Applied Multivariate Statistical Analysis

Chapter 2. A Short Excursion into Matrix Algebra

2.6. Geometrical Aspects—Proofs of Theorems

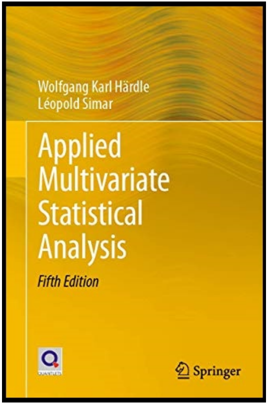

Table of contents

Theorem 2.7

Theorem 2.7. Let $\mathcal{A} = \mathcal{A}(p \times p)$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ and with corresponding orthonormal eigenvectors $\gamma_1, \gamma_2, \ldots, \gamma_p$.

(i) The *principal axes* of iso-distance curve E_d with center x_0 are in the direction of $\pm\gamma_i$ where $i = 1, 2, \ldots, p$.

(ii) The *half-lengths of the axes* are $\sqrt{d^2/\lambda_i}$ where $i = 1, 2, \ldots, p$.

(iii) The rectangle surrounding the ellipsoid E_d is defined by the inequalities:

$$x_{0i} - \sqrt{d^2 a^{ii}} \leq x_i \leq x_{0i} + \sqrt{d^2 a^{ii}} \text{ where } i = 1, 2, \ldots, p$$

where a^{ii} is the (i, i) element of $\mathcal{A}^{-1}$ and x_{0i} is the ith component of x_0. By the *rectangle surrounding the ellipsoid* E_d we mean the rectangle whose sides are parallel to the coordinate axes.

Theorem 2.7 (continued 1)

Proof. We take (i) and (ii) as definitions. We assume $x_0 = 0$ (so that $x_{0i} = 0$ for $i = 1, 2, \ldots, p$) and then the general result holds be translating both E_d and the rectangle surrounding it.

We need to find the coordinates of the tangency points of the ellipsoid and the rectangle surrounding it. With x as such a point (well, *vector* actually), we need x such that

$$\max_{x^\top \mathcal{A} x = d^2} e_j^\top x \text{ for all } j = 1, 2, \ldots, p$$

where $e_j^\top$ is the jth column of the identity matrix $\mathcal{I}_p$ (i.e., the jth standard basis vector for $\mathbb{R}^p$).

Theorem 2.7 (continued 1)

Proof. We take (i) and (ii) as definitions. We assume $x_0 = 0$ (so that $x_{0i} = 0$ for $i = 1, 2, \ldots, p$) and then the general result holds be translating both E_d and the rectangle surrounding it.

We need to find the coordinates of the tangency points of the ellipsoid and the rectangle surrounding it. With x as such a point (well, *vector* actually), we need x such that

$$\max_{x^\top \mathcal{A} x = d^2} e_j^\top x \text{ for all } j = 1, 2, \ldots, p$$

where $e_j^\top$ is the jth column of the identity matrix $\mathcal{I}_p$ (i.e., the jth standard basis vector for $\mathbb{R}^p$). Since x is on E_d (and the center of E_d is 0), then $\|x\|_{\mathcal{A}} = \sqrt{x^\top \mathcal{A} x} = \sqrt{d^2} = d$ or $x^\top \mathcal{A} x = d^2$. Now $e_j^\top x$ is $\|e_j\|\|x\| = \|x\|$ times the cosine of the angle θ between e_j and x (as seen in Linear Algebra; see Section 1.2. The Norm and Dot Product). Now the maximum occurs when $\theta = 0$ and x and e_j are parallel (and the minimum occurs when $\theta = \pi$ and x and e_j are anti-parallel; this corresponds to the other "side" of the rectangle surrounding E_d). [Hmm. . .]

Theorem 2.7 (continued 1)

Proof. We take (i) and (ii) as definitions. We assume $x_0 = 0$ (so that $x_{0i} = 0$ for $i = 1, 2, \ldots, p$) and then the general result holds be translating both E_d and the rectangle surrounding it.

