}_{-S})$，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ψ_1，ψ_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：$\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z}$

证:(1)略。（2）给定$\mathbf{z}$时的条件概率密度

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

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

注意：$\mathbf{z}$是常数

性质6*将多元正态分布的条件分布用逆矩阵$\Omega = \Sigma^{-1}$表达了出来。从多元正态概率密度函数的形式来看，这是比较自然的：记$\Omega = \Sigma^{-1} = (\omega_{ij})$,则我们可以展开$\mathbf{x}\sim N_p(0,\Sigma)$指数上的二次型：

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

则我们可以利用引理1方便地研究$f(\mathbf{x})$的分解和条件独立性质，再次得到性质6*($q=1$情形)，并得到条件相关系数的表达

定理1. 假设 $\mathbf{x}\sim N_p(\mu,\Sigma)$, 记$\Omega = \Sigma^{-1} = (\omega_{ij})$, 则

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

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

$$\rho_{x_i,x_j|\mathbf{x}_{-(i,j)}} = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}, i\neq j.$$

特别地，$\omega_{ij}=0 \Leftrightarrow x_i \perp\!\!\!\perp x_j|\mathbf{x}_{-(i,j)}$（条件独立 $\Leftrightarrow$ 条件不相关）

注: 非正态情形下，偏相关系数$\rho_{ij\bullet -(ij)} = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}$

定理1的证明:

(1). 不妨假设$\boldsymbol{\mu}=0$，给定$\mathbf{x}_{-i}$条件下，x_i的条件概率密度

$$p(x_i|\mathbf{x}_{-i}) \propto p(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\omega_{ii}x_i^2-\left(\sum_{j\neq i}\omega_{ij}x_j\right)x_i\right)$$

指数上是x_i的二次函数 $\Rightarrow$ $x_i|\mathbf{x}_{-i}\sim N_1\left(-\sum_{j\neq i}\omega_{ij}x_j/\omega_{ii}\,,1/\omega_{ii}\right)$。

$N(\mu,\sigma^2)$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μ/σ^2.

(2). 给定$\mathbf{x}_{-(i,j)}=(x_u,u\neq i,j)$时，$x_i$和$x_j$的联合密度

$$p\left(x_i,x_j\middle|\mathbf{x}_{-(i,j)}\right)\propto p(x_i,x_j,\mathbf{x}_{-(i,j)})=p(\mathbf{x})$$
$$\propto \exp\left(-\frac{1}{2}\omega_{ii}x_i^2-\frac{1}{2}\omega_{jj}x_j^2-\omega_{ij}x_ix_j+ax_i+bx_j+c\right)\quad(*)$$

其中一次项的系数a,b,c只与$\mathbf{x}_{-(i,j)}$有关，是常数 。

将(*)指数上以矩阵向量表示

$$p\left(x_i, x_j \middle| \mathbf{x}_{-(i,j)}\right) \propto \exp\left(-\frac{1}{2}(x_i, x_j)\begin{pmatrix}\omega_{ii} & \omega_{ij}\\ \omega_{ji} & \omega_{jj}\end{pmatrix}\begin{pmatrix}x_i\\ x_j\end{pmatrix} + ax_i + bx_j + c\right)$$

$$\Rightarrow \begin{pmatrix}x_i\\ x_j\end{pmatrix}|\mathbf{x}_{-(i,j)} \sim N_2\left(*, \begin{pmatrix}\omega_{ii} & \omega_{ij}\\ \omega_{ji} & \omega_{jj}\end{pmatrix}^{-1}\right),$$

该正态分布的协方差矩阵

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

x_i, x_j的相关系数（给定$\mathbf{x}_{-(i,j)}$条件下）$\rho_{x_i,x_j|\mathbf{x}_{-(i,j)}} = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}$.

