第七讲 Wishart分布II

2024.3.25

- 数据矩阵$X_{n\times p}$每一行分布可能复杂，但每一列都是球对称的（由于样本的独立性）。
- 矩阵乘积 AB中A作为线性变换, 它作用在B的每一列上

Recap: 逐行排列数据，数据矩阵的列有简单结构

多元正态数据矩阵：

假设$\mathbf{x}_1,\ldots,\mathbf{x}_n$iid $\sim N_p(\boldsymbol{\mu},\Sigma)$，样本$\mathbf{x}_1,\ldots,\mathbf{x}_n$逐行排列而成数据矩阵：

$$X_{n\times p}=(\mathbf{x}_1,\ldots,\mathbf{x}_n)^{\mathsf{T}}=\begin{pmatrix}\mathbf{x}_1^{\mathsf{T}}\\ \vdots\\ \mathbf{x}_n^{\mathsf{T}}\end{pmatrix}\overset{\text{按列}}{=}(\mathbf{x}_{(1)},\ldots,\mathbf{x}_{(p)})$$

- 按行看：每一个行向量都是椭球正态分布

$$\mathbf{x}_i\sim N_p(\boldsymbol{\mu},\Sigma),\ i=1,\ldots,n$$

- 按列看：每一个列向量都是球对称正态分布

$$\mathbf{x}_{(j)}\sim N_n(\mathbf{1}_n\mu_j,\sigma_{jj}\boldsymbol{I}_n),\ j=1,\ldots,p$$

数据矩阵$X_{n\times p}$所有元素联合服从np元的多元正态。虽然每一行都是椭球正态（结构复杂），但每一列都是球对称正态（简单，由于样本的独立性）。

Cochran定理：假设$\mathbf{z}_1,\ldots,\mathbf{z}_m \text{iid} \sim N_p(\mathbf{0},\Sigma), Z_{m\times p}=(\mathbf{z}_1,\ldots.,\mathbf{z}_m)^{\top}$。
若A是一个$m\times m$投影矩阵，则

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

$Z^{\top}AZ=(AZ)^{\top}AZ$，$Z_{m\times p}=(\mathbf{z}_1,\ldots.,\mathbf{z}_m)^{\top}$

问题：A是投影阵，AZ如何投影？

$\mathbf{z}_1,\ldots,\mathbf{z}_m$ iid $\sim N_p(\mathbf{0},\Sigma)$不是球对称正态，为什么$Z^{\top}AZ$与卡方有关？

虽然Z是样本$\mathbf{z}_1,\ldots,\mathbf{z}_m$按行排列而成，但我们需要逐列看待矩阵$Z$：

记Z的各列$Z=(\mathbf{z}_{(1)},\ldots,\mathbf{z}_{(p)})$，$AZ=A(\mathbf{z}_{(1)},\ldots,\mathbf{z}_{(p)})=(A\mathbf{z}_{(1)},\ldots,A\mathbf{z}_{(p)})$

即Z的各列投影后按列排列组成AZ，Z的列$\mathbf{z}_{(1)} \sim N_n(\mathbf{0},\sigma_{11}I_n)$，

是球对称正态，$\mathbf{z}_{(1)}^{\top}A\mathbf{z}_{(1)}=\|A\mathbf{z}_{(1)}\|^2 \sim \sigma_{11}\chi_r^2$，$\Sigma=(\sigma_{ij})$。

矩阵乘积 AB 中 A 作为线性变换，它作用在 B 的每一列上。

多元正态+矩阵拉直和Kronecker乘积

正态数据矩阵

$\mathbf{x}_1,\dots,\mathbf{x}_n \;\; iid \sim N_p(\boldsymbol{\mu},\Sigma)$, $\mathbf{x}_1,\dots,\mathbf{x}_n$的联合概率密度

$$p(\mathbf{x}_1,\dots,\mathbf{x}_n)=\frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^{\mathsf T}\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})\right)$$

$X=(\mathbf{x}_1,\dots,\mathbf{x}_n)^{\mathsf T}$作为一个$n\times p$随机矩阵($assemble$),其概率密度函数?

容易验证指数上, $\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^{\mathsf T}\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})=\sum_{i=1}^{n}tr[(\mathbf{x}_i-\boldsymbol{\mu})^{\mathsf T}\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})]$

$$=\sum_{i=1}^{n}tr\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^{\mathsf T}=tr[\Sigma^{-1}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^{\mathsf T}]=\mathrm{tr}(\Sigma^{-1}(X-M)^{\mathsf T}(X-M))$$

正态随机矩阵的分布

X的概率密度即$\mathbf{x}_1,\dots,\mathbf{x}_n$的联合概率密度

$$p(X)=p(\mathbf{x}_1,\dots,\mathbf{x}_n)=\frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}}\exp\left(-\frac{1}{2}tr\left\{\Sigma^{-1}(X-M)^{\mathsf T}(X-M)\right\}\right),$$

其中$M=\mathbf{1}_n\boldsymbol{\mu}^{\mathsf T}$，我们称$X$服从矩阵多元正态，记作$X\sim MN_{n\times p}(M,\Sigma)$。

但如何表达X元素之间的协方差$\{\mathrm{cov}(x_{ij},x_{kl}),i,k=1,\dots,n;\ j,l=1,\dots,p\}$?

定义1（矩阵拉直和Kronecker乘积）：

矩阵$A=(\mathbf{a}_1,\ldots,\mathbf{a}_p)$拉直/向量化：$\mathrm{vec}(A)=\begin{pmatrix}\mathbf{a}_1\\ \vdots\\ \mathbf{a}_p\end{pmatrix}$.

矩阵A,B的Kronecker乘积$A\otimes B=(a_{ij}B)$，其中$A=(a_{ij})$.

