Seth Temple
sdtemple@uw.edu

Notes

1 Lectures

This section records key facts presented in lectures in roughly chronological order.

Singular value decomposition. Let $\mathbf{X} \in \mathbb{R}^{n \times p}$. We can write $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, where

- $\mathbf{U}$ is an orthogonal $n \times n$ matrix
- $\mathbf{V}$ is an orthogonal $p \times p$ matrix
- $D_{ij} = 0$ for all $i \neq j$ and in non-decreasing order $D_{ii} \geq 0$ for all $i \leq \min(n, p)$.

Some facts about SVDs are

- A singular value decomposition is unique up to the signs of columns of $\mathbf{U}$ and $\mathbf{V}$
- All matrices have SVDs whereas only symmetric matrices have spectral decompositions
- We can construct compact SVDs.

Subspace. A *subspace* is contained in a larger vector space and is a vector space itself. *Vector spaces* are closed under addition and scalar multiplication. An *orthogonal complement* of a subspace of a vector space is the set of all vectors in the vector space orthogonal to every vector in the subspace. We can decompose $\mathbf{Y} = \mathbf{Y}_{\mathcal{V}} + \mathbf{Y}_{\mathcal{V}^\perp}$. $\hat{\mathbf{Y}} \in \mathbf{Y}_{\mathcal{V}}$ and $\hat{\mathbf{e}} \in \mathbf{Y}_{\mathcal{V}^\perp}$.

Generalized inverse. Let $\mathbf{F} \in \mathbb{R}^{n \times p}$. Then generalized inverse $\mathbf{F}^-$ satisfies $\mathbf{F}\mathbf{F}^-\mathbf{F} = \mathbf{F}$.

- Every matrix has a generalized inverse.
- A matrix can have more than 1 generalized inverse.
- The inverse of an invertible matrix is unique and is a generalized inverse.

Pseudoinverse. For any matrix $\mathbf{F}$, $\exists$ a unique Moore-Penrose inverse $\mathbf{F}^+$ satisfying

- $\mathbf{F}^+$ is a generalized inverse of $\mathbf{F}$
- $\mathbf{F}$ is a generalized inverse of $\mathbf{F}^+$
- $\mathbf{F}\mathbf{F}^+$ and $\mathbf{F}^+\mathbf{F}$ are symmetric

This pseudoinverse is often implemented in computer programs.

Estimability. Consider model $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ where $\mathbb{E}[\varepsilon|\mathbf{X}] = \mathbf{0}$. $a^T\beta$ is estimable if a is in the row space of $\mathbf{X}$.

- For $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{Y}$, $a^T\hat{\beta}$ is unbiased estimator of $a^T\beta$. If $\text{Var}(\varepsilon|\mathbf{X}) = \sigma^2\mathbf{I}_n$, then $\text{Var}(a^T\hat{\beta}|\mathbf{X}) = \sigma^2 a^T(\mathbf{X}^T\mathbf{X})^{-}a$ (exercise 8).
- $a^T\hat{\beta}$ is BLUE if $a^T\beta$ is estimable (Gauss-Markov theorem).
- There are connections to identifiability, defined as $\theta \neq \theta_0 \implies f_\theta \neq f_{\theta_0}$.

Rank deficiency.

- Reduce to full rank.
  - Best. Easiest. Most common.
  - If $\mathbf{X} = \begin{bmatrix}\mathbf{Z}_1 & \mathbf{Z}_2\end{bmatrix}$, columns of $\mathbf{Z}_1$ are linearly independent, and columns of $\mathbf{Z}_2$ are linear combinations of columns of $\mathbf{Z}_1$, then $\hat{\beta} = \begin{bmatrix}(\mathbf{Z}_1^T\mathbf{Z}_1)^{-1}\mathbf{Z}_1^T\mathbf{Y} \\ \mathbf{0}\end{bmatrix}$.
- Use a generalized inverse ($\hat{\beta}$ still satisfies normal equations).
- Impose identifiability constraints.
  - $\mathbf{H}\beta = \mathbf{0}_s$ is an identifiability constraint if
    1. The rows of $\mathbf{H}$ are linearly independent of $\mathbf{X}$
    2. $\text{rank}\left(\begin{bmatrix}\mathbf{X} \\ \mathbf{H}\end{bmatrix}\right) = p$.
  - $\text{rank}(\mathbf{H}) = p - \text{rank}(\mathbf{X})$.
  - $\hat{\beta} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{Z}$, where $\mathbf{W} = \begin{bmatrix}\mathbf{X} \\ \mathbf{H}\end{bmatrix}$, $\mathbf{Z} = \begin{bmatrix}\mathbf{Y} \\ \mathbf{0}\end{bmatrix}$, and $\mathbf{H}$ corresponds to an identifiability constraint, is a unique solution to constrained least squares.

