Regression

The Linear Model

Bruce M. Boghosian

Tufts UNIVERSITY | School of Arts and Sciences

Department of Mathematics

Tufts University

Outline

Bruce M. Boghosian

Least squares with linear algebra

- Some of you have learned an alternative linear algebraic approach to the method of least squares.
- Write the equations as follows

$$\begin{aligned} a + bx_1 &= y_1 \\ a + bx_2 &= y_2 \\ &\vdots \\ a + bx_n &= y_n \end{aligned}$$

- Cast in rectangular matrix form $Ax = y$, where

$$\underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{y}$$

- n equations for two unknowns – generally no solution

Least squares with linear algebra (continued)

- We can not generally solve $Ax = y$.
- So instead we minimize

$$\begin{aligned}\|Ax - y\|^2 =& (Ax - y)^T (Ax - y)\\ =& (x^T A^T - y^T)(Ax - y)\\ =& x^T A^T Ax - 2x^T A^T y + y^T y\end{aligned}$$

- Minimize with respect to components of x to obtain

$$2A^T Ax - 2A^T y = 0.$$

- Result is

$$\left(A^T A\right) x = A^T y$$

- Note that $A^T A$ is a 2×2 square matrix

Least squares with linear algebra (continued)

- Write out the result $\left(A^T A\right) x = A^T y$

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- But the above is equivalent to

$$\begin{pmatrix} n & \sum_i^n x_i \\ \sum_i^n x_i & \sum_i^n x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_i^n y_i \\ \sum_i^n x_i y_i \end{pmatrix}$$

which is exactly what we derived last time.

Motivation

- We studied least-squares fits from a geometric perspective.
- Our points (x_i, y_i) were given.
- The coordinates x_i and y_i were numbers.
- We found polynomial that minimizes least-squares error L.

Motivation (continued)

- More likely, horizontal and/or vertical coordinates are r.v.s.
- Example:
  - We might consider (x_i, Y_i), so Y_i is a r.v.
  - For each given x_i there is a distribution for Y_i.
- Example:
  - We might consider (X_i, Y_i), so both are r.v.s.
  - For each i, there is a bivariate distribution of X_i and Y_i.

- **Def.:** Let $f_{Y|x}(y)$ be the pdf of r.v. Y for a given value x, and let $E(Y|x)$ denote the expected value associated with $f_{Y|x}(y)$. The function

$$y = E(Y|x)$$

is called the *regression curve* of Y on x.

Example of regression curve

- Suppose that, for $0 \le x \le 1$ and $0 \le y \le 1$,

$$f_{Y|x}(y) = 2\left(\frac{1-x-y+3xy}{1+x}\right).$$

- Note that this is normalized,

$$\int_0^1 dy\ f_{Y|x}(y) = 2\int_0^1 dy\ \left(\frac{1-x-y+3xy}{1+x}\right) = 1.$$

- Also note that

$$E(Y|x) = \int_0^1 dy\ f_{Y|x}(y)y = 2\int_0^1 dy\ \left(\frac{1-x-y+3xy}{1+x}\right)y = \frac{x+1/3}{x+1}.$$

Example of regression curve (continued)

- For $0 \leq x \leq 1$ and $0 \leq y \leq 1$, we have

$$f_{Y|x}(y) = 2\left(\frac{1-x-y+3xy}{1+x}\right) \quad \text{and} \quad E(Y|x) = \frac{x+1/3}{x+1}.$$

- This means, for example, that when $x = 1/2$, we have

$$f_{Y|1/2}(y) = \frac{2}{3}(1+y) \quad \text{and} \quad E(Y|1/2) = \frac{5}{9}.$$

- One can check that, as expected,

$$E(Y|1/2) = \int_0^1 dy\ f_{Y|1/2}(y)y = \int_0^1 dy\ (2y)y = \frac{2}{3}.$$

Definition of The Linear Model

- Special case:
  - $f_{Y|x}(y)$ is a normal pdf for all x.
  - The standard deviation of $f_{Y|x}(y)$, σ, is the same for all x.
  - The means of the conditional Y distributions are collinear,

$$y = E(Y|x) = \beta_0 + \beta_1 x.$$

  - All conditional distributions represent independent r.v.s.
- β_0, β_1, and σ^2 are unknown and should be estimated.

