Spectral Theory and Fisher-Cochran (Lab 5)

BST 235: Advanced Regression and Statistical Learning
Alex Levis, Fall 2019

1 Motivation and Spectral Theorem

In our discussion of linear models, we saw that in order to perform inference on the regression parameters $\boldsymbol{\beta}(P)$, it would help to be able to characterize the distribution of quadratic forms. In particular, we saw that

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}(P)\|^2 = \boldsymbol{\epsilon}^T A(\mathbb{X})\boldsymbol{\epsilon},$$

where $A(\mathbb{X}) = \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-2}\mathbb{X}^T \in \mathbb{R}^{n\times n}$ is symmetric, and $\boldsymbol{\epsilon} \mid \mathbb{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$. The motivation for spectral decomposition was as follows: if $V = [v_1 \cdots v_n] \in \mathbb{R}^n$ orthogonal (i.e., $V^TV = I_n$), $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ diagonal, and

$$A(\mathbb{X}) = V\Lambda V^T = \sum_{i=1}^{n} \lambda_i v_i v_i^T,$$

then $V^T\boldsymbol{\epsilon} \mid \mathbb{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$ would imply

$$\boldsymbol{\epsilon}^T A(\mathbb{X})\boldsymbol{\epsilon} \mid \mathbb{X} \sim \sigma^2 \sum_{i=1}^{n} \lambda_i \chi^2_{1,i}(0), \text{ where } \chi^2_{1,i}(0) \perp\!\!\!\perp \chi^2_{1,j}(0) \text{ for all } i \neq j.$$

Well, it turns out that $A(\mathbb{X})$ being symmetric guarantees the existence of such a decomposition!

Theorem 1 (Spectral Theorem)**.** Let $A \in \mathbb{R}^{m\times m}$ be a symmetric real matrix. Then there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ (the eigenvalues of A), and an orthonormal basis $v_1, \ldots, v_m \in \mathbb{R}^m$ of $\mathbb{R}^m$, such that

$$A = \sum_{i=1}^{m} \lambda_i v_i v_i^T.$$

Equivalently, for such A, there exists $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^{m\times n}$, $V \in \mathbb{R}^{m\times m}$ with $V^TV = I_m$ such that

$$A = V\Lambda V^T.$$

The correspondence between these equivalent versions is that $V = [v_1 \cdots v_m]$.

Exercise 1. Let $A \in \mathbb{R}^{m\times m}$ be a symmetric matrix with eigenvalues $\lambda_1, \ldots, \lambda_m$. Show that

$$\text{tr}(A) = \sum_{i=1}^{m} \lambda_i.$$

Exercise 2. Let $D = \operatorname{diag}(d_1, \ldots, d_m) \in \mathbb{R}^{m \times m}$. Show that $\operatorname{rank}(D) = \sum_{i=1}^{m} \mathbb{1}(d_i \neq 0)$. That is, the rank of a diagonal matrix is the number of its non-zero diagonal elements.

Exercise 3. Recall from last lab that if $B \in \mathbb{R}^{m \times n}$, and $C \in \mathbb{R}^{n \times n}$ is invertible, then

$$\operatorname{rank}(BC) = \operatorname{rank}(B).$$

Use this and the previous exercise to show that for any symmetric matrix A, its rank is equal to the number of its non-zero eigenvalues.

2 Matrix Square Root

Let $A \in \mathbb{R}^{m \times m}$. We say A is *positive semi-definite* if for all $\mathbf{x} \in \mathbb{R}^m$,

$$\mathbf{x}^T A \mathbf{x} \geq 0.$$

We call A (strictly) *positive definite* if for all $\mathbf{x} \in \mathbb{R}^m$ with $\mathbf{x} \neq \mathbf{0}$,

$$\mathbf{x}^T A \mathbf{x} > 0.$$

Exercise 4. Let $A \in \mathbb{R}^{m \times m}$ be symmetric. Show that A is positive (semi-)definite if and only if the eigenvalues of A are all positive (non-negative).

Let now $A \in \mathbb{R}^{m\times m}$ be a symmetric, positive semi-definite matrix, and let $A = V\Lambda V^T$ be its spectral decomposition, with $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$ the matrix of real non-negative eigenvalues. For such a matrix, by Exercise 4 we can define the *square root* of A as

$$A^{1/2} := V\Lambda^{1/2}V^T,$$

where $\Lambda^{1/2} = \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_m})$. As desired, we have

$$A^{1/2}A^{1/2} = V\Lambda^{1/2}V^TV\Lambda^{1/2}V^T = V\Lambda^{1/2}\Lambda^{1/2}V^T = V\Lambda V^T = A.$$

For an invertible symmetric matrix $A = V\Lambda V^T$, by Exercise 3 we know that all eigenvalues are non-zero, so the inverse of A must be given by $A^{-1} = V\Lambda^{-1}V^T$, where $\Lambda^{-1} = \operatorname{diag}\left(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_m}\right)$. Therefore when A is symmetric, positive semi-definite and invertible (thus strictly positive definite), we can even define

$$A^{-1/2} := V\Lambda^{-1/2}V^T.$$

This matrix is both the square root of A^{-1} and the inverse of $A^{1/2}$. These ideas will be useful in our study of general linear hypothesis testing in the coming weeks.

