1. Expectations and covariance matrices of random vectors

Definition 1 (随机向量的期望与协方差) 对 $\forall\ S \in \mathbb{R}^p,\ T \in \mathbb{R}^q$，记：

$$E(S) = [E(S_{[1]}), E(S_{[2]}), \cdots, E(S_{[p]})]^T$$

为向量 S 的期望 (expectation)

$$\mathrm{cov}(S,T) = E\{(S-E(S))(T-E(T))^T\} \in \mathbb{R}^{p \times q}$$

为向量 S 和 T 的协方差矩阵，特别地，记

$$\mathrm{cov}(T) = \mathrm{cov}(T,T)$$

为 T 的协方差矩阵 (covariance matrix)

Proposition 1 (期望、协方差性质)

$C_1 \in \mathbb{R}^{c_1 \times p}$ $\quad C_2 \in \mathbb{R}^{c_2 \times q}$ $\quad S \in \mathbb{R}^p$ $\quad T \in \mathbb{R}^q$

(a) $E(C_1 S) = C_1 E(S)$

(b) 若 $p = q$ 则 $E(S+T) = E(S) + E(T)$

(c) $\mathrm{cov}(S,T) = [\mathrm{cov}(T,S)]^T$

(d) $\mathrm{cov}(T)$ 是半正定的 (positive semi-definite, PSD)

(e) $\mathrm{cov}(C_1 S, C_2 T) = C_1 \cdot \mathrm{cov}(S,T) \cdot C_2^T \in \mathbb{R}^{c_1 \times c_2}$

(f) 若 $c_1 = c_2$ 则

$$\mathrm{cov}(C_1 S + C_2 T) = \mathrm{cov}(C_1 S) + \mathrm{cov}(C_2 T) + \mathrm{cov}(C_1 S, C_2 T) + \mathrm{cov}(C_2 T, C_1 S)$$

Proof $\mathrm{cov}(T) = E\{(T-E(T))(T-E(T))^T\}$

$\therefore\ v^T\cdot \mathrm{cov}(T)\cdot v = E\{v^T\cdot(T-E(T))\cdot(T-E(T))^T\cdot v\}$

注：$v^T(T-E(T)) \in \mathbb{R}$ 为标量

$\therefore\ v^T \mathrm{cov}(T)\cdot v = E([v^T(T-E(T))]^2) \geqslant 0 \quad \square$

2. Multivariate Normal Distributions (MNDs)

Definition 2 (特征函数 characteristic function)

$$\varphi_X(t) = E\{\exp(it^TX)\} \quad \forall\, t \in \mathbb{R}^n \quad i^2=-1$$

如果随机向量 X 满足如下特征函数：

$$\varphi_X(t) = E\{\exp(it^TX)\} = \exp(i\mu^Tt - \tfrac{1}{2}\cdot t^T\Sigma t)$$

其中 $t\in\mathbb{R}^n$ $X\in\mathbb{R}^n$ $i^2=-1$。则称 X 服从 n维正态分布 (n-dimentional normal distribution)，期望为 $\mu\in\mathbb{R}$，协方差矩阵为 $\Sigma\in\mathbb{R}^{n\times n}$，记作 $X\sim N(\mu,\Sigma)$

Remark：Σ 为半正定 (PSD) 即可

Theorem 1 (X 服从正态 $\Leftrightarrow$ X 所有线性组合服从正态) 下面二者等价：

(a) 随机向量 X 服从多元正态分布 (n维正态分布)

(b) 对 $\forall\ v\in\mathbb{R}^n$，随机变量 $v^TX\in\mathbb{R}$ 均服从正态分布

$$(a) \Leftrightarrow (b)$$

Lemma 1 $\forall$ 随机向量 $X, Y \in \mathbb{R}^n$，有

$X \overset{d}{=} Y$ 等价于 $v^T X \overset{d}{=} v^T Y$ 对 $\forall\ v \in \mathbb{R}^n$ 成立

Proof $\Rightarrow$ 已知 $X \overset{d}{=} Y$，有

$$\varphi_X(t) = \varphi_Y(t) \quad \forall\ t \in \mathbb{R}^n$$

有 $\varphi_{v^T X}(b) = E[\exp\{i(bv^T)X\}] = \varphi_X(bv)$

$$= \varphi_Y(bv)$$

$$= E[\exp\{i(bv^T)Y\}]$$

$$= \varphi_{v^T Y}(b)$$

$\therefore\ v^T X \overset{d}{=} v^T Y$ 对 $\forall\ v \in \mathbb{R}^n$ $\square$

