2014.3.6

第六讲．多元正态分布

1. 多元正态分布（中文课本第二章）

定义：若n维随机向量$\mathbf{x}$的密度为

$$f(\mathbf{x})=\frac{1}{(2\pi)^{n/2}|\Sigma|^{-1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),\quad \mathbf{x}\in \mathrm{R}^n$$

其中参数$\boldsymbol{\mu}\in \mathrm{R}^n$，$\Sigma$为$n\times n$正定矩阵，则$\mathbf{x}$服从多元正态分布，记为$\mathbf{x}\sim N_n(\boldsymbol{\mu},\Sigma)$。特别地，$\mathbf{x}\sim N_n(0,\mathrm{I}_n)\Leftrightarrow \mathrm{x}_1,..,\mathrm{x}_\mathrm{n}\, iid\sim N(0,1)$

定理1(不相关等价于独立). 若$\mathbf{x}\sim N_n(\boldsymbol{\mu},\Sigma)$，分块：

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

其中$\mathbf{x}_1$，$\boldsymbol{\mu}_1$长度为$k<n$，Σ_{11}为$k\times k$矩阵. 则$\mathbf{x}_1$与$\mathbf{x}_2$独立 $\Leftrightarrow$ $\Sigma_{12}=0$.

定理2：$\mathbf{x}\sim N_n(\boldsymbol{\mu},\Sigma)$,则$\mathbf{y}=\Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\sim N_n(0,I_n)$

证明：令$\mathbf{y}=\Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu})$,该变换的$Jacobi$行列式为$\mathrm{J}=\left|\frac{\partial\mathbf{x}}{\partial\mathbf{y}}\right|=\left|\Sigma^{1/2}\right|=\left|\Sigma\right|^{1/2}$

得$\mathbf{y}$的密度函数,$f(\mathbf{y})=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-\frac{y_i^2}{2}}=\frac{1}{(2\pi)^{n/2}}e^{-\frac{\mathbf{y}'\mathbf{y}}{2}}$,按定义为$N_n(0,I_n)$的密度函数。所以$\mathbf{y}=\Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\sim N_n(0,I_n)\Leftrightarrow y_1,...,y_\mathrm{n}\, iid\sim N(0,1)$

推论：$\mathbf{x}\sim N_n(\boldsymbol{\mu},\Sigma)$,则$\mathrm{E}(\mathbf{x})=\boldsymbol{\mu}$, $\mathrm{var}(\mathbf{x})=\Sigma$

证明：令$\mathbf{y}=\Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\sim N_n(0,I_n)\Leftrightarrow y_1,...,y_\mathrm{n}\, iid\sim N(0,1)$

则$E(\mathbf{y})=0,\mathrm{var}(\mathbf{y})=I_n$。由$\mathbf{x}=\boldsymbol{\mu}+\Sigma^{1/2}\mathbf{y}\Rightarrow$

$\mathrm{E}(\mathbf{x})=\boldsymbol{\mu}+\Sigma^{1/2}E(\mathbf{y})=\boldsymbol{\mu}$, $\mathrm{var}(\mathbf{x})=\Sigma^{1/2}\,\mathrm{var}(\mathbf{y})\Sigma^{1/2}=\Sigma.$

定理3：$\mathbf{x}\sim N_n(\boldsymbol{\mu},\Sigma)\Leftrightarrow$

矩母函数 $E\exp(\mathbf{t}'\mathbf{x})=\exp\left(\mathbf{t}'\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right)$,任意$\mathbf{t}\in \mathrm{R}^n$

证明：令$\mathbf{y}=\Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\sim N_n(0,I_n)$

$\Leftrightarrow y_1,...,y_\mathrm{n}$ iid $\sim N(0,1)$

已知对任意$u\in R$, $E\exp(uy_i)=\exp(u^2/2)$,所以对$\mathbf{s}\in R^n$

$E\exp(\mathbf{s}'\mathbf{y})=\prod E\exp(s_iy_i)=\exp(\mathbf{s}'\mathbf{s}/2)$,

