Math 4/896: Seminar in Mathematics
Topic: Inverse Theory

Instructor: Thomas Shores
Department of Mathematics

AvH 10

Outline

1. Chapter 2: Linear Regression
   - A Motivating Example
   - Solutions to the System
   - Statistical Aspect of Least Squares

The Example

To estimate the mass of a planet of known radius while on the (airless) surface:

Observe a projectile thrown from some point and measure its altitude.

From this we hope to estimate the acceleration a due to gravity and then use Newton's laws of gravitation and motion to obtain from

$$\frac{GMm}{R^2} = ma$$

that

$$M = \frac{aR^2}{G}.$$

Equations of Motion

A Calculus I problem:

What is the vertical position $y(t)$ of the projectile as a function of time?

Just integrate the constant acceleration twice to obtain

$$y(t) = m_1 + m_2 t - \frac{1}{2} a t^2.$$

We follow the text and write at time t_k, we have observed value d_k and

$$d_k = y(t_k) = m_1 + t\, m_2 - \frac{1}{2} t^2 m_3$$

where $k = 1, 2, \ldots, m$. This is a system of m equations in 3 unknowns. Here

- m_1 is initial y-displacement
- m_2 is initial velocity
- m_3 is acceleration due to gravity.

A Calculus I problem:

What is the vertical position $y(t)$ of the projectile as a function of time?

Just integrate the constant acceleration twice to obtain

$$y(t) = m_1 + m_2 t - \frac{1}{2} a t^2 .$$

We follow the text and write at time t_k, we have observed value d_k and

$$d_k = y(t_k) = m_1 + t\, m_2 - \frac{1}{2} t^2 m_3$$

where $k = 1, 2, \ldots, m$. This is a system of m equations in 3 unknowns. Here

- m_1 is initial y-displacement
- m_2 is initial velocity
- m_3 is acceleration due to gravity.

A Calculus I problem:

What is the vertical position $y(t)$ of the projectile as a function of time?

Just integrate the constant acceleration twice to obtain

$$y(t) = m_1 + m_2 t - \frac{1}{2} a t^2.$$

We follow the text and write at time t_k, we have observed value d_k and

$$d_k = y(t_k) = m_1 + t\, m_2 - \frac{1}{2} t^2 m_3$$

where $k = 1, 2, \ldots, m$. This is a system of m equations in 3 unknowns. Here

- m_1 is initial y-displacement
- m_2 is initial velocity
- m_3 is acceleration due to gravity.

Matrix Form

Matrix Form of the System:

$$\text{(Linear Inverse Problem):} \quad G\mathbf{m} = \mathbf{d}.$$

Here $G = \begin{bmatrix} 1 & t_1 & -\frac{t_1^2}{2} \\ 1 & t_2 & \frac{t_2^2}{2} \\ \vdots & \vdots & \vdots \\ 1 & t_m & \frac{t_m^2}{2} \end{bmatrix}$, $\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}$ and $\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix}$, but we shall examine the problem in the more general setting where G is $m \times n$, $\mathbf{m}$ is $n \times 1$ and $\mathbf{d}$ is $m \times 1$.

A Specific Problem

The exact solution:

$$\mathbf{m} = [10, 100, 9.8]^T = (10, 100, 9.8)\,.$$

Spacial units are meters and time units are seconds. It's easy to simulate an experiment. We will do so assuming an error distribution that is independent and normally distributed with mean $\mu = 0$ and standard deviation $\sigma = 16$.

```
>randn('state',0)
>m = 10
>sigma = 16
>mtrue = [10,100,9.8]'
>G = [ones(m,1), (1:m)', -0.5*(1:m)'.^2]
>datatrue = G*mtrue;
>data = datatrue + sigma*randn(m,1);
```

Outline

Solution Methods

We could try

- The most naive imaginable: we only need three data points. Let's use them to solve for the three variables.
  Let's really try it with Matlab and plot the results for the exact data and the simulated data.
- A better idea: We are almost certain to have error. Hence, the full system will be inconsistent, so we try a calculus idea: minimize the sum of the norm of the residuals. This requires development. The basic problem is to find the **least squares solution** $m = m_{L_2}$ such that

$$\text{(Least Squares Problem):} \quad \|\mathbf{d} - G\mathbf{m}_{L_2}\|_2^2 = \min_{\mathbf{m}} \|\mathbf{d} - G\mathbf{m}\|_2^2$$

Solution Methods

We could try

- The most naive imaginable: we only need three data points. Let's use them to solve for the three variables.
  Let's really try it with Matlab and plot the results for the exact data and the simulated data.
- A better idea: We are almost certain to have error. Hence, the full system will be inconsistent, so we try a calculus idea: minimize the sum of the norm of the residuals. This requires development. The basic problem is to find the **least squares solution** $m = m_{L_2}$ such that

$$\text{(Least Squares Problem)}: \;\; \|\mathbf{d} - G\mathbf{m}_{L_2}\|_2^2 = \min_{\mathbf{m}} \|\mathbf{d} - G\mathbf{m}\|_2^2$$

Key Results for Least Squares

Theorem

The least squares problem has a solution for any $m \times n$ matrix G and data $\mathbf{d}$, namely any solution to the ***normal equations***

$$G^T G\mathbf{m} = G^T \mathbf{d}$$

Proof sketch:

- Show product rule holds for products of matrix functions.
- Note $f(\mathbf{m}) = \|\mathbf{d} - G\mathbf{m}\|^2$ is a nonnegative quadratic function in $\mathbf{m}$, so must have a minimum
- Find the critical points of f by setting $\nabla f(\mathbf{m}) = \mathbf{0}$.