We need to find the coordinates of the tangency points of the ellipsoid and the rectangle surrounding it. With x as such a point (well, *vector* actually), we need x such that

$$\max_{x^\top \mathcal{A} x = d^2} e_j^\top x \text{ for all } j = 1, 2, \ldots, p$$

where $e_j^\top$ is the jth column of the identity matrix $\mathcal{I}_p$ (i.e., the jth standard basis vector for $\mathbb{R}^p$). Since x is on E_d (and the center of E_d is 0), then $\|x\|_{\mathcal{A}} = \sqrt{x^\top \mathcal{A} x} = \sqrt{d^2} = d$ or $x^\top \mathcal{A} x = d^2$. Now $e_j^\top x$ is $\|e_j\|\|x\| = \|x\|$ times the cosine of the angle θ between e_j and x (as seen in Linear Algebra; see Section 1.2. The Norm and Dot Product). Now the maximum occurs when $\theta = 0$ and x and e_j are parallel (and the minimum occurs when $\theta = \pi$ and x and e_j are anti-parallel; this corresponds to the other "side" of the rectangle surrounding E_d). [Hmm. . .]

Theorem 2.7 (continued 2)

Proof (continued). That is, we want to maximize (and minimize) $e_j^\top x$ subject to $x^\top \mathcal{A} x = d^2$. We define the scalar valued functions of vector x of $f(x) = e_j^\top x$ and $g(x) = x^\top \mathcal{A} x - d^2$. Recall that the Method of Lagrange Multipliers allows us to find the extrema of f subject to the constraint $g(x) = 0$ by finding vector x and scalar λ such that $\frac{\partial f(x)}{\partial x} = \lambda \frac{\partial g(x)}{\partial x}$ and $g(x) = 0$; see my online notes for Calculus 3 (MATH 2110) on Section 14.8. Lagrange Multipliers where this is considered for $x \in \mathbb{R}^3$.

Now

$$\frac{\partial f(x)}{\partial x} = \frac{\partial(e_j^\top x)}{\partial x} = e_j \text{ and } \frac{\partial g(x)}{\partial x} = \frac{\partial(x^\top \mathcal{A} x - d^2)}{\partial x} = 2\mathcal{A}x,$$

where the second equality holds by Exercise 2.5 (which is worked as an example in Section 2.4. Derivatives).

Theorem 2.7 (continued 2)

Proof (continued). That is, we want to maximize (and minimize) $e_j^\top x$ subject to $x^\top \mathcal{A} x = d^2$. We define the scalar valued functions of vector x of $f(x) = e_j^\top x$ and $g(x) = x^\top \mathcal{A} x - d^2$. Recall that the Method of Lagrange Multipliers allows us to find the extrema of f subject to the constraint $g(x) = 0$ by finding vector x and scalar λ such that $\frac{\partial f(x)}{\partial x} = \lambda \frac{\partial g(x)}{\partial x}$ and $g(x) = 0$; see my online notes for Calculus 3 (MATH 2110) on Section 14.8. Lagrange Multipliers where this is considered for $x \in \mathbb{R}^3$.

Now

$$\frac{\partial f(x)}{\partial x} = \frac{\partial (e_j^\top x)}{\partial x} = e_j \text{ and } \frac{\partial g(x)}{\partial x} = \frac{\partial (x^\top \mathcal{A} x - d^2)}{\partial x} = 2\mathcal{A}x,$$

where the second equality holds by Exercise 2.5 (which is worked as an example in Section 2.4. Derivatives).