附录列举了若干Kronecker乘积的性质，此处略过。

拉直$X^\top$

X各行的联合分布

用$\mathrm{vec}(X^\top)$的分布表示X的分布：

$$\mathrm{vec}(X^\top)\sim N_{np}\left(\mathbf{1}_n\otimes\boldsymbol{\mu}, I_n\otimes\Sigma\right)$$

将$X^\top=(\mathbf{x}_1,\ldots,\mathbf{x}_n)$拉直：$\mathrm{vec}(X^\top)=\mathrm{vec}(\mathbf{x}_1,\ldots,\mathbf{x}_n)=\begin{pmatrix}\mathbf{x}_1\\ \vdots\\ \mathbf{x}_n\end{pmatrix}$

则$E(\mathrm{vec}(X^\top))=\begin{pmatrix}\boldsymbol{\mu}\\ \vdots\\ \boldsymbol{\mu}\end{pmatrix}=\mathbf{1}_n\otimes\boldsymbol{\mu}$，$\mathrm{var}(\mathrm{vec}(X^\top))=\begin{pmatrix}\Sigma & 0 & \cdots & 0\\ 0 & \Sigma & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \Sigma\end{pmatrix}=I_n\otimes\Sigma$

拉直X

X各列的联合分布

用vec(X)的分布表示X的分布：

$$\text{vec}(X) \sim N_{np}(\boldsymbol{\mu}\otimes\mathbf{1}_n, \Sigma\otimes I_n)$$

记$X_{n\times p}=(\mathbf{x}_{(1)},\ldots,\mathbf{x}_{(p)})$，将$X$拉直为$vec(X)=\begin{pmatrix}\mathbf{x}_{(1)}\\ \vdots \\ \mathbf{x}_{(p)}\end{pmatrix}$。

$\Sigma=(\sigma_{ij})$。由于样本独立，X的第j列球对称：

$$\mathbf{x}_{(j)} \sim N_n(\mathbf{1}_n\mu_j, \sigma_{jj}I_n),, \quad \text{cov}(\mathbf{x}_{(j)},\mathbf{x}_{(k)})=\sigma_{jk}I_n$$

则 $$vec(X)=\begin{pmatrix}\mathbf{x}_{(1)}\\ \vdots \\ \mathbf{x}_{(p)}\end{pmatrix} \sim N_{np}\left(\begin{pmatrix}\mathbf{1}_n\mu_1\\ \vdots \\ \mathbf{1}_n\mu_p\end{pmatrix}, \begin{pmatrix}\sigma_{11}I_n & \sigma_{12}I_n & \cdots & \sigma_{1p}I_n\\ \sigma_{21}I_n & \sigma_{22}I_n & \cdots & \sigma_{2p}I_n\\ & & \ddots & \\ \sigma_{p1}I_n & \sigma_{p2}I_n & \cdots & \sigma_{pp}I_n\end{pmatrix}\right)$$

$$=N_{np}(\boldsymbol{\mu}\otimes\mathbf{1}_n, \Sigma\otimes I_n)$$

引理4.

(1) 对任何 $A_{k\times n}, X_{n\times p}$, $\mathrm{vec}(AX) = \left(I_p \otimes A\right)\mathrm{vec}(X)$.

(2) $\mathrm{vec}(AXB) = (B^{\top} \otimes A)\mathrm{vec}(X)$.

(3) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.

证明：只证（1）。记$X = (\mathbf{x}_{(1)},\dots,\mathbf{x}_{(p)})$, $AX = (A\mathbf{x}_{(1)},\dots,A\mathbf{x}_{(p)})$

$$\Rightarrow \mathrm{vec}(AX) = \begin{pmatrix} A\mathbf{x}_{(1)} \\ \vdots \\ A\mathbf{x}_{(p)} \end{pmatrix} = \begin{pmatrix} A & & \\ & \ddots & \\ & & A \end{pmatrix} \begin{pmatrix} \mathbf{x}_{(1)} \\ \vdots \\ \mathbf{x}_{(p)} \end{pmatrix} = (I_p \otimes A) vec(X)$$

引理5. 假设$\mathbf{x}_1,\dots,\mathbf{x}_n$ $iid \sim N_p\left(\boldsymbol{\mu},\Sigma\right) \Leftrightarrow vec(X) \sim N_{np}(\boldsymbol{\mu} \otimes \mathbf{1}_n, \Sigma \otimes I_n)$,

其中$X_{n\times p} = (\mathbf{x}_1,\dots,\mathbf{x}_n)^{\top}$。若$A$是任一$k \times n$矩阵，则

$$\mathrm{vec}(AX) \sim N_{kp}(\boldsymbol{\mu} \otimes A\mathbf{1}_n, \Sigma \otimes AA^{\top})$$

证明1：由$vec(X) \sim N_{np}(\boldsymbol{\mu} \otimes \mathbf{1}_n, \Sigma \otimes I_n)$

$$\Rightarrow \mathrm{vec}(AX) = (I_p \otimes A) vec(X) \sim N_{kp}\left((I_p \otimes A)(\boldsymbol{\mu} \otimes \mathbf{1}_n), (I_p \otimes A)(\Sigma \otimes I_n)(I_p \otimes A)^{\top}\right)$$

$$= N_{kp}(\boldsymbol{\mu} \otimes A\mathbf{1}_n, \Sigma \otimes AA^{\top})$$

证明2：（更直观的证明，不利用拉直和Kronecker乘积的性质）

记$X=(\mathbf{x}_{(1)},\ldots,\mathbf{x}_{(p)}),\ AX=(A\mathbf{x}_{(1)},\ldots,A\mathbf{x}_{(p)})$

$$由vec(X)=\begin{pmatrix}\mathbf{x}_{(1)}\\ \vdots\\ \mathbf{x}_{(p)}\end{pmatrix}\sim N_{np}\left(\begin{pmatrix}\mathbf{1}_n\mu_1\\ \vdots\\ \mathbf{1}_n\mu_p\end{pmatrix},\begin{pmatrix}\sigma_{11}I_n & \cdots & \sigma_{1p}I_n\\ \vdots & & \vdots\\ \sigma_{p1}I_n & \cdots & \sigma_{pp}I_n\end{pmatrix}\right)=N_{np}(\boldsymbol{\mu}\otimes\mathbf{1}_n,\Sigma\otimes I_n)$$

$$\Rightarrow \mathrm{vec}(AX)=\begin{pmatrix}A\mathbf{x}_{(1)}\\ \vdots\\ A\mathbf{x}_{(p)}\end{pmatrix}\sim N_{kp}\left(\begin{pmatrix}A\mathbf{1}_n\mu_1\\ \vdots\\ A\mathbf{1}_n\mu_p\end{pmatrix},\begin{pmatrix}\sigma_{11}AA^{\mathsf{T}} & \cdots & \sigma_{1p}AA^{\mathsf{T}}\\ \vdots & & \vdots\\ \sigma_{p1}AA^{\mathsf{T}} & \cdots & \sigma_{pp}AA^{\mathsf{T}}\end{pmatrix}\right)$$

$$=N_{kp}(\boldsymbol{\mu}\otimes A\mathbf{1}_n,\Sigma\otimes AA^{\mathsf{T}})$$

总之，各种分布描述等价：

$\mathbf{x}_1,\ldots,\mathbf{x}_n \;\; iid \sim N_p(\boldsymbol{\mu},\Sigma)$ （样本分布）

