Fall 2013 Statistics 151 (Linear Models) : Lecture Three

Derek Bean

05 September 2013

1 Linear Algebra Review, cont'd

- Result: for matrix A, $\text{rank}(A) + \dim(\mathcal{K}(A)) =$ no. of columns in A.
  - **Definition:** Matrix A is *full rank* if $\text{rank}(A) =$ no. of columns in A (i.e. $\dim(\mathcal{K}(A)) = 0$ and $\mathcal{K}(A) = \{0\}$).
- Invertible matrices
- **Result:** $\mathcal{C}(A^T A) = \mathcal{C}(A^T)$.
  - Implies A is full-rank if and only if inner product matrix $A^T A$ is invertible
  - **Proof:**
    1. $\mathcal{C}(A^T A) \subseteq \mathcal{C}(A^T)$.
       If $v \in \mathcal{C}(A^T A)$ then $v = (A^T A)w$ for some w. Putting $x = Aw$ we see $v = A^T x$, which implies $v \in \mathcal{C}(A^T)$.
    2. $\mathcal{C}(A^T) \subseteq \mathcal{C}(A^T A)$.
       Pick $v \in \mathcal{C}(A^T)$. Then $v = A^T w$ for some vector w. Decompose $w = w' + w''$ where:

$$\begin{aligned} w' &\in \mathcal{K}(A^T) \\ w'' &\in \mathcal{K}(A^T)^\perp = \mathcal{C}(A). \end{aligned}$$

       Since w' is in the null space of A we have $v = A^T w''$. But $w'' = Ax$ for some vector x. So $v = A^T Ax \Rightarrow v \in \mathcal{C}(A^T A)$.

2 Back to the normal equations

Recall that least squares estimates $\hat{\beta}_{LS}$ satisfy:

$$X^T X \hat{\beta}_{LS} = X^T Y.$$

Since $\mathcal{C}(X^T X) = \mathcal{C}(X^T)$, $\hat{\beta}_{LS}$ must always exist.

1. $\hat{\beta}_{LS} = \left(X^T X\right)^{-1} X^T Y$ is the unique least-squares estimate if $X^T X$ is invertible (i.e. X is full rank)
2. If X is not full-rank, there are an infinite number of solutions. For instance, if $\hat{\beta}_{LS}$ is a least squares estimate and k is in the null space of X (we can choose k to be nonzero) then $\hat{\beta}'_{LS} = \hat{\beta}_{LS} + k$ is another least squares estimate (since it also satisfies the normal equations).

3 Identifiability in the linear model

Identifiability is a property of parameters in statistical models. Clearly, it is beneficial to know which parameters we may be able to estimate from the data, and which parameters cannot be estimated from the data. Identifiability helps make this determination. The idea is to study the mapping between parameters and statistical models. We want to know if the mapping is one-to-one. To understand why, suppose θ is a parameter and $\mathbb{P}(x, \theta)$ is a statistical model (a probability distribution). Clearly if $\theta_1 = \theta_2$, then $\mathbb{P}(x, \theta_1) = \mathbb{P}(x, \theta_2)$ for all x. However, if there exist some $\theta_1 \neq \theta_2$ but $\mathbb{P}(x, \theta_1) = \mathbb{P}(x, \theta_2)$, then data distributed according to $\mathbb{P}(\cdot, \theta_1)$ will have the same distribution as data distributed according to $P(\cdot, \theta_2)$. Therefore it is impossible to detect, from the data, whether the parameter of the underlying distribution is θ_1 or θ_2. In such models, then, it is not always possible to estimate the parameter θ.

Linear models are particularly simple: for β to be identifiable it suffices that the mean parameter $\mu(\beta) = \mathbb{E}(Y) = X\beta$ is a one-to-one function of β. This is true as long as $\text{Cov}(\epsilon) = \Sigma$ for some symmetric, positive semi-definite matrix Σ (that does not depend on β!).

Definition (Identifiability of β in the linear model): The parameter β is *identifiable* if $\forall \beta_1, \beta_2$ such that $\mu(\beta_1) = \mu(\beta_2)$ it must be true that $\beta_1 = \beta_2$.

- If X is full-rank, β is identifiable.

  If β_1 and β_2 satisfy $X\beta_1 = X\beta_2$, then $X(\beta_1 - \beta_2) = 0$. Therefore $\beta_1 - \beta_2 \in \mathcal{K}(X)$. But X full rank means $\mathcal{K}(X) = \{0\}$. Therefore $\beta_1 - \beta_2 = 0$ i.e. $\beta_1 = \beta_2$.

- If X is not full-rank, β is not identifiable.

  If X is not full rank then there exists $k \in \mathcal{K}(X)$ with $k \neq 0$. Pick a parameter β_1. Set $\beta_2 = \beta_1 + k$. Then:

  $$X\beta_2 = X(\beta_1 + k) = X\beta_1 + Xk = X\beta_1.$$

  But $\beta_1 \neq \beta_2$.

3.1 Identifiable functions of β

While β itself may not be identifiable, there are functions of β which are identifiable (i.e. determined by the mean parameter $X\beta$, in one-to-one correspondence with $X\beta$, etc.).

Definiton (identifiable functions of β): We say $g(\beta)$ is an $i$dentifiable function of β if $\forall \beta_1$, β_2 such that $\mu(\beta_1) = \mu(\beta_2)$, we must have $g(\beta_1) = g(\beta_2)$.

Theorem (characterization of identifiable functions of β): $g(\beta)$ is an identifiable function of β if and only if g depends on β only through $\mu(\beta)$, i.e. there exists a function g_* such that:

$$g(\beta) = g_*(\mu(\beta)) = g_*(X\beta).$$

3.2 Connection between identifiability and least squares estimation.

- If X is full rank, $X^T X$ is invertible and

  $$\hat{\beta}_{LS} = (X^T X)^{-1} X^T Y$$

  is the unique least squares estimate of the identifiable parameter β; $g(\hat{\beta}_{LS})$ is the unique least squares estimate of $g(\beta)$ for any function g.

- If X is not full rank, there is no unique least squares estimate - which one should we use to estimate β? Moreover, the parameter β is not identifiable, so it is not even clear what parameter a least squares solution $\hat{\beta}_{LS}$ would be estimating!

  However, if $\hat{\beta}_{LS}^{(1)}$ and $\hat{\beta}_{LS}^{(2)}$ are two least squares estimates, then $X\hat{\beta}_{LS}^{(1)} = X\hat{\beta}_{LS}^{(2)}$ (this takes a little more linear algebra to prove). Therefore, the identifiable functions of β have unique least squares solutions. Indeed if $g(\beta) = g_*(X\beta)$, then for any least squares estimate $\hat{\beta}_{LS}$, $g(\hat{\beta}_{LS}) = g_*(X\hat{\beta}_{LS})$ is the unique least squares estimate of $g(\beta)$.

  There is a converse for *linear* functions of β, i.e. $g(\beta) = \lambda^T\beta$ for a set of weights λ. If $\hat{\beta}_{LS}^{(1)}$ and $\hat{\beta}_{LS}^{(2)}$ are two least squares estimates which satisfy $\lambda^T\hat{\beta}_{LS}^{(1)} = \lambda^T\hat{\beta}_{LS}^{(2)}$, then $\lambda = X^T\rho$ for some set of weights ρ. That is, $g(\beta) = \lambda^T\beta = \rho^T X\beta = g_*(X\beta)$, so g is an identifiable linear function of β.