

STATS305A - Lecture 7

John Duchi
Scribed by Michael Howes

10/12/21

Contents

1	Announcements	1
2	Fisherian Testing	1
2.1	Setting	1
2.2	P-values	2
2.3	Sampling/M-tests	2
2.4	Aside: power of a test	3
3	ANOVA (Analysis of variance)	3
3.1	ANOVA Decomposition	4
3.2	Testing differences	4
3.3	An issue	5

1 Announcements

- Etude 1 due 5pm Thursday sharp
 - On Thursday solutions to the etude will be posted at around 5pm.
 - Students grade their own etudes.
 - Students upload a new etude with corrections made to the original submission. You must upload a revised etude. This second submission is due 5pm Friday.
- Homework 2 will be out soon.

2 Fisherian Testing

2.1 Setting

We propose a null H_0 . We collect data and compute a statistic T_n which is just some function of our data. Under the null H_0 , T_n follows *some* distribution. Call this distribution T (this is not a T -distribution, just the distribution of T_n when the null holds).

We next pick a *level* $\alpha \in (0, 1)$ at which to reject H_0 . We choose a *rejection region* R where

$$\mathbb{P}_{H_0}(T \in R) \leq \alpha.$$

That is, under the null H_0 , T_n is unlikely to fall in R . We then reject if $T_n \in R$.

2.2 P-values

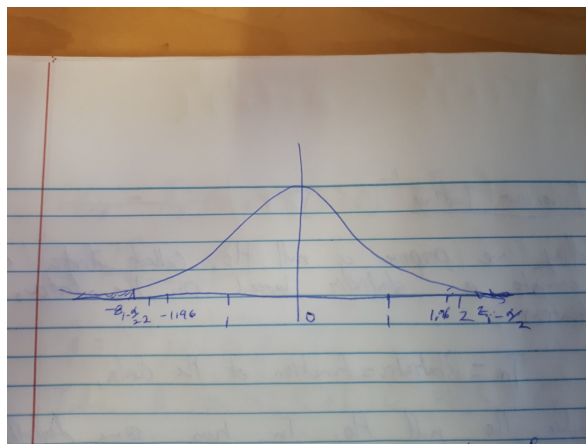
Often our rejection regions are nested. That is if R_α denotes the rejection region at level α , then

$$R_{\alpha_0} \subsetneq R_{\alpha_1}, \text{ if } \alpha_0 < \alpha_1.$$

For example if $T \sim N(0, 1)$ under H_0 , then we often choose

$$R_\alpha = (-\infty, -z_{1-\alpha/2}) \cup (z_{1+\alpha/2}, \infty),$$

where $z_{1-\alpha}$ is the $1 - \alpha$ quantile of a standard normal $Z \sim N(0, 1)$, that is $\mathbb{P}(Z \geq z_{1-\alpha}) = \alpha$. See below



The *p-value* of a statistic T_n is:

- The smallest level α at which we can reject the test ie $p = \inf\{\alpha : T_n \in R_\alpha\}$.
- Equivalently, the probability under H_0 of seeing a sample as strange/extreme as what we have observed.

Example 1. Consider the example above when H_0 implies that T_n is normally distributed. Set $t_n := T_n$ (the observed value), then our p-value is

$$p = \mathbb{P}_{H_0}(|T| \geq |T_n|).$$

Example 2. Assume that $Y = X\beta + \mathbf{1}\beta_0 + \varepsilon$ where $X \in \mathbb{R}^{n \times (d-1)}$ and $Z = [1, X]$. Let $H = Z(Z^T Z)^{-1}Z^T$ = projection onto full model. Let $H_0 = \frac{1}{n}\mathbf{1}\mathbf{1}^T$ = projection onto range of the submodel $Y = \mathbf{1}\beta_0 + \varepsilon$. Our null hypothesis is that the submodel is true. Define $S_n^2 = \frac{1}{n-d} \|(I - H)Y\|_2^2$. Under the null hypothesis we saw that

$$T_n := \frac{\frac{1}{d-1} \|(H - H_0)Y\|_2^2}{\frac{1}{n-d} \|(I - H)Y\|_2^2} \sim F_{d-1, n-d}.$$

We reject the null if T_n is large which means the larger model explains the data better than the smaller model.

2.3 Sampling/M-tests

Suppose that we decide to estimate β via

$$\hat{\beta} = \arg \min_b \sum_{i=1}^n |Y_i - X_i^T b|.$$

Where we assume $Y = X\beta + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$ where we assume σ^2 is known. We wish to test the null hypothesis $\beta = 0$.

Under the null hypothesis we can compute $\varepsilon^{(1)}, \dots, \varepsilon^{(N)}$ independent samples from $\mathcal{N}(0, \sigma^2 I)$ where N is big. We then define $Y^{(i)} := \varepsilon^{(i)}$ and

$$\hat{\beta}^{(i)} = \arg \min_b \sum_{j=1}^n |Y_j^{(i)} - X_j b|.$$

The samples $\hat{\beta}^{(i)}$ gives us an approximation to the distribution of $\hat{\beta}$ assuming that the null is true.

Abstractly all we need is some region in \mathbb{R}^d that contains a $1 - \alpha$ fraction of all $\hat{\beta}^{(i)}$. Then we reject the null if $\hat{\beta}$ is out of that region. Here is one way to do this.

