第六讲 Wishart分布

2023.3.18

随机矩阵的分布

Recap

假设所有n个晶格点 的 spin (± 1 或0/1) 联合分布为Boltzman分布：

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

注：统计学的指数族分布尝试以Boltzman分布统一概率分布

其中$H(\mathbf{x})$是能量函数，(homogeneous) Ising 模型假设

$$H(\mathbf{x}) = -J\sum_{i\sim j} x_i x_j - h\sum_i x_i，\quad i\sim j\text{表示近邻}$$

这实际上假设了局部马氏性即相邻的格点存在直接交互作用（HC定理）。

简言之，Ising模型是一个多元伯努利（或二值）分布：

$$p(x_1,\ldots,x_n) = \frac{1}{Z}\exp\left\{\left(J\sum_{i\sim j} x_i x_j + h\sum_i x_i\right)/T\right\},\ x_i = -1,1$$

其中配分函数(partition)

$$Z = \sum_{\mathbf{x}\in\{-1,1\}^n}\exp\left\{\left(J\sum_{i\sim j} x_i x_j + h\sum_i x_i\right)/T\right\} = \sum_{\mathbf{x}\in\{-1,1\}^n}\prod_{i\sim j}(\ldots)$$

是2^n项之和，每一项是乘积（多项式），如下式

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.

我们只考察$n=1,2$情况

$n=1$情形： $p(x)=e^{\theta x}/\mathrm{Z}(\theta)$，$x=0,1$

$p(0)+p(1)=1 \Rightarrow$配分函数$\mathrm{Z}(\theta)=1+e^{\theta}$，

$$p(x)=\frac{1}{1+e^{\theta}}e^{\theta x}=\left(\frac{e^{\theta}}{1+e^{\theta}}\right)^{x}\left(\frac{1}{1+e^{\theta}}\right)^{1-x}=p^{x}(1-p)^{x}, p=\frac{e^{\theta}}{1+e^{\theta}}$$

这是伯努利分布。 $\mathrm{Z}(\theta)=0 \Rightarrow \theta=i\pi$

$n=2$： $p(x_1,x_2)=\frac{\exp(\theta_1(x_1+x_2)+\theta_2 x_1 x_2)}{\mathrm{Z}(\theta)}$， $x_1,x_2=0,1$

		x_2: 0	1
x_1	0	$1/\mathrm{Z}$	e^{θ_1}/Z
	1	e^{θ_1}/Z	$e^{2\theta_1+\theta_2}/\mathrm{Z}$

配分函数$\mathrm{Z}(\theta)=1+2e^{\theta_1}+e^{2\theta_1+\theta_2}=1+2z_1+z_1^2 z_2$

只有一本离散多元统计的专著(基本限于n=2,3情形):
Bishop , Fienberg, Holland (1974) Discrete Multivariate Analysis: Theory and Practice.

样本均值和样本方差

多元样本

假设样本$\mathbf{x}_1,\dots,\mathbf{x}_n \in R^p$，记数据矩阵：$X=(\mathbf{x}_1,\dots,\mathbf{x}_n)^\top \in R^{n\times p}$

样本均值：$\bar{\mathbf{x}}=\sum_{i=1}^{n}\mathbf{x}_i/n = X^\top\mathbf{1}/n,$

中心化矩阵：$X_c=(\mathbf{x}_1-\bar{\mathbf{x}},\dots,\mathbf{x}_n-\bar{\mathbf{x}})^\top = X-\mathbf{1}\bar{\mathbf{x}}^\top = P_{\mathbf{1}^\perp}X,$

样本方差：$S=\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top/(n-1) = X_c^\top X_c/(n-1)$

$$= X^\top P_{\mathbf{1}^\perp}X/(n-1)$$

$P_{\mathbf{1}} = \mathbf{1}\mathbf{1}^\top/n$

$P_{\mathbf{1}^\perp} = I_n - \mathbf{1}\mathbf{1}^\top/n$

验证：$(n-1)S=\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top = X_c^\top X_c = X^\top P_{\mathbf{1}^\perp}X = X^\top(I_n-\mathbf{1}\mathbf{1}^\top/n)X$。

一元情形$(p=1)$：$x_1,\dots,x_n$ iid，$\mathbf{x}=(x_1,\dots,x_n)^\top \in R^n$

样本均值：$\bar{x}=\sum_{i=1}^{n}x_i/n=\mathbf{x}^\top\mathbf{1}/n$，样本方差：$s^2=\sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1)=\mathbf{x}^\top P_{\mathbf{1}^\perp}\mathbf{x}/(n-1).$

样本均值和样本方差的无偏性

$\mathbf{x}_1,\ldots,\mathbf{x}_n \in R^p \ iid \sim (\boldsymbol{\mu},\Sigma)$ (总体未必正态),即$E(\mathbf{x}_i)=\boldsymbol{\mu}, \mathrm{var}(\mathbf{x}_i)=\Sigma$

样本均值：$\bar{\mathbf{x}}=\sum_{i=1}^n \mathbf{x}_i/n$，样本方差：$S=\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathrm{T}}/(n-1)$