$\Leftarrow$ 已知 $\forall\ v \in \mathbb{R}^n$ 有 $v^T X \overset{d}{=} v^T Y$ 有

取 $b = 1$：

$$\varphi_{v^T X}(1) = \varphi_{v^T Y}(1) \quad \forall\ v \in \mathbb{R}^n$$

有 $\varphi_X(v) = E[\exp\{i(v^T X)\}] = \varphi_{v^T X}(1)$

$$= \varphi_{v^T Y}(1)$$

$$= E[\exp\{i(v^T Y)\}]$$

$$= \varphi_Y(v)$$

$\therefore\ X \overset{d}{=} Y$ $\square$

Proof Theorem 1

(a) $\Rightarrow$ (b) 记 $X \sim N(\mu, \Sigma)$

有 $\varphi_{\underbrace{v^T X}_{\text{scaler}}}(\underbrace{b}_{\text{scaler}}) = E\{\exp(i b v^T X)\}$

$$= \underset{\text{vector}}{\varphi_X}(\underset{\text{vector}}{bv})$$

$$= \exp\{ i(v^T\mu)^T b - \frac{1}{2} b^2 (v^T \Sigma v) \}$$

$\therefore \quad v^T X \sim N(v^T\mu, \ v^T \Sigma v)$

(b) $\Rightarrow$ (a) 记 $E(X) = \mu \quad cov(X) = \Sigma$

故 $E(v^T X) = v^T\mu \qquad var(v^T X) = v^T \Sigma v$

取任一个 $Y \in \mathbb{R}^n \quad Y \sim N(\mu, \Sigma)$

由 (a) $\Rightarrow$ (b) 有 $v^T Y \sim N(v^T\mu, v^T\Sigma v) \qquad \forall v \in \mathbb{R}^n$

加之 (b) 表示 $\forall \ v \in \mathbb{R}^n \quad v^T X \sim N(v^T\mu, v^T\Sigma v)$

得出 $v^T Y \overset{d}{=} v^T X$ 对 $\forall \ v \in \mathbb{R}^n$ 成立

由 lemma 知 $X \overset{d}{=} Y \sim N(\mu, \Sigma)$ $\square$

Proposition 2 (密度函数) 随机向量 $X \in \mathbb{R}^n$ 服从 $X \sim N(\mu, \Sigma)$

则有 $E(X) = \mu \quad cov(X) = \Sigma$，且当 Σ 正定时，X 的概率密度函数为：

$$f(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \cdot \exp\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\} \qquad x \in \mathbb{R}^n$$

Remark 正态随机向量的分布完全取决于 μ 和 Σ

Proof: 对特征函数求导

$$\left.\frac{d\varphi_X(t)}{dt}\right|_{t=0} = E\left(\frac{d\exp(it^T X)}{dt}\right)\Bigg|_{t=0}$$

$$= E\left(iX \cdot \exp(it^T X) \right)\Big|_{t=0} \qquad (代入\ t=0)$$

$$= i \cdot E(X)$$

$$\therefore\ E(X) = i^{-1} \cdot \frac{d\varphi_X(t)}{dt}\Big|_{t=0}$$

(可递推得： $E(X^m) = i^{-m} \cdot \frac{d^m \varphi_X(t)}{dt^m}\Big|_{t=0}$)

回到原题： $E(X) = i^{-1} \cdot \frac{d\varphi_X(t)}{dt}\Big|_{t=0}$

$$= i^{-1} \cdot \frac{d\ \exp(i\mu^T t - \frac{1}{2} t^T \Sigma t)}{dt}\Big|_{t=0}$$

$$= i^{-1} \cdot (i\mu - 0)$$

$$= \mu \qquad \square$$

$$\mathrm{cov}(X) = E(XX^T) - EX \cdot EX^T$$

$$= i^{-2} \cdot \frac{d^2 \varphi_X(t)}{dt\,dt^T}\Big|_{t=0} - \mu\mu^T$$

$$= \Sigma + \mu\mu^T - \mu\mu^T$$

$$= \Sigma \qquad \square$$

计算概率密度，采用傅利叶逆变换：

$$f(x) = (2\pi)^{-n} \cdot \int_{\mathbb{R}^n} \varphi_X(t) \cdot \exp(-it^T X)\ dt$$

$$f(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} \exp(i\mu^T t - \frac{1}{2} t^T \Sigma t - it^T X)\, dt$$

$$= \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \cdot \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}$$

