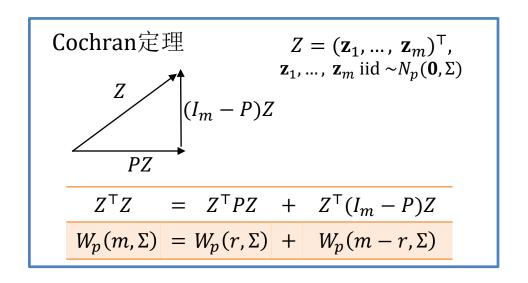
第八讲 Wishart分布III

2025.3.26



Wishart 分布的性质

我们下面讨论Wishart分布的一些性质,这些性质都可以从定义导出(不需要概率密度)。

定理1:
$$若W \sim W_p(m,\Sigma)$$
, $A \not\equiv q \times p$ 矩阵,则 $AWA^{\mathsf{T}} \sim W_q(m,A\Sigma A^{\mathsf{T}})$ 。

证:
$$W = \sum_{i=1}^{m} \mathbf{z}_{i} \mathbf{z}_{i}^{\mathsf{T}}$$
, 其中 $\mathbf{z}_{1},...,\mathbf{z}_{m}$ iid ~ $N_{p}(\mathbf{0},\Sigma)$, 因为 $A\mathbf{z}_{i}$ ~ $N_{q}(\mathbf{0},A\Sigma A^{\mathsf{T}})$,

按照定义,
$$AWA^{\mathsf{T}} = \sum_{i=1}^{m} A\mathbf{z}_{i} (A\mathbf{z}_{i})^{\mathsf{T}} \sim W_{q}(m, A\Sigma A^{\mathsf{T}}).$$

推论1:

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

反之,若
$$U \sim W_p(m, I_p)$$
,则 $\Sigma^{1/2}U\Sigma^{1/2} \sim W_p(m, \Sigma)$.

(2)对任何
$$\mathbf{t} \in R^p, \mathbf{t}^\mathsf{T} W \mathbf{t} \sim W_1(m, \mathbf{t}^\mathsf{T} \Sigma \mathbf{t}) \stackrel{d}{=} (\mathbf{t}^\mathsf{T} \Sigma \mathbf{t}) \chi_m^2, \quad 即 \frac{\mathbf{t}^\mathsf{T} W \mathbf{t}}{\mathbf{t}^\mathsf{T} \Sigma \mathbf{t}} \sim \chi_m^2$$

引理1. 假设 $W \sim W_p(m, \Sigma)$,若 $m \geq p$,则P(W > 0) = 1.

证明:假设
$$W = Z^{\mathsf{T}}Z = \sum_{i=1}^{m} \mathbf{z}_{i} \mathbf{z}_{i}^{\mathsf{T}} \sim W_{p}(m, \Sigma),$$
其中 $\mathbf{z}_{1},...,\mathbf{z}_{m}$ $iid \sim N_{p}(\mathbf{0}, \Sigma), Z = (\mathbf{z}_{1},...,\mathbf{z}_{m})^{\mathsf{T}},$
则 $P(W$ 不可逆 $) = P(rank(W) < p) = P(rank(Z) < p)$

$$= P(\mathbf{z}_{1},...,\mathbf{z}_{m}$$
中任取 p 个都线性相关 $)$

$$= \binom{m}{p} P(\mathbf{z}_{1},...,\mathbf{z}_{p}$$
线性相关 $)$

$$= \binom{m}{p} E\{P(\mathbf{z}_{1} \in L(\mathbf{z}_{2},...,\mathbf{z}_{p}) | \mathbf{z}_{2},...,\mathbf{z}_{p})\} = 0,$$
因为 \mathbf{z}_{1} 与 $\mathbf{z}_{2},...,\mathbf{z}_{p}$ 独立。