若$\omega_{ij} = 0$, 则$\rho_{x_i,x_j|\mathbf{x}_{-(i,j)}} = 0$, x_i, x_j条件独立。

这也可以从(*)容易看出：当$\omega_{ij} = 0$，（*）中二次型无交叉项，从而 $p\left(x_i, x_j \middle| \mathbf{x}_{-(i,j)}\right) = \varphi_1(x_i)\varphi_2(x_j) \Leftrightarrow x_i, x_j$条件独立。证毕

条件(偏)相关系数矩阵

协方差矩阵$\Sigma_{p\times p}=(\sigma_{ij}),\Omega_{p\times p}=\Sigma^{-1}=(\omega_{ij})$，偏相关系数矩阵

$$R_p=(\rho_{ij\bullet-(ij)})=\left(-\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\right)=2I_p-C^{-1/2}\Omega C^{-1/2},C=diag(\Omega).$$

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

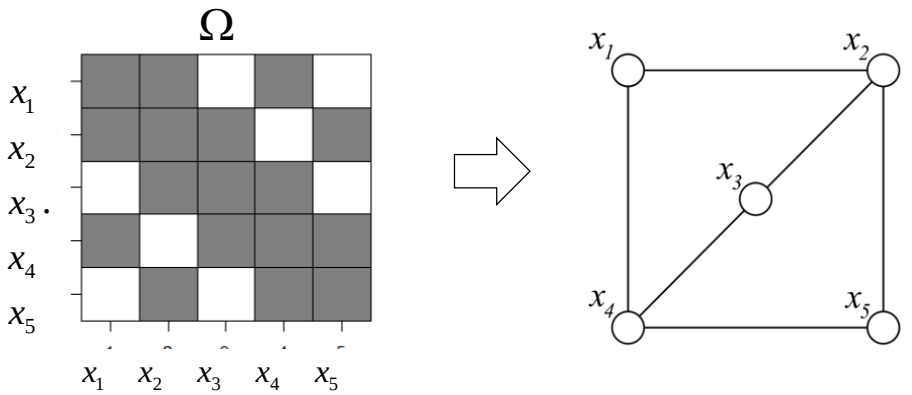
高斯图模型基于精度矩阵或条件相关系数矩阵的0元素，将变量之间的关系以无向图(undirected graph/network)表达出来。

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：

$$i\sim j\ (\text{即}x_i\text{和}x_j\text{以边相连})\Leftrightarrow\omega_{ij}\neq 0\Leftrightarrow x_i\text{和}x_j\text{的条件相关系数}\neq 0$$



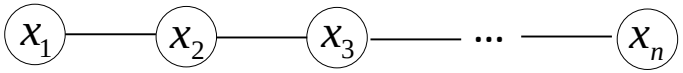
空白处为0

例3. 假设$\mathbf{x}=(x_1,\ldots,x_n)^{\mathsf{T}} \sim N_n(0,\Sigma), -1<\rho<1$

$$\Sigma=\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix},$$

$$\Sigma^{-1}=(1-\rho^2)\begin{pmatrix} 1 & -\rho & & & & \\ -\rho & 1+\rho^2 & -\rho & & & \\ & -\rho & 1+\rho^2 & \ddots & & \\ & & \ddots & \ddots & & \\ & & & & 1+\rho^2 & -\rho \\ & & & & -\rho & 1 \end{pmatrix}$$

图表示：



这称为马氏链或AR1时间序列模型

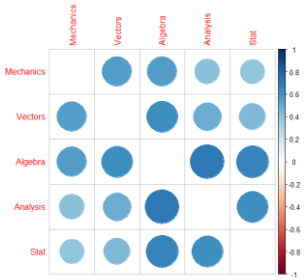
注：马氏链的常规定义：$P(x_i \mid x_{i-1}, x_{i-2},\ldots,x_1)=P(x_i \mid x_{i-1})$,

$\Leftrightarrow$ 马氏随机场的定义：$P(x_i \mid \mathbf{x}_{-i})=P(x_i \mid x_{i-1}, x_{i+1})$

例4. 88个学生5门数学成绩相关系数矩阵如下,我们关心：统计课成绩与力学成绩的相关系数是0.39，学习力学对统计有帮助吗？我们假设5门课程成绩联合服从多元正态分布。