$\Leftrightarrow X = \begin{pmatrix} \mathbf{x}_1^{\top} \\ \vdots \\ \mathbf{x}_n^{\top} \end{pmatrix} \sim MN_{n\times p}(M,\Sigma)$ （整体矩阵X）

$\Leftrightarrow \mathrm{vec}(X^{\top}) \sim N_{np}(\mathbf{1}_n \otimes \boldsymbol{\mu}, I_n \otimes \Sigma)$ （X的各行联合分布）

$\Leftrightarrow vec(X) \sim N_{np}(\boldsymbol{\mu} \otimes \mathbf{1}_n, \Sigma \otimes I_n)$ （X的各列联合分布）

$\Rightarrow \mathrm{vec}(AX) \sim N_{kp}(\boldsymbol{\mu} \otimes A\mathbf{1}_n, \Sigma \otimes AA^{\top})$， $A: k \times n$

大多数时间我们将使用最基本的分布表达：

$$\mathbf{x}_1,\ldots,\mathbf{x}_n \;\; iid \sim N_p(\boldsymbol{\mu},\Sigma)$$

非必要时，不采用Kronecker记号（因为不直观）。

但前述对数据矩阵X的各种理解观点（整体、行、列）都有其重要性。

Wishart分布定义：设$\mathbf{z}_1,\ldots,\mathbf{z}_m \ iid \sim N_p(\mathbf{0},\Sigma)$，记$Z=(\mathbf{z}_1,\ldots,\mathbf{z}_m)^{\mathsf{T}}$，

则$p\times p$随机矩阵$W=Z^{\mathsf{T}}Z=\sum_{i=1}^{m}\mathbf{z}_i\mathbf{z}_i^{\mathsf{T}} \sim W_p(m,\Sigma)$。

显然$E(W)=m\Sigma$，W/m是Σ的一个无偏估计。

注1：由引理6，$W=(n-1)S$，S是基于$\mathbf{x}_1,\ldots,\mathbf{x}_n \ iid\sim N_p(\boldsymbol{\mu},\Sigma)$的样本方差矩阵，$m=n-1$ 。

注2：$Z=(\mathbf{z}_{(1)},\ldots,\mathbf{z}_{(p)}),\Sigma=\left(\sigma_{ij}\right)$。$Z$的每列球对称：

$$\mathbf{z}_{(j)}\sim N_m(0,\sigma_{jj}I_m),\ \left\|\mathbf{z}_{(j)}\right\|^2\sim\sigma_{jj}\chi_m^2,\quad j=1,\ldots,p$$

每两列$\mathbf{z}_{(j)}$，$\mathbf{z}_{(k)}$的协方差矩阵也是球形的:

$$cov(\mathbf{z}_{(j)},\mathbf{z}_{(k)})=E\left(\mathbf{z}_{(j)}\mathbf{z}_{(k)}^{\mathsf{T}}\right)=\sigma_{jk}I_m$$

Wishart 矩阵W:

$$W=Z^{\mathsf{T}}Z=\begin{pmatrix}\mathbf{z}_{(1)}^{\mathsf{T}}\mathbf{z}_{(1)} & \cdots & \mathbf{z}_{(1)}^{\mathsf{T}}\mathbf{z}_{(p)}\\ \vdots & \ddots & \vdots\\ \mathbf{z}_{(p)}^{\mathsf{T}}\mathbf{z}_{(1)} & \cdots & \mathbf{z}_{(p)}^{\mathsf{T}}\mathbf{z}_{(p)}\end{pmatrix}\stackrel{d}{=}\begin{pmatrix}\sigma_{11}\chi_m^2 & \cdots & \\ \vdots & \ddots & \vdots\\ & \cdots & \sigma_{pp}\chi_m^2\end{pmatrix}$$

理解为对角线上是若干相关的scaled χ_m^2随机变量的集成(assemble) 。

样本协方差矩阵

引理6. 若$\mathbf{x}_1,\dots,\mathbf{x}_n\ iid{\sim}N_p(\boldsymbol{\mu},\Sigma)$, 则$S=\frac{1}{n-1}\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^\top$ 可以表示为

$$S=\frac{1}{n-1}\sum_{i=1}^{n-1}\mathbf{z}_i\mathbf{z}_i^\top=\frac{1}{n-1}Z^\top Z,$$

其中$\mathbf{z}_1,\dots,\mathbf{z}_{n-1}\ iid{\sim}N_p(\mathbf{0},\Sigma)$, $Z=(\mathbf{z}_1,\dots,\mathbf{z}_{n-1})^\top$.

证：首先$(n-1)S=\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top-n\bar{\mathbf{x}}\bar{\mathbf{x}}^\top$, $X=(\mathbf{x}_1,\dots,\mathbf{x}_n)^\top$.

