Cochran's Theorem

Paweł Polak

October 9, 2017

Linear Regression Models - Lecture 6

Importance of Cochran's Theorem

- Cochran's theorem tells us about the distributions of partitioned sums of squares of normally distributed random variables.
- Traditional linear regression analysis relies upon making statistical claims about the distribution of sums of squares of normally distributed random variables (and ratios between them)
- In the Simple Normal Regression model:

$$\frac{SSE}{\sigma^2} = \frac{\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}{\sigma^2} \sim \chi^2\left(N-2\right)$$

Where does this come from?

Outline

- Establish the fact that the multivariate Gaussian sum of squares is $\chi^2(N)$ distributed.
- Provide intuition for Cochran's theorem.
- Prove a lemma in support of Cochran's theorem.
- Prove Cochran's theorem.
- Connect Cochran's theorem back to matrix linear regression.

χ^2 Distribution

Theorem

Suppose Z_i are i.i.d. $N(0,1)$, for $i = 1, \dots, N$, then

$$\sum_{i=1}^{N} Z_i^2 \sim \chi^2(N).$$

Proof.

- $Z_i^2 \sim \chi^2(1)$ with the moment generating functions (MGF)

$$m_{Z_i^2}(t) := \mathbb{E}[e^{tZ_i^2}] = (1-2t)^{-1/2} \quad \text{for} \quad t < 1/2.$$

- If $Y_1; \dots; Y_N$ are *i.i.d.* random variables with MGF $m_{Y_1}(t); \dots; m_{Y_N}(t)$, then the moment generating function for $U = Y_1 + \dots + Y_N$ is

$$m_U(t) = m_{Y_1}(t) m_{Y_2}(t) \dots m_{Y_N}(t).$$

- MGF fully characterizes the distribution.
- The MGF for $\chi^2(N)$ is $(1-2t)^{-N/2}$.
- So $\sum_{i=1}^{N} Z_i^2 \sim \chi^2(N)$.

Quadratic Forms and Cochran's Theorem

- Quadratic forms of normal random variables are of great importance in many branches of statistics
  - Least Squares
  - ANOVA
  - Regression Analysis
- General idea: Split the sum of the squares of observations into a number of quadratic forms where each corresponds to some cause of variation.
- The conclusion of Cochran's theorem is that, under the assumption of normality, the various quadratic forms are independent and χ^2 distributed.
- This fact is the foundation upon which many statistical tests rest.

Preliminaries: A Common Quadratic Form

- Let
$$\mathbf{X} \sim \mathrm{N}\left(\boldsymbol{\mu}, \boldsymbol{\Lambda}\right).$$

- Consider the quadratic form that appears in the exponent of the normal density
$$(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\mathbf{X} - \boldsymbol{\mu}).$$
- In the special case of $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Lambda} = \mathbf{I}$, this reduces to $\mathbf{X}^T\mathbf{X}$ for which we just proved that it is $\chi^2(N)$ distributed.
- In the general case we know that
$$(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1/2} \sim \mathrm{N}\left(\mathbf{0}, \mathbf{I}\right).$$

- Therefore
$$(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(N).$$

Cochrane's Theorem

Theorem

Let $X_1, \ldots, X_N$ be i.i.d. $N(0, \sigma^2)$ distributed random variables, and suppose that

$$\sum_{i=1}^{N} X_i^2 = Q_1 + Q_2 + \ldots + Q_K,$$

where $Q_1, Q_2, \ldots, Q_K$ are positive semi-definite quadratic forms in $X_1, X_2, \ldots, X_K$, i.e.,