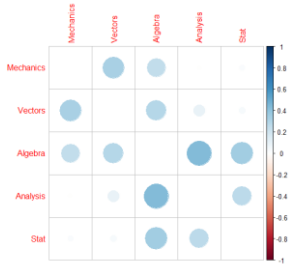
相关性系数矩阵

	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



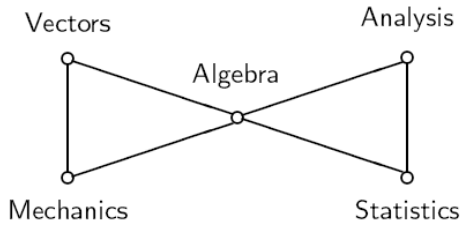
条件相关性系数矩阵

	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5门课的关系：



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

概率图模型：马氏随机场

图graph

图由节点(node,vertex)和它们之间的连线(边,edge)组成。节点i,j之间有边，记作$i\sim j$ 或$i \to j$。

概率图模型

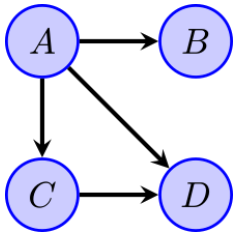
概率图模型(probabilistic graphical model)或图模型是一种概率模型（多元概率分布），它用图（graph）表示随机变量之间的条件相依性。图模型主要包括

- 马氏随机场(MRF, Markov random fields)
- 贝叶斯网络(BN, Bayesian network)。

贝叶斯网络

定义3.贝叶斯网络用有向图的节点表示相依随机变量$x_1,\dots,x_n$,(即随机向量$\mathbf{x}$),每个节点的分布仅依赖于其上一级父节点(parents nodes), 代表因果传递关系。所有变量的联合分布：

$$p(\mathbf{x}) = p(x_1,\dots,x_n) = \prod_{i=1}^{n} p(x_i|pa(x_i))$$



$$p(A,B,C,D) = p(A)p(B|A)p(C|A)p(D|A,C)$$

$p(\bullet)$表示概率密度(连续)或概率 (离散).

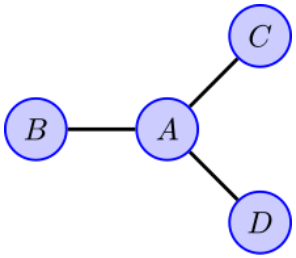
BN主要应用条件概率计算公式，与一般的概率计算无异。
下面我们主要考虑MRF

马氏随机场

定义4. 马氏随机场MRF用无向图(graph) 表示多个变量之间的条件相依和条件独立关系。图的节点是随机变量，条件相依的节点/变量之间连线构成图的边，条件独立的变量之间不连线。

高斯图模型是最重要的连续马氏随机场。

例如，下图表示给定变量A时，变量B, C, D条件独立



联合分布仅与cliques有关（HC定理）：

$$p(A, B, C, D) = f_1(AB)\, f_2(AC)\, f_3(AD)$$

Cliques 是全联通的最大子集：
$\{A, B\}, \{A, C\}, \{A, D\}$

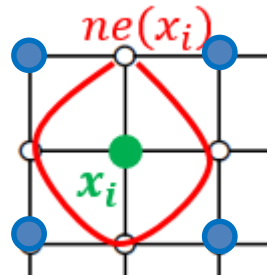
局部马氏性

对任一节点i (或x_i)，定义其邻域为所有与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，MRF图模型表达了局部马氏性质：

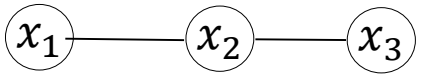
$$x_i \perp\!\!\!\perp \mathbf{x}_{-(i,ne(i))} | \mathbf{x}_{ne(i)} \Leftrightarrow p(x_i|\mathbf{x}_{-i}) = p\left(x_i|\mathbf{x}_{ne(i)}\right)$$



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

给定x_i的邻域$ne(x_i)$，x_i与不相邻的$\mathbf{x}_{-(i,ne(i))}$蓝点独立，蓝点通过邻域$ne(x_i)$对x_i发生影响。

例如，假设随机变量x_1, x_2, x_3图表示如下：



$$\Leftrightarrow x_1 \perp\!\!\!\perp x_3 | x_2 \Leftrightarrow P(x_3|x_1, x_2) = P(x_3|x_2)$$