Consistency. The Gauss-Markov theorem is a result that holds for finite samples. We now discuss under which conditions we have asymptotically (weakly) consistent $\hat{\beta}$.

- An estimator $\hat{\theta}$ is consistent for θ if

$$\lim(P(|\hat{\theta} - \theta| < \varepsilon)) = 1,$$

  or, equivalently,

$$\lim(P(|\hat{\theta} - \theta| \geq \varepsilon)) = 0.$$

  Note that $|\hat{\theta} - \theta|$ is a random quantity and $P(\cdot)$ is a deterministic quantity.

- We often argue consistency using Chebyshev's inequality:

$$P\left(\frac{|X-\mu|}{\sigma} \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2},$$

  where X is a random variable with $\mathbb{E}[X] = \mu$ and $\sigma^2 < \infty$, and this inequality holds for any $\varepsilon > 0$.

- $\lim a_n = a$ if for all $\varepsilon > 0$ there exists m such that, for all $n > m$,

$$|a_n - a| < \varepsilon.$$

- Suppose we have a linear model with a full rank design matrix. If $\lambda_{\min}(\mathbf{X}'\mathbf{X}) \to \infty$, then $\hat{\beta} \xrightarrow{p} \beta$.

2 Exercises

This section records the facts presented the in-class exercises in chronological order.

1. Any solution $\hat{\beta}$ to $\arg\min_{\beta} (\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta)$ satisfies that $\mathbf{X}^T\mathbf{X}\hat{\beta} = \mathbf{X}^T\mathbf{Y}$.
2. Let $\mathbf{A} \in \mathbb{R}^{s\times s}$, $\text{rank}(\mathbf{A}) = s$, and $\mathbf{B} \in \mathbb{R}^{s\times t}$. Then, $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{B})$.
3. (a) The columns of $\mathbf{U}$ in the SVD of $\mathbf{X}$ are the eigenvectors of $\mathbf{X}\mathbf{X}^T$.
   (b) The columns of $\mathbf{V}$ in the SVD of $\mathbf{X}$ are the eigenvectors of $\mathbf{X}^T\mathbf{X}$.
   (c) The diagonal elements of $\mathbf{D}$ in the SVD of $\mathbf{X}$ are the square roots of the eigenvalues of $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}\mathbf{X}^T$.
4. (a) $\text{rank}(\mathbf{X}'\mathbf{X}) = \text{rank}(\mathbf{X})$. (Full rank $\mathbf{X}$ is a sufficient condition for LSE to be unique.)
   (b) If $\text{rank}(\mathbf{X}) = p \leq n$, then $\mathbf{X}'\mathbf{X}$ is positive definite. (Full rank $\mathbf{X}$ is sufficient condition for SSE to be strictly convex.)
5. Let $\mathbf{P}_{\mathbf{X}}$ be the projection matrix onto $\mathbf{X}$ where $\mathbf{X} \in \mathbb{R}^{n\times p}$.
   (a) $\mathbf{P}_{\mathbf{X}}$ can be written $\mathbf{U}\mathbf{A}\mathbf{U}'$ using SVD.
   (b) $\mathbf{P}_{\mathbf{X}}$ has eigenvalue 1 of multiplicity p and eigenvalue 0 of multiplicity $n - p$.
   (c) $\text{rank}(\mathbf{P}_{\mathbf{X}}) = p$.
6. Every matrix has a generalized inverse.
7. If $\mathbf{G}$ and $\mathbf{H}$ are generalized inverses of $\mathbf{X}'\mathbf{X}$, then $\mathbf{XGX}' = \mathbf{XHX}'$.