Theorem 2 (closedness under linear operations) 已知 $X \sim N(\mu, \Sigma)$ 对

$\forall$ 确定的 $A \in \mathbb{R}^{s\times n}$ 和 $b \in \mathbb{R}^n$ ，均有

$$AX + b \sim N(A\mu + b, A\Sigma A^T)$$

Proof: $\varphi_{AX+b}(t) = E[\exp\{it^T(AX+b)\}]$ X is random

$$= E[\exp\{i(A^Tt)^TX\}] \cdot \exp(it^Tb)$$

$$= \varphi_X(A^Tt) \cdot \exp(it^Tb)$$

$\because X \sim N(\mu, \Sigma)$

$\therefore$ 原式 $= \exp\{i\mu^TA^Tt - \frac{1}{2}t^TA\Sigma A^Tt + it^Tb\}$

$$= \exp\{i(A\mu + b)^Tt - \frac{1}{2}t^T(A\Sigma A^T)t\}$$

$\therefore AX + b \sim N(A\mu + b, A\Sigma A^T)$

下面讨论独立性、边际分布、条件分布

Proposition 3 (独立性 independence) 随机向量 X_1 和 X_2 ，构造 $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ 及任意 $t = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$ 则有

$$X_1 \perp X_2 \iff \varphi_X(t) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2)$$

Remark: 如果 X_1, X_2 形状相同，则有若 $X_1 \perp X_2$ ，则有

$$\varphi_{X_1+X_2}(t_1) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_1)$$

反之不成立

Proof: $\Rightarrow$ 已知 $X_1 \perp X_2$

对于 $X \perp Y$ 有 $E(XY) = EX \cdot EY$

$\therefore E\{\exp(it_1^T X_1 + it_2^T X_2)\} = E\{\exp(it_1^T X_1) \cdot \exp(it_2^T X_2)\}$

$= E\{\exp(it_1^T X_1)\} \cdot E\{\exp(it_2^T X_2)\}$

即 $\varphi_X(t) = \varphi_{X_1}(t_1) \cdot \varphi_{X_2}(t_2)$ $\square$

$\Leftarrow$ 已知 $\varphi_X(t) = \varphi_{X_1}(t_1) \cdot \varphi_{X_2}(t_2)$

取一随机向量 $\tilde{X}_2 \overset{d}{=} X_2$ 且 $\tilde{X}_2 \perp X_1$

记 $\tilde{X} = \begin{bmatrix} X_1 \\ \tilde{X}_2 \end{bmatrix}$ 由 $\varphi_X(t) = \varphi_{X_1}(t_1) \varphi_{X_2}(t_2)$

加之 $\tilde{X}_2 \overset{d}{=} X_2 \Rightarrow \varphi_{X_2}(t_2) = \varphi_{\tilde{X}_2}(t_2)$

故 $\varphi_X(t) = \varphi_{X_1}(t_1) \varphi_{X_2}(t_2) = \varphi_{X_1}(t_1) \cdot \varphi_{\tilde{X}_2}(t_2)$

而由 $\Rightarrow \tilde{X}_2 \perp X_1$ $\therefore$ $\varphi_{X_1}(t_1) \cdot \varphi_{\tilde{X}_2}(t_2) = \varphi_{\tilde{X}}(t)$

故 $\varphi_X(t) = \varphi_{\tilde{X}}(t) \Rightarrow X \overset{d}{=} \tilde{X}$

而 $X_1 \perp \tilde{X}_2 \Rightarrow X_1 \perp X_2$ $\square$

Theorem 3 (边际分布 marginal distribution) 已知 X 是 X_1, X_2 的联合分布 (joint distribution) 服从如下联合正态分布:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

则有 X_1, X_2 的边际分布:

$$X_1 \sim N(\mu_1, \Sigma_{11}) \qquad X_2 \sim N(\mu_2, \Sigma_{22})$$

更进一步,如果 $\Sigma_{12} = \Sigma_{21}^T = 0$,则有 $X_1 \perp X_2$

Remark：边际分布 $X_1 \sim N(\mu_1, \Sigma_{11})$　$X_2 \sim N(\mu_2, \Sigma_{22})$

且 $cov(X_1, X_2) = 0$

无法推出二者独立

Proof：由题 $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$　记 $t = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$　且 $X \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$