Key Results

Theorem

If $m \times n$ matrix G has full column rank, then the least squares solution is unique, and is given by

$$\mathbf{m}_{L_2} = \left(G^T G\right)^{-1} G^T \mathbf{d}$$

Proof sketch:

- Show $G^T G$ has zero kernel, hence is invertible.
- Plug into normal equations and solve.

Least Squares Experiments:
Use Matlab to solve our specific problem with experimental data and plot solutions. Then let's see why the theorems are true. There remains:

Problem:

How good is our least squares solution? Can we trust it? Is there a better solution?

Key Results

Theorem

If $m \times n$ matrix G has full column rank, then the least squares solution is unique, and is given by

$$\mathbf{m}_{L_2} = \left(G^T G\right)^{-1} G^T \mathbf{d}$$

Proof sketch:

- Show $G^T G$ has zero kernel, hence is invertible.
- Plug into normal equations and solve.

Least Squares Experiments:
Use Matlab to solve our specific problem with experimental data and plot solutions. Then let's see why the theorems are true. There remains:

Problem:

How good is our least squares solution? Can we trust it? Is there a better solution?

Key Results

Theorem

If $m \times n$ matrix G has full column rank, then the least squares solution is unique, and is given by

$$\mathbf{m}_{L_2} = \left(G^T G\right)^{-1} G^T \mathbf{d}$$

Proof sketch:

- Show $G^T G$ has zero kernel, hence is invertible.
- Plug into normal equations and solve.

Least Squares Experiments:
Use Matlab to solve our specific problem with experimental data and plot solutions. Then let's see why the theorems are true. There remains:

Problem:

How good is our least squares solution? Can we trust it? Is there a better solution?

Outline

1. Chapter 2: Linear Regression
   - A Motivating Example
   - Solutions to the System
   - Statistical Aspect of Least Squares

Quality of Least Squares

View the ProbStatLectures notes regarding point estimation. Then we see why this fact is true:

Theorem

Suppose that the error of ith coordinate of the residual is normally distributed with mean zero and standard deviation σ_i. Let $W = \operatorname{diag}(1/\sigma_1, \ldots, 1/\sigma_m)$ and $G_W = WG$, $\mathbf{d}_W = W\mathbf{d}$. Then the least squares solution to the scaled inverse problem

$$G_W \mathbf{m} = \mathbf{d}_W$$

is a ***maximum liklihood estimator*** *to the parameter vector.*

An Example

Let's generate a problem as follows

```
>randn('state',0)
>m = 10
>sigma = blkdiag(8*eye(3),16*eye(3),24*eye(4)]
>mtrue = [10,100,9.8]'
>G = [ones(m,1), (1:m)', -0.5*(1:m)'.^2]
>datatrue = G*mtrue;
>data = datatrue + sigma*randn(m,1);
>G = [ones(m,1), (1:m)', -0.5*(1:m)'.^2]
>datatrue = G*mtrue;
>data = datatrue + sigma*randn(m,1);
% compute the least squares solution without
% reference to sigma, then do the scaled least squares
% and compare....also do some graphs
```

Quality of Least Squares

A very nontrivial result which we assume:

Theorem

Let G have full column rank and $\mathbf{m}$ *the least squares solution for the scaled inverse problem. The statistic*

$$\|\mathbf{d} - G\mathbf{m}\|_2^2 = \sum_{i=1}^{m} \left(d_i - \left(Gm_{L_2}\right)_i\right)^2 / \sigma_i^2$$

in the random variable $\mathbf{d}$ *has a chi-square distribution with* $\nu = m - n$ *degrees of freedom.*

This provided us with a statistical assessment (the chi-square test) of the quality of our data. We need the idea of the p-**value** of the test, the probability of obtaining a larger chi-square value than the one actually obtained:

$$p = \int_{\chi^2_{obs}}^{\infty} f_{\chi^2}(x)\, dx.$$

Interpretation of p

As a random variable, the p-value is uniformly distributed between zero and one. This can be very informative:

1. "Normal sized" p: we probably have an acceptable fit
2. Extremely small p: data is very unlikely, so model $G\mathbf{m} = \mathbf{d}$ may be wrong or data may have larger errors than estimated.
3. Extremely large p (i.e., very close to 1): fit to model is almost exact, which may be too good to be true.

Interpretation of p

As a random variable, the p-value is uniformly distributed between zero and one. This can be very informative:

1. "Normal sized" p: we probably have an acceptable fit
2. Extremely small p: data is very unlikely, so model $G\mathbf{m} = \mathbf{d}$ may be wrong or data may have larger errors than estimated.
3. Extremely large p (i.e., very close to 1): fit to model is almost exact, which may be too good to be true.

Interpretation of p

As a random variable, the p-value is uniformly distributed between zero and one. This can be very informative:

1. "Normal sized" p: we probably have an acceptable fit
2. Extremely small p: data is very unlikely, so model $G\mathbf{m} = \mathbf{d}$ may be wrong or data may have larger errors than estimated.
3. Extremely large p (i.e., very close to 1): fit to model is almost exact, which may be too good to be true.

Uniform Distributions

Reason for uniform distribution:

Theorem

Let X have a continuous c.d.f. $F(x)$ such that $F(x)$is strictly increasing where $0 < x < 1$. Then the r.v. $Y = F(X)$ is uniformly distributed on the interval $(0, 1)$

Proof sketch:

- Calculate $P(Y \leq y)$ using fact that F has an inverse function F^{-1}.
- Use the fact that $P(X \leq x) = F(x)$ to prove that $P(Y \leq y) = y$.