命题1. 设$\mathbf{x}_1,\ldots,\mathbf{x}_n \ iid \sim (\boldsymbol{\mu},\Sigma)$,即$E(\mathbf{x}_i)=\boldsymbol{\mu}, \mathrm{var}(\mathbf{x}_i)=\Sigma$，则

(1) $E(\bar{\mathbf{x}})=\boldsymbol{\mu}, \mathrm{var}(\bar{\mathbf{x}})=\Sigma/n.$

(2) $E(S)=\Sigma.$

证: (1)$E(\bar{\mathbf{x}})=E(\sum_{i=1}^n \mathbf{x}_i/n)=\sum_{i=1}^n E\mathbf{x}_i/n=\boldsymbol{\mu}$,

$$\mathrm{var}(\bar{\mathbf{x}})=\mathrm{var}(\sum_{i=1}^n \mathbf{x}_i/n)=\sum_{i=1}^n \mathrm{var}\,\mathbf{x}_i/n^2=\Sigma/n$$

(2)$(n-1)S=\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathrm{T}}=\sum_{i=1}^n(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^{\mathrm{T}}-n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})^{\mathrm{T}}$

因为$\mathrm{var}(\mathbf{x}_i)=E(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^{\mathrm{T}}=\mathrm{var}(\mathbf{x}_i)=\Sigma$, $E(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})^{\mathrm{T}}=\mathrm{var}(\bar{\mathbf{x}})=\Sigma/n$

所以$(n-1)E(S)=n\Sigma-\Sigma=(n-1)\Sigma \Rightarrow E(S)=\Sigma.$

多元正态分布参数的极大似然估计

命题2. 设$\mathbf{x}_1,\ldots,\mathbf{x}_n$ $iid \sim N_p(\boldsymbol{\mu},\Sigma)$, $\Sigma > 0, n > p$ ，则$\boldsymbol{\mu}$和Σ的极大似然估计分别是样本均值$\bar{\mathbf{x}}$和$\tilde{S} = (n-1)S/n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf{T}}$。

证1：样本$\mathbf{x}_1,\ldots,\mathbf{x}_n$的联合概率密度函数（参数的似然函数）：

$$L(\boldsymbol{\mu},\Sigma)=\prod_{i=1}^{n} f(\mathbf{x}_i)=\prod_{i=1}^{n}\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})\right),$$

对数似然函数（令$\Omega=\Sigma^{-1}$，略去常数）：

$$\log L=-\frac{n}{2}\log|\Sigma|-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})$$

$$=\frac{n}{2}\log|\Omega|-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\boldsymbol{\mu})^{\mathsf{T}}\Omega(\mathbf{x}_i-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\boldsymbol{\mu})$$

$$=\frac{n}{2}\log|\Omega|-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf{T}}\Omega(\mathbf{x}_i-\bar{\mathbf{x}})-\frac{1}{2}\sum_{i=1}^{n}(\bar{\mathbf{x}}-\boldsymbol{\mu})^{\mathsf{T}}\Omega(\bar{\mathbf{x}}-\boldsymbol{\mu})$$

$$=\frac{n}{2}\log|\Omega|-\frac{1}{2}tr\left(\Omega\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf{T}}\right)-\frac{1}{2}\sum_{i=1}^{n}(\bar{\mathbf{x}}-\boldsymbol{\mu})^{\mathsf{T}}\Omega(\bar{\mathbf{x}}-\boldsymbol{\mu})$$

$$=\frac{n}{2}\log|\Omega|-\frac{n}{2}tr\left(\Omega\tilde{S}\right)-\frac{1}{2}\sum_{i=1}^{n}(\bar{\mathbf{x}}-\boldsymbol{\mu})^{\mathsf{T}}\Omega(\bar{\mathbf{x}}-\boldsymbol{\mu}),$$

显然，$\boldsymbol{\mu}$的最优解（MLE）$\hat{\boldsymbol{\mu}}=\bar{\mathbf{x}}$。

为了求解Ω的MLE，下面只需极大化（略去$n/2$):

$$\psi(\Omega)=\log|\Omega|-tr\left(\Omega\tilde{S}\right)=\log|\Omega\tilde{S}|-tr\left(\Omega\tilde{S}\right)-\log|\tilde{S}|$$

令$A=\Omega\tilde{S}$，记其所有特征根为$\lambda_1,\ldots,\lambda_p>0$

$$\psi(\Omega)=\log|A|-tr(A)+c=\sum\left(\log\lambda_i-\lambda_i\right)$$

$$\partial\psi/\partial\lambda_i=\frac{1}{\lambda_i}-1=0\Rightarrow\lambda_i=1\Rightarrow A=\Omega\tilde{S}=I_p$$

$$\Rightarrow\Omega\text{的最优解}(MLE):\ \hat{\Omega}=\tilde{S}^{-1},\ \hat{\Sigma}=\tilde{S}.$$