注意到：
$$\begin{aligned}\varphi_{X_1}(t_1) &= E\{\exp(it_1^T X_1)\} \\ &= E\{\exp(i(t_1^T, 0^T)X)\} \\ &= \varphi_X\left(\begin{bmatrix} t_1 \\ 0 \end{bmatrix}\right) \\ &= \exp(i\mu_1^T t_1 - \tfrac{1}{2} t_1^T \Sigma_{11} t_1) \quad \text{对 } \forall\, t_1 \in \mathbb{R}^m\end{aligned}$$

故 $X_1 \sim N(\mu_1, \Sigma_{11})$　同理 $X_2 \sim N(\mu_2, \Sigma_{22})$　□

更进一步，若 $\Sigma_{21} = \Sigma_{12}^T = 0$　有
$$\begin{aligned}\varphi_X(t) &= \exp(i\mu_1^T t_1 + i\mu_2^T t_2 - \tfrac{1}{2} t_1^T \Sigma_{11} t_1 - \tfrac{1}{2} t_2^T \Sigma_{22} t_2) \\ &= \exp(i\mu_1^T t_1 - \tfrac{1}{2} t_1^T \Sigma_{11} t_1) \cdot \exp(i\mu_2^T t_2 - \tfrac{1}{2} t_2^T \Sigma_{22} t_2) \\ &= \varphi_{X_1}(t_1) \cdot \varphi_{X_2}(t_2)\end{aligned}$$

由 Proposition 3 知 $X_1 \perp X_2$

Theorem 4（条件分布 conditional distribution）已知 X 是 X_1, X_2 的联合分布（joint distribution）服从如下联合正态分布：
$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

且 Σ_{22} 可逆，则

$$X_1 | X_2 \sim N(\mu_1 + \Sigma_{12}\cdot\Sigma_{22}^{-1}\cdot(X_2-\mu_2),\ \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

即给定 X_2 时 X_1 的分布.

Remark 如上的假定可视为 X_1 (response), X_2 (covariate) 的线性模型

$$X_1 = a_0 + b_0 X_2 + \varepsilon$$

其中: $a_0 = \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2$

$b_0 = \Sigma_{12}\Sigma_{22}^{-1}$

随机误差 $\varepsilon \sim N(0,\ \Sigma_{11}-\Sigma_{12}\cdot\Sigma_{22}^{-1}\Sigma_{21})$

Proof: 记 $Z = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$ 且记 $A = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 则

$$\begin{bmatrix} X_2 \\ Z \end{bmatrix} = \begin{bmatrix} O_{(n-m)\times m} & I_{n-m} \\ I_m & -\Sigma_{12}\Sigma_{22}^{-1} \end{bmatrix}\cdot X \sim N\left(\begin{bmatrix} \mu_2 \\ \mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2 \end{bmatrix},\ \begin{bmatrix} \Sigma_{22} & O \\ O & A \end{bmatrix}\right)$$

由 Theorem 3 副对角矩阵为 0 故 $Z \perp X_2$

且 $Z \sim N(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2,\ A)$

故 $Z|X_2 = Z \sim N(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2,\ A)$

$\therefore\ X_1 | X_2 = Z + \Sigma_{12}\Sigma_{22}^{-1}X_2 \,|\, X_2$

而 X_2 给定, $\Sigma_{12}\Sigma_{22}^{-1}X_2$ 为常数

$\therefore\ X_1|X_2 \sim N(\mu_1 + \Sigma_{12}\cdot\Sigma_{22}^{-1}\cdot(X_2-\mu_2),\ \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

3. Quadratic forms of random vectors

Definition 3（二次型） 对 $\forall A \in \mathbb{R}^{n\times n}$，标量 $X^T A X$ 称为矩阵 A 的二次型（$X \in \mathbb{R}^n$）

Lemma 1（迹的交换） 对于 $\forall C_1 \in \mathbb{R}^{p\times q}$ 和 $C_2 \in \mathbb{R}^{q\times p}$ 均有

$$tr(C_1C_2) = tr(C_2C_1)$$

Remark：只能交换 2 个矩阵，$tr(ABC) \neq tr(CBA)$

Proof：$C_1 \in \mathbb{R}^{p\times q}$　$C_2 \in \mathbb{R}^{q\times p}$

易知 $(C_1C_2)[i,j] = \sum_{k=1}^{q} C_1[ik]\, C_2[kj]$

$(C_2C_1)[i,j] = \sum_{k=1}^{p} C_2[ik]\, C_1[kj]$

$$\therefore tr(C_1C_2) = \sum_{i=1}^{p} (C_1C_2)[ii]$$

$$= \sum_{i=1}^{p}\sum_{k=1}^{q} C_1[ik]\, C_2[ki] = \sum_{k=1}^{q}\sum_{i=1}^{p} C_1[ik]\, C_2[ki]$$