$$Q_i = \boldsymbol{X}^T \mathbf{A}_i \boldsymbol{X}; i = 1, 2, \ldots, K$$

Set $r_i = rank(A_i)$. If $r_1 + r_2 + \ldots + r_K = N$, then

(1) *$Q_1, Q_2, \ldots, Q_K$ are independent, and*

(2) *$Q_i \sim \sigma^2 \chi^2(r_i)$.*

Several Linear Algebra Results

- $\mathbf{X}$ be a normal random vector. The components of $\mathbf{X}$ are independent, if and only if they are uncorrelated.
- Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, then $\mathbf{Y} = \mathbf{C}^T\mathbf{X} \sim \mathrm{N}\left(\mathbf{C}^T\boldsymbol{\mu}, \mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}\right)$:
  - We can find an orthogonal matrix $\mathbf{C}$, i.e., $\mathbf{C}^T = \mathbf{C}^{-1}$, such that $\mathbf{D} = \mathbf{C}^T\boldsymbol{\Lambda}\mathbf{C}$ is a diagonal matrix. (Eigen Value Decomposition for Semi Positive Definite Matrix).
  - The components of $\mathbf{Y}$ will be independent and $\mathrm{Var}(Y_k) = \lambda_k$, where $\lambda_1, \ldots \lambda_K$; λ_k are the eigenvalues of $\boldsymbol{\Lambda}$.

Lemma

Lemma

Let $X_1, X_2, \ldots, X_N$ be real numbers. Suppose that $\sum_{i=1}^{N} X_i^2$ can be split into a sum of positive semi-definite quadratic forms, that is

$$\sum_{i=1}^{N} X_i^2 = Q_1 + Q_2 + \ldots + Q_K,$$

where $Q_i = \mathbf{X}^T \mathbf{A}_i \mathbf{X}$ with $\mathbf{A}_i$ square $N \times N$ positive semidefinite matrices with $\mathrm{rank}(\mathbf{A}_i) = r_i$. If $\sum_{i=1}^{N} r_i = N$, then there exists an orthogonal matrix $\mathbf{C}$ such that, with $\mathbf{X} = \mathbf{CY}$, we have

$$\begin{aligned} Q_1 &= Y_1^2 + Y_2^2 + \ldots + Y_{r_1}^2 \\ Q_2 &= Y_{r_1+1}^2 + Y_{r_1+2}^2 + \ldots + Y_{r_2}^2 \\ &\vdots \\ Q_K &= Y_{N-r_K+1}^2 + Y_{N-r_K+2}^2 + \ldots + Y_N^2 \end{aligned}$$

Remark

- Different quadratic forms contain different $\mathbf{Y}$ -variables and that the number of terms in each Q_i equals that rank, r_i , of Q_i
- The Y_i^2 end up in different sums, we'll use this to prove independence of the different quadratic forms.
- Just prove for $K = 2$ case, the general case can be obtained by induction.

Proof of Lemma for $K = 2$ (part 1)

Proof of Lemma for $K = 2$ (part 1):

- For $K = 2$, we have $\sum_{i=1}^{N} X_i^2 = \mathbf{Q}$, where $\mathbf{Q} = Q_1 + Q_2 = \mathbf{X}^T\mathbf{A}_1\mathbf{X} + \mathbf{X}^T\mathbf{A}_2\mathbf{X}$.
- By assumption there exists an orthogonal matrix $\mathbf{C}$ such that $\mathbf{C}^T\mathbf{A}_1\mathbf{C} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix with eigenvalues of $\mathbf{A}_1$.
- Since $rank(\mathbf{A}_1) = r_1$, r_1 eigenvalues are positive and $N - r_1$ eigenvalues are 0.
- Suppose without loss of generality, the first r_1 eigenvalues are positive.
- Define $\mathbf{Y} = \mathbf{C}^T\mathbf{X}$, then we have $\mathbf{X} = \mathbf{C}\mathbf{Y}$ and $\mathbf{X}^T\mathbf{X} = \mathbf{Y}^T\mathbf{C}^T\mathbf{C}\mathbf{Y} = \mathbf{Y}^T\mathbf{Y}$.
- Therefore, $\sum_{i=1}^{N} Y_i^2 = \mathbf{Q}$ and

$$\mathbf{Q} = \underbrace{\mathbf{X}^T\mathbf{A}_1\mathbf{X}}_{=\mathbf{X}^T\mathbf{C}\mathbf{D}\mathbf{C}^T\mathbf{X}} + \mathbf{X}^T\mathbf{A}_2\mathbf{X} = \mathbf{Y}^T\mathbf{D}^T\mathbf{Y} + \mathbf{X}^T\mathbf{A}_2\mathbf{X} = \sum_{i=1}^{r_1} \lambda_i \mathbf{Y}_i^2 + \mathbf{Y}^T\mathbf{C}^T\mathbf{A}_2\mathbf{C}\mathbf{Y}.$$