说明：
(1) $n>p$时， $P(\tilde{S}>0)=1$（待证），
(2) $A=\Omega\tilde{S}=\Omega^{1/2}\Omega^{1/2}\tilde{S}$与$\Omega^{1/2}\tilde{S}\Omega^{1/2}$有相同特征根，后者正定，特征根都大于0

引理1（矩阵、向量导数）：

(1) A对称，则 $\dfrac{\partial \mathbf{x}^{\mathrm{T}} A\mathbf{x}}{\partial \mathbf{x}} = 2A\mathbf{x}$

(2) X，A为$p \times p$矩阵，则 $\dfrac{\partial \log|X|}{\partial X} = (X^{-1})^{\mathrm{T}}$，$\dfrac{\partial \mathrm{tr}(AX)}{\partial X} = A^{\mathrm{T}}$.

证2：不限制Ω对称，$\psi(\Omega) = \log|\Omega| - tr\left(\Omega\tilde{S}\right)$对$\Omega$求导，令之为0:

$$\frac{\partial \psi(\Omega))}{\partial \Omega} = \frac{\partial \log|\Omega|}{\partial \Omega} - \frac{\partial tr\left(\Omega\tilde{S}\right)}{\partial \Omega} = (\Omega^{-1})^{\mathrm{T}} - \tilde{S}^{\mathrm{T}} = 0$$

$\Rightarrow \hat{\Sigma} = \hat{\Omega}^{-1} = \tilde{S}$（得到的解$\tilde{S}$对称，所以$\tilde{S}$是最优解）。

极大化目标函数

$$\psi(\Omega) = \log|\Omega| - tr\Omega S$$

求解Ω 或Σ。我们将多次用到。

Glasso 极大化该目标函数的L1惩罚。

我们以后将考虑多元正态总体的统计推断，其中需要样本均值和样本方差矩阵的概率分布, 以及涉及两者的二次型的分布。

一元推断主要利用如下结论：

假设$x_1,\dots,x_n$ iid $\sim N(\mu,\sigma^2)$,

随机向量$\mathbf{x}=(x_1,\dots,x_n)^\top \sim N(\mathbf{1}\mu,\sigma^2 I_n)$,

(1) $\sqrt{n}(\bar{x}-\mu)/\sigma \sim N(0,1)$;

(2) $(n-1)s^2=\mathbf{x}^\top P_{\mathbf{1}^\perp}\mathbf{x}=\mathbf{x}_c^\top\mathbf{x}_c \sim \sigma^2\chi^2_{n-1}$

(3) $\bar{x}$与s^2独立.

(4) $n(\bar{x}-\mu)^2/s^2 \sim F_{1,n-1}$

多元$(p>1)$情形需要考虑类似问题：

假设正态：$\mathbf{x}_1,\dots,\mathbf{x}_n \ iid \sim N_p(\boldsymbol{\mu},\Sigma)$,

随机矩阵$X=(\mathbf{x}_1,\dots,\mathbf{x}_n)^\top$

(1) $\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu},\Sigma/n)$

(2) $(n-1)S=X^\top P_{\mathbf{1}^\perp}X=X_c^\top X_c$的分布？

(3) $\bar{\mathbf{x}}$与S独立？

(4) $\sqrt{n}(\bar{\mathbf{x}}-\boldsymbol{\mu})^\top S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})$的分布？！

其中的关键是

$(n-1)S=X_c^\top X_c \sim$ Wishart分布

这是今天的主要内容。

Wishart 分布

回忆卡方定义：若$z_1, \dots, z_m$ iid $\sim N(0, \sigma^2)$，则 $\sum_{i=1}^m z_i^2 \sim \sigma^2 \chi_m^2$
等价地，记$m \times 1$向量 $\mathbf{z} = (z_1, \dots, z_m)^\top \sim N_m(\mathbf{0}, \sigma^2 I_m)$，则其模长平方

$$\mathbf{z}^\top \mathbf{z} = \|\mathbf{z}\|^2 \sim \sigma^2 \chi_m^2$$

对多元情形我们可以问类似的问题：

若$\mathbf{z}_1, \dots, \mathbf{z}_m$ iid $\sim N_p(0, \Sigma)$，则它们的“平方和” $\sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i^\top$服从什么分布？

等价地，记$m \times p$矩阵$Z = (\mathbf{z}_1, \dots, \mathbf{z}_m)^\top = \begin{pmatrix} \mathbf{z}_1^\top \\ \vdots \\ \mathbf{z}_m^\top \end{pmatrix}$，其“模长平方”

$$Z^\top Z = \sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i^\top \text{服从什么分布？}$$

$Z^\top Z = \sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i^\top$ 作为一个$p \times p$对称随机矩阵, 它的概率分布指的是其所有$p(p+1)/2$个不同元素（对角线+上三角）的联合分布。苏格兰数学家John Wishart在1928年求出了该分布的密度函数，后被称为Wishart分布，是卡方或gamma分布的多元拓展。