令H是一个$n\times n$正交阵，$H=\begin{pmatrix}H_1\\ \frac{1}{\sqrt{n}}\mathbf{1}_n^\top\end{pmatrix}$, 注意$H\mathbf{1}_n=(0,\dots,0,\sqrt{n})^\top$,

则$Y=HX=\begin{pmatrix}H_1\\ \frac{1}{\sqrt{n}}\mathbf{1}_n^\top\end{pmatrix}X=\begin{pmatrix}H_1X\\ \sqrt{n}\bar{\mathbf{x}}^\top\end{pmatrix}\triangleq\begin{pmatrix}\mathbf{y}_1^\top\\ \vdots\\ \mathbf{y}_n^\top\end{pmatrix}$, $\mathbf{y}_n=\sqrt{n}\bar{\mathbf{x}}$, 则一定有

$$\mathbf{y}_1,\dots,\mathbf{y}_{n-1}\ \text{iid}\sim N_p(\mathbf{0},\Sigma),\ \ \mathbf{y}_n\sim N_p(\sqrt{n}\boldsymbol{\mu},\Sigma),\qquad (*)$$

这是因为由引理5， $vec(X)\sim N_{np}(\boldsymbol{\mu}\otimes\mathbf{1}_n,\Sigma\otimes I_n)$，

$$vec(Y)=vec(HX)\sim N_{np}(\boldsymbol{\mu}\otimes H\mathbf{1}_n,\Sigma\otimes HH^\top)=N_{np}(\boldsymbol{\mu}\otimes H\mathbf{1}_n,\Sigma\otimes I_n)$$

注意$H\mathbf{1}_n=(0,\dots,\sqrt{n})^\top\Rightarrow\mathbf{y}_1,\dots,\mathbf{y}_{n-1}\ \text{iid}\sim N_p(\mathbf{0},\Sigma)$, $\mathbf{y}_n\sim N_p(\sqrt{n}\boldsymbol{\mu},\Sigma)$.

（*）也可直观得到：观察Y的列

(1) $Y=(\mathbf{y}_{(1)},\dots,\mathbf{y}_{(p)})=(H\mathbf{x}_{(1)},\dots,H\mathbf{x}_{(p)})$联合正态；

(2) $\mathbf{x}_{(j)}\sim N_n\left(\mathbf{1}_n\mu_j,\sigma_{jj}I_n\right)$

$$\Rightarrow H\mathbf{x}_{(j)}\sim N_n(H\mathbf{1}_n\mu_j,\sigma_{jj}I_n),$$

其中均值$H\mathbf{1}_n\mu_j=(0,\dots,0,\sqrt{n}\mu_j)^\top$;

(3) $\mathrm{cov}(H\mathbf{x}_{(j)},H\mathbf{x}_{(k)})=H\mathrm{cov}(\mathbf{x}_{(j)},\mathbf{x}_{(k)})H^\top=H\left(\sigma_{jk}I_n\right)H^\top=\sigma_{jk}I_n$.

(1)-(3)表明Y的各行都是p元正态，协方差矩阵都是Σ；前$n-1$行均值都是$\mathbf{0}$，最后一行的均值为$\sqrt{n}\boldsymbol{\mu}$，这就是(*)

另外，$Y^\top Y=X^\top H^\top HX=X^\top X$，所以

$$\begin{aligned}(n-1)S&=\textstyle\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top-n\bar{\mathbf{x}}\bar{\mathbf{x}}^\top=X^\top X-\mathbf{y}_n\mathbf{y}_n^\top=Y^\top Y-\mathbf{y}_n\mathbf{y}_n^\top\\&=\textstyle\sum_{i=1}^n\mathbf{y}_i\mathbf{y}_i^\top-\mathbf{y}_n\mathbf{y}_n^\top=\sum_{i=1}^{n-1}\mathbf{y}_i\mathbf{y}_i^\top\end{aligned}$$

令$Z=(\mathbf{z}_1,\dots,\mathbf{z}_{n-1})^\top=(\mathbf{y}_1,\dots,\mathbf{y}_{n-1})^\top$即可。

Cochran定理

第6讲定理2(I)： 假设$\mathbf{z}_1,\ldots,\mathbf{z}_m \sim N_p(\mathbf{0},\Sigma), Z_{m\times p}=(\mathbf{z}_1,\ldots.,\mathbf{z}_m)^{\mathsf{T}}$。
若A是一个$m\times m$对称幂等的常数矩阵，则

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

我们熟知（第6讲命题3）：

若A对称幂等，向量 $\mathbf{x} \sim N_m(0,\sigma^2 I_m)$, 则$\|A\mathbf{x}\|^2=\mathbf{x}^{\mathsf{T}}A\mathbf{x}\sim\sigma^2\chi_r^2$

Cochran定理与此类似：

以列向量表示Z: $Z=(\boldsymbol{z}_{(1)},\ldots,\boldsymbol{z}_{(p)})$，$\boldsymbol{z}_{(j)}\sim N_m(0,\sigma_{jj}I_m)$, 由第6讲命题3
对任一秩r投影矩阵A

$$A\,\boldsymbol{z}_{(j)}\sim N_m(0,\sigma_{jj}A),\quad \left\|A\,\boldsymbol{z}_{(j)}\right\|^2\sim\sigma_{jj}\chi_r^2$$

$AZ=(A\boldsymbol{z}_{(1)},\ldots,A\boldsymbol{z}_{(p)})$是$Z$的各列的投影变换，每列都是球对称正态，每列模长平方都服从卡方分布。

（hw3.4，第六讲P16新增了如下推论）：

$\boxed{\text{第6讲推论2：若}\mathbf{x}_1,\ldots,\mathbf{x}_n \ \ iid \sim N_p(\boldsymbol{\mu},\Sigma),\text{则}(n-1)S \sim W_p(n-1,\Sigma)}$

证明1：由引理6立得。

证明2：　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，$(n-1)S \sim W_p(n-1,\Sigma)$.