引理2. $\mathbf{z}_1,...,\mathbf{z}_m$ iid $\sim N_p(\mathbf{0},\Sigma),\ Z=(\mathbf{z}_1,...,\mathbf{z}_m)^\mathsf{T}$,常数矩阵A,B都是m列,若 $AB^\mathsf{T}=0$,则 $AZ \sqcup BZ$ 。

推论2 (p = 1): 假设 $\mathbf{z} \sim N_m(0, \sigma^2 I_m)$, 若 $AB^{\mathsf{T}} = 0$,则 $A\mathbf{z} \perp B\mathbf{z}$ 。

证明思路: $AB^{\mathsf{T}} = 0$ 说明 A,B的行向量正交。若A,B具有 形式:

$$A = (A_1, 0), B = (0, B_1),$$
 (*)

证明1 (Schmidt): 对 $A_{k \times m}$ 的行向量(假设行满秩)实施Schmidt正交化:

$$A_{k \times m} = T_{1k \times k} H_{1k \times m}$$
,其中 T_1 是可逆下三角矩阵, $H_1 H_1^{\mathsf{T}} = I_k$,

将
$$H_1$$
补成正交方阵 $H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} \Rightarrow A = (T_1, 0) \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}$ 。

记
$$Y = HZ \stackrel{d}{=} Z$$
, Y 的行向量 iid $\sim N_p(\mathbf{0}, \Sigma)$, 划分 $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, $Y_1 \perp Y_2$.

$$AZ = (T_1, 0)Y = (T_1, 0) {Y_1 \choose Y_2} = T_1Y_1,$$
 仅与 Y_1 有关;

$$BZ = BH^{\mathsf{T}}Y \triangleq CY = (C_1, C_2) \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \not\exists \, \oplus C = BH^{\mathsf{T}}$$

$$\Rightarrow BZ = (C_1, C_2) {Y_1 \choose Y_2} = C_2 Y_2$$
仅与 Y_2 有关,故与 $AZ = T_1 Y_1$ 独立。

证明2 (svd): 假设 $rank(A_{k \times m}) = r$, 其奇异值分解(svd):

$$A = UDV^{\top}, U \in \mathcal{O}(k), V \in \mathcal{O}(m), D_{k \times m} = \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix}$$

$$记Y = V^{\mathsf{T}}Z = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} 与 Z 同分布, Y_1 \perp Y_2$$

(1)
$$AZ = UDV^{\mathsf{T}}Z = UDY = U\begin{pmatrix} \Lambda_r Y_1 \\ 0 \end{pmatrix};$$

(2)
$$BZ = BVY \triangleq CY = (C_1, C_2) {Y_1 \choose Y_2} = C_1Y_1 + C_2Y_2, \sharp \oplus C = BV,$$

$$\Rightarrow C_1^\top = 0 \Rightarrow BZ = C_1Y_1 + C_2Y_2 = C_2Y_2$$

所以*AZ*业*BZ*。

Cochran定理是第7讲引理4的多元版本:

引理4. 假设 $\mathbf{x} \sim N_m(0, I_m)$, P是秩为r的 $m \times m$ 投影矩阵(对称幂等矩阵),则 $\mathbf{x}^{\mathsf{T}} P \mathbf{x} \sim \chi_r^2$, $\mathbf{x}^{\mathsf{T}} (I_m - P) \mathbf{x} \sim \chi_{m-r}^2$,两者独立

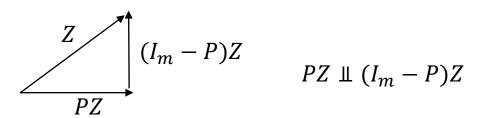
Cochran 定理

定理2 (Cochran) 假设 $\mathbf{z}_1, ..., \mathbf{z}_m$ iid $\sim N_p(\mathbf{0}, \Sigma), Z = (\mathbf{z}_1, ..., \mathbf{z}_m)^{\mathsf{T}},$ 若P是 $m \times m$ 对称幂等常数矩阵,r = rank(P),则

$$Z^{\mathsf{T}}PZ \sim W_p(r,\Sigma), \ Z^{\mathsf{T}}(I_m-P)Z \sim W_p(m-r,\Sigma),$$