Wishart 分布

定义(Wishart，1928). 设$\mathbf{z}_1,\dots,\mathbf{z}_m \;\; iid \sim N_p(\mathbf{0},\Sigma)$，记$Z=(\mathbf{z}_1,\dots,\mathbf{z}_m)^{\mathsf{T}}$，则$p\times p$随机矩阵$W=Z^{\mathsf{T}}Z=\sum_{i=1}^{m}\mathbf{z}_i\mathbf{z}_i^{\mathsf{T}}$的分布称为自由度为$m$，参数为$\Sigma$的Wishart分布，记为$W\sim W_p(m,\Sigma)$。

$\Sigma=I_p$时，$W_p(m,I_p)$简记为$W_p(m)$，标准的Wishart分布。

注1：$p=1$时，$z_1,\dots,z_m \;\; iid \sim N_1(\mathbf{0},\sigma^2), W=\sum_{i=1}^{m}z_i^2\sim\sigma^2\chi_m^2=W_1(m,\sigma^2)$，

标准的情形即$\sigma^2=1$时，$W_1(m)=\chi_m^2$.

注2：显然，若$W\sim W_p(m,\Sigma)$，则$E(W)=m\Sigma$.

注3：Wishart是随机矩阵的分布，也称为Wishart ensemble.

统计物理、概率论中一般称随机矩阵或其分布为ensemble(系综、集成)，最著名的是GOE: Gaussian orthogonal ensemble (另外还有GUE,GSE)。

一个$n\times n$对称实随机矩阵H~GOE $\Leftrightarrow H=(G+G^{\mathsf{T}})/\sqrt{2n}$，其中$G=(g_{ij}), g_{ij}, i,j=1,\dots n, iid\ \sim N(0,1)$，$H$的概率密度函数为

$$p(H)=\frac{1}{Z}\exp\left(-\frac{n}{4}tr(H^2)\right).$$

正像使用卡方分布及其性质的时候，我们一般并不需要知道其具体的概率密度。我们首先讨论Wishart分布的一些性质，这些性质都可以从定义导出。

定理1：若$W \sim W_p(m,\Sigma)$，A是$q\times p$矩阵，则$AWA^\top \sim W_q(m, A\Sigma A^\top)$。

证：$W=\sum_{i=1}^{m}\mathbf{z}_i\mathbf{z}_i^\top$，其中$\mathbf{z}_1,\ldots,\mathbf{z}_m \ iid \sim N_p(\mathbf{0},\Sigma)$,

$A\mathbf{z}_i \sim N_q(\mathbf{0}, A\Sigma A^\top)$,

按照定义，$AWA^\top=\sum_{i=1}^{m}A\mathbf{z}_i(A\mathbf{z}_i)^\top \sim W_q(m, A\Sigma A^\top)$.

推论1:(1)W的标准化：$\Sigma^{-1/2}W\Sigma^{-1/2}\sim W_p(m, I_p)=W_p(m)$;

(2)对任何$\mathbf{t}\in R^p, \mathbf{t}^\top W\mathbf{t}\sim W_1(m, \mathbf{t}^\top\Sigma\mathbf{t})\overset{d}{=}(\mathbf{t}^\top\Sigma\mathbf{t})\chi_m^2$，即$\dfrac{\mathbf{t}^\top W\mathbf{t}}{\mathbf{t}^\top\Sigma\mathbf{t}}\sim\chi_m^2$。

回忆：多元标准正态的二次型

命题3（标准多元正态的二次型）假设$\mathbf{x}\sim N(0,I_n)$,

(1) 假设P是对称幂等矩阵(即投影矩阵),则$\mathbf{x}^{\mathsf T}P\mathbf{x}\sim\chi_r^2, r=rank(P)$.

(2) 假设A,B的列数都是n(行数任意), 若$AB^{\mathsf T}=0$, 则$A\mathbf{x}$与$B\mathbf{x}$独立。

证明（1）因为$P_{n\times n}$是秩为r的对称幂等阵，存在正交矩阵$H_{n\times n}$使得

$P=H^{\mathsf T}\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}H$，故$\mathbf{y}=H\mathbf{x}\sim N_n(\mathbf{0},I_n)$, $\mathbf{y}$的前r个分量组成的$\mathbf{y}_1\sim N_r(\mathbf{0},I_r)$

$$\mathbf{x}^{\mathsf T}P\mathbf{x}=\mathbf{x}^{\mathsf T}H^{\mathsf T}\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}H\mathbf{x}=\mathbf{y}^{\mathsf T}\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}\mathbf{y}=(\mathbf{y}_1^{\mathsf T},\mathbf{y}_2^{\mathsf T})\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{pmatrix}=\mathbf{y}_1^{\mathsf T}\mathbf{y}_1\sim\chi_r^2.$$

(2)由多元正态Hsu定义

$$\begin{pmatrix}A\mathbf{x}\\ B\mathbf{x}\end{pmatrix}=\begin{pmatrix}A\\ B\end{pmatrix}\mathbf{x}\sim N\left(\mathbf{0},\begin{pmatrix}AA^{\mathsf T} & AB^{\mathsf T}\\ BA^{\mathsf T} & BB^{\mathsf T}\end{pmatrix}\right)=N\left(\mathbf{0},\begin{pmatrix}AA^{\mathsf T} & 0\\ 0 & BB^{\mathsf T}\end{pmatrix}\right)$$