- Then, rearranging the terms we have

$$\sum_{i=1}^{r_1} (1 - \lambda_i)\, Y_i^2 + \sum_{i=r_1+1}^{N} Y_i^2 = \mathbf{Y}^T\mathbf{C}^T\mathbf{A}_2\mathbf{C}\mathbf{Y}$$

□

Proof of Lemma for $K = 2$ (part 2)

Proof of Lemma for $K = 2$ (part 2):

- Recall from the last slide

$$\sum_{i=1}^{r_1} (1 - \lambda_i) \, Y_i^2 + \sum_{i=r_1+1}^{N} Y_i^2 = \mathbf{Y}^T \mathbf{C}^T \mathbf{A}_2 \mathbf{C} \mathbf{Y}$$

- Since $rank(\mathbf{A}_2) = r_2 = N - r_1$
  - Therefore, the above holds if and only if

$$\lambda_1 = \lambda_2 = \ldots = \lambda_{r_1} = 1$$

  - and

$$Q_1 = \sum_{i=1}^{r_1} Y_i^2, \quad Q_2 = \sum_{i=r_1+1}^{N} Y_i^2.$$

$\square$

About Lemma

- This lemma is about real numbers, not random variables
- It says that if $\sum_{i=1}^{N} X_i^2$ can be split into a sum of positive semi-definite quadratic forms, then there is a orthogonal transformation $\mathbf{X} = \mathbf{CY}$ such that each of the quadratic forms have nice properties: Each Y_i appears in only one resulting sum of squares, which leads to the independence of the sum of squares.

Proof of Cochrane's Theorem:

- Using the Lemma, $Q_1, \ldots, Q_k$ can be written using different Y_i's, therefore, they are independent.
- Furthermore, $Q_1 = \sum_{i=1}^{r_1} Y_i^2 \sim \sigma^2 \chi^2(r_1)$.
- Other Q_i's are the same.

$\square$

Applications

- Sample variance is independent from sample mean.
- Recall $SSTO = (N-1)s^2(Y)$,
  $SSTO = \sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{N} Y_i^2 - \frac{(\sum_{i=1}^{N} Y_i)^2}{N}$
- Rearrange the term and express in matrix format

$$\sum_{i=1}^{N} Y_i^2 = \sum_{i=1}^{N}(Y_i - \bar{Y})^2 + \frac{(\sum_{i=1}^{N} Y_i)^2}{N}$$

$$\mathbf{Y}^T\mathbf{I}\mathbf{Y} = \mathbf{Y}^T(\mathbf{I} - \frac{1}{N}\mathbf{J})\mathbf{Y} + \mathbf{Y}^T(\frac{1}{N}\mathbf{J})\mathbf{Y}$$

- We know $\mathbf{Y}^T\mathbf{I}\mathbf{Y} \sim \sigma^2\chi^2(N)$, $rank(\mathbf{I} - \frac{1}{N}\mathbf{J}) = N-1$ (see the next slide), and $rank(\frac{1}{N}\mathbf{J}) = 1$.
- As results

$$\sum_{i=1}^{N}(Y_i - \bar{Y})^2 \sim \sigma^2\chi^2(N-1)$$

$$\frac{(\sum_{i=1}^{N} Y_i)^2}{N} \sim \sigma^2\chi^2(1)$$

Rank of $\mathbf{I} - \frac{1}{N}\mathbf{J}$

- $$rank(\mathbf{I} - \frac{1}{N}\mathbf{J}) \geq rank(\mathbf{I}) - rank(\frac{1}{N}\mathbf{J}) = N - 1$$
- Since $(\mathbf{I} - \frac{1}{N}\mathbf{J})\mathbf{1} = \mathbf{0}$, we have
$$rank(\mathbf{I} - \frac{1}{N}\mathbf{J}) \leq N - 1$$
Therefore
$$rank(\mathbf{I} - \frac{1}{N}\mathbf{J}) = N - 1$$

