Chapter 11
Asymptotic Evaluations
F Statistic

Outline

Basic Set-up

For linear models, we begin with a general structure

$$y = X\beta + \epsilon.$$

- y is a matrix whose rows form a series of multivariate measurements, the response variables,
- X is a matrix of explanatory variables,
- β is a matrix of parameters, and
- ϵ is a matrix containing residuals (i.e., errors or noise).

If the residuals have a multivariate normal distribution, then least squares estimation is a maximum likelihood estimation procedure for the β.

Likelihood Ratio Test

Assume that $\beta \in \mathbb{R}^m$ and that X is a $n \times m$ matrix of rank $m < n$. Let $Y_1, \ldots, Y_n$ are independent normally distributed random variables with mean vector $\mu = X\beta$. Then, the likelihood ratio test of the hypothesis

$$H_0 : A\beta = 0 \quad \text{versus} \quad H_1 : A\beta \neq 0.$$

where A is a $r \times m$ matrix has critical region

$$C = \{\mathbf{y}; F(\mathbf{y}) \geq F_0\}.$$

F is given by

$$F(\mathbf{y}) = \frac{\sum_{k=1}^{n}(y_k - \widehat{\widehat{\mu}_k})^2 - \sum_{k=1}^{n}(y_k - \widehat{\mu}_k)^2}{\sum_{k=1}^{n}(y_k - \widehat{\mu}_k)^2} \frac{n-m}{r}.$$

Likelihood Ratio Test

For the expression

$$\sum_{k=1}^{n} (y_k - \mu_k)^2,$$

- the vector $\widehat{\mu}$ is the minimum value under the restriction $\mu = X\beta$, and
- The vector $\widehat{\widehat{\mu}}$ is the minimum value under the pair of restrictions $\mu = X\beta$ and $A\beta = 0$.

Proof. The likelihood function

$$L(\beta, \sigma^2 | \mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp -\frac{1}{2\sigma^2} \sum_{k=1}^{n} (y_k - \mu_k)^2 = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp -\frac{1}{2\sigma^2} (\mathbf{y} - \mu)^T (\mathbf{y} - \mu)$$

Likelihood Ratio Test

The likelihood ratio

$$\Lambda(\mathbf{x},\mathbf{y}) = \frac{\sup\{L(\beta,\sigma^2|\mathbf{x},\mathbf{y}); \mu = X\beta, A\beta = 0\}}{\sup\{L(\beta,\sigma^2|\mathbf{x},\mathbf{y}); \mu = X\beta\}}$$

For the numerator, let $\hat{\beta}$ be the maximum likelihood estimator for the parameter β and let $\widehat{\mu} = X\hat{\beta}$. Then, the the maximum likelihood estimator for σ^2 is

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{k=1}^{n}(y_k - \widehat{\mu}_k)^2 = \frac{1}{n}(\mathbf{y} - \widehat{\mu})^T(\mathbf{y} - \widehat{\mu})$$

Therefore

$$L(\hat{\beta}, \widehat{\sigma^2}|\mathbf{x},\mathbf{y}) = \frac{\exp -\frac{n}{2}}{(2\pi\widehat{\sigma^2})^{n/2}}.$$

Likelihood Ratio Test

Similarly, for the denominator, let $\widehat{\widehat{\mu}} = X\hat{\hat{\beta}}$ and $\widehat{\widehat{\sigma^2}}$ be the corresponding maximum likelihood estimates when the null hypothesis is true. Then,

$$L(\hat{\hat{\beta}}, \widehat{\widehat{\sigma^2}}|\mathbf{x}, \mathbf{y}) = \frac{\exp -\frac{n}{2}}{(2\pi\widehat{\widehat{\sigma^2}})^{n/2}}.$$