故$A\mathbf{x}$，$B\mathbf{x}$独立。

注：多元情形有类似结论（命题4，定理2）。

命题4. 假设$\mathbf{z}_1,\ldots,\mathbf{z}_m \sim N_p(\mathbf{0},\Sigma), Z=(\mathbf{z}_1,\ldots.,\mathbf{z}_m)^\top$,
若矩阵A,B列数都是m，且$AB^\top=0$,则AZ,BZ独立。

命题4已更新，
不要求A对称和
$\Sigma=I_p$。

证明:不妨假设$\Sigma=I_p$（否则令$\tilde{Z}=Z\Sigma^{-1/2}$,只需证$A\tilde{Z} \perp\!\!\!\perp B\tilde{Z}$）。

假设$A_{k\times m}$秩r,由SVD, $A=U\begin{pmatrix}\Lambda_r & 0\\ 0 & 0\end{pmatrix}_{k\times m}V^\top$,其中$U_{k\times k},V_{m\times m}$正交, Λ_r对角可逆。

记$Y=V^\top Z=\begin{pmatrix}Y_1\\ Y_2\end{pmatrix}$, Y_1为$r\times p$, Y_2为$(m-r)\times p$, 则Y_1与Y_2独立，

Z所有元素 iid $\sim N(0,1)$
$\Rightarrow Y=H^\top Z$也是。

$$\Rightarrow AZ=U\begin{pmatrix}\Lambda_r & 0\\ 0 & 0\end{pmatrix}V^\top Z=U\begin{pmatrix}\Lambda_r & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}Y_1\\ Y_2\end{pmatrix}=U\begin{pmatrix}\Lambda_r Y_1\\ 0\end{pmatrix}, \text{仅与}Y_1\text{有关。}$$

划分$BV=(C_1,C_2)$，由$0=AB^\top=U\begin{pmatrix}\Lambda_r & 0\\ 0 & 0\end{pmatrix}V^\top B^\top=U\begin{pmatrix}\Lambda_r & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}C_1^\top\\ C_2^\top\end{pmatrix}$

$$=U\begin{pmatrix}\Lambda_r C_1^\top\\ 0\end{pmatrix}\Rightarrow \Lambda_r C_1^\top=0\Rightarrow C_1=0, BV=(0,C_2)$$

$$\Rightarrow BZ=BVY=(0,C_2)\begin{pmatrix}Y_1\\ Y_2\end{pmatrix}=C_2Y_2\text{仅与}Y_2\text{有关，所以}AZ \perp\!\!\!\perp BZ\text{。}$$

Cochran 定理I

定理2(I). 假设$\mathbf{z}_1,\ldots,\mathbf{z}_m \sim N_p(\mathbf{0},\Sigma), Z_{m\times p}=(\mathbf{z}_1,\ldots,\mathbf{z}_m)^{\top}$。
若A是一个$m\times m$对称幂等的常数矩阵，则

$$Z^{\top}AZ \sim W_p(r,\Sigma),\quad r=rank(A)。$$

证明：令$\mathbf{y}_i=\Sigma^{-1/2}(\mathbf{z}_i-\boldsymbol{\mu})\sim N(\mathbf{0},I_p), i=1,\ldots,m$。记$Y=(\mathbf{y}_1,\ldots,\mathbf{y}_m)^{\top}$
$=(\Sigma^{-1/2}\mathbf{z}_1,\ldots,\Sigma^{-1/2}\mathbf{z}_m)^{\top}=X\Sigma^{-1/2}$, 所以$X^{\top}AX=\Sigma^{1/2}Y^{\top}AY\Sigma^{1/2}$,
下面只需证明$Y^{\top}AY\sim W_p(r,I_p)$, 其中$Y$的所有元素$iid\sim N(0,1)$.

因为$A_{n\times n}$是秩为r的对称幂等阵，存在正交矩阵$H_{m\times m}$使得$A=H\begin{pmatrix} I_r & 0 \\ 0 & 0\end{pmatrix}H^{\top}$,

所以$Y^{\top}AY=Y^{\top}H\begin{pmatrix} I_r & 0 \\ 0 & 0\end{pmatrix}H^{\top}Y=\tilde{Y}^{\top}\begin{pmatrix} I_r & 0 \\ 0 & 0\end{pmatrix}\tilde{Y}$, 其中$\tilde{Y}=H^{\top}Y$。

Y的所有元素$iid\sim N(0,1)$，故$\tilde{Y}$也是。

划分$\tilde{Y}=\begin{pmatrix}\tilde{Y}_1 \\ \tilde{Y}_2\end{pmatrix}$，其中$\tilde{Y}_1$是$r\times p$, $\tilde{Y}_2$是$(m-r)\times p$矩阵，