Theorem 2.7 (continued 3)

Proof (continued). So we need to solve the system

$$\frac{\partial f(x)}{\partial x} = \lambda \frac{\partial g(x)}{\partial x} \text{ or } e_j = 2\lambda \mathcal{A} x \tag{2.36}$$

$$g(x) = 0 \text{ or } x^\top \mathcal{A} x - d^2 = 0 \tag{2.37}$$

From (2.36) we have $x = \frac{1}{2\lambda}\mathcal{A}^{-1}e_j$ or, with $\mathcal{A}^{-1} = (a^{ij})$ we have componentwise that $x_i = \frac{1}{2\lambda}a^{ij}$ for $i = 1, 2, \ldots, p$. Multiplying both sides of (2.36) on the left by $x^\top$ gives

$$x^\top(e_j - 2\lambda\mathcal{A}x) = 0 \text{ or } x_j = 2\lambda x^\top \mathcal{A}x \text{ or } x_j = 2\lambda d^2 \text{ for } j = 1, 2, \ldots, p$$

where the last equality holds by (2.37). Hence $x_j = \frac{1}{2\lambda}a^{jj} = 2\lambda d^2$ or $4\lambda^2 = a^{jj}/d^2$ or $2\lambda = \pm\sqrt{a^{jj}/d^2}$.

Theorem 2.7 (continued 3)

Proof (continued). So we need to solve the system

$$\frac{\partial f(x)}{\partial x} = \lambda \frac{\partial g(x)}{\partial x} \text{ or } e_j = 2\lambda \mathcal{A} x \tag{2.36}$$

$$g(x) = 0 \text{ or } x^\top \mathcal{A} x - d^2 = 0 \tag{2.37}$$

From (2.36) we have $x = \frac{1}{2\lambda}\mathcal{A}^{-1}e_j$ or, with $\mathcal{A}^{-1} = (a^{ij})$ we have componentwise that $x_i = \frac{1}{2\lambda}a^{ij}$ for $i = 1, 2, \ldots, p$. Multiplying both sides of (2.36) on the left by $x^\top$ gives

$$x^\top(e_j - 2\lambda\mathcal{A}x) = 0 \text{ or } x_j = 2\lambda x^\top \mathcal{A} x \text{ or } x_j = 2\lambda d^2 \text{ for } j = 1, 2, \ldots, p$$

where the last equality holds by (2.37). Hence $x_j = \frac{1}{2\lambda}a^{jj} = 2\lambda d^2$ or $4\lambda^2 = a^{jj}/d^2$ or $2\lambda = \pm\sqrt{a^{jj}/d^2}$.

Theorem 2.7 (continued 4)

Proof (continued). For the maximum of $e_j^\top x$, we choose $2\lambda = \sqrt{a^{jj}/d^2}$ and for the minimum of $e_j^\top x$ we choose $2\lambda = -\sqrt{a^{jj}/d^2}$. Since componentwise, $x_i = \frac{1}{2\lambda} a^{ij}$ as shown above, then with $i = j$ we have the maximum of $e_j^\top x$ satisfies

$$x_j = \frac{1}{\sqrt{a^{jj}/d^2}} a^{jj} = \sqrt{\frac{d^2}{a^{jj}}} a^{jj} = \sqrt{d^2 a^{jj}} \text{ for } j = 1, 2, \ldots, p.$$

Similarly, the the minimum of $e_j^\top x$ satisfies $x_j = -\sqrt{d^2 a^{jj}}$ for $j = 1, 2, \ldots, p$. So (iii) holds when $x = 0$ and so holds in general by translation, as described above. □

Theorem 2.8

Theorem 2.8. Let $\mathcal{X} = \mathcal{X}(n \times p)$, $\mathcal{P} = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top$, and $\mathcal{Q} = \mathcal{I}_n - \mathcal{P}$. Then

(i) $\{x = \mathcal{P}b \mid b \in \mathbb{R}^n\} \subseteq C(\mathcal{X})$,

(ii) if $y = \mathcal{Q}b$ then $y^\top x = 0$ for all $x \in C(\mathcal{X})$.