且两者独立。

Cochran定理图示



$$Z$$
 = PZ + $(I_m - P)Z$
 Z^TZ = Z^TPZ + $Z^T(I_m - P)Z$
 $W_p(m, \Sigma)$ = $W_p(r, \Sigma)$ + $W_p(m - r, \Sigma)$

证明:因为P是对称幂等矩阵,存在正交矩阵 $H \in \mathcal{O}(m)$

$$P = H \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} H^{\top} \Rightarrow I_m - P = H \begin{pmatrix} 0 & 0 \\ 0 & I_{m-r} \end{pmatrix} H^{\top}$$

记 $Y = H^{\mathsf{T}}Z$, 它与Z同分布,其行向量 iid $\sim N_p(\mathbf{0}, \Sigma)$,划分 $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, $Y_1 \coprod Y_2$,

$$Z^{\mathsf{T}}PZ = Y^{\mathsf{T}} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Y = Y_1^{\mathsf{T}} Y_1 \sim W_p(r, \Sigma)$$

$$Z^{\mathsf{T}}(I_m - P)Z = Y^{\mathsf{T}} \begin{pmatrix} 0 & 0 \\ 0 & I_{m-r} \end{pmatrix} Y = Y_2^{\mathsf{T}} Y_2 \sim W_p(m - r, \Sigma)$$

 $\perp \!\!\! \perp Z^{\mathsf{T}} P Z \perp \!\!\! \perp Z^{\mathsf{T}} (I_m - P) Z_{\circ}$

样本协方 差矩阵的 分布

引入Wishart分布的目的是为了考察样本协方差矩阵S的分布,由Cochran定理可知(n-1)S服从Wishart分布。

样本:
$$\mathbf{x}_{1},...,\mathbf{x}_{n} \in R^{p}$$

样本矩阵: $X = (\mathbf{x}_{1},...,\mathbf{x}_{n})^{\mathsf{T}}$
样本协方差矩阵:
$$S = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_{i} - \bar{\mathbf{x}}) (\mathbf{x}_{i} - \bar{\mathbf{x}})^{\mathsf{T}} = X^{\mathsf{T}} (I_{n} - P_{1}) X$$

定理3. 假设 $\mathbf{x}_1, ..., \mathbf{x}_n$ iid $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), S$ 为样本协方差矩阵,则 $W = (n-1)S \sim W_p(n-1, \boldsymbol{\Sigma})$

证明:
$$\diamondsuit \mathbf{z}_i = \mathbf{x}_i - \mathbf{\mu} \sim N_p(\mathbf{0}, \Sigma)$$
 $X = (\mathbf{x}_1, ..., \mathbf{x}_n)^{\mathsf{T}}, \ Z = (\mathbf{z}_1, ..., \mathbf{z}_n)^{\mathsf{T}} = X - \mathbf{1}\mathbf{\mu}^{\mathsf{T}}$
 $(n-1)S = X^{\mathsf{T}}(I_n - P_1)X = Z^{\mathsf{T}}(I_n - P_1)Z$
因为 $rank(I_n - P_1) = tr(I_n - P_1) = n - 1$,
由Cochran定理, $(n-1)S \sim W_p(n-1, \Sigma)$.