一般，这种直线链状图表示适用于下标有次序的情形（比如下标是时间），这称为马氏链；

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。更常见的是反问题：

依据实际背景（时间或空间结构）假设局部马氏性

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

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、自洽的马氏随机场或多元分布？

例5（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

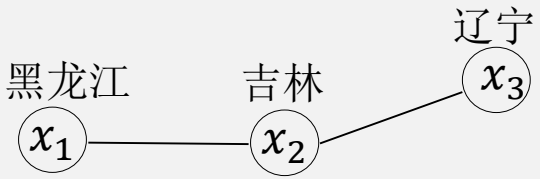
$$y|x \sim N(ax, \sigma^2), x|y \sim N(by, \tau^2), \quad \Leftrightarrow$$

$$y = ax + \varepsilon_1,\ \varepsilon_1 \sim N(0, \sigma^2),\ \varepsilon_1 \perp\!\!\!\perp x$$
$$x = by + \varepsilon_2,\ \varepsilon_2 \sim N(0, \tau^2),\ \varepsilon_2 \perp\!\!\!\perp y$$

它们是否矛盾，(x, y)是否有一个恰当的二元分布？

由HC定理，如果$ab < 1, a/\sigma^2 = b/\tau^2$，则$(x, y)$服从二元正态

例5（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：

$$x_1 = \beta_{12}x_2 + \varepsilon_1,\ \ \varepsilon_1 \sim N(0, \kappa_1^{-1}),$$
$$x_2 = \beta_{21}x_1 + \beta_{23}x_3 + \varepsilon_2,\ \varepsilon_2 \sim N(0, \kappa_2^{-1}),$$
$$x_3 = \beta_{32}x_2 + \varepsilon_3,\ \varepsilon_3 \sim N(0, \kappa_3^{-1}),$$

何时这些方程是合理的/相容的（可唯一确定一个MRF）？

由Hammersley − Clifford定理，条件是

$$\Omega = \begin{pmatrix} \kappa_1 & -\kappa_1\beta_{12} & 0 \\ -\kappa_2\beta_{21} & \kappa_2 & -\kappa_2\beta_{23} \\ 0 & -\kappa_3\beta_{32} & \kappa_3 \end{pmatrix} > 0$$ （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？

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

定理2 (Hammersley-Clifford定理的特殊情形). 假设$\mathbf{x} = (x_1, \ldots, x_p)^{\mathsf{T}}$ 服从边际正态回归模型（条件正态）：

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

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

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 $K = diag(1/\kappa_1, \ldots 1/\kappa_p\,)$,

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

HC定理给出了MRF的联合分布形式

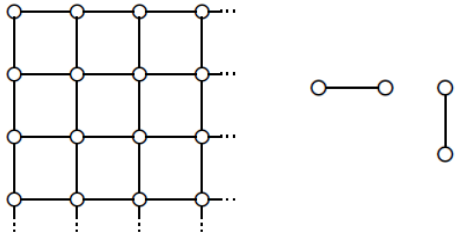
定理3 (Hammersley-Clifford定理). 若马氏随机场的联合分布$p(\mathbf{x}) > 0$（未必正态）, 则具有如下Boltzmann分布形式

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

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$p(\mathbf{x}) > 0$,即何时这才是一个真正的分布？参见例3，定理3

例如，如果$\mathbf{x} = (x_i, i = 1,2,..)$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

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

这称为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

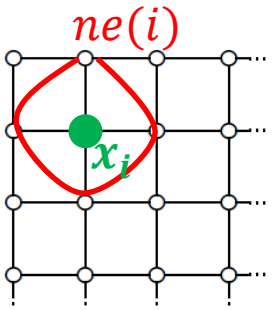
$$P(X_1 = x_1, \ldots, X_n = x_n) = \frac{1}{Z}\exp\left(-a - \sum_i b_i x_i - \sum_{i\sim j} d_{ij} x_i x_j\right), x_i = \pm 1$$