ANOVA

$$SSTO = \mathbf{Y}^T[\mathbf{I} - \frac{1}{N}\mathbf{J}]\mathbf{Y}$$

$$SSE = \mathbf{Y}^T[\mathbf{I} - \mathbf{H}]\mathbf{Y}$$

$$SSR = \mathbf{Y}^T[\mathbf{H} - \frac{1}{N}\mathbf{J}]\mathbf{Y}$$

- Under the null hypothesis, when $\beta = 0$, we know

$$SSTO \sim \sigma^2\chi^2(N-1)$$

- From linear algebra: $rank(\mathbf{I} - \mathbf{H}) = N - P$ and $rank(\mathbf{H} - \frac{1}{N}\mathbf{J}) = P - 1$.
- Then we have

$$SSE \sim \sigma^2\chi^2(N-P)$$

$$SSR \sim \sigma^2\chi^2(P-1)$$

- As a byproduct, $MSE = SSE/(N-P)$ is an unbiased estimator of σ^2, since the mean of $\chi^2(N-P) = N - P$.

Rank of $\mathbf{I} - \mathbf{H}$

We have

$$\begin{aligned} trace(\mathbf{H}) &= trace(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) \\ &= trace(\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}) \\ &= trace(\mathbf{I}_P) \\ &= P \end{aligned}$$

Then

$$\begin{aligned} rank(\mathbf{I} - \mathbf{H}) &= trace(\mathbf{I} - \mathbf{H}) \\ &= trace(\mathbf{I}) - trace(\mathbf{H}) \\ &= N - P \end{aligned}$$

Rank of $\mathbf{H} - \frac{1}{N}\mathbf{J}$

- First, since we have $\mathbf{H1} = \mathbf{1}$ (This amounts to do a multiple linear regression with the response always equal to 1 and therefore, the fitted value is still 1 because we can just use the constant to perfectly fit the model), then it is straightforward to check that $\mathbf{H} - \frac{1}{N}\mathbf{J}$ is an idempotent and symmetric matrix.

$$\begin{aligned}
\mathbf{H} - \frac{1}{N}\mathbf{J} &= (\mathbf{H} - \frac{1}{N}\mathbf{J})(\mathbf{H} - \frac{1}{N}\mathbf{J}) \\
&= \mathbf{HH} - \frac{2}{N}\mathbf{JH} + \frac{1}{N^2}\mathbf{JJ} \\
&= \mathbf{H} - \frac{2}{N}\mathbf{J} + \frac{N}{N^2}\mathbf{J} \\
&= \mathbf{H} - \frac{1}{N}\mathbf{J}
\end{aligned}$$

- Then, we have

$$rank(\mathbf{H} - \frac{1}{N}\mathbf{J}) = trace(\mathbf{H}) - trace(\frac{1}{N}\mathbf{J}) = P - 1$$

Some Results on Idempotent Matrices

- All eigenvalues of an idempotent matrix are equal to 0 or 1.

$$\mathbf{H}\mathbf{v} = \lambda\mathbf{v}.$$

Multiply both sides by $\mathbf{H}$ from the left to get

$$\mathbf{H}\mathbf{H}\mathbf{v} = \lambda\mathbf{H}\mathbf{v}$$

Since $\mathbf{H}$ is idempotent,

$$\mathbf{H}\mathbf{v} = \lambda\mathbf{H}\mathbf{v}$$

Combining we get

$$\lambda\mathbf{H}\mathbf{v} = \lambda\lambda\mathbf{v} = \lambda\mathbf{v}$$

So $\lambda = 0$ or $\lambda = 1$.

- Rank of an idempotent matrix is equal to its trace.

$$trace(\mathbf{H}) = \sum_{i=1}^{N} h_{i,i} = \sum_{i=1}^{N} \lambda_i = rank(\mathbf{H})$$