Fisher问题:

假设**z** ~ $N(0, I_m)$,若**z**^TA**z** ~ χ_r^2 ,**z**^T**z** - **z**^TA**z** = **z**^T $(I_m - A)$ **z** ~ χ_{m-r}^2 , 矩阵A必须是幂等/投影矩阵吗?Cochran问题是其多元情形

Cochran 定理II

Cochran定理II: 假设 $\mathbf{z}_1,...,\mathbf{z}_m \sim N_p(\mathbf{0},\Sigma), Z = (\mathbf{z}_1,....,\mathbf{z}_m)^\mathsf{T},$ 若A是一个 $m \times m$ 对称常数矩阵, r = rank(A),则 $A 对称幂等 \Leftrightarrow Z^\mathsf{T}AZ \perp\!\!\!\perp Z^\mathsf{T}(I_m - A)Z$ $\Leftrightarrow Z^\mathsf{T}AZ \sim W_p(r,\Sigma), \ Z^\mathsf{T}(I_m - A)Z \sim W_p(m - r,\Sigma).$

Cochran定理II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 $A = H \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} H^{\mathsf{T}}, H$ 正交, $\Lambda = diag(\lambda_1, ..., \lambda_r), \lambda_i \neq 0$

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

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.

矩阵向量化 和Kronecker 乘积

矩阵向量化将矩阵拉直成向量,方便处理一些复杂运算,主要方便于描述随机矩阵的方差和协方差。但拉直会失去(打乱)矩阵的结构,因此我们只在描述随机矩阵的协方差结构时才使用该记号。

$$\square$$
 矩阵 $Y = (\mathbf{y}_1, ..., \mathbf{y}_n)$ 拉直: $\operatorname{vec}(Y) = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}$

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

$$\square X = \begin{pmatrix} \mathbf{x}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{x}_n^{\mathsf{T}} \end{pmatrix} = (\mathbf{x}_1, ..., \mathbf{x}_n)^{\mathsf{T}} = (\mathbf{x}_{(1)}, ..., \mathbf{x}_{(n)})$$

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

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

$$\Leftrightarrow \operatorname{vec}(X) \sim N_{np}(\mu \otimes \mathbf{1}, \Sigma \otimes I_n)$$

□ 常用性质:

$$\operatorname{vec}(AXB) = (B^{\mathsf{T}} \otimes A)\operatorname{vec}(X), \ (A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

参见附录

Wishart 分块

定理4. 假设 $W \sim W_p(m, \Sigma), m \geq p$, 划分

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

其中 W_{11} , Σ_{11} 是 $q \times q$ 方阵, $1 \leq q < p$

- (1) $W_{11} \sim W_q(m, \Sigma_{11}), W_{22} \sim W_{p-q}(m, \Sigma_{22});$
- (2) $W_{11 \bullet 2} = W_{11} W_{12}W_{22}^{-1}W_{21} \sim W_q(m p + q, \Sigma_{11 \bullet 2}),$ $\coprod W_{11 \bullet 2} \coprod \{W_{12}, W_{22}\}.$
- (3) 当q=1时 $W_{12}|W_{22}\sim N_{p-1}(\Sigma_{12}\Sigma_{22}^{-1}W_{22},W_{22}\sigma^2),\sigma^2=\Sigma_{11\bullet 2}$

q > 1时,给定 W_{22} 时, W_{12} 所有元素服从q(p-q)-元正态分布: $vec(W_{12})|W_{22}\sim N_{q(p-q)}(vec(\Sigma_{12}\Sigma_{22}^{-1}W_{22}),W_{22}\otimes\Sigma_{11•2})$

参见Bilodeau & Brenner, 2009, P92, Proposition 7.9

注:为什么考察 $W_{11•2}$?这是由生成W的多元正态随机向量条件化/去相关化所需的。我们可应用定理4递归求W, $\det(W)$ 的分布。

证明:

(1) 取 $A = (I_q, 0)$,由定理1, $W_{11} = AWA^{\mathsf{T}} \sim W_q(m, \Sigma_{11})$.