所以$E\exp(\mathbf{t}'\mathbf{x})=\mathrm{E}\exp(\mathbf{t}'(\boldsymbol{\mu}+\Sigma^{1/2}\mathbf{y}))$

$=\exp(\mathbf{t}'\boldsymbol{\mu})E\exp((\Sigma^{1/2}\mathbf{t})'\mathbf{y})=\exp(\mathbf{t}'\boldsymbol{\mu}+\mathbf{t}'\Sigma\mathbf{t}/2)$

定理4：若$\mathbf{x}\sim N_n(\boldsymbol{\mu},\Sigma)$，A为任一$k\times n$矩阵（行满秩），$\mathbf{b}\in \mathrm{R}^k$常数向量，则 $\mathrm{A}\mathbf{x}+\mathbf{b}\sim N_k(\mathrm{A}\boldsymbol{\mu}+\mathbf{b},\mathrm{A}\Sigma\mathrm{A}')$

证明：对任何$\mathbf{t}\in R^k,\mathbf{s}=A'\mathbf{t}\in R^n$

$$E\exp(\mathbf{t}'(A\mathbf{x}+\mathbf{b}))=\exp(\mathbf{t}'\mathbf{b})E\exp(\mathbf{t}'A\mathbf{x})=\exp(\mathbf{t}'\mathbf{b})E\exp(\mathbf{s}'\mathbf{x})$$
$$=\exp(\mathbf{t}'\mathbf{b}+\mathbf{s}'\boldsymbol{\mu}+\mathbf{s}'\Sigma\mathbf{s}/2)=\exp(\mathbf{t}'(\mathbf{b}+A\boldsymbol{\mu})+\mathbf{t}'(A\Sigma A')\mathbf{t}/2)$$

所以$A\mathbf{x}+\mathbf{b}\sim N(A\boldsymbol{\mu}+\mathbf{b},A\Sigma A')$

推论：$\mathbf{x}\sim N_n(0,\mathrm{I_n}),A_{k\times n},B_{m\times n}$为对行满秩矩阵，若$\mathrm{AB}'=0$，则$A\mathbf{x}$与$\mathrm{B}\mathbf{x}$独立。

证明：注意$k+m\le n$，$\begin{pmatrix}A\\B\end{pmatrix}$行满秩，所以$\begin{pmatrix}A\\B\end{pmatrix}\mathbf{x}\sim N\left(0,\begin{pmatrix}AA' & 0\\ 0 & BB'\end{pmatrix}\right)$

定理5（条件分布）：若$\mathbf{x}\sim N_n(\boldsymbol{\mu},\Sigma)$，分块：

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

其中$\mathbf{x}_1$，$\boldsymbol{\mu}_1$长度为$k<n$，Σ_{11}为$k\times k$矩阵. 则

(1) $\mathbf{x}_1\sim N_k(\boldsymbol{\mu}_1,\Sigma_{11})$，$\mathbf{x}_2\sim N_{n-k}(\boldsymbol{\mu}_2,\Sigma_{22})$;

(2) $\mathbf{x}_1\mid\mathbf{x}_2\sim N_k\left(\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2),\Sigma_{11\bullet 2}\right)$

该结果说明如下重要事实：

(a) $\mathrm{E}(\mathbf{x}_1\mid\mathbf{x}_2)=\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)$是$\mathbf{x}_2$的线性函数;

(b) $\mathrm{var}(\mathbf{x}_1\mid\mathbf{x}_2)=\Sigma_{11\bullet 2}$，不依赖于$\mathbf{x}_2$的具体值.