则$Y^{\top}AY=\tilde{Y}^{\top}\begin{pmatrix} I_r & 0 \\ 0 & 0\end{pmatrix}\tilde{Y}=(\tilde{Y}_1^{\top},\tilde{Y}_2^{\top})\begin{pmatrix} I_r & 0 \\ 0 & 0\end{pmatrix}\begin{pmatrix}\tilde{Y}_1 \\ \tilde{Y}_2\end{pmatrix}=\tilde{Y}_1^{\top}\tilde{Y}_1\sim W_p(r,I_p)$。证毕

引入Wishart分布的目的是为了考察样本协方差矩阵的分布，由Cochran定理可得（hw3.4）：

推论2：设$\mathbf{x}_1,\ldots,\mathbf{x}_n \ \ iid \sim N_p\left(\boldsymbol{\mu},\Sigma\right),$，则$W=(n-1)S\sim W_p(n-1,\Sigma)$。

证明：令$\mathbf{z}_i=\mathbf{x}_i-\boldsymbol{\mu}\sim N_p(\mathbf{0},\Sigma)$

$X=(\mathbf{x}_1,\ldots,\mathbf{x}_n)^\top$，$Z=(\mathbf{z}_1,\ldots,\mathbf{z}_n)^\top=X-\mathbf{1}\boldsymbol{\mu}^\top$

$$(n-1)S=X^\top(I_n-P_{\mathbf{1}})X=Z^\top(I_n-P_{\mathbf{1}})Z$$

因为$rank(I_n-P_{\mathbf{1}})=tr(I_n-P_{\mathbf{1}})=n-1,$

由Cochran定理，$(n-1)S\sim W_p(n-1,\Sigma)$.

反问题

Fisher问题：

我们已知：

$$\mathbf{z} \sim N(0, I_m) \Rightarrow \mathbf{z}^\top \mathbf{z} \sim \chi_m^2,\ \text{且对任何投影阵} A, rank(A) = r$$

$$\mathbf{z}^\top A\mathbf{z} \sim \chi_r^2, \mathbf{z}^\top \mathbf{z} - \mathbf{z}^\top A\mathbf{z} = \mathbf{z}^\top (I_m - A)\mathbf{z} \sim \chi_{m-r}^2,$$

卡方二次型中间的矩阵必须是幂等/投影矩阵吗?

Cochran问题：

对于多元情形，若$Z^\top AZ, Z^\top (I_m - A)Z$都是Wishart分布，$A$必须是幂等矩阵吗?

$$\mathrm{rank}(A) + rank(I_m - A) = m.$$

Cochran定理II 的矩阵版本

Cochran定理的矩阵版本：

假设A是m阶对称矩阵，则

A幂等 $\Leftrightarrow A(I_m - A) = 0 \Leftrightarrow rank(A) + rank(I_m - A) = m$

证明：假设$rank(A) + rank(I_m - A) = m$, 设$r = rank(A)$,

假设A有谱分解$A = H\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} H^{\top}$, H正交，$\Lambda = diag(\lambda_1, \ldots, \lambda_r), \lambda_i \neq 0$

则$B = I - A = H\begin{pmatrix} I_r - \Lambda & 0 \\ 0 & I_{m-r} \end{pmatrix} H^{\top}$,

因为$rank(B) = m - r$，所以$I_r - \Lambda$必为0，$\Lambda = I_r$,

所以A是幂等矩阵，B也是，且$AB = 0$.

Cochran 定理II

定理2（II）假设$\mathbf{z}_1,\dots,\mathbf{z}_m \sim N_p(\mathbf{0},\Sigma), Z=(\mathbf{z}_1,\dots,\mathbf{z}_m)^\top$，若$A$是一个$m\times m$对称常数矩阵，$r=rank(A)$，则

$$\begin{aligned}
&A\text{对称幂等}\\
&\Leftrightarrow Z^\top AZ \perp\!\!\!\perp Z^\top(I_m-A)Z\\
&\Leftrightarrow Z^\top AZ\sim W_p(r,\Sigma),\ Z^\top(I_m-A)Z\sim W_p(m-r,\Sigma)。
\end{aligned}$$

类似于代数版本的Cochran定理可证。

Wishart分布$W_2(m, I_2)$的概率密度

为了求Wishart概率密度，首先介绍三个引理

引理1(参见Bilodeau & Brenner,2009, $P27$, Proposition 2.13)

(1) 随机元(随机向量、随机矩阵)x, y独立 $\Leftrightarrow$ 对任何可测(可积）函数f, g

$Ef(x)g(y) = Ef(x)Eg(y)$.

(2) $x \overset{d}{=} y \Leftrightarrow$ 对任何可测函数f, $Ef(x) = Ef(y)$

对于通常的随机变量，f, g取任何示性函数就足够了，即

独立性：对任何s, t, $P(x < s, y < t) = P(x < s)P(y < t)$,

同分布：对任何t，$P(x < t) = P(y < t)$.

引理2. u，v是两个随机变量，若$u\,|\,v$的条件分布仅与v的某个函数$\varphi(v)$有关，则$u\,|\,\varphi(v)\overset{d}{=}u\,|\,v$.