 W_{11} 与 W_{22} 未必独立,除非 $\Sigma_{12} = 0$

(2) 假设
$$W = Z^{\mathsf{T}}Z, Z = \begin{pmatrix} \mathbf{z}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{z}_m^{\mathsf{T}} \end{pmatrix}, \mathbf{z}_1, \dots, \mathbf{z}_m \ iid \sim N_p(\mathbf{0}, \Sigma),$$

划分 $\mathbf{z}_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix}$,其中 \mathbf{x}_i 是 $q \times 1$ 的,则

$$Z_{m \times p} = \begin{pmatrix} \mathbf{z}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{z}_m^{\mathsf{T}} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^{\mathsf{T}} & \mathbf{y}_1^{\mathsf{T}} \\ \vdots & \vdots \\ \mathbf{x}_m^{\mathsf{T}} & \mathbf{y}_m^{\mathsf{T}} \end{pmatrix} \triangleq (X_{m \times q}, Y_{m \times (p-q)})$$

$$\Rightarrow \ W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} = Z^\top Z = \begin{pmatrix} X^\top \\ Y^\top \end{pmatrix} (X,Y) = \begin{pmatrix} X^\top X & X^\top Y \\ Y^\top X & Y^\top Y \end{pmatrix},$$

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

不能直接对 $W_{11•2} = X^{\mathsf{T}}(I_m - P_Y)X$ 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,但此时X的分布会有变化。

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

记 $\mathbf{v}_i = \mathbf{x}_i - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{y}_i$, $\mathbf{v}_i | \mathbf{y}_i \sim N_q(\mathbf{0}, \Sigma_{11 \bullet 2})$ 不依赖于 \mathbf{y}_i , 故 $\mathbf{v}_i \perp \mathbf{y}_i$. 记

$$V = \begin{pmatrix} \mathbf{v}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{v}_m^{\mathsf{T}} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{x}_m^{\mathsf{T}} \end{pmatrix} - \begin{pmatrix} \mathbf{y}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{y}_m^{\mathsf{T}} \end{pmatrix} \Sigma_{22}^{-1} \Sigma_{21} = X - Y \Sigma_{22}^{-1} \Sigma_{21}$$

所以

$$W_{11 \bullet 2} = X^{\mathsf{T}} (I_m - P_Y) X = (V + Y \Sigma_{22}^{-1} \Sigma_{21})^{\mathsf{T}} (I_m - P_Y) (V + Y \Sigma_{22}^{-1} \Sigma_{21})$$
$$= V^{\mathsf{T}} (I_m - P_Y) V$$

注意 $V \perp Y$ (给定Y时不改变V 的分布),由Cochran定理,

$$W_{11•2}|Y = V^{\mathsf{T}}(I_m - P_Y)V|Y \sim W_q(m - (p - q), \Sigma_{11•2}),$$

该分布与Y无关,所以W_{11•2} LY, 且

$$W_{11•2} \sim W_q(m - (p - q), \Sigma_{11•2}).$$

显然,
$$W_{11 \bullet 2} \perp Y \Rightarrow W_{11 \bullet 2} \perp Y^{\mathsf{T}} Y = W_{22}$$

改写 $W_{12} = X^{\mathsf{T}}Y = (V + Y\Sigma_{22}^{-1}\Sigma_{21})^{\mathsf{T}}Y = V^{\mathsf{T}}Y + \Sigma_{12}\Sigma_{22}^{-1}Y^{\mathsf{T}}Y$ 给定Y时, $I_m - P_Y$,Y, $Y^{\mathsf{T}}Y$ 都是常数矩阵,且 $(I_m - P_Y)Y = 0$,由引理 2, $Y^{\mathsf{T}}V \perp (I_m - P_Y)V \mid Y$,所以

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

但 $W_{11•2}$ $\perp Y$, 所以由第7讲引理3,

$$W_{11 \bullet 2} \perp \{ W_{12}, W_{22} \}$$

(3) $V = X - Y\Sigma_{22}^{-1}\Sigma_{21}$, 其各行 \mathbf{v}_i iid $\sim N_q(\mathbf{0}, \Sigma_{11 \bullet 2})$,