显然$p(\mathbf{x})$是一个恰当的多元分布（与正态相像）。

伊辛模型 Ising model

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，平面晶格上有若干小磁针,每个磁针有两个自旋状态spin：$x_i = -1\ or\ 1$，其中每个格点的邻域$ne(i)$由与其直接连结的4个点组成。假设每个磁针只受临近磁针影响，即 $(x_1, \ldots., x_n)^\top$是一个马氏随机场：

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



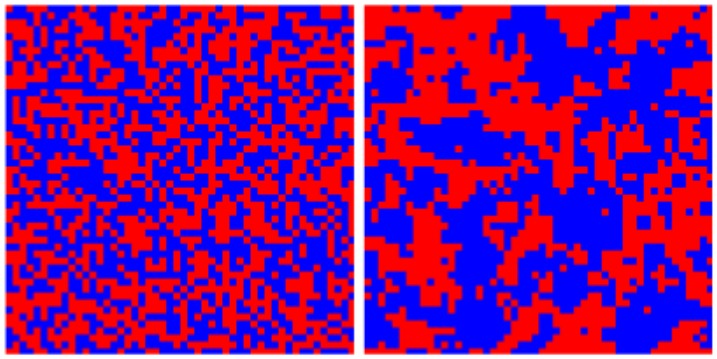
一致性假设下, 所有磁针的联合分布(Boltzmann分布、Gibbs分布):

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

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代表相互作用，h代表外部磁场， T代表温度。 $J > 0$ 时可磁化：

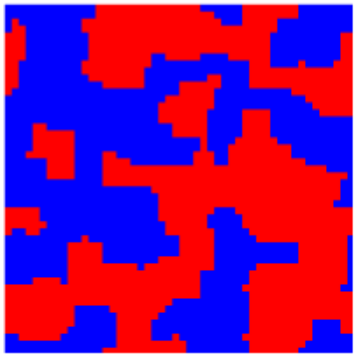
- 若温度较低，相邻磁针趋于同号（能量低，概率大），显现磁性。
- 当温度升高，到一定程度时磁性会突然消失，称为相变。

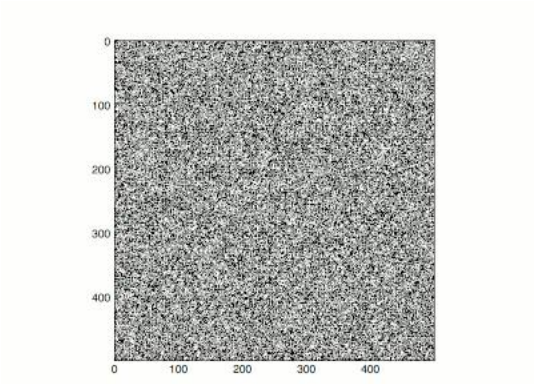
磁性
消失



相变

磁性





一个著名的例子：应用Ising模型研究社会意见的趋同和分裂。

配分与相变

Ising模型是一个多元二值分布，考虑简单情形($\theta = \frac{J}{T}, h = 0$)：

$$p(x_1, \dots, x_n) = \frac{1}{Z}\exp(\theta \sum_{i\sim j} x_i x_j)\,,\ x_i = -1,1$$

其中配分函数(partition)

$$Z = \sum_{\mathbf{x}\in\{-1,1\}^n} \exp(\theta \sum_{i\sim j} x_i x_j) = \sum_{\mathbf{x}\in\{-1,1\}^n} \prod_{i\sim j}(\dots)$$

是2^n项之和。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 单位圆定理（ circle theorem）证明了零点在复平面圆周上（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; Lee&Yang, 1952）；

3d Ising: open

Lee, T. D.; Yang, C. N.(1952), Statistical Theory of Equations of State and Phase Transitions. II. Lattice Gas and Ising Model, Physical Review, 87 (3): 410–419.
Yang, C. N. (1952), "The spontaneous magnetization of a two-dimensional Ising model", Physical Review, Series II, 85 (5): 808–816.

简化版单位圆定理

若 $a \in [-1,1]$, 则多项式

$$p_n(z) = \sum_{k=0}^{n} \binom{n}{k} a^{k(n-k)} z^k$$

的所有零点都分布在单位圆周 $|z| = 1$ 上.