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

Instructor: Thomas Shores
Department of Mathematics

AvH 10

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

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

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 $P(Y \leq y) = y$.

Application: One can use this to generate random samples for X.

An Example

Let's resume our experiment from above. Open the script Lecture8.m and have a look. Then run Matlab on it and resume calculations.

> % now set up for calculating the p-value of the test under both scenarios.

```
>chiobs1 = norm(data - G*mapprox1)^2
>chiobs2 = norm(W*(data - G*mapprox2))^2
>help chis_pdf
>p1 = 1 - chis_cdf(chiobs1,m-n)
>p2 = 1 - chis_cdf(chiobs2,m-n)
% How do we interpret these results?
% Now put the bad estimate to the real test
```

How do we interpret these results?

More Conceptual Tools

Examine and use the MVN theorems of ProbStatLectures to compute the expectation and variance of the r.v. $\mathbf{m}$, where $\mathbf{m}$ is the modified least squares solution, G has full column rank and $\mathbf{d}$ is a vector of independent r.v.'s.

- Each entry of $\mathbf{m}$ is a linear combination of independent normally distributed variables, since

$$\mathbf{m} = \left(G_W^T G_W\right)^{-1} G_W^T \mathbf{d}_W .$$

- The weighted data $\mathbf{d}_W = W\mathbf{d}$ has covariance matrix I.
- Deduce that $\operatorname{Cov}(\mathbf{m}) = \left(G_W^T G_W\right)^{-1}$.
- Note simplification if variances are constant: $\operatorname{Cov}(\mathbf{m}) = \sigma^2 \left(G^T G\right)^{-1}$.

Conceptual Tools

Next examine the mean of $\mathbf{m}$ and deduce from the facts that

$$E[\mathbf{d}_W] = W\mathbf{d}_{true} \text{ and } G_W \mathbf{m}_{true} = \mathbf{d}_{true}$$

and MVN facts that

- $E[\mathbf{m}] = \mathbf{m}_{true}$
- Hence, modified least squares solution is an **unbiased estimator** of $\mathbf{m}_{true}$.
- Hence we can construct a confidence interval for our experiment:

$$\mathbf{m} \pm 1.96 \cdot \text{diag}(\text{Cov}(\mathbf{m}))^{1/2}$$

- What if the (constant) variance is unknown? Student's t to the rescue!

How do we interpret these results?

Next examine the mean of $\mathbf{m}$ and deduce from the facts that

$$E\left[\mathbf{d}_W\right] = W\mathbf{d}_{true} \text{ and } G_W\mathbf{m}_{true} = \mathbf{d}_{true}$$

and MVN facts that

- $E\left[\mathbf{m}\right] = \mathbf{m}_{true}$
- Hence, modified least squares solution is an **unbiased estimator** of $\mathbf{m}_{true}$.
- Hence we can construct a confidence interval for our experiment:

$$\mathbf{m} \pm 1.96 \cdot \operatorname{diag}\left(\operatorname{Cov}\left(\mathbf{m}\right)\right)^{1/2}$$

- What if the (constant) variance is unknown? Student's t to the rescue!

How do we interpret these results?

Next examine the mean of $\mathbf{m}$ and deduce from the facts that

$$E[\mathbf{d}_W] = W\mathbf{d}_{true} \text{ and } G_W\mathbf{m}_{true} = \mathbf{d}_{true}$$

and MVN facts that

- $E[\mathbf{m}] = \mathbf{m}_{true}$
- Hence, modified least squares solution is an **unbiased estimator** of $\mathbf{m}_{true}$.
- Hence we can construct a confidence interval for our experiment:
$$\mathbf{m} \pm 1.96 \cdot \text{diag}(\text{Cov}(\mathbf{m}))^{1/2}$$
- What if the (constant) variance is unknown? Student's t to the rescue!

How do we interpret these results?

Conceptual Tools

Next examine the mean of $\mathbf{m}$ and deduce from the facts that

$$E[\mathbf{d}_W] = W\mathbf{d}_{true} \text{ and } G_W\mathbf{m}_{true} = \mathbf{d}_{true}$$

and MVN facts that

- $E[\mathbf{m}] = \mathbf{m}_{true}$
- Hence, modified least squares solution is an **unbiased estimator** of $\mathbf{m}_{true}$.
- Hence we can construct a confidence interval for our experiment:
$$\mathbf{m} \pm 1.96 \cdot \mathrm{diag}(\mathrm{Cov}(\mathbf{m}))^{1/2}$$
- What if the (constant) variance is unknown? Student's t to the rescue!

How do we interpret these results?

Outliers

These are discordant data, possibly due to other error or simply bad luck. What to do?

- Use statistical estimation to discard the outliers.
- Use a different norm from $\|\cdot\|_2$. The 1-norm is an alternative, but this makes matters much more complicated! Consider the optimization problem

$$\|\mathbf{d} - G\mathbf{m}_{L_2}\|_1 = \min_{\mathbf{m}} \|\mathbf{d} - G\mathbf{m}\|_1$$

How do we interpret these results?

A Motivating Example: Integral Equations

Contanimant Transport

Let $C(x,t)$ be the concentration of a pollutant at point x in a linear stream, time t, where $0 \le x < \infty$ and $0 \le t \le T$. The defining model

$$\begin{aligned} \frac{\partial C}{\partial t} &= D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x} \\ C(0,t) &= C_{in}(t) \\ C(x,t) &\to 0,\ x \to \infty \\ C(x,0) &= C_0(x) \end{aligned}$$

Solution

Solution:

$$C(x, T) = \int_0^T C_{in}(t)\, f(x, T - t)\, dt,$$

where

$$f(x, \tau) = \frac{x}{2\sqrt{\pi D \tau^3}} e^{-(x - v\tau)^2/(4D\tau)}$$

The Inverse Problem

Problem:

Given simultaneous measurements at time T, to estimate the contaminant inflow history.

Some Methods

More generally

Problem:

Given the IFK

$$d(s) = \int_a^b g(x,s)\, m(x)\, dx$$

and a finite sample of values $d(s_i)$, to estimate parameter $m(x)$.

Methods we discuss at the board:

1. Quadrature
2. Representers
3. Orthogonal representers

Some Methods

More generally

Problem:

Given the IFK

$$d(s) = \int_a^b g(x, s)\, m(x)\, dx$$

and a finite sample of values $d(s_i)$, to estimate parameter $m(x)$.

Methods we discuss at the board:

1. Quadrature
2. Representers
3. Orthogonal representers

Some Methods

More generally

Problem:

Given the IFK

$$d(s) = \int_a^b g(x,s)\, m(x)\, dx$$

and a finite sample of values $d(s_i)$, to estimate parameter $m(x)$.

Methods we discuss at the board:

1. Quadrature
2. Representers
3. Orthogonal representers

Some Methods

More generally

Problem:

Given the IFK

$$d(s) = \int_a^b g(x, s)\, m(x)\, dx$$

and a finite sample of values $d(s_i)$, to estimate parameter $m(x)$.

Methods we discuss at the board:

1. Quadrature
2. Representers
3. Orthogonal representers