Some Theoretical Results in Chapter 3

STOR 664, Fall 2012

9/10/2012
\( A^T \): transpose of matrix \( A \)

\( \|y\| \): norm of vector \( y \)

\( \langle x, y \rangle \): inner product of vectors \( x \) and \( y \)
Figure 1

\[ e = y - \hat{y} = y - H \hat{y} \]

- "Plane" \( P = \{ \beta, x_1 + \cdots + \beta_p \hat{x}_p : \beta_1, \ldots, \beta_p \in \mathbb{R} \} \)

\( \hat{x}_1, \ldots, \hat{x}_p \): column vectors of the design matrix \( X \)

- \( H = X(X^t X)^{-1} X^t \): the projection (hat) matrix
Consider the regression model
\[ y = X\beta + \epsilon \]
where \( y \) is a random vector \((n \times 1)\), \( X \) is a known constant matrix \((n \times p)\), \( \beta \) is an unknown parameter \((p \times 1)\), and \( \epsilon \) is a random vector \((n \times 1)\) whose components are uncorrelated RV’s with mean 0 and an unknown variance \( \sigma^2 \).

**G-M Thm :** For any \( a \in \mathbb{R}^p \), \( a^t \hat{\beta} \) is the BLUE of \( a^t \beta \), where 
\[ \hat{\beta} = (X^tX)^{-1}X^ty \] is the LSE of \( \beta \).
G-M Thm continued

Proof:
1st, checking unbiasedness: \( E(a^t \hat{\beta}) = a^t E \hat{\beta} = a^t \beta \).
Next, suppose \( b^t y \) is a linear unbiased estimator of \( a^t \beta \), then

\[
a^t \beta = E(b^t y) = b^t E(X \beta + \epsilon) = b^t X \beta, \quad \forall \beta \in \mathbb{R}^p,
\]

which implies \( a^t = b^t X \). Moreover, \( \text{Var}(b^t y) = \|b\|^2 \sigma^2 \).
Let \( H = X(X^t X)^{-1} X^t \) be the projection hat matrix. We have

\[
\text{Var}(a^t \hat{\beta}) = \text{Var}(a^t (X^t X)^{-1} X^t y) = \|a^t (X^t X)^{-1} X^t\|^2 \sigma^2 = \|b^t H\|^2 \sigma^2
\]

\[
= \|H^t b\|^2 \sigma^2 = \|H b\|^2 \sigma^2 \leq \|b\|^2 \sigma^2 = \text{Var}(b^t y),
\]

i.e. \( a^t \hat{\beta} \) has the minimum variance among all linear unbiased estimators for \( a^t \beta \).
Let $e = y - \hat{y}$ be the residual vector. Assume the error vector $\epsilon$ has iid $N(0, \sigma^2)$ components. Then $\frac{\|e\|^2}{\sigma^2}$ follows a (central) $\chi^2$-distribution with degree of freedom $n - p$.

Proof:

\[
e = (I - H) y = (I - H) (X\beta + \epsilon) = (X\beta - HX\beta) + (I - H) \epsilon = (I - H) \epsilon.
\]

Fact:

$I - H$ is an idempotent matrix, i.e. $(I - H)^2 = I - H$ whose eigenvalues are zeros and ones, and

\[
rank(I - H) = trace(I - H) = n - p.
\]
\[ \|e\|^2 = \langle (I - H) \, \epsilon, (I - H) \, \epsilon \rangle = \epsilon^t (I - H)^2 \epsilon = \epsilon^t (I - H) \, \epsilon \]
\[ = \epsilon^t \, U^t \Lambda \, U \, \epsilon, \]

where \( U \) is an orthogonal matrix, and \( \Lambda \) is a diagonal matrix whose main diagonal entries are the eigenvalues of \( I - H \):

\[ \lambda_1 = \cdots \lambda_{n-p} = 1; \quad \lambda_{n-p+1} = \cdots \lambda_n = 0. \]

Note that \( U\epsilon \sim N(0, \sigma^2 I) \) (\( n \)-variate normal distribution). Therefore,

\[ \epsilon^t \, U^t \Lambda \, U \, \epsilon = z_1^2 + \cdots + z_{n-p}^2 \]

where \( z_1, \ldots, z_{n-p} \) are iid \( N(0, \sigma^2) \) RV’s. Hence

\[ \frac{\|e\|^2}{\sigma^2} \sim \chi^2_{n-p}. \]