Consequently, the likelihood ratio test,

$$\lambda_0 \geq \Lambda(\mathbf{x}, \mathbf{y}) = \left(\frac{\widehat{\sigma^2}}{\widehat{\widehat{\sigma^2}}}\right)^{n/2} \qquad \lambda_0^{-2/n} - 1 \leq \frac{\widehat{\widehat{\sigma^2}}}{\widehat{\sigma^2}} - 1$$

$$\begin{aligned} F(\mathbf{y}) &= (\lambda_0^{-2/n} - 1)\frac{n-m}{r} \leq \left(\frac{\widehat{\widehat{\sigma^2}}}{\widehat{\sigma^2}} - 1\right)\frac{n-m}{r} \\ &= \frac{\sum_{k=1}^n (y_k - \widehat{\widehat{\mu}}_k)^2 - \sum_{k=1}^n (y_k - \widehat{\mu}_k)^2}{\sum_{k=1}^n (y_k - \widehat{\mu}_k)^2} \frac{n-m}{r}. \end{aligned}$$

Test Statistic

We will now show that the F statistic is a constant times the ratio of independent χ^2 random variables.

Property 1. Let $\xi_1, \ldots, \xi_m$ be the columns of X. Then these vectors are linearly independent. In other words, $\mathcal{L}$, the span of $\xi_1, \ldots, \xi_m$ has dimension m.

This follows from the assumption that X has rank m.

Property 2. $\mu \in \mathcal{L}$.

This follows from $\mu = X\beta = \beta_1\xi_1 + \cdots + \beta_m\xi_m$.

Property 3. When H_0 holds, $\mu = \tilde{\beta}_1\eta_1 + \cdots + \tilde{\beta}_{m-r}\eta_{m-r}$ where $\eta_i \in \mathcal{L}$ and $\tilde{\beta}_i$ is one of the components of β. Call $\mathcal{L}_0$ the linear span of the linearly independent vectors $\eta_1, \ldots, \eta_{m-r}$.

The restriction $A\beta = 0$ results in r independent homogenous linear restrictions on the β_i's. Use this to eliminate r of the components of β and let the η_i be the linear combination of the ξ_i resulting from this elimination.

Test Statistic

Property 4, $\mathcal{L}_0 \subset \mathcal{L}$. Choose an orthonormal basis $\alpha_1, \ldots, \alpha_n$ so that $\alpha_1, \ldots, \alpha_m$ is an orthonormal basis for $\mathcal{L}$ and $\alpha_1, \ldots, \alpha_{m-r}$ is an orthonormal basis for $\mathcal{L}_0$. Let P be a matrix whose k-th row is α_k. Then $PP^T = I$, the identity matrix.

$$(PP^T)_{k\ell} = \alpha_k^T \alpha_\ell = \begin{cases} 0 & \text{if } k \neq \ell, \\ 1 & \text{if } k = \ell. \end{cases}$$

Property 5. $\mathbf{y} = \sum_{k=1}^n z_k \alpha_k$ for some scalars $z_1, \ldots, z_m$ and $\mu = \sum_{k=1}^m \nu_k \alpha_k \in \mathcal{L}$.

The follows because $\alpha_1, \ldots, \alpha_n$ is a basis and $\alpha_1, \ldots, \alpha_m$ is a basis for $\mathcal{L}$.

Test Statistic

Property 6. $(\mathbf{y}-\mu)^T(\mathbf{y}-\mu) = \sum_{k=1}^{m}(z_k-\nu_k)^2 + \sum_{k=m+1}^{n} z_j^2$.

$$\mathbf{y}-\mu = \sum_{k=1}^{m}(z_k-\nu_k)\alpha_k + \sum_{k=m+1}^{n} z_k\alpha_k$$

and $(\mathbf{y}-\mu)^T(\mathbf{y}-\mu)$ is the square of the norm of this vector.

Property 7. $\hat{\mu} = \sum_{k=1}^{m} z_k\alpha_k$ and $(\mathbf{y}-\hat{\mu})^T(\mathbf{y}-\hat{\mu}) = \sum_{k=m+1}^{n} z_j^2$.

To minimize $(\mathbf{y}-\mu)^T(\mathbf{y}-\mu)$ over all $\mu \in \mathcal{L}$, make the choice $\nu_k = z_k$.