交换 i,k 记号 $= \sum_{i=1}^{q}\sum_{k=1}^{p} C_1[ki]\, C_2[ik] = \sum_{i=1}^{q}\sum_{k=1}^{p} C_2[ik]\, C_1[ki]$

$$= \sum_{i=1}^{q} (C_2C_1)[ii]$$

$$= tr(C_2C_1) \quad \square$$

Theorem 5（二次型期望） 已知 $A \in \mathbb{R}^{n\times n}$，随机向量 $X \in \mathbb{R}^n$，已知

$E(X) = \mu$　　$cov(X) = \Sigma$　　则有

$$E(X^TAX) = \mu^T A \mu + tr(A\Sigma)$$

Proof: $E\{(X-\mu)^TA(X-\mu)\}$ $\because (X-\mu)^TA(X-\mu)$ 是标量

$$\therefore = tr(E\{(X-\mu)^TA(X-\mu)\})$$

$$= E\{tr((X-\mu)^TA(X-\mu))\}$$ $\because$ 交换顺序 Lemma 1

$$\therefore = E\{tr(A(X-\mu)^T(X-\mu))\}$$

$$= E\{tr(A(X-\mu)(X-\mu)^T)\}$$

$$= tr(A\cdot E\{(X-\mu)(X-\mu)^T\})$$

$$= tr(A\Sigma)$$

而 $E(X^TAX) = E\{(X-\mu+\mu)^TA(X-\mu+\mu)\}$

$$= E\{(X-\mu)^TA(X-\mu)\} + 2E\{(X-\mu)^TA\mu\} + \mu^TA\mu$$

其中 $2E\{(X-\mu)^TA\mu\} = 2A\mu\cdot(EX-\mu) = 0$

$$\therefore E(X^TAX) = tr(A\Sigma) + \mu^TA\mu \quad \square$$

下面，讨论正态分布的二次型相关

Theorem 6 $A\in\mathbb{R}^{n\times n}$ 是实对称矩阵，$B\in\mathbb{R}^{p\times n}$ 且 $BA=0$，如果随机向量 $X\sim N(\mu,\sigma^2\cdot I)$ $\sigma\in\mathbb{R}^+$。则有 $BX\perp X^TAX$

Theorem 7 $A, B\in\mathbb{R}^{n\times n}$ 均为实对称阵，满足 $AB=0$，如果随机向量 $X\sim N(\mu,\sigma^2\cdot I)$ $\sigma\in\mathbb{R}^+$。则有 $X^TAX\perp X^TBX$

见作业 HW5

4. Chi-Squre distribution (χ^2)

Definition 4 若随机变量 $X_1, X_2, \cdots, X_n \sim N(0,1)$ 且相互独立。记 $Y = X_1^2 + X_2^2 + \cdots + X_n^2$，服从自由度为 n 的 χ^2 分布 (chi-square)

$$Y \sim \chi^2(n)$$

Proposition 4 ☆

(a) 正态随机向量的二次型：

$X \in \mathbb{R}^n$ 且 $X \sim N(\mu, \Sigma)$ 且 Σ 可逆，则有

$$(X-\mu)^T \cdot \Sigma^{-1} \cdot (X-\mu) \sim \chi^2(n)$$

(b) χ^2 分布的期望和方差：

若 $Y \sim \chi^2(n)$ 则

$$E(Y) = n \qquad var(Y) = 2n$$

(c) χ^2 分布的独立可加性：

若 $Y_1 \sim \chi^2(n_1)$ $Y_2 \sim \chi^2(n_2)$ 且 $Y_1 \perp Y_2$ 则有

$$Y_1 + Y_2 \sim \chi^2(n_1+n_2)$$

下面，讨论正态随机向量二次型 & 幂等矩阵 (idempotent)