定理5的证明：

(1)令$A=(I_k,0),\mathbf{x}_1=A\mathbf{x}\sim N(A\boldsymbol{\mu},A\Sigma A')=N(\boldsymbol{\mu}_1,\Sigma_{11})$

(2)正交化，令 $\mathbf{y}=\begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{pmatrix}=\begin{pmatrix}\mathrm{I}_k & -\Sigma_{12}\Sigma_{22}^{-1}\\ 0 & \mathrm{I}_{n-k}\end{pmatrix}\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}\triangleq \mathrm{A}\mathbf{x}$

$$\sim \mathrm{N}(\mathrm{A}\boldsymbol{\mu},\ \mathrm{A}\Sigma\mathrm{A}')=N\left(\begin{pmatrix}\boldsymbol{\mu}_1-\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2\\ \boldsymbol{\mu}_2\end{pmatrix},\begin{pmatrix}\Sigma_{11.2} & 0\\ 0 & \Sigma_{22}\end{pmatrix}\right)$$

所以$\mathbf{y}_1\sim N_k(\boldsymbol{\mu}_1-\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2,\Sigma_{11\bullet 2})$，$\mathbf{y}_2\sim N_{n-k}(\boldsymbol{\mu}_2,\Sigma_{22})$，且两者独立。

所以：$\mathbf{x}_1=\mathbf{x}_1-\Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2+\Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2=\mathbf{y}_1+\Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2$

给定$\mathbf{x}_2=\mathbf{u}$时，$\mathbf{x}_1\sim N_k\left(\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\mathbf{u}-\boldsymbol{\mu}_2),\Sigma_{11\bullet 2}\right)$

方差分解公式

一般地，对任意随机向量$\mathbf{x}_1,\mathbf{x}_2$，有方差分解

$$\mathrm{var}(\mathbf{x}_1)\ =\ \mathrm{var}\,E(\mathbf{x}_1\mid\mathbf{x}_2)\ +\ \mathrm{E}\,\mathrm{var}(\mathbf{x}_1\mid\mathbf{x}_2)$$

$\mathbf{x}_1$的方差 = 回归能解释的部分 + 回归不能解释的部分

条件期望$E(\mathbf{x}_1\mid\mathbf{x}_2)$ 称为回归函数.

证明：令$\hat{\mathbf{x}}_1=E(\mathbf{x}_1\mid\mathbf{x}_2)$，$\mathbf{x}_1^{\perp}=\mathbf{x}_1-E(\mathbf{x}_1\mid\mathbf{x}_2)$，

$\mathbf{x}_1^{\perp}$与$\mathbf{x}_2$不相关，特别地$\mathbf{x}_1^{\perp}$与$\hat{\mathbf{x}}_1$不相关，

$\mathbf{x}_1=\hat{\mathbf{x}}_1+\mathbf{x}_1^{\perp}$可以称为"正交"分解，两边取方差

$$\mathrm{var}(\mathbf{x}_1)=\mathrm{var}(\hat{\mathbf{x}}_1)+\mathrm{var}(\mathbf{x}_1^{\perp})=\mathrm{var}\,E(\mathbf{x}_1\mid\mathbf{x}_2)+\mathrm{var}(\mathbf{x}_1-E(\mathbf{x}_1\mid\mathbf{x}_2))$$

而$\mathrm{var}(\mathbf{x}_1-E(\mathbf{x}_1\mid\mathbf{x}_2))=E(\mathbf{x}_1-E(\mathbf{x}_1\mid\mathbf{x}_2))(\mathbf{x}_1-E(\mathbf{x}_1\mid\mathbf{x}_2))'$

$$=E[E(\mathbf{x}_1-E(\mathbf{x}_1\mid\mathbf{x}_2))(\mathbf{x}_1-E(\mathbf{x}_1\mid\mathbf{x}_2))'\mid\mathbf{x}_2]=E[\mathrm{var}(\mathbf{x}_1\mid\mathbf{x}_2)]$$