8. For $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ and $\varepsilon|\mathbf{X} \sim (\mathbf{0}, \sigma^2\mathbf{I}_n)$, if $a^T\beta$ is estimable, then $\text{var}(a^T\hat{\beta}|\mathbf{X}) = \sigma^2 a^T(\mathbf{X}^T\mathbf{X})^- a$ where $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{Y}$.

9. Gauss-Markov theorem for full rank $\mathbf{X}$. $a^T\hat{\beta}$ is unique UMVUE for $a^T\beta$.

10. Using Chebyshev's inequality, we show that

$$P(|Y_n - \mu| \geq \delta) \leq \frac{\sigma_n^2}{\delta^2}$$

where $Y_1, \ldots, Y_n$ is a sequence of random variables with indexed variances and common expectation. If $\lim \sigma_n^2 = 0$, then $Y_n \xrightarrow{p} \mu$. We use this exercise to say that, if our estimator's variance goes to zero as the sample gets asymptotically large, then the estimator is asymptotically (weakly) consistent for μ.

3 Homeworks

This section records the facts presented in homeworks in roughly chronological order.

1. For any matrix $\mathbf{A}$, $\mathbf{A}\mathbf{A}' = \mathbf{0}$ implies $\mathbf{A} = 0$.

2. Projection matrices.

   (a) For any matrix $\mathbf{A}$, $\mathbf{P_A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ is a projection matrix onto $\mathcal{C}(\mathbf{A})$.

   (b) $\mathbf{P_A}\mathbf{A} = \mathbf{A}$.

   (c) $\text{rank}(\mathbf{P_A}) = \text{rank}(\mathbf{A})$.

3. Given two OLS estimates of β, $\mathbf{X}\hat{\beta}_1 = \mathbf{X}\hat{\beta}_2$.

4. Consider models $\mathbb{E}[\mathbf{Y}|\mathbf{X}] = \mathbf{1}\alpha_0 + \mathbf{W}\alpha$ and $\mathbb{E}[\mathbf{Y}|\mathbf{X}] = \mathbf{1}\beta_0 + \mathbf{X}\beta$. Suppose $\mathbf{W}$, a column centered version of design matrix $\mathbf{X}$, has full rank $p < n$. Then least squares estimates of α and β are unique and $\hat{\alpha} = \hat{\beta}$.

5. Let $\mathbf{P}$ be a $n \times n$ projection matrix and $\mathbf{R}$ be a $n \times n$ orthogonal matrix.

   - $\mathbf{P}$ is positive semidefinite.
   - If $\text{rank}(\mathbf{P}) = r$, then $\mathbf{P}$ has eigenvalue 1 with multiplicity r and eigenvalue 0 with multiplicity $n - r$.
   - $\mathbf{R}$ has real eigenvalues ± 1.

6. The (unique) least squares estimate is unbiased when the design matrix is full rank.

7. In simple linear regression, $\hat{\beta}_0$ and $\hat{\beta}_1$ are uncorrelated if and only if $\bar{x} = 0$.

4 Potpourri

Lemmas.

- Suppose $\mathbf{AX'X} - \mathbf{BX'X} = \mathbf{0}$. Then $\mathbf{AX'} = \mathbf{BX'}$.
- $\text{trace}(\mathbf{P}) = \text{rank}(\mathbf{P})$ for any projection matrix $\mathbf{P}$.
- Expected value of the residuals is $\mathbf{0}$.
- For our standard LM setup, $\frac{1}{n-\text{rank}(\mathbf{X})}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta})$ is unbiased estimator of $\hat{\sigma}^2$.

Aside. The only full rank projection matrix is the identity matrix.

STAT 512 Facts.

- $\mathbb{E}[\mathbf{Z}^T\mathbf{AZ}] = \text{trace}(\mathbf{A}\text{Var}(\mathbf{Z})) + \mathbb{E}[\mathbf{Z}]^T\mathbf{A}\mathbb{E}[\mathbf{Z}]$.
-