We could compute $q_{1-\alpha}^{sim} = 1 - \alpha$ quartile of $\|\hat{\beta}^{(i)}\|_2$. Thus we have

$$\mathbb{P}_{H_0} \left(\|\hat{\beta}\|_2 \geq q_{1-\alpha}^{sim} \right) \leq \alpha + \frac{1}{N}.$$

This is because under H_0 , $\|\hat{\beta}^{(i)}\|_2$ has the same distribution as $\|\hat{\beta}\|_2$. If we define $t_i = \|\hat{\beta}^{(i)}\|_2$ and $T_n = \|\hat{\beta}\|_2$, then $\|\hat{\beta}\|_2 > q_{1-\alpha}^{sim}$ if and only if $T_n > t_i$ for a $1 - \alpha$ fraction of $\{t_1, \dots, t_n\}$. And $T_n > t_i$ for a $1 - \alpha$ fraction of $\{t_1, \dots, t_n\}$ occurs with probability $\alpha \pm \frac{1}{N}$.

The p-value in this case is

$$\begin{aligned} & \inf\{\alpha : \|\hat{\beta}\|_2 > t_i \text{ for a } 1 - \alpha \text{ fraction of } \{t_i\}_{i=1}^N\} \\ &= \inf\{\alpha : \|\hat{\beta}\|_2 > q_{1-\alpha}^{sim}\} \\ &\approx \text{the fraction of } \{t_i\}_{i=1}^N \text{ such that } \|\hat{\beta}\|_2 \leq t_i. \end{aligned}$$

2.4 Aside: power of a test

Choose an alternative H_1 such as $Y = X\beta + \varepsilon$, $\|\beta\| \geq \sqrt{\frac{d}{n}}$ or $Y = X\beta + \varepsilon$, $\beta_1 > \frac{1}{\sqrt{n}}$.

The *power* of a test is also written as β but now β is a probability not a parameter in a model. The power is defined as

$$\beta := \mathbb{P}_{H_1}(T_n \text{ rejects}).$$

Once could try to maximise the power of a test while keeping the level α constant but this is subtle and complicated. The rejection region with the most power will depend on H_1 . More often in practice we focus on the level and choose the rejection region in a way that reflects the null/alternative hypotheses.

3 ANOVA (Analysis of variance)

Suppose we have the model

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

where $i = 1, \dots, k$ are different groups and $Y_{i,j}$ for $j = 1, \dots, n_i$ are different samples from group i . We call μ the population mean and α_i the group effect. We are interested in testing/estimating the differences $\alpha_i - \alpha_j$. The structure of this model allows us to write cleaner/more direct computations and tests.

3.1 ANOVA Decomposition

Let $\bar{Y}_{\bullet\bullet} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{i,j}$ be the global mean ($N = n_1 + n_2 + \dots + n_k$). Let $\bar{Y}_{i\bullet} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{i,j}$ be the mean for group i . Define

$$\begin{aligned} SST &:= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_{\bullet\bullet})^2 \quad (\text{total sum of squares.}) \\ SSB &:= \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 \quad (\text{between groups sum of squares.}) \\ &= \sum_{i=1}^k n_i (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 \\ SSW &:= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_{i\bullet})^2 \quad (\text{within groups sum of squares.}) \end{aligned}$$

We then have

$$\begin{aligned} SST &= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_{\bullet\bullet})^2 \\ &= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_{i\bullet} + \bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 \\ &= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_{i\bullet})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 \\ &= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_{i\bullet})^2 + \sum_{i=1}^k n_i (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 \\ &= SSW + SSB \end{aligned}$$

This is called the ANOVA decomposition.

Suppose our null is $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$, $\varepsilon_{i,j} \sim N(0, \sigma^2)$. Then (exercise) under H_0 :

- $SSW \perp SSB$.
- $SSW \sim \sigma^2 \chi_{N-k}^2$.
- $SSB \sim \sigma^2 \chi_{k-1}^2$.

Thus $\frac{1}{N-k} \frac{SSB}{SSW} \sim F_{k-1, N-k}$.

We reject H_0 when the above statistic is large which means the between group differences are large relative to the within-group differences.

3.2 Testing differences

Often we may care more about differences in between the mean treatments. For example we may be dosing a drug at different levels $i = 1, \dots, k$. We care more about $\alpha_i - \alpha_j$ more than just “is there a difference in treatment from $\alpha_1 = \alpha_2 = \dots = \alpha_k = 0$?”

This gives rise to testing *contrasts* which are vectors $c \in \mathbb{R}^k$ satisfying $c^T \mathbf{1} = 0$. For example $c = e_i - e_j$. If $\hat{\alpha}$ is any least squares solution then $\frac{c^T \hat{\alpha}}{\sqrt{c^T (X^T X)^{\dagger} c}} \sim T_{n-(d-1)}$ (T -distribution), where

$$S^2 = \frac{1}{n-(d-1)} \sum_{i=1}^n (Y_i - X_i^T \hat{\beta})^2.$$

Exercise: If $c = e_i - e_j$ (ie $c^T \beta = \alpha_i - \alpha_j$), then $c^T (X^T X)^{\dagger} c = \frac{1}{n_i} + \frac{1}{n_j}$.

3.3 An issue

If we look at all pairs of differences $\alpha_i - \alpha_j$ for $i < j$. Then we are doing $\frac{k(k-1)}{2}$ tests. If we reject the nulls at a level α , then we “expect” to have roughly $\alpha \frac{k^2}{2}$ false rejections. Note that $\alpha \frac{k^2}{2} \gg \alpha$. This is the issue of *multiple testing*.