Parameter estimation in The Linear Model

- We have $f_{Y|x}(y) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{y-\beta_0-\beta_1 x}{\sigma}\right)^2\right]$
- Likelihood is then

$$L = \prod_{i=1}^{n} f_{Y_i|x_i}(y_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{y_i-\beta_0-\beta_1 x_i}{\sigma}\right)^2\right]$$
$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-\beta_0-\beta_1 x_i}{\sigma}\right)^2\right]$$

- Log likelihood is then

$$-2\ln L = n\ln(2\pi) + n\ln\left(\sigma^2\right) + \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2$$

Maximize log likelihood

- Find maximum with respect to β_0, β_1 and σ^2

$$0 = \frac{\partial(-2\ln L)}{\partial\beta_0} = \frac{2}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(-1)$$

$$0 = \frac{\partial(-2\ln L)}{\partial\beta_1} = \frac{2}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(-x_i)$$

$$0 = \frac{\partial(-2\ln L)}{\partial\sigma^2} = \frac{n}{\sigma^2} - \frac{2}{\sigma^4}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

- Result for estimators is (where $\hat{Y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i$)

$$\hat{\beta}_1 = \frac{n\sum_i^n x_i Y_i - \left(\sum_i^n x_i\right)\left(\sum_i^n Y_i\right)}{n\left(\sum_i^n x_i^2\right) - \left(\sum_i^n x_i\right)^2}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{x}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i^n\left(Y_i - \hat{Y}_i\right)^2$$

A theorem from probability

Bruce M. Boghosian

- Let $Y_1, \ldots, Y_n$ be any set of independent random variables with means $\mu_1, \ldots, \mu_n$ and variances $\sigma_1^2, \ldots, \sigma_n^2$, respectively. Let $a_1, \ldots, a_n$ be any set of constants. Then $Y = a_1 Y_1 + \cdots + a_n Y_n$ is normally distributed with mean $\mu = \sum_i^n a_i \mu_i$ and variance $\sigma^2 = \sum_i^n a_i^2 \sigma_i^2$.
- Proof is straightforward using moment-generating functions.

Theorems about Linear Model estimators

Bruce M. Boghosian

- **Thm.:** $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ are both normally distributed.
- **Thm.:** $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ are both unbiased,

$$E\left(\hat{\boldsymbol{\beta}}_0\right) = \hat{\beta}_0$$
$$E\left(\hat{\boldsymbol{\beta}}_1\right) = \hat{\beta}_1$$

- **Thm.:** The variances of the estimators are

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_0\right) = \frac{\sigma^2 \sum_i^n x_i^2}{n \sum_i^n (x_i - \overline{x})^2}$$
$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_1\right) = \frac{\sigma^2}{n \sum_i^n (x_i - \overline{x})^2}$$

Proof of normal distribution

- **Pf. (for $\hat{\beta}_1$):** Note that

$$\begin{aligned}\hat{\beta}_1 &= \frac{n\sum_i^n x_i Y_i - \left(\sum_i^n x_i\right)\left(\sum_i^n Y_i\right)}{n\left(\sum_i^n x_i^2\right) - \left(\sum_i^n x_i\right)^2} = \frac{\sum_i^n x_i Y_i - \left(\frac{1}{n}\sum_i^n x_i\right)\left(\sum_i^n Y_i\right)}{\left(\sum_i^n x_i^2\right) - n\left(\frac{1}{n}\sum_i^n x_i\right)\left(\frac{1}{n}\sum_i^n x_i\right)} \\ &= \frac{\sum_i^n x_i Y_i - \overline{x}\left(\sum_i^n Y_i\right)}{\left(\sum_i^n x_i^2\right) - n\overline{x}^2} = \frac{\sum_i^n \left(x_i - \overline{x}\right) Y_i}{\left(\sum_i^n x_i^2\right) - n\overline{x}^2}\end{aligned}$$

- This is a linear combination of normally distributed r.v.s, and thus normally distributed.