将条件期望/条件分布看作投影，若u在v空间上的投影$P(u|v)$恰好落在v的子空间$\varphi(v)$上，则$P(u|v)$可看成是u直接在$\varphi(v)$上的投影，即$P(u|v)=P(u|\varphi(v))$.

证明：对任何给定的u的函数f，由条件期望的平滑性质/tower性质，

$$E[E(f(u)\,|\,v)\,|\,\varphi(v)]=E(f(u)\,|\,\varphi(v)) \qquad (*)$$

但因给定v时，u的条件分布仅与$\varphi(v)$有关，所以

$E(f(u)\,|\,v)$仅与$\varphi(v)$有关，记作$E(f(u)\,|\,v)=h(\varphi(v))$,

（*）时右端$=E[E(f(u)\,|\,v)\,|\,\varphi(v)]=E[h(\varphi(v))\,|\,\varphi(v)]=h(\varphi(v))=E(f(u)\,|\,v)$。

所以$E(f(u)\,|\,v)=E(f(u)\,|\,\varphi(v))$，即$u\,|\,\varphi(v)\overset{d}{=}u\,|\,v$.

引理3. 若 $x \perp\!\!\!\perp y \mid z$, 且$x \perp\!\!\!\perp z$，则 $x \perp\!\!\!\perp y$

证明：任取f, g。由$x \perp\!\!\!\perp z \Rightarrow E(f(x)|z) = Ef(x)$

由$x \perp\!\!\!\perp y \mid z \Rightarrow E(f(x)g(y)|z) = E(f(x)|z) \times E(g(y)|z)$。

所以$Ef(x)g(y) = E(Ef(x)g(y)|z) = E\{[Ef(x)|z][Eg(y)|z]\}$

$= E\{Ef(x)\,[Eg(y)|z]\} = Ef(x)\,E\{Eg(y)|z\} = Ef(x)Eg(y) \Rightarrow x \perp\!\!\!\perp y.$

注：也可如下形式化地证明（p表示概率或分布）：

$$p(x,y) = \int p(x,y|\mathrm{z})\,p(z)\mathrm{dz} = \int p(x|\mathrm{z})\,p(y|\mathrm{z})p(z)\,\mathrm{dz}$$
$$= \int p(x)\,p(y|z)p(z)\mathrm{d}z = p(x)\int p(y|z)p(z)\mathrm{d}z = p(x)p(y)$$

如何求$W_p(m,\Sigma)$的概率密度（即W的对角元和上三角元素的联合概率）？

我们下面考虑简单情况：$p=2, \Sigma = I_2$

$W_2(m)$的概率密度

命题5. 假设$\mathbf{z}_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}, i=1,\dots,n$ iid $\sim N_2(0, I_2)$，则$W = \sum_{i=1}^{m} \mathbf{z}_i \mathbf{z}_i^{\mathsf{T}}$

$$= \begin{pmatrix} \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i y_i \\ \sum_{i=1}^{m} x_i y_i & \sum_{i=1}^{m} y_i^2 \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \sim W_2(p, I_2)\text{的联合概率密度为}$$

$$p(W) = p(w_{11}, w_{12}, w_{22}) = c_m (w_{11}w_{22} - w_{12}^2)^{(m-3)/2} \exp\left(-\frac{1}{2}(w_{11}+w_{22})\right)$$

$$= c_m |W|^{(m-3)/2} \exp\left(-\frac{1}{2} tr(W)\right),\ 1/c_m = 2^m \sqrt{\pi}\,\Gamma(m/2)\Gamma((m-1)/2)\text{。}$$

评注：显然，$w_{11} \sim \chi_m^2$，$w_{22} \sim \chi_m^2$，两者独立，但w_{12}与w_{11}，w_{22}不独立。

容易知道，$w_{12} = \sum_{i=1}^{m} x_i y_i \mid \mathbf{x} \sim N(0, \sum_{i=1}^{m} x_i^2) = N(0, w_{11}) \overset{\text{引理1}}{\Rightarrow} w_{12} \mid w_{11} \sim N(0, w_{11})$

但这还不足以导出联合分布。

命题5的证明：

记$\mathbf{x}=(x_1,\dots,x_n)^{\mathsf T}\sim N_n(0,I_n)$, $\mathbf{y}=(y_1,\dots,y_n)^{\mathsf T}\sim N_n(0,I_n)$,

$$W=\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}=\begin{pmatrix} \sum_{i=1}^m x_i^2 & \sum_{i=1}^m x_i y_i \\ \sum_{i=1}^m x_i y_i & \sum_{i=1}^m y_i^2 \end{pmatrix}=\begin{pmatrix} \mathbf{x}^{\mathsf T}\mathbf{x} & \mathbf{x}^{\mathsf T}\mathbf{y} \\ \mathbf{x}^{\mathsf T}\mathbf{y} & \mathbf{y}^{\mathsf T}\mathbf{y} \end{pmatrix},$$