对于正态分布，定理5中，由$\mathbf{x}_1 \mid \mathbf{x}_2 \sim N_k\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11\bullet 2}\right)$

$$\operatorname{var} E(\mathbf{x}_1 \mid \mathbf{x}_2) = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$\operatorname{var}(\mathbf{x}_1 \mid \mathbf{x}_2) = \Sigma_{11.2}\text{(常数矩阵)} \Rightarrow \mathrm{E}(\operatorname{var}(\mathbf{x}_1 \mid \mathbf{x}_2)) = \operatorname{var}(\mathbf{x}_1 \mid \mathbf{x}_2).$$

$$\begin{array}{ccccc} \operatorname{var}(\mathbf{x}_1) & = & \operatorname{var} E(\mathbf{x}_1 \mid \mathbf{x}_2) & + & \mathrm{E}\operatorname{var}(\mathbf{x}_1 \mid \mathbf{x}_2) \\ \Sigma_{11} & = & \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & + & \Sigma_{11\bullet 2} \end{array}$$

2. 高斯图模型

条件相关系数

定义：给定随机向量z时，随机变量y, x的条件相关系数为

$$\rho_{yx|\mathbf{z}} = \frac{\operatorname{cov}(y, x \mid \mathbf{z})}{\sqrt{\operatorname{var}(x \mid \mathbf{z}) \operatorname{var}(y \mid \mathbf{z})}}$$

注意条件相关系数$\rho_{yx|\mathbf{z}}$可能依赖于$\mathbf{z}$。

但在正态情形下，$\rho_{yx|\mathbf{z}}$不依赖于z的特定取值，而且$\rho_{yx|\mathbf{z}} = \rho_{yx\bullet\mathbf{z}}$

定理6. 多元正态假设下，条件相关系数即为偏相关系数

证明：假设多元正态：

$$\begin{pmatrix} y \\ x \\ \mathbf{z} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_y \\ \mu_x \\ \mu_\mathbf{z} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} & \Sigma_{y\mathbf{z}} \\ \Sigma_{xy} & \Sigma_{xx} & \Sigma_{x\mathbf{z}} \\ \Sigma_{\mathbf{z}y} & \Sigma_{\mathbf{z}x} & \Sigma_{\mathbf{zz}} \end{pmatrix}\right) \Rightarrow$$

$$\left.\begin{pmatrix} y \\ x \end{pmatrix}\right|_{\mathbf{z}} \sim N\left(\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix} + \begin{pmatrix} \Sigma_{y\mathbf{z}} \\ \Sigma_{x\mathbf{z}} \end{pmatrix}\Sigma_{\mathbf{zz}}^{-1}(\mathbf{z} - \mu_\mathbf{z}), \Omega\right)$$

其中 $\Omega = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} - \begin{pmatrix} \Sigma_{y\mathbf{z}} \\ \Sigma_{x\mathbf{z}} \end{pmatrix}\Sigma_{\mathbf{zz}}^{-1}\left(\Sigma_{\mathbf{z}y}, \Sigma_{\mathbf{z}x}\right)$

$$= \begin{pmatrix} \Sigma_{yy} - \Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{z}y} & \Sigma_{yx} - \Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{z}x} \\ \Sigma_{xy} - \Sigma_{x\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{z}y} & \Sigma_{xx} - \Sigma_{x\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{z}x} \end{pmatrix}$$

则 条件相关系数 $\rho_{y,x|z} = \dfrac{\operatorname{cov}(y, x \mid \mathbf{z})}{\sqrt{\operatorname{var}(x \mid \mathbf{z}) \operatorname{var}(y \mid \mathbf{z})}}$

$$= \frac{\Sigma_{yx} - \Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{z}x}}{\sqrt{\Sigma_{yy} - \Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{z}y}}\sqrt{\Sigma_{xx} - \Sigma_{x\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{z}x}}} = \rho_{y,x\bullet z}$$

例. 高中男生的数学成绩**x**与语文成绩**y**的相关系数为**0.6**,女生的相关系数也是**0.6**. 假设男女生数学成绩平均来看没有差异, 但女生的语文成绩要好于男生, 如果男女生混合在一起, 得到的相关系数是大于、等于还是小于**0.6**?

