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

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

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

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

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 $V^T V = I_m$ 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

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

Let A have spectral decomposition $A = U\Lambda U^T = \sum_{i=1}^m \lambda_i u_i u_i^T$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m)$ and $U = [u_1 \cdots u_m]$ is orthogonal. Then

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

Equivalently, $\text{tr}(A) = \text{tr}\left(\sum_{i=1}^m \lambda_i u_i u_i^T\right) = \sum_{i=1}^m \lambda_i \text{tr}(u_i u_i^T) = \sum_{i=1}^m \lambda_i \text{tr}(u_i^T u_i) = \sum_{i=1}^m \lambda_i$, since $u_i^T u_i = \|u_i\|^2 = 1$, for all i.

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

Recall that $\mathrm{rank}(D) = \dim(\mathcal{C}(D))$, and for $e_j^{(m)}$ the j-th canonical basis vector in $\mathbb{R}^m$,

$$\mathcal{C}(D) = \mathscr{L}(d_1 e_1^{(m)}, \ldots, d_m e_m^{(m)}) = \left\{ \sum_{i:d_i\neq 0} \alpha_i e_i^{(m)} \,\middle|\, \alpha_i \in \mathbb{R}, \forall i \text{ such that } d_i \neq 0 \right\},$$

since when $d_i = 0$, the i-th column does not contribute to the span. Therefore,

$$\mathrm{rank}(D) = \dim(\mathcal{C}(D)) = \sum_{i=1}^m \mathbb{1}(d_i \neq 0),$$

as the column space is the span of the $\sum_{i=1}^m \mathbb{1}(d_i \neq 0)$ linearly independent vectors.

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

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

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.

Let $A = U\Lambda U^T$ be the spectral decomposition of A. Then as U and U^T are invertible,

$$\mathrm{rank}(A) = \mathrm{rank}(U\Lambda U^T) = \mathrm{rank}(U\Lambda) = \mathrm{rank}(\Lambda U^T) = \mathrm{rank}(\Lambda).$$

But since Λ is a diagonal matrix with diagonal entries given by the eigenvalues $\lambda_1, \ldots, \lambda_m$ of A, by Exercise 2 we obtain $\mathrm{rank}(A) = \mathrm{rank}(\Lambda) = \sum_{i=1}^m \mathbb{1}(\lambda_i \neq 0)$.

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).

We will show the result for positive semi-definite matrices. Let $A = U\Lambda U^T = \sum_{i=1}^m \lambda_i u_i u_i^T$ be the spectral decomposition of A, where $U = [u_1 \cdots u_m]$. Suppose A is positive semi-definite, then for any $i \in \{1, \ldots, m\}$,

$$0 \leq u_i^T A u_i = (U^T u_i)^T \Lambda U^T u_i = \{e_i^{(m)}\}^T \Lambda e_i^{(m)} = \lambda_i,$$

as $U^T u_i = [u_i^T u_1 \cdots u_i^T u_m]^T = e_i^{(m)}$. Conversely, suppose $\lambda_1, \ldots, \lambda_m \geq 0$. Then we can see that $\mathbf{x}^T A\mathbf{x} = \sum_{i=1}^m \lambda_i (\mathbf{x}^T u_i)^2 \geq 0$, since $(\mathbf{x}^T u_i)^2 \geq 0$.

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

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

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 $\Lambda^{-1} = \text{diag}\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_m}\right)$. 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 $\mathbf{Z}^T A\mathbf{Z} \sim \chi^2_r(0)$ if and only if A is idempotent and $\text{rank}(A) = r$.

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

Exercise 5. Recall the classic probability theory result: if $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, 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.

Let $Z_i = \frac{X_i - \mu}{\sigma}$, for all $i = 1, \ldots, n$, and let $\overline{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$. We can first center, as $\overline{X} = \sigma\overline{Z} + \mu$ and $\frac{n-1}{\sigma^2}S^2 = \sum_{i=1}^n (Z_i - \overline{Z})^2$ — it suffices to show $\overline{Z} \perp\!\!\!\perp \sum_{i=1}^n (Z_i - \overline{Z})^2$, where $Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. To that end, let $\mathbf{Z} = [Z_1 \cdots Z_n]^T \sim \mathcal{N}(\mathbf{0}, I_n)$, and let $\mathbf{1}_n$ denote a column of 1's in $\mathbb{R}^n$. Then $\overline{Z} = \frac{1}{n}\mathbf{1}_n^T\mathbf{Z}$ and $\sum_{i=1}^n (Z_i - \overline{Z})^2 = \|\mathbf{Z} - \overline{Z}\mathbf{1}_n\|^2 = \|\mathbf{Z} - \mathbf{1}_n(\mathbf{1}_n^T\mathbf{1}_n)^{-1}\mathbf{1}_n^T\mathbf{Z}\|^2 = \mathbf{Z}^T(I_n - P_1)\mathbf{Z}$, where $P_1 = \mathbf{1}_n(\mathbf{1}_n^T\mathbf{1}_n)^{-1}\mathbf{1}_n^T$ is symmetric and idempotent. Since