3 Fisher-Cochran Theorem

In the linear model, the sample least squares coefficients $\widehat{\boldsymbol{\beta}}$ are a linear function of the data $\mathbf{Y}$, and in the homoscedastic setting the maximum likelihood estimate of the variance $\widehat{\sigma}^2$ is a quadratic form in $\mathbf{Y}$. Under an assumption of normality for the data, we seek to understand the distribution of linear functions and quadratic forms of multivariate normal random vectors. This is the content of the Fisher-Cochran theorem, which we proved in class.

Theorem 2 (Fisher-Cochran)**.** Let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_m)$ be standard multivariate normal.

(i) Let $A \in \mathbb{R}^{m\times m}$ be symmetric. Then $\mathbf{Z}^TA\mathbf{Z} \sim \chi^2_r(0)$ if and only if A is idempotent and $\operatorname{rank}(A) = r$.

(ii) Let $A_1, A_2 \in \mathbb{R}^{m\times m}$ be symmetric and idempotent. Then $\mathbf{Z}^TA_1\mathbf{Z} \perp\!\!\!\perp \mathbf{Z}^TA_2\mathbf{Z}$ if and only if $A_1A_2 = \mathbf{0}_{m\times m}$.

(iii) Let $A \in \mathbb{R}^{m\times m}$ be symmetric and idempotent, and let $B \in \mathbb{R}^{k\times m}$. Then $\mathbf{Z}^TA\mathbf{Z} \perp\!\!\!\perp B\mathbf{Z}$ if and only if $AB^T = \mathbf{0}_{m\times k}$.

Exercise 5. Recall the classic probability theory result: if $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, then $\overline{X} \perp\!\!\!\perp S^2$, where $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$. Use the (much more powerful) Fisher-Cochran theorem to prove this fact.

Exercise 6. In proving the 'only if' component of part (ii) in Fisher-Cochran, we needed to argue that if $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_m)$, $A_1, A_2 \in \mathbb{R}^{m\times m}$ symmetric and idempotent, and $\mathbf{Z}^T A_1 \mathbf{Z} \perp\!\!\!\perp \mathbf{Z}^T A_2 \mathbf{Z}$, then $A_1 A_2 = \mathbf{0}_{m\times m}$. We first said that by independence

$$\mathbf{Z}^T(A_1 + A_2)\mathbf{Z} = \mathbf{Z}^T A_1 \mathbf{Z} + \mathbf{Z}^T A_2 \mathbf{Z} \sim \chi^2_{\text{rank}(A_1)+\text{rank}(A_2)}(0).$$

Since $A_1 + A_2$ is symmetric, we used Fisher-Cochran (i) to deduce that $A_1 + A_2$ is idempotent and

$$\text{rank}(A_1 + A_2) = \text{rank}(A_1) + \text{rank}(A_2). \tag{1}$$

In class, we followed a direct but tricky argument to show that $A_1 A_2 = 0$. Now, we outline a linear algebraic argument to come to the same conclusion. Take for granted the following fact: if V is a finite-dimensional vector space, and $U, W \subseteq V$ are linear subspaces, then

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W),$$

where $U + W = \{u + w \,|\, u \in U, w \in W\}$.

(a) Show that $\text{rank}(A_1 + A_2) \leq \dim(\mathcal{C}(A_1) + \mathcal{C}(A_2))$. By the above fact, this will imply

$$\text{rank}(A_1 + A_2) \leq \text{rank}(A_1) + \text{rank}(A_2) - \dim(\mathcal{C}(A_1) \cap \mathcal{C}(A_2)).$$

(b) Combine (1) and (a) to compute $\dim(\mathcal{C}(A_1) \cap \mathcal{C}(A_2))$.

(c) Take for granted that $\mathcal{C}(A_1) = [\mathcal{C}(A_1) \cap \mathcal{C}(A_2)] \oplus \left[\mathcal{C}(A_1) \cap \mathcal{C}(A_2)^{\perp}\right]$.
(Bonus: I am having trouble figuring this out — if you have a proof, let me know!)

(d) Argue from (b) and (c) that $\mathcal{C}(A_1) \subseteq \mathcal{C}(A_2)^{\perp} = \mathcal{N}(A_2^T) = \mathcal{N}(A_2)$.

(e) Conclude that $A_2 A_1 = A_1 A_2 = \mathbf{0}_{m\times m}$.