Proof of unbiasedness

Bruce M. Boghosian

- **Pf. (for $\hat{\beta}_1$):** Using the same form for $\hat{\beta}_1$ used above,

$$
\begin{aligned}
E\left(\hat{\beta}_1\right) &= E\left(\frac{\sum_i^n (x_i - \overline{x})\, Y_i}{\left(\sum_i^n x_i^2\right) - n\overline{x}^2}\right) \\
&= \frac{\sum_i^n (x_i - \overline{x})\, E\left(Y_i\right)}{\left(\sum_i^n x_i^2\right) - n\overline{x}^2} \\
&= \frac{\sum_i^n (x_i - \overline{x})\left(\beta_0 + \beta_1 x_i\right)}{\left(\sum_i^n x_i^2\right) - n\overline{x}^2} \\
&= \beta_1.
\end{aligned}
$$

Proof of expressions for variances

- **Pf. (for $\hat{\beta}_1$):** Using the same form for $\hat{\beta}_1$ used above,

$$\begin{aligned}
\mathrm{Var}\left(\hat{\beta}_1\right) &= \mathrm{Var}\left(\frac{\sum_i^n \left(x_i - \overline{x}\right) Y_i}{\left(\sum_i^n x_i^2\right) - n\overline{x}^2}\right) \\
&= \sum_i^n \left(\frac{\left(x_i - \overline{x}\right)}{\left(\sum_j^n x_j^2\right) - n\overline{x}^2}\right)^2 \mathrm{Var}\left(Y_i\right) \\
&= \frac{\sigma^2}{n \sum_i^n \left(x_i - \overline{x}\right)^2}
\end{aligned}$$

More theorems about Linear Model estimators

- **Thm.:** Let $(x_1, Y_1), \ldots, (x_n, Y_n)$ satisfy the assumptions of the Linear Model. Then
  - $\hat{\beta}_1$, $\overline{\mathbf{Y}}$, and $\hat{\boldsymbol{\sigma}}^2$ are mutually independent.
  - $\frac{n\hat{\boldsymbol{\sigma}}^2}{\sigma^2}$ has a χ^2 distribution with $n-2$ degrees of freedom.
- **Corr.:** Let $\hat{\boldsymbol{\sigma}}^2$ be the MLE for σ^2 in a simple linear model. Then
  - $\left(\frac{n}{n-2}\right)\hat{\boldsymbol{\sigma}}^2$ is an unbiased estimator for σ^2.
  - The random variables $\hat{\mathbf{Y}}^2$ and $\hat{\boldsymbol{\sigma}}^2$ are independent.

More theorems about Linear Model estimators

- Calculating $\sum_i^n (y_i - \hat{y}_i)^2 = \sum_i^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$

$$= \sum_i^n (y_i - \overline{y}_i)^2 - \hat{\beta}_1^2 \sum_i^n (x_i - \overline{x}_i)^2$$

$$= \sum_i^n y_i^2 - \frac{1}{n} \sum_i^n y_i - \frac{\left[\sum_i^n x_i y_i - \frac{1}{n}\left(\sum_i^n x_i\right)\left(\sum_i^n y_i\right)\right]^2}{\left(\sum_i^n x_i^2\right) - \frac{1}{n}\left(\sum_i^n x_i\right)}$$

$$= \sum_i^n y_i^2 - \hat{\beta}_0 \sum_i^n y_i - \hat{\beta}_1 \sum_i^n x_i y_i.$$

Drawing inferences about $\hat{\beta}_1$

Bruce M. Boghosian

- Let $(x_1, Y_1), \ldots, (x_n, Y_n)$ satisfy the assumptions of the Linear Model, and let

$$S^2 = \frac{1}{n-2} \sum_i^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2$$

- Then the following has a Student T distribution with $n-2$ df.

$$T_{n-2} = \frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{\sum_i^n (x_i - \overline{x})^2}}$$

Summary

- We reexamined the application of linear algebra to least-squares fitting.
- We have examined the least-squares method with random variables for the dependent variable.
- We have defined and motivated the Linear Model.
- We have defined regression curves.
- We worked out ML estimators for the three parameters of the linear model.
- We showed that $\hat{\beta}_1$ is Student T distributed with $n-2$ degrees of freedom.