$$(I_n - P_1)\mathbf{1}_n = \mathbf{1}_n - \mathbf{1}_n(\mathbf{1}_n^T\mathbf{1}_n)^{-1}\mathbf{1}_n^T\mathbf{1}_n = \mathbf{1}_n - \mathbf{1}_n = \mathbf{0},$$

$\overline{Z} \perp\!\!\!\perp \sum_{i=1}^n (Z_i - \overline{Z})^2$ by Fisher Cochran part (iii).

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 $A_1A_2 = 0$. 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 $U + W = \{u + w \mid u \in U, w \in W\}$.

(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

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

See that $\mathcal{C}(A_1+A_2) \subseteq \mathcal{C}(A_1)+\mathcal{C}(A_2)$, since $\forall x \in \mathbb{R}^n$, $(A_1+A_2)x = A_1x+A_2x \in \mathcal{C}(A_1)+\mathcal{C}(A_2)$. The result follows by taking the dimension of the two spaces.

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

We obtain $\dim(\mathcal{C}(A_1) \cap \mathcal{C}(A_2)) \leq 0$, so we must have $\dim(\mathcal{C}(A_1) \cap \mathcal{C}(A_2)) = 0$. As a bonus exercise, try showing this fact directly in the following steps: i) $\forall x \in \mathcal{C}(A_j)$, $x = A_j x$, for $j = 1, 2$; ii) $\forall x \in \mathcal{C}(A_1 + A_2)$, $x = (A_1 + A_2)x$; iii) $\mathcal{C}(A_1) \cap \mathcal{C}(A_2) \subseteq \mathcal{C}(A_1 + A_2)$; iv) If $x \in \mathcal{C}(A_1) \cap \mathcal{C}(A_2)$, then $2x = A_1x + A_2x = (A_1 + A_2)x = x$, so $x = \mathbf{0}$.

(c) Argue that $\mathcal{C}(A_1) \subseteq \mathcal{N}(A_2)$.

Combining (b) with the original vector space sum dimension formula gives

$$\dim(\mathcal{C}(A_1) + \mathcal{C}(A_2)) = \text{rank}(A_1) + \text{rank}(A_2) = \text{rank}(A_1 + A_2) = \dim(\mathcal{C}(A_1 + A_2)).$$

But since, $\mathcal{C}(A_1 + A_2) \subseteq \mathcal{C}(A_1) + \mathcal{C}(A_2)$, we must have $\mathcal{C}(A_1 + A_2) = \mathcal{C}(A_1) + \mathcal{C}(A_2)$. Next, let $x \in \mathcal{C}(A_1)$, and note by this vector space equality that $x \in \mathcal{C}(A_1 + A_2)$. As these are projection matrices,

$$A_1x = x = (A_1 + A_2)x = A_1x + A_2x \implies A_2x = \mathbf{0},$$

meaning $x \in \mathcal{N}(A_2)$. Hence, $\mathcal{C}(A_1) \subseteq \mathcal{N}(A_2)$.

(d) Conclude that $A_2A_1 = A_1A_2 = \mathbf{0}_{m\times m}$.

For any $x \in \mathbb{R}^m$, $A_2A_1x = \mathbf{0}$, since $\mathcal{C}(A_1) \subseteq \mathcal{N}(A_2)$ by part (c). It follows that $A_2A_1 = \mathbf{0}_{m\times m}$, and also $A_1A_2 = (A_2A_1)^T = \mathbf{0}^T_{m\times m} = \mathbf{0}_{m\times m}$.

The first version of this exercise used "$\mathcal{C}(A_1) = [\mathcal{C}(A_1) \cap \mathcal{C}(A_2)] \oplus [\mathcal{C}(A_1) \cap \mathcal{C}(A_2)^{\perp}]$", a false statement. Note that this decomposition does not hold in general. A situation where this decomposition works is when one subspace is contained in the other — see Exercise 1 of Lab 6.