Statistics 210A Lecture 19 Notes

Daniel Raban

October 28, 2021

1 General Linear Model for Gaussian Hypothesis Tests

1.1 Recap: Canonical linear model for Gaussian hypothesis tests

Last time, we the χ^2 distribution: if $Z_1, \dots, Z_d \overset{\text{iid}}{\sim} N(0,1)$, then $V = \|Z\|^2 \sim \chi^2_d$. We also had the t distribution, where if $Z \sim N(0,1) \amalg V \sim \chi^2_d$, then $Z/\sqrt{V/d} \sim t_d$. We also had the F-distribution, where if $V_1 \sim \chi^2_{d_1} \amalg V_2 \sim \chi^2_{d_2}$, then $\frac{V_1/d_1}{V_2/d_2} \sim F_{d_1,d_2}$.

In our canonical linear model, we had

$$Z = \begin{bmatrix} Z_0 \\ Z_1 \\ Z_r \end{bmatrix} \sim N_n \left(\begin{bmatrix} \mu_0 \\ \mu_1 \\ \mu_r \end{bmatrix}, \sigma^2 I_n \right)$$

where Z_i has dimension d_i with $n - d_0 - d_1 = d_r$. Here, $\mu_0 \in \mathbb{R}^{d_0}$, $\mu_1 \in \mathbb{R}^{d_1}$, $\mu_r \in \mathbb{R}^{d_r}$. We are interesting in testing $H_0 : \mu_1 = 0$ vs $H_1 : \mu_1 \neq 0$. We ended up with 4 cases last time:

	σ^2 known	σ^2 unknown
$d_1 = 1$	$Z_1/\sigma \overset{H_0}{\sim} N(0,1)$	$Z_1/\widehat{\sigma} \overset{H_0}{\sim} t_{n-d}$
$d_1 \geq 1$	$\|Z_1\|^2/\sigma^2 \overset{H_0}{\sim} \chi^2_{d_1}$	$\frac{\|Z_!\|^2/d_1}{\|Z_r\|^2/(n-d)} = \frac{\|Z_1\|^2/d_1}{\widehat{\sigma}^2} \overset{H_0}{\sim} F_{d_1,n-d}$

where $\widehat{\sigma}^2 = \frac{\|Z_r\|^2}{d_r}$.

1.2 General linear model for testing Gaussian means

In the general linear model for testing Gaussian means, we have $Y \sim N_n(\theta, \sigma^2 I_n)$ with $\sigma^2 > 0$. We want to test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta \setminus \Theta_0$, where $\Theta_0 \subseteq \Theta \subseteq \mathbb{R}^n$ are subspaces. Denote $dd_0 = \dim(\Theta_0)$ and $d = \dim(\Theta) = d_0 + d_1$.

Let

$$Q = \begin{bmatrix} Q_0 & Q_1 & Q_r \end{bmatrix},$$

where Q_0 is an orthonormal basis for Θ_0, Q_1 is an orthonormal basis for $\Theta \cap \theta_0^\perp$, and Q_r is an orthonormal basis for $\mathbb{R}^n \cap \Theta^\perp$. Then

$$Z = Q^\top Y \sim N\left(\begin{bmatrix} Q_0^\top \theta \\ Q_1^\top \theta \end{bmatrix}, \sigma^2 I_n\right).$$

In this basis, we are testing $H_0 : Q_1^\top \theta = 0$ vs $H_1 : Q_1^\top \theta \neq 0$.

1.3 Linear regression

Let $x_i \in \mathbb{R}^d$ be fixed, and let $Y_i = x_i^\top \beta + \varepsilon_i$, where $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$. Then $Y \sim N(X\beta, \sigma^2 I_n)$, where

$$X = \begin{bmatrix} - & x_1 & - \\ & \vdots & \\ - & x_d & - \end{bmatrix} = \begin{bmatrix} | & & | \\ X_1 & \cdots & X_n \\ | & & | \end{bmatrix} \in \mathbb{R}^{n \times d}$$

Assume that X has full column rank. Our model is to estimate $\theta = \mathbb{E}[Y] = X\beta$, where $\theta \in \operatorname{span}(X_1, \ldots, X_d)$. Our null hypothesis is $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{d_1} = 0$, where $1 \leq d_1 \leq d$. This is the same as $\theta \in \operatorname{span}(X_{d_1+1}, \ldots, X_d)$ (or $\{0\}$ if $d_1 = d$). In this model, we have $Q_0 = \operatorname{Proj}_{\operatorname{span}(x_{d_1+1}, \ldots, x_d)}$ and $Q_1 = \operatorname{Proj}_{\operatorname{span}(x_1, \ldots, x_d) \cap \Theta_0^\perp}$.

We have

$$\begin{aligned} \|Z_r\| &= \|Y - \operatorname{Proj}_\Theta Y\|^2 \\ &= \|Y - X\widehat{\beta}_{\text{OLS}}\|^2 \\ &= \sum_{i=1}^n (Y_i - x_i^\top \beta)^2, \end{aligned}$$

the residual sum of squares (RSS). Here,

$$\widehat{\beta}_{\text{OLS}} = (X^\top X)^{-1} X^\top Y = \operatorname*{arg\,min}_{\beta \in \mathbb{R}^d} \|Y - X\beta\|^2 = \operatorname*{arg\,min}_{\theta \in \Theta} \|Y - \theta\|^2.$$

Note that

$$\|Z_1\|^2 + \|Z_r\|^2 = \|Y - \operatorname{Proj}_{\Theta_0}(Y)\| = \text{RSS}_0 .$$