给定Y时, $W_{12} = X^{\mathsf{T}}Y = (V + Y\Sigma_{22}^{-1}\Sigma_{21})^{\mathsf{T}}Y = V^{\mathsf{T}}Y + \Sigma_{12}\Sigma_{22}^{-1}Y^{\mathsf{T}}Y$ 一定服从多元正态,但具体形式用矩阵拉直和Kronecker乘积比较容易描述:

$$vec(V^{\top}) \sim N_{mq}(\mathbf{0}, I_m \otimes \Sigma_{11 \bullet 2})$$

$$\Rightarrow vec(V^{\top}Y) = (Y^{\top} \otimes I_q) vec(V^{\top})$$

$$\Rightarrow vec(V^{\top}Y)|Y \sim N_{(p-q)q}(\mathbf{0}, (Y^{\top}Y \otimes \Sigma_{11 \bullet 2}))$$

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

 W_{12} 代表两部分变量**x**,**y**之间的相关性或相似性,它将是典则相关分析的主要考察对象。其分布在 H_0 : $\Sigma_{12} = 0$ 时的分布是什么?

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

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

(1) 由定理4,
$$W_{11•2} = W_{11} - \mathbf{W}_{12}W_{22}^{-1}\mathbf{W}_{21} \sim W_1(m-p+1,\sigma_{11•2}) = \sigma_{11}\chi_{m-p+1}^2$$
.

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

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

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

Wishart分 布的概率 密度

定理5. $m \ge p$ 时, $W \sim W_p(m, \Sigma)$ 的概率密度函数为

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

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

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

由定理3, $W_{11\bullet 2} \sim W_1(m-p+1,I) = \chi^2_{m-p+1}, \mathbf{W}_{21} \mid W_{22} \sim N(0,W_{22}), W_{22} \sim W_{p-1}(m)$ 所以W的密度

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

因此我们得到了 $W_p(m)$ 和 $W_{p-1}(m)$ 密度的递归表达,结合 $w_{pp} \sim \chi_m^2$,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4, $W_{11\bullet 2} \sim \chi^2_{m-p+1}$, $\mathbf{W}_{21} \mid W_{22} \sim N_{p-1}(\mathbf{0}, W_{22})$, $W_{22} \sim W_{p-1}(m)$,

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

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

$$(w_{11\bullet 2}, \mathbf{w}_{21}, W_{22})$$
的联合密度 $p(w_{11\bullet 2}, \mathbf{w}_{21}, W_{22}) = p(w_{11\bullet 2}) p(\mathbf{w}_{21} | W_{22}) p(W_{22})$

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

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

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

注意|W|=W_{11•2}|W₂₂|

变换 $(w_{11}, \mathbf{w}_{21}, W_{22}) \rightarrow (w_{11}, \mathbf{w}_{21}, W_{22})$ 的Jacobian J = 1, 所以(*)式就是 (w_1, \mathbf{w}_2, W_2) 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:
$$p_{W_p(m,\Sigma)}(W) = p_{W_p(m)}(\Sigma^{-1/2}W\Sigma^{-1/2}) |\Sigma|^{-(p+1)/2}$$
。

至此,我们证明了定理4.

其中X,Y是 $p \times p$ 对称矩阵,B是 $p \times p$ 可逆矩阵,则 $J(X \to Y) = |B|^{p+1}$

Wishart 有关的 二次型

定理6: 假设 $W \sim W_p(m)$, $m \geq p$, 则对 \forall 常数向量 $\mathbf{t} \in S^{p-1}$, 有

- (1) $\mathbf{t}^{\mathsf{T}}W\mathbf{t} \sim \chi_m^2$;
- $(2) \frac{1}{\mathbf{t}^{\mathsf{T}} W^{-1} \mathbf{t}} \sim \chi_{m-p+1}^2.$