Proof. **(i)** For $x = \mathcal{P}b$ where $b \in \mathbb{R}^n$, we have $x = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top b$. Define $a = (\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top b \in \mathbb{R}^p$ and then $\mathcal{X}a = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top b = x$; so x is in the column space $C(\mathcal{X})$. Since x is an arbitrary element of $\{x = \mathcal{P}b \mid b \in \mathbb{R}^n\}$, then (i) follows.

Theorem 2.8

Theorem 2.8. Let $\mathcal{X} = \mathcal{X}(n \times p)$, $\mathcal{P} = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top$, and $\mathcal{Q} = \mathcal{I}_n - \mathcal{P}$. Then

(i) $\{x = \mathcal{P}b \mid b \in \mathbb{R}^n\} \subseteq C(\mathcal{X})$,

(ii) if $y = \mathcal{Q}b$ then $y^\top x = 0$ for all $x \in C(\mathcal{X})$.

Proof. (i) For $x = \mathcal{P}b$ where $b \in \mathbb{R}^n$, we have $x = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top b$. Define $a = (\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top b \in \mathbb{R}^p$ and then $\mathcal{X}a = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top b = x$; so x is in the column space $C(\mathcal{X})$. Since x is an arbitrary element of $\{x = \mathcal{P}b \mid b \in \mathbb{R}^n\}$, then (i) follows.

(ii) Since $\mathcal{Q} = \mathcal{I}_n - \mathcal{P}$, then $y = \mathcal{Q}b = (\mathcal{I}_n - \mathcal{P})b = b - \mathcal{P}b$. Since $x \in C(\mathcal{X})$ then $x = \mathcal{X}a$ for some $a \in \mathbb{R}^p$. Then

$$y^\top x = (b - \mathcal{P}b)^\top \mathcal{X}a = (b^\top - b^\top \mathcal{P}^\top)\mathcal{X}a = b^\top \mathcal{X}a - b^\top \mathcal{P}\mathcal{X}a$$

$$= b^\top \mathcal{X}a - b^\top \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top \mathcal{X}a = b^\top \mathcal{X}a - b^\top \mathcal{X}(\mathcal{I}_n)a = 0,$$

as claimed. □

Theorem 2.8

Theorem 2.8. Let $\mathcal{X} = \mathcal{X}(n \times p)$, $\mathcal{P} = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top$, and $\mathcal{Q} = \mathcal{I}_n - \mathcal{P}$. Then

(i) $\{x = \mathcal{P}b \mid b \in \mathbb{R}^n\} \subseteq C(\mathcal{X})$,

(ii) if $y = \mathcal{Q}b$ then $y^\top x = 0$ for all $x \in C(\mathcal{X})$.

Proof. (i) For $x = \mathcal{P}b$ where $b \in \mathbb{R}^n$, we have $x = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top b$. Define $a = (\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top b \in \mathbb{R}^p$ and then $\mathcal{X}a = \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top b = x$; so x is in the column space $C(\mathcal{X})$. Since x is an arbitrary element of $\{x = \mathcal{P}b \mid b \in \mathbb{R}^n\}$, then (i) follows.

(ii) Since $\mathcal{Q} = \mathcal{I}_n - \mathcal{P}$, then $y = \mathcal{Q}b = (\mathcal{I}_n - \mathcal{P})b = b - \mathcal{P}b$. Since $x \in C(\mathcal{X})$ then $x = \mathcal{X}a$ for some $a \in \mathbb{R}^p$. Then

$$y^\top x = (b - \mathcal{P}b)^\top \mathcal{X}a = (b^\top - b^\top \mathcal{P}^\top)\mathcal{X}a = b^\top \mathcal{X}a - b^\top \mathcal{P}\mathcal{X}a$$

$$= b^\top \mathcal{X}a - b^\top \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{X}^\top \mathcal{X}a = b^\top \mathcal{X}a - b^\top \mathcal{X}(\mathcal{I}_n)a = 0,$$

as claimed. □