The F-statistic is

$$\frac{\|Z_1\|^2/(d - d_0)}{\|Z_r\|^2/(n - d)} = \frac{(\text{RSS}_0 - \text{RSS})/(d - d_0)}{\text{RSS}/(n - d)} \overset{\text{H}_0}{\sim} F_{d - d_0, n - d}.$$

If $d = 1$, let $X_0 = \begin{bmatrix} X_2 & \cdots & X_d \end{bmatrix} \in \mathbb{R}^{d_0 \times n}$. Then let

$$X_{1\perp} = X_1 - \operatorname{Proj}_{\Theta_0}(X_1)$$

$$
\begin{aligned}
&= X_1 - X_0 \underbrace{(X_0^\top X_0)^{-1} X_0^\top X_1}_{\gamma} \\
&= X_1 - X_0\gamma
\end{aligned}
$$

To make X_1 special, write $\theta = X\beta = X_{1\perp}\beta_1 + X_0(\beta_{-1} + \gamma\beta_1) = X_{1\perp}\beta_1 + X_0\delta$. Then

$$
\begin{aligned}
\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\delta} \end{bmatrix} &= ((X_{1\perp} X_0)^\top (X_{1\perp} X_0))^{-1} (X_{1\perp} X_0)^\top Y \\
&= \begin{bmatrix} (X_{1\perp}^\top X_{1\perp})^{-1} & 0 \\ 0 & (X_0^\top X_0)^{-1} \end{bmatrix} \begin{bmatrix} X_{1\perp}^\top Y \\ X_0^\perp Y \end{bmatrix},
\end{aligned}
$$

so

$$
\widehat{\beta}_1 = \frac{X_{1\perp}^\top Y}{\|X_{1\perp}\|^2} = \frac{Z_1}{\|X_{1\perp}\|}.
$$

Here $Q = [q_1] = \frac{X_{1\perp}}{\|X_{1\perp}\|}$, so $Z_1 = q_1^\top Y = \frac{X_{1\perp}^\top Y}{\|X_{1\perp}\|}$.

The variance of the OLS estimator is

$$
\mathrm{Var}(\widehat{\beta}_1) = \mathrm{Var}\left(\frac{Z_1}{\|X_{1\perp}\|}\right) = \frac{\sigma^2}{\|X_{1\perp}\|^2}.
$$

So the standard error of $\widehat{\beta}_1$ is

$$
\mathrm{s.e.}(\widehat{\beta}_1) = \frac{\sigma}{\|X_{1\perp}\|}.
$$

The t-statistic is

$$
\frac{q_1^\top Y}{\sqrt{\mathrm{RSS}/(n-d)}} = \frac{\widehat{\beta}_1}{\widehat{\sigma}/\|X_{1\perp}\|} = \frac{\widehat{\beta}_1}{\widehat{\mathrm{s.e.}}(\widehat{\beta}_1)}.
$$

1.4 One way ANOVA (fixed effect)

ANOVA is short for "analysis of variates."

Our model has $Y_{k,i} \overset{\text{ind}}{\sim} \mu_k + \varepsilon_{k,i}$, where $\varepsilon_{k,i} \overset{\text{iid}}{\sim} N(0, \sigma^2)$ with $k = 1, \ldots, m$ and $i = 1, \ldots, n$. We want to test $H_0 : \mu_1 = \cdots = \mu_m = \mu$ for any $\mu \in \mathbb{R}$. Then the null has dimension $d_0 = 1$, and the whole model has dimension $d = m$. The residual dimension is $d_r = m(n-1)$.

If we concatenate everything into 1 long vector,

$$
Q_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad Q_1 = \text{basis for orthogonal complement of } \mathbf{1}_{mn}.
$$

Denote

$$
\overline{Y}_k = \frac{1}{n}\sum_i Y_{k,i}. S_k^2 = \frac{1}{n-1}\sum_i (Y_{k,i} - \overline{Y}_k)^2,
$$

$$\overline{Y} = \frac{1}{mn}\sum_k\sum_i Y_{k,i}, \qquad S_0^2 = \frac{1}{mn-1}\sum_k\sum_i (Y_{k,i} - \overline{Y})^2.$$

Then

$$\mathrm{RSS} = \sum_k\sum_i (Y_{k,i} - \overline{Y}_k)^2 = (n-1)\sum_k S_k^2 = \|Y\|^2 - n\sum_k \overline{Y}_k^2,$$

$$\mathrm{RSS}_0 \sum_k\sum_i (Y_{k,i} - \overline{Y})^2 = (mn-1)S_0^2 = \|Y\|^2 - mn\overline{Y}^2.$$

The F-statistic is

$$\frac{(\mathrm{RSS}_0 - \mathrm{RSS})/(m-1)}{\mathrm{RSS}\,/(m(n-1))} = \frac{\frac{n}{m-1}(\sum_k \overline{Y}_k^2 - m\overline{Y}^2)}{\frac{1}{m(n-1)}\sum_k\sum_i (Y_{k,i} - \overline{Y}_i)^2} = \frac{\text{between variance}}{\text{within variance}}.$$

An equivalent test statistic would be

$$\frac{\mathrm{RSS}_0 - \mathrm{RSS}}{\mathrm{RSS}_0} = \frac{\|Z_1\|^2}{\|Z_1\|^2 + \|Z_r\|^2}.$$

This is asking "by what percentage does the residual sum of squares goes down?'" or "what fraction of the variance is explained by adding these extra variables to the model?" We reject when the residual variance goes down by a larger than expected percentage.