证明:首先,因为 $m \ge p$, W可逆。

(1)
$$W \sim W_p(m, I_p) \Rightarrow \mathbf{t}^\mathsf{T} W \mathbf{t} \sim W_1(m, \mathbf{t}^\mathsf{T} I_p \mathbf{t}) = W_1(m, 1) = \chi_m^2$$

(2) 的证明思路: 如果
$$\mathbf{t} = \mathbf{e_1} = (1,0,...,0)^{\mathsf{T}}$$
, 那么 $\mathbf{t}^{\mathsf{T}}W^{-1}\mathbf{t} = w_{11\bullet2} \sim W_1(m-p+1,1) = \chi_{m-p+1}^2$ $W^{-1} = \begin{pmatrix} w_{11\bullet2}^{-1} & * \\ * & * \end{pmatrix}$ 对于一般的 \mathbf{t} , 我们只需旋转使得 $H\mathbf{t} = \mathbf{e_1}$ (而 $HW \stackrel{d}{=} W$,分布不变)

(2) 注意到对任何正交矩阵H,

$$V = HWH^{\mathsf{T}} \sim W_p(m, HH^{\mathsf{T}}) = W_p(m), \ \mathbf{t}^{\mathsf{T}}W^{-1}\mathbf{t} = \mathbf{t}^{\mathsf{T}}H^{\mathsf{T}}(HWH^{\mathsf{T}})^{-1}H\mathbf{t},$$
 取正交矩阵 H 的第一行为 \mathbf{t}^{T} ,即 $H = \begin{pmatrix} \mathbf{t}^{\mathsf{T}} \\ * \end{pmatrix}$,则 $\mathbf{u} = H\mathbf{t} = \begin{pmatrix} \mathbf{t}^{\mathsf{T}}\mathbf{t} \\ * \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}$ 故 $\mathbf{t}^{\mathsf{T}}W^{-1}\mathbf{t} = \mathbf{t}^{\mathsf{T}}H^{\mathsf{T}}(HWH^{\mathsf{T}})^{-1}H\mathbf{t} = \mathbf{u}^{\mathsf{T}}(HWH^{\mathsf{T}})^{-1}\mathbf{u} = \mathbf{u}^{\mathsf{T}}V^{-1}\mathbf{u}.$

划分
$$V = \begin{pmatrix} v_{11} & \mathbf{v}_{12} \\ \mathbf{v}_{21} & V_{22} \end{pmatrix}$$
,其中 v_{11} 是 1×1 ,则 $V^{-1} = \begin{pmatrix} v_{11 \bullet 2}^{-1} & * \\ * & * \end{pmatrix}$,所以
$$\mathbf{u}^{\mathsf{T}} V^{-1} \mathbf{u} = (1, \mathbf{0}^{\mathsf{T}}) \begin{pmatrix} v_{11 \bullet 2}^{-1} & * \\ * & * \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} = v_{11 \bullet 2}^{-1}$$

由定理4, $1/\mathbf{t}^{\mathsf{T}}W^{-1}\mathbf{t} = v_{11\bullet 2} \sim W_1(m-p+1,1) = \chi^2_{m-p+1}$.

推论4. 假设 $W \sim W_p(m)$, $\mathbf{z} \sim N_p(\mathbf{0}, I_p)$, $m \geq p$, 假设 $W \perp \mathbf{z}$, 则 $\mathbf{z}^{\mathsf{T}} \mathbf{z}/\mathbf{z}^{\mathsf{T}} W^{-1} \mathbf{z} \sim \chi^2_{m-p+1}$.