正态随机矩阵的独立性

第6讲命题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。

注：不要求A对称（虽然具体应用中A都是对称且幂等），也不要求Σ是单位阵。

直观解释：AZ,BZ作为Z的变换，并不是样本$\mathbf{z}_1,\ldots,\mathbf{z}_m$的线性变换，而是Z的各列的线性变换：

$$AZ=A(\mathbf{z}_{(1)},\ldots,\mathbf{z}_{(1)})=(A\mathbf{z}_{(1)},\ldots,A\mathbf{z}_{(1)}),$$

$$BZ=B(\mathbf{z}_{(1)},\ldots,\mathbf{z}_{(1)})=(B\mathbf{z}_{(1)},\ldots,B\mathbf{z}_{(1)}),$$

由第6讲命题3(向量情形)，$AB^\top=0$,

$$\operatorname{cov}(A\mathbf{z}_{(j)},B\mathbf{z}_{(k)})=A\operatorname{cov}(\mathbf{z}_{(j)},\mathbf{z}_{(k)})B^\top=\sigma_{jk}AB^\top=0$$

$$\Rightarrow A\mathbf{z}_{(j)} \perp\!\!\!\perp B\mathbf{z}_{(j)}$$

这解释了为什么不必要求$\Sigma=I_p$

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的SVD：

$$A = U\begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix}_{k\times m} V^\top, U_{k\times k}, V_{m\times m}\text{正交，}\Lambda_r\text{对角可逆。}$$

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(引理5)

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

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

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

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

Wishart分布的概率密度（一般情形）

$$\boxed{\text{引理7. 假设} W \sim W_p(m,\Sigma)\text{，若} m \ge p\text{,则} P(W>0)=1.}$$

证明：假设$W = Z^{\top}Z = \sum_{i=1}^{m} \mathbf{z}_i \mathbf{z}_i^{\top} \sim W_p(m,\Sigma)$,

其中$\mathbf{z}_1,\dots,\mathbf{z}_m \ iid \sim N_p(\mathbf{0},\Sigma)$, $Z = (\mathbf{z}_1,\dots,\mathbf{z}_m)^{\top}$,

则$P(W\text{不可逆}) = P(rank(W) < p) = P(rank(Z) < p)$

$= P(\mathbf{z}_1,\dots,\mathbf{z}_m\text{中任取}p\text{个都线性相关})$

$= \binom{m}{p} P(\mathbf{z}_1,\dots,\mathbf{z}_p\text{线性相关})$

$= \binom{m}{p} E\left\{P(\mathbf{z}_1 \in L(\mathbf{z}_2,\dots,\mathbf{z}_p) \mid \mathbf{z}_2,\dots,\mathbf{z}_p)\right\} = 0$,因为$\mathbf{z}_1$与$\mathbf{z}_2,\dots,\mathbf{z}_p$独立。

Wishart 分块

定理3：假设$W \sim W_p(m,\Sigma), \Sigma > 0, m \ge p$，划分

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

其中W_{11},Σ_{11}为$q \times q$矩阵，W_{22},Σ_{22}为$(p-q)\times(p-q)$矩阵，则

(1) $W_{11} \sim W_q(m,\Sigma_{11})$，$W_{22} \sim W_{p-q}(m,\Sigma_{22})$。

(2) $W_{11\bullet 2} = W_{11} - W_{12}W_{22}^{-1}W_{21} \sim W_q(m-p+q,\Sigma_{11\bullet 2})$，且$W_{11\bullet 2}$与$W_{22},W_{21}$独立。

(3) $q=1$时，$W_{21} \mid W_{22} \sim N_{p-1}(W_{22}\Sigma_{22}^{-1}\Sigma_{21}, W_{22}\sigma^2)$，$\sigma^2 = \Sigma_{11\bullet 2}$。

(4) $q>1$时，给定W_{22}时，W_{21}所有元素联合服从$q(p-q)$元正态：

$$vec(W_{21}^{\mathrm{T}}) \mid W_{22} \sim N_{q(p-q)}(vec(\Sigma_{12}\Sigma_{22}^{-1}W_{22}), W_{22} \otimes \Sigma_{11\bullet 2})$$

参见Bilodeau & Brenner, 2009, P92, Proposition 7.9

证：(1)取$A=(I_q,0)$,则由定理1，$W_{11}=AWA^{\top}\sim W_q(m,\Sigma_{11})$。

(2) 假设$W\overset{d}{=}Z^{\top}Z$，其中$\mathbf{z}_1,\ldots,\mathbf{z}_m\ \ iid\sim N_p(\mathbf{0},\Sigma)$，$Z=(\mathbf{z}_1,\ldots,\mathbf{z}_m)^{\top}$。

划分$\mathbf{z}_i=\begin{pmatrix}\mathbf{x}_i\\ \mathbf{y}_i\end{pmatrix}\begin{matrix}q\times 1\\ (p-q)\times 1\end{matrix}$，划分$Z_{m\times p}=\begin{pmatrix}\mathbf{z}_1^{\top}\\ \vdots\\ \mathbf{z}_m^{\top}\end{pmatrix}=\begin{pmatrix}\mathbf{x}_1^{\top} & \mathbf{y}_1^{\top}\\ \vdots & \vdots\\ \mathbf{x}_m^{\top} & \mathbf{y}_m^{\top}\end{pmatrix}\overset{\Delta}{=}(X_{m\times q},Y_{m\times(p-q)})$。

由$Z=(X,Y)$,得

$$W=\begin{pmatrix}W_{11} & W_{12}\\ W_{21} & W_{22}\end{pmatrix}=\begin{pmatrix}X^{\top}\\ Y^{\top}\end{pmatrix}(X,Y)=\begin{pmatrix}X^{\top}X & X^{\top}Y\\ Y^{\top}X & Y^{\top}Y\end{pmatrix}=\begin{pmatrix}\sum\mathbf{x}_i\mathbf{x}_i^{\top} & \sum\mathbf{x}_i\mathbf{y}_i^{\top}\\ \sum\mathbf{y}_i\mathbf{x}_i^{\top} & \sum\mathbf{y}_i\mathbf{y}_i^{\top}\end{pmatrix}$$

$$W_{11\bullet 2}=W_{11}-W_{12}W_{22}^{-1}W_{21}=X^{\top}X-X^{\top}Y(Y^{\top}Y)^{-1}Y^{\top}X=X^{\top}(I_m-P_Y)X$$

不能直接对$W_{11\bullet 2}=X^{\top}(I_m-P_Y)X$应用Cochran定理，这是因为$I_m-P_Y$不是常数矩阵，故下面考虑$Y$给定时的情形。给定$Y$的条件下，$X$各行的分布仍是正态，但均值和方差都有变化。

$$\mathbf{z}_i=\begin{pmatrix}\mathbf{x}_i\\ \mathbf{y}_i\end{pmatrix}\sim N_p\left(0,\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right)\quad\Rightarrow\quad\mathbf{x}_i\,|\,\mathbf{y}_i\sim N_q(\Sigma_{12}\Sigma_{22}^{-1}\mathbf{y}_i,\Sigma_{11\bullet 2})$$