$z=1,0$分别代表男女，$\rho_{xy\bullet z}=0.6$

$\rho_{xz}=0$，故$0.6=\rho_{xy\bullet z}=\dfrac{\rho_{xy}}{\sqrt{1-\rho_{yz}^2}}\geq\rho_{xy}$

定理7:
设 $\mathbf{x}\sim N_n(\boldsymbol{\mu},\Sigma)$, 若$\Sigma^{-1}$的$(i,j)$元素$\omega_{ij}$为0，或等价地相关系数矩阵之逆的$(i,j)$元素为0，则给定其它变量时，$x_i$与$x_j$条件独立。

证明1：由偏相关系数与协方差矩阵逆之间的关系知，

$\omega_{ij}=0\Leftrightarrow\rho_{ij|\text{其它}}=0$

（由于正态分布不相关即独立）

$\Leftrightarrow$ x_i与x_j条件独立。

证明2：令$\Omega=-\Sigma^{-1}=\left(\omega_{ij}\right)_{i,j=1,\dots,n}$，不妨设$\omega_{12}=0$

联合密度

$f(\mathbf{x})=\mathrm{C}\times\exp(\omega_{11}x_1^2+\omega_{22}x_2^2+x_1h_1+x_2h_2+g)$

其中C是常数，h_1,h_2,g都只是$(x_3,\dots,x_n)$的函数。

给定$x_3,\dots,x_n$(为常数)时，x_1,x_2的联合密度

$f(x_1,x_2\mid x_3,\dots,x_n)\propto f(\mathbf{x})$

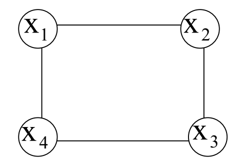
$=\mathrm{C}\times\mathrm{e}^g\times\exp(\omega_{11}x_1^2+\omega_{22}x_2^2+x_1h_1+x_2h_2)$

为独立二元正态密度的乘积，所以x_1,x_2条件独立。

高斯图模型：
以节点代表多元正态随机向量的各个分量，若若$\left(\Sigma^{-1}\right)_{ij}=0$，即$i,j$分量条件独立，则它们之间无连线。

例如：$(x_1,x_2,x_3,x_4)'\sim N_4(\mu,\Sigma)$,

$$\Sigma^{-1}=\begin{pmatrix} * & * & 0 & * \\ * & * & * & 0 \\ 0 & * & * & * \\ * & 0 & * & * \end{pmatrix}\quad\Longleftrightarrow$$



例1. 88个学生5门数学成绩(联合正态)相关系数矩阵如下，

	Mechanics	Vectors	Algebra	Analysis	Statistics
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
Statistics	0.39	0.44	0.66	0.61	1.00

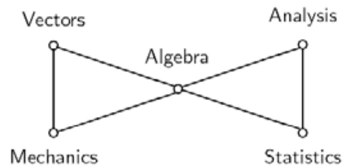
我们关心的问题比如有：统计课成绩与力学成绩的相关系数是**0.39**， 学好力学对学好统计有帮助吗？为此我们研究它们之间的偏/条件相关性。

偏相关性系数矩阵如下：

	Mechanics	Vectors	Algebra	Analysis	Statistics
Mechanics	1.0	0.3	0.2	0.0	0.0
Vectors	0.3	1.0	0.3	0	0.0
Algebra	0.2	0.3	1.0	0.4	0.3
Analysis	0.0	0	0.4	1.0	0.3
Statistics	0.0	0.0	0.3	0.3	1.0

给定其他课程成绩条件下，统计与力学程是条件独立的！

以高斯图模型表示5门课成绩的相依关系：



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

例2(一阶自回归模型). 若$\mathbf{x} \sim N_n(\boldsymbol{\mu}, \Sigma)$,其中

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

$$因为\Sigma^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix},所以\rho_{x_i, x_j|其它} = 0，当|i-j|>1;$$

图模型表示： $x_1 — x_2 — x_3 — \cdots — x_{n-1} — x_n$