Lemma 2 (幂等矩阵的特征值) 若矩阵 C 满足 $CC = C^2 = C$ 则它的特征值一定为 0 或 1

Theorem 8 已知 $X \sim N(\mu, I)$，有一实对称阵 $A \in R^{n\times n}$ 满足 $A^2 = A$ $\mu^T A \mu = 0$，记 $r = rank(A) > 0$，则有

$$X^T A X \sim \chi^2(r)$$

见作业 HW5

因为A是实对称矩阵，故A可以被分解为$A = PDP^T$。这里P是一个正交矩阵而$D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$，其中$\{\lambda_i : i = 1, \dots, n\}$是$A$的特征值。根据引理 2，对角矩阵$D$的对角线元素必为0或1。因为$\mathrm{rank}(D) = \mathrm{rank}(A) = r$，故0和1的代数重数分别为$n - r$和$r$。不失一般性地，我们将$D$写为

$$D = \mathrm{diag}(I_r, 0_{n-r}). \tag{6}$$

以上I_r和0_{n-r}分别表示r阶单位矩阵和$(n-r)$阶零矩阵。另外，由于$A = A^2$且A对称，所以$0 = \mu^T A \mu = \mu^T A^2 \mu = (A\mu)^T (A\mu)$。这表明$A\mu = 0$，故

$$\begin{aligned} X^T A X = (X - \mu + \mu)^T A (X - \mu + \mu) &= (X - \mu)^T A (X - \mu) + 2(X - \mu)^T A\mu + \mu^T A \mu \\ &= (X - \mu)^T A (X - \mu). \end{aligned}$$

令$Y = P^T(X - \mu)$。因为P正交，我们有$Y \sim N(0, I)$，即$\{Y_{[i]}; i = 1, \dots, n\}$是相互独立的标准正态随机变量，这里$Y_{[i]}$表示$Y$的第$i$个元素。我们进而知道

$$X^T A X = (X - \mu)^T A (X - \mu) = (X - \mu)^T P D P^T (X - \mu) = Y^T D Y = \sum_{i=1}^{r} Y_{[i]}^2 \sim \chi^2(r).$$

以上第四步用到了(6)。

若$\max\{\mathrm{rank}(A),\mathrm{rank}(B)\}=n$，则$AB=0$表明$A=0$或$B=0$。此时$X^TAX$与$X^TBX$中有一个为0，故$X^TAX\perp X^TBX$。

考虑$\max\{\mathrm{rank}(A),\mathrm{rank}(B)\}<n$的情形。记$r=\mathrm{rank}(A)<n$。因为$A$是实对称矩阵，故存在正交矩阵$P\in\mathbb{R}^{n\times n}$和满秩对角矩阵$\Lambda\in\mathbb{R}^{r\times r}$使得$A=P\mathrm{diga}(\Lambda,0)P^T$，即$\mathrm{diga}(\Lambda,0)=P^TAP$。记$Y=P^TX=(Y_1^T,Y_2^T)^T$，其中$Y_1\in\mathbb{R}^r$、$Y_2\in\mathbb{R}^{n-r}$。因为$P$正交，故$Y\sim N(P^T\mu,\sigma^2I)$。这表明

独立

$$Y_1\perp Y_2. \tag{3}$$

此外，我们有

$$X^TAX=X^TP\mathrm{diga}(\Lambda,0)P^TX=Y^T\mathrm{diga}(\Lambda,0)Y=Y_1^T\Lambda Y_1. \tag{4}$$

记$G=P^TBP=\begin{pmatrix}G_1 & G_2\\ G_2^T & G_3\end{pmatrix}$，这里$G_1\in\mathbb{R}^{r\times r}$、$G_2\in\mathbb{R}^{r\times(n-r)}$、$G_3\in\mathbb{R}^{(n-r)\times(n-r)}$。由$AB=0$可知

$$0=P^TABP=P^TAPP^TBP=\mathrm{diga}(\Lambda,0)G=\begin{pmatrix}\Lambda G_1 & \Lambda G_2\\ 0 & 0\end{pmatrix}.$$

故$\Lambda G_1=\Lambda G_2=0$。结合$\Lambda$满秩的事实，这表明$G_1=G_2=0$，即$G=\mathrm{diag}(0,G_3)$。所以

$$X^TBX=X^TPP^TBPP^TX=Y^TGY=Y^T\mathrm{diag}(0,G_3)Y=Y_2^TG_3Y_2. \tag{5}$$

结合(3)、(4)、(5)，我们有$X^TAX\perp X^TBX$。