$\mathbf{x}_i \mid \mathbf{y}_i \sim N_q(\Sigma_{12}\Sigma_{22}^{-1}\mathbf{y}_i, \Sigma_{11\bullet 2})$，记去相关化$\mathbf{v}_i \triangleq \mathbf{x}_i - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{y}_i$，则

$$\mathbf{v}_i \sim N_q(0, \Sigma_{11\bullet 2})，且\mathbf{v}_i与\mathbf{y}_i独立.$$

记$V = (\mathbf{v}_1, \ldots, \mathbf{v}_m)^\top = X - Y\Sigma_{22}^{-1}\Sigma_{21} \Rightarrow X = V + Y\Sigma_{22}^{-1}\Sigma_{21}$,所以

$$\begin{aligned} W_{11\bullet 2} &= X^\top (I_m - P_Y) X = (V + Y\Sigma_{22}^{-1}\Sigma_{21})^\top (I_m - P_Y)(V + Y\Sigma_{22}^{-1}\Sigma_{21}) \\ &= V^\top (I_m - P_Y) V。\end{aligned}$$

因为

- V各行为iid $N_q(0, \Sigma_{11\bullet 2})$随机向量,与$Y$独立;
- 给定Y时,$(I_m - P_Y)$是对称幂等常数矩阵，$\operatorname{rank}(I_m - P_Y) = m - p + q$，

所以由*Cochran*定理（定理2）

$$W_{11\bullet 2} \mid Y = V^\top (I_m - P_Y) V \mid Y \sim W_q(m - p + q, \Sigma_{11\bullet 2})$$

上述分布与给定的Y无关，故$W_{11\bullet 2} \sim W_q(m - p + q, \Sigma_{11\bullet 2})$,与$Y$独立,

与$W_{22} = Y^\top Y$也独立.

下面证明$W_{11\bullet 2}$与$\{W_{22}, W_{21}\}$独立

改写$W_{21}=Y^{\top}X=Y^{\top}(V+Y\Sigma_{22}^{-1}\Sigma_{21})=Y^{\top}V+Y^{\top}Y\Sigma_{22}^{-1}\Sigma_{21}$

而$W_{11\bullet 2}=V^{\top}(I_m-P_Y)V,$

给定Y时,(I_m-P_Y)和Y都是常数矩阵, 且$(I_m-P_Y)Y=0,$

注意V与Y独立,

由上一讲命题4, 给定Y时, $(I_m-P_Y)V$与$Y^{\top}V$独立。

$\Rightarrow$ 给定Y时,

$W_{11\bullet 2}=[(I_m-P_Y)V]^{\top}(I_m-P_Y)V$ 与$W_{21}=Y^{\top}V+Y^{\top}Y\Sigma_{22}^{-1}\Sigma_{21}$条件独立

即$W_{11\bullet 2}\perp\!\!\!\perp\{W_{12}, W_{22}\}\mid Y$, 但$W_{11\bullet 2}\perp\!\!\!\perp Y$,

$\Rightarrow W_{11\bullet 2}\perp\!\!\!\perp\{W_{12}, W_{22}\}$ （引理3 ）

（3）$k=1$情形：X为$m\times 1$向量,Y为$m\times(p-1)$矩阵。

$V=(\mathbf{v}_1,\ldots,\mathbf{v}_m)^\top = X - Y\Sigma_{22}^{-1}\Sigma_{21}$为$m\times 1$向量，各分量 iid $\sim N_1(0,\Sigma_{11\bullet 2})$,

且与Y独立，即$V\sim N_m(0,\Sigma_{11\bullet 2}I_m)$。因为

$$W_{21}=Y^\top X = Y^\top(V+Y\Sigma_{22}^{-1}\Sigma_{21}) = Y^\top V + Y^\top Y\Sigma_{22}^{-1}\Sigma_{21},$$

则$W_{21}\,|\,Y\sim N(Y^\top Y\Sigma_{22}^{-1}\Sigma_{21}, Y^\top Y\Sigma_{11\bullet 2})$仅与$W_{22}=Y^\top Y$有关，

所以由引理2，$W_{21}\,|\,W_{22}\sim N(W_{22}\Sigma_{22}^{-1}\Sigma_{21}, W_{22}\Sigma_{11\bullet 2})$。

（4）一般k情形(可忽略，后面我们将只需要$k=1$情形)：

$V=(\mathbf{v}_1,\ldots,\mathbf{v}_m)^\top = X - Y\Sigma_{22}^{-1}\Sigma_{21}$，其各行为 iid $N(0,\Sigma_{11\bullet 2})$

随机向量，且与Y独立，即$vec(V^\top)\sim N(0, I_m\otimes\Sigma_{11\bullet 2})$

$W_{21}=Y^\top X = Y^\top(V+Y\Sigma_{22}^{-1}\Sigma_{21}) = Y^\top V + Y^\top Y\Sigma_{22}^{-1}\Sigma_{21}$

$vec(V^\top Y)=(Y^\top\otimes I_q)vec(V^\top)$，则

$vec(W_{21}{}^\top)\,|\,Y\sim N(vec(\Sigma_{12}\Sigma_{22}^{-1}Y^\top Y), Y^\top Y\otimes\Sigma_{11\bullet 2})$仅与$W_{22}=Y^\top Y$有关，

所以$vec(W_{21}{}^\top)\,|\,W_{22}\sim N(vec(\Sigma_{12}\Sigma_{22}^{-1}W_{22}), W_{22}\otimes\Sigma_{11\bullet 2})$.