> $\{w_{11},w_{12},w_{22}\}$三个元素不独立，我们首先将它们转换为$\{w_{11\bullet 2},w_{12},w_{22}\}$，它们几乎相互独立，容易求出分布，然后转换回$\{w_{11},w_{12},w_{22}\}$

向量$\mathbf{x}$对$\mathbf{y}$做Schmidt正交化：令$\mathbf{x}^{\perp}=(I_m-P_{\mathbf{y}})\mathbf{x}$，

则$\mathbf{x}^{\perp\mathsf T}\mathbf{x}^{\perp}=\mathbf{x}^{\mathsf T}(I_m-P_{\mathbf{y}})\mathbf{x}=\mathbf{x}^{\mathsf T}\mathbf{x}-\mathbf{x}^{\mathsf T}\mathbf{y}(\mathbf{y}^{\mathsf T}\mathbf{y})^{-1}\mathbf{y}^{\mathsf T}\mathbf{x}=w_{11}-w_{12}w_{22}^{-1}w_{21}\hat{=}w_{11\bullet 2}$

则$w_{11\bullet 2}\mid\mathbf{y}=\mathbf{x}^{\mathsf T}(I_m-P_{\mathbf{y}})\mathbf{x}\mid\mathbf{y}\sim\chi^2_{m-1}$，该分布与$\mathbf{y}$无关

故$w_{11\bullet 2}$与$\mathbf{y}$独立,进而与$w_{22}=\mathbf{y}^{\mathsf T}\mathbf{y}$独立。

下面证明$w_{11\bullet 2}$与w_{12}也独立。

因为$(I_m - P_{\mathbf{y}})\mathbf{y} = 0, \mathbf{x} \sim N_m(0, I_m)$, 以及$\mathbf{x} \perp\!\!\!\perp \mathbf{y}$，由命题3，
在$\mathbf{y}$给定时，$(I_m - P_{\mathbf{y}})\mathbf{x}$与$w_{12} = \mathbf{y}^\top \mathbf{x}$ 条件独立，进而
$w_{11\bullet 2} = \mathbf{x}^\top (I_m - P_{\mathbf{y}})\mathbf{x}$与$w_{12}$ 条件独立：

$$w_{11\bullet 2} \perp\!\!\!\perp w_{12} \mid \mathbf{y}$$

但因

$$w_{11\bullet 2} | \mathbf{y} = \mathbf{x}^\top (I_m - P_{\mathbf{y}})\mathbf{x} | \mathbf{y} \sim \chi^2_{m-1},$$
$$w_{12} | \mathbf{y} = \mathbf{y}^\top \mathbf{x} \ | \mathbf{y} \sim N_m(\mathbf{0}, \mathbf{y}^\top \mathbf{y}) = N_m(\mathbf{0}, w_{22}),$$

条件分布仅与w_{22}有关，所以由引理2

$$w_{11\bullet 2} \perp\!\!\!\perp w_{12} \mid w_{22},$$

其中$w_{11\bullet 2} \perp\!\!\!\perp w_{22}$， 再由引理2，$w_{11\bullet 2} \perp\!\!\!\perp w_{12}$。

$w_{11\bullet 2} \sim \chi^2_{m-1}$与$w_{22}, w_{12}$独立，$w_{22} \sim \chi^2_m$，$w_{12} \mid w_{22} \sim N(0, w_{22})$，
所以W的联合概率密度：

$$p(W) = p(w_{11}, w_{12}, w_{22}) = p(w_{11\bullet 2})\, p(w_{12}, w_{22}) = p(w_{11\bullet 2})\, p(w_{12} \mid w_{22})\, f(w_{22}) = \ldots$$

推论2. 条件同命题5，定义Galton样本相关系数 $r=\dfrac{w_{12}}{\sqrt{w_{11}w_{22}}}$，则

$$\sqrt{m-1}\,\frac{r}{\sqrt{1-r^2}}\sim t_{m-1}.$$

证明：由命题5, $w_{11\bullet 2}\sim\chi^2_{m-1}$，且与$w_{22},w_{12}$独立。

$w_{12}\,|\,w_{22}\sim N(0,\ w_{22})$,

$\Rightarrow w_{12}/\sqrt{w_{22}}\,|\,w_{22}\sim N(0,1)$，该分布与条件无关，所以

$w_{12}/\sqrt{w_{22}}\sim N(0,1)$

$$\Rightarrow\sqrt{m-1}\,\frac{r}{\sqrt{1-r^2}}=\sqrt{m-1}\,\frac{\dfrac{w_{12}}{\sqrt{w_{11}w_{22}}}}{\sqrt{1-\dfrac{w_{12}^2}{w_{11}w_{22}}}}=\sqrt{m-1}\,\frac{w_{12}/\sqrt{w_{22}}}{\sqrt{w_{11}-w_{12}^2/w_{22}}}$$

$$=\frac{w_{12}/\sqrt{w_{22}}}{\sqrt{w_{11\bullet 2}/(m-1)}}\sim t_{m-1}.$$