证明: 令 $\mathbf{u} = \mathbf{z} / \|\mathbf{z}\|$, 它与W独立,由定理6,给定z条件下

$$\frac{\mathbf{z}^{\mathsf{T}}\mathbf{z}}{\mathbf{z}^{\mathsf{T}}W^{-1}\mathbf{z}} = \frac{1}{\mathbf{u}^{\mathsf{T}}W^{-1}\mathbf{u}} \bigg|_{\mathbf{z}} \sim \chi_{m-p+1}^{2},$$

该分布与**z**无关,所以 $\frac{\mathbf{z}^{\mathsf{T}}\mathbf{z}}{\mathbf{z}^{\mathsf{T}}W^{-1}\mathbf{z}} \sim \chi_{m-p+1}^{2}$,且与**z**独立。

附: 矩阵拉直/向量化

□ 矩阵
$$Y = (\mathbf{y}_1, ..., \mathbf{y}_n)$$
拉直: $\operatorname{vec}(Y) = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}$

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

$$\square X = \begin{pmatrix} \mathbf{x}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{x}_n^{\mathsf{T}} \end{pmatrix} = (\mathbf{x}_1, ..., \mathbf{x}_n)^{\mathsf{T}} = (\mathbf{x}_{(1)}, ..., \mathbf{x}_{(n)})$$

$$\mathbf{x}_1, ..., \mathbf{x}_n \text{ iid } \sim N_p(\mathbf{\mu}, \Sigma)$$

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

$$\Leftrightarrow \operatorname{vec}(X) \sim N_{np}(\mu \otimes \mathbf{1}, \Sigma \otimes I_n)$$

□ 常用性质:

$$\operatorname{vec}(AXB) = (B^{\mathsf{T}} \otimes A)\operatorname{vec}(X)$$
,
 $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$

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)
$$(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$$

(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, $\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$, and $\mathbf{B}\mathbf{u} = \gamma \mathbf{u}$, then $\mathbf{v} \otimes \mathbf{u}$ is an eigenvector of $\mathbf{A} \otimes \mathbf{B}$ corresponding to the eigenvalue $\lambda \gamma$.

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

(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
$$A > 0$$
 and $B > 0$, then $A \otimes B > 0$.

参见Bilodeau & Brenner, P74, Lemma 6.1.

附录2: Wishart矩阵的特征根分布

$$W \sim W_p(m)$$
,谱分解: $W = V \Lambda V^{\mathsf{T}}$, $\Lambda = diag(\lambda_1, ..., \lambda_p)$ 。 W 的概率密度

$$p(W) = c|W|^{(m-p-1)/2} \exp\left(-\frac{trW}{2}\right)$$
$$= \frac{1}{2^{mp/2}\Gamma_p(m/2)} |\Lambda|^{(m-p-1)/2} \exp\left(-\frac{tr\Lambda}{2}\right) \triangleq h(\Lambda)$$

仅依赖于特征根Λ,

范德蒙行列式 Vandermonde determinant

Alexandre-Théophile Vandermonde: 法 国18世纪音乐家、 数学家、化学家。

$$p(W)(dW) = h(\Lambda)J(W \to V, \Lambda)(dV)(d\Lambda)$$

$$= \frac{2^{p}\pi^{p^{2}/2}}{2^{mp/2}\Gamma_{p}(m/2))\Gamma_{p}(p/2)}|\Lambda|^{(m-p-1)/2}\exp\left(-\frac{tr\Lambda}{2}\right)\prod_{1\leq i\leq j\leq p}(\lambda_{i}-\lambda_{j})(d\Lambda)\times\frac{1}{|\mathcal{O}(p)|}(dV)$$

定理A1: 假设 $W \sim W_p(m, I_p)$, $m \geq p$, 则W的特征根 $\lambda_1, ..., \lambda_p > 0$ 的联合概率密度

$$p(\lambda_1, \dots, \lambda_p) = \frac{\pi^{p^2/2}}{2^{mp/2} \Gamma_p(m/2)) \Gamma_p(p/2)} |\Lambda|^{(m-p-1)/2} \exp\left(-\frac{tr\Lambda}{2}\right) \prod_{1 \le i < j \le p} |\lambda_i - \lambda_j|,$$
 另外 $V \sim U(\mathcal{O}(m)).$