推论3：假设$W \sim W_p(m,\Sigma)$，划分$W=\begin{pmatrix} w_{11} & \mathbf{w}_{12} \\ \mathbf{w}_{21} & W_{22} \end{pmatrix}, \Sigma=\begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \Sigma_{22} \end{pmatrix}$，其中$w_{11},\sigma_{11}$为标量，$W_{22}$，$\Sigma_{22}$为$(p-1)\times(p-1)$矩阵，定义决定系数$R^2=\dfrac{\mathbf{w}_{12}W_{22}^{-1}\mathbf{w}_{21}}{w_{11}}$，则当$\boldsymbol{\sigma}_{12}=0$时，

$$\frac{m-p+1}{p-1}\times\frac{R^2}{1-R^2} \sim F_{p-1,m-p+1}\text{。}$$

证明：$R^2/(1-R^2)=\dfrac{\mathbf{w}_{12}W_{22}^{-1}\mathbf{w}_{21}}{w_{11\bullet 2}}, \sigma_{11\bullet 2}=\sigma_{11}$

(1) 由定理3，$w_{11\bullet 2}=w_{11}-\mathbf{w}_{12}W_{22}^{-1}\mathbf{w}_{21} \sim W_1(m-p+1,\sigma_{11\bullet 2})=\sigma_{11}\chi^2_{m-p+1}$.

(2) $\mathbf{w}_{21}\,|\,W_{22} \sim N_{p-1}(W_{22}\Sigma_{22}^{-1}\Sigma_{21}, W_{22}\sigma_{11})=N_{p-1}(\mathbf{0},W_{22}\sigma_{11}) \Rightarrow \mathbf{w}_{21}^{\mathsf{T}}W_{22}^{-1}\mathbf{w}_{21} \sim \sigma_{11}\chi^2_{p-1}$

两者独立（因为$w_{11\bullet 2}$与$\mathbf{w}_{21},W_{22}$独立），所以

$$\frac{m-p+1}{p-1}\times\frac{R^2}{1-R^2}=\frac{\mathbf{w}_{12}W_{22}^{-1}\mathbf{w}_{21}/(p-1)}{w_{11\bullet 2}/(m-p+1)} \sim F_{p-1,m-p+1}.$$

注1：$p=2$时， $R^2=\frac{w_{12}}{\sqrt{w_{11}w_{22}}}=r^2$(相关系数平方),

$$(m-1)\times\frac{r^2}{1-r^2}\sim F_{1,m-1}\Leftrightarrow\sqrt{m-1}\frac{r}{\sqrt{1-r^2}}\sim t_{m-1},$$

这是推论2 的结果。

注2：若$W=\sum_{i=1}^m\boldsymbol{z}_i\,\boldsymbol{z}_i^\top$， $\boldsymbol{z}_1,\ldots,$ $\boldsymbol{z}_m\ iid\sim N_p(\boldsymbol{0},\Sigma)$.

记$\boldsymbol{z}_i=(z_{i1},z_{i2},\ldots.,z_{ip})^\top$, 考虑线性回归

$$z_{i1}\sim z_{i2}+\ldots.+z_{ip}$$

其决定系数$R^2=\frac{\mathbf{w}_{12}W_{22}^{-1}\mathbf{w}_{21}}{w_{11}}$

第6讲命题5和习题3.5求出了 $W_2(m, I_2)$的 密度，一般情况类似。

Wishart分布的概率密度

定理4. $m \ge p$时，$W \sim W_p(m,\Sigma)$的概率密度函数为

$$p_{W_p(m,\Sigma)}(W) = \frac{|\Sigma|^{-m/2}|W|^{(m-p-1)/2}}{2^{mp/2}\Gamma_p(m/2)}\exp\left(-\frac{1}{2}tr(W\Sigma^{-1})\right),$$

其中$\Gamma_p(x) = \pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(x-\frac{i-1}{2}\right),\ x > (p-1)/2.$

证明大概：首先假设$\Sigma = I_p$. 划分$W = \begin{pmatrix} w_{11} & \mathbf{w}_{21}^{\mathrm{T}} \\ \mathbf{w}_{21} & W_{22} \end{pmatrix}$,其中$v_{11}$为$1\times 1$,

由定理3, $w_{11\bullet 2} \sim W_1(m-p+1, I) = \chi^2_{m-p+1}, \mathbf{w}_{21} | W_{22} \sim N(0, W_{22}), W_{22} \sim W_{p-1}(m)$

所以W的密度

$$f_p(W) = f(w_{11\bullet 2}, \mathbf{w}_{21}, W_{22}) = f(w_{11\bullet 2}, \mathbf{w}_{21} | W_{22}) f_{p-1}(W_{22})$$
$$= f(w_{11\bullet 2}) f(\mathbf{w}_{21} | W_{22}) f_{p-1}(W_{22})$$

因此我们得到了$W_p(m)$和$W_{p-1}(m)$密度的递归表达，结合$w_{pp} \sim \chi^2_m$,

可推导出$f_p(V)$。进一步，通过变换$W \to \Sigma^{1/2}W\Sigma^{1/2}$得到$W_p(m，\Sigma)$的密度。

细节参见附录.

证明：首先，假设$\Sigma = I_p$。归纳法，显然$p=1$时成立（卡方）。

划分$W = \begin{pmatrix} w_{11} & \mathbf{w}_{12} \\ \mathbf{w}_{21} & W_{22} \end{pmatrix}$，左上角$w_{11}$为标量.

由定理3，$w_{11\bullet2} \sim \chi^2_{m-p+1}, \mathbf{w}_{21} \mid W_{22} \sim N_{p-1}(\mathbf{0}, W_{22}), W_{22} \sim W_{p-1}(m)$,

假设$(p-1)\times(p-1)$矩阵W_{22}的密度$p(W_{22})$具有形式

$$p(W_{22}) = \frac{1}{2^{m(p-1)/2}\Gamma_{p-1}(m/2)} |W_{22}|^{(m-p)/2} \exp(-trW_{22}/2)$$

$(w_{11\bullet2}, \mathbf{w}_{21}, W_{22})$的联合密度$p(w_{11\bullet2}, \mathbf{w}_{21}, W_{22}) = p(w_{11\bullet2})p(\mathbf{w}_{21} \mid W_{22})p(W_{22})$

$$= \frac{1}{2^{(m-p+1)/2}\Gamma((m-p+1)/2)} w_{11\bullet2}{}^{(m-p+1)/2-1} \exp(-w_{11\bullet2}/2)$$

$$\times \frac{1}{(2\pi)^{(p-1)/2}|W_{22}|} \exp(-\mathbf{w}_{21}^{\top} W_{22}^{-1} \mathbf{w}_{21}) \times \frac{1}{2^{m(p-1)/2}\Gamma_{p-1}(m/2)} |W_{22}|^{(m-p)/2} \exp(-trW_{22}/2)$$

$$= \frac{1}{2^{mp/2}\Gamma_p(m/2)} |W|^{(m-p-1)/2} \exp\left(-\frac{1}{2}tr(W)\right) \qquad （*）$$