Property 8. $\hat{\hat{\mu}} = \sum_{k=1}^{m-r} z_k\alpha_k$ and $(\mathbf{y}-\hat{\hat{\mu}})^T(\mathbf{y}-\hat{\hat{\mu}}) = \sum_{k=m-r+1}^{n} z_k^2$

This uses an argument very similar to checking Property 7.

Test Statistic

Combining we find:

Property 9.

$$F = \frac{\sum_{k=m-r+1}^{n} z_k^2 - \sum_{k=m+1}^{n} z_k^2}{\sum_{k=m+1}^{n} z_k^2} \frac{n-m}{r} = \frac{\sum_{k=m-r+1}^{m} z_m^2}{\sum_{k=m+1}^{n} z_k^2} \frac{n-m}{r}.$$

Property 10. Let $Z = PY$, then $Z = (Z_1 \dots, Z_n)$ are n independent normal random variables with mean vector $\nu = P\mu$ and variance σ^2.

For a linear transformation, the Jacobian is simply the determinant. Thus, the density

$$f_Z(\mathbf{z}) = |\det(P^{-1})| f_Y(P^{-1}\mathbf{z}).$$

Because $PP^T = I$, $P^{-1} = P^T$ and $|\det(P^{-1})| = |\det(P)| = 1$.

Test Statistic

Recall that $Y_1, \ldots, Y_n$ are n independent normal random variables with mean vector $\mu = X\beta$ and variance σ^2. Therefore,

$$f_Y(\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp -\frac{1}{2\sigma^2}(\mathbf{y}-\mu)^T(\mathbf{y}-\mu).$$

Also, $(\mathbf{y}-\mu)^T(\mathbf{y}-\mu)$

$$= (P^T\mathbf{z} - P^T\nu)^T(P^T\mathbf{z} - P^T\nu) = (\mathbf{z} - \nu^T)PP^T(\mathbf{z}-\nu) = (\mathbf{z}-\nu)^T(\mathbf{z}-\nu).$$

Property 11 Under the null hypothesis, $\mu \in \mathcal{L}_0$ and so $\nu_k = 0$ for all $k \geq m - r + 1$. Consequently, each of the $Z_j, \ldots, Z_n$ has mean zero and therefore

$$\frac{\sum_{k=m-r+1}^{m} Z_j^2}{\sum_{k=m+1}^{n} Z_j^2}$$

is the ratio of independent χ^2 random variables. The numerator has r degrees of freedom and the denominator has $n - m$ degrees of freedom.

Test Statistic

Recall that a χ^2_ℓ random variable has mean ℓ. The constant factor was chosen so that both the numerator and the denominator each has mean 1.

Property 12. The F statistic can be written in the more compact form

$$F = \frac{\sum_{k=m-r+1}^{m}(\hat{\mu}_k - \hat{\hat{\mu}}_k)^2}{\sum_{k=m+1}^{n}(y_k - \hat{\mu}_k)^2}\frac{n-m}{r}.$$

In this form, we see that the F statistic is the ratio of the between group variance and the within group variance. If this is significantly large, then the between group variance dominates and we reject the null hypothesis. From Properties 7 and 8,

$$\hat{\mu} = \sum_{k=1}^{m} z_k \alpha_k \quad \text{and} \quad \hat{\hat{\mu}} = \sum_{k=1}^{m-r} z_k \alpha_k.$$

$$\hat{\mu} - \hat{\hat{\mu}} = \sum_{k=m-r+1}^{m} z_k \alpha_k \quad \text{and} \quad \sum_{k=m-r+1}^{m} (\hat{\mu}_k - \hat{\hat{\mu}}_k)^2 = \sum_{k=m-r+1}^{m} z_k^2.$$

Test Statistic

The F distribution is said to have two degrees of freedom parameters, n_1 for the numerator and n_2 for the denominator. The density is found using very much the same strategy in finding the density for the t-distribution

$$f_{n_1,n_2}(x) = \frac{1}{B(n_1/2, n_2/2)} \left(\frac{n_1\, x}{n_1\, x + n_2}\right)^{n_1/2} \left(1 - \frac{n_1\, x}{n_1\, x + n_2}\right)^{n_2/2} \frac{1}{x}.$$

Here $B(\alpha, \beta)$ is the beta function.