注意$|W| = w_{11\bullet2}|W_{22}|$

变换$(w_{11\bullet 2},\mathbf{w}_{21},W_{22})\to(w_{11},\mathbf{w}_{21},W_{22})$的Jacobian $\mathrm{J}=1$,

所以（*）式就是$(w_{11},\mathbf{w}_{21},W_{22})$的联合概率密度，即$W$的概率密度。

若$W\sim W_p(m,\Sigma)$，我们已知$V=\Sigma^{-1/2}W\Sigma^{-1/2}\sim W_p(m)$,

上一页已经证明了V的概率密度

$$p_{W_p(m)}(V)=\frac{1}{2^{mp/2}\Gamma_p(m/2)}|V|^{(m-p-1)/2}\exp\left(-\frac{1}{2}tr(V)\right),$$

则$W=\Sigma^{1/2}V\Sigma^{1/2}$的Jacobian：$J(V\to W)=|\Sigma|^{-(p+1)/2}$，$W$的密度：

$$p_{W_p(m,\Sigma)}(W)=p_{W_p(m)}(\Sigma^{-1/2}W\Sigma^{-1/2})|\Sigma|^{-(p+1)/2}\text{。}$$

至此，我们证明了定理4.

设X是$n\times n$对称矩阵，A是$n\times n$常数矩阵，

变换$Y=AXA^{\mathrm{T}}$的Jacobian $J=|A|^{n+1}$.

参见Bilodeau and Brenner (2009) P30

问题：从矩阵正态密度出发，通过变换求解Wishart分布？

我们已知(P2)，若$\mathbf{z}_1,\ldots,\mathbf{z}_m\ iid \sim N_p(0,\Sigma)$，则$Z=(\mathbf{z}_1,\ldots,\mathbf{z}_m)^\top$的联合概率密度

$$p(Z)=\frac{1}{(2\pi)^{mp/2}|\Sigma|^{m/2}}\exp\left(-\frac{1}{2}tr\left(\Sigma^{-1}Z^\top Z\right)\right).$$

$Z\to W=Z^\top Z$的Jacobian：$J=J(Z\to W)=?$

Wishart矩阵的特征根分布

$m\geq p$时，$W\sim W_p(m)$的特征根$\lambda_1,\ldots,\lambda_p>0$的联合概率密度函数

$$p(\lambda_1,\ldots,\lambda_p)=c_{mp}\exp\left(-\frac{1}{2}\sum_i\lambda_i\right)\left(\prod_{i=1}^p\lambda_i\right)^{(m-p-1)/2}\prod_{i<j}|\lambda_i-\lambda_j|$$

附：矩阵拉直/向量化

- 矩阵$Y=(\mathbf{y}_1,\ldots,\mathbf{y}_n)$拉直： $\mathrm{vec}(Y)=\begin{pmatrix}\mathbf{y}_1\\ \vdots\\ \mathbf{y}_n\end{pmatrix}$,
- 矩阵A,B的Kronecker乘积：$A\otimes B=\left(a_{ij}B\right)$.
- $X=\begin{pmatrix}\mathbf{x}_1^\top\\ \vdots\\ \mathbf{x}_n^\top\end{pmatrix}=(\mathbf{x}_1,\ldots,\mathbf{x}_n)^\top=(\mathbf{x}_{(1)},\ldots,\mathbf{x}_{(n)})$

  $\mathbf{x}_1,\ldots,\mathbf{x}_n\ \mathrm{iid}\sim N_p(\boldsymbol{\mu},\Sigma)$

  $\Leftrightarrow \mathrm{vec}(X^\top)\sim N_{np}(\mathbf{1}\otimes\boldsymbol{\mu},I_n\otimes\Sigma)$

  $\Leftrightarrow \mathrm{vec}(X)\sim N_{np}(\boldsymbol{\mu}\otimes\mathbf{1},\Sigma\otimes I_n)$
- 常用性质：

$$\mathrm{vec}(AXB)=(B^\top\otimes A)\mathrm{vec}(X)\ ,$$
$$(A\otimes B)(C\otimes D)=(AC)\otimes(BD)$$

Lemma 6.1 *The Kronecker product satisfies the following:*

(i) $(a\mathbf{A}) \otimes (b\mathbf{B}) = ab(\mathbf{A} \otimes \mathbf{B})$, $a, b \in \mathbb{R}$

(ii) $(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C})$

(iii) $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$

(iv) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$,

(v) $(\mathbf{AB}) \otimes (\mathbf{CD}) = (\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D})$

(vi) $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$, *whenever* $\mathbf{A}$ *and* $\mathbf{B}$ *are nonsingular.*

(vii) *If* $\mathbf{v} \neq \mathbf{0}$ *and* $\mathbf{u} \neq \mathbf{0}$ *are eigenvectors of* $\mathbf{A}$ *and* $\mathbf{B}$, *respectively,* $\mathbf{Av} = \lambda\mathbf{v}$, *and* $\mathbf{Bu} = \gamma\mathbf{u}$, *then* $\mathbf{v} \otimes \mathbf{u}$ *is an eigenvector of* $\mathbf{A} \otimes \mathbf{B}$ *corresponding to the eigenvalue* $\lambda\gamma$.

(viii) $\mathrm{tr}\,(\mathbf{A} \otimes \mathbf{B}) = (\mathrm{tr}\,\mathbf{A})(\mathrm{tr}\,\mathbf{B})$

(ix) $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^q\ |\mathbf{B}|^p$, $\mathbf{A} \in \mathbb{R}_p^p$, $\mathbf{B} \in \mathbb{R}_q^q$

(x) *If* $\mathbf{A} > \mathbf{0}$ *and* $\mathbf{B} > \mathbf{0}$, *then* $\mathbf{A} \otimes \mathbf{B} > \mathbf{0}$.

参见Bilodeau & Brenner, *P*74, Lemma 6.1.