GEOL/PHYS 6670
Geophysical Inverse Theory

Prof. Anne Sheehan
Anne.Sheehan @ colorado.edu
Office: Benson Rm 440A

Last Time

Moment Tensor Example

Linear versus nonlinear

Probability

Python tutorial

Today

Library, literature search info – Phil White

Menke Ch 2 continued (probability – variance and covariance)

Start Menke Ch 3 – L2 Norm and Least Squares

Phil White, Earth Sciences Librarian

Literature searches and accessing journal articles via the CU library

Philip B. White
Earth, Environment and Geospatial Librarian **|** Jerry Crail Johnson Earth Sciences & Map Library
University of Colorado Boulder Libraries **|** 184 UCB **|** Boulder, CO 80309 **|** 303-735-8278
Philip.White@Colorado.EDU **|** https://orcid.org/0000-0001-6087-8286 **|** http://www.colorado.edu/libraries
(he/him)

Homework 2 – due now ($Gm = d$ setup)
+ exploring LinkedIn Learning

Homework 3 – will be posted later today
library exercise – look up another paper
coding practice – simple exercises

LinkedIn Learning – what you learned

Technical
Programming Foundations: Fundamentals by Annyce Davis
Learning Matlab by Steven Moser
Learning Python with Joe Marini
Python Quick start by Laranya Vijayan and Madecraft
Data Science Foundations: Fundamentals by Barton Poulson
Illustrator 2021 Essential Training by Tony Harmer
Artificial Intelligence Foundations: Machine Learning by Doug Rose.

Professional Development
Master Confident Presentations by Chris Croft
The Six Habits of High Performers by Pete Mockaitis
Sleep is your superpower by Nancy H. Rothstein

DEI
Uncovering your authentic self at work by Kenji Yoshino

Gm=d review

Suppose that you determine the height of 50 objects by measuring the first, and then stacking the second on top of the first and measuring their combined height, stacking the third on top of the first two and measuring their combined height, and so forth.

(A) Identify the data and model parameters in this problem. How many of each are there?
(B) Write down the matrix **G** in the equation **d** = **Gm** that relates the data to the model parameters.

Take a few minutes to work on this on your own

Gm=d review

Suppose that you determine the height of 50 objects by measuring the first, and then stacking the second on top of the first and measuring their combined height, stacking the third on top of the first two and measuring their combined height, and so forth.

(A) Identify the data and model parameters in this problem. How many of each are there?

The 50 heights of the objects are the model parameters (h1, h2, h3,…h50). The 50 heights of the stack, as measured by a ruler, are the data (s1, s2, s3, … s50)

(B) Write down the matrix **G** in the equation **d** = **Gm** that relates the data to the model parameters.

Gm=d review

Suppose that you determine the height of 50 objects by measuring the first, and then stacking the second on top of the first and measuring their combined height, stacking the third on top of the first two and measuring their combined height, and so forth.

(A) Identify the data and model parameters in this problem. How many of each are there?

The 50 heights of the objects are the model parameters. The 50 heights of the stack, as measured by a ruler, are the data

(B) Write down the matrix **G** in the equation **d** = **Gm** that relates the data to the model parameters.

m = (h1, h2, h3, … h50)

d = (s1, s2, s3, … s50)

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & \cdots & 0 \\ & & & & & & \ddots & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

More Menke Ch 2 - Probability

Probability is relevant to data analysis because all data contains noise, and noise can be understood using concepts drawn from probability.

probability density function (p.d.f.) describes the behavior of the random variable. It is the idealization of a histogram of the realizations, in the limit of an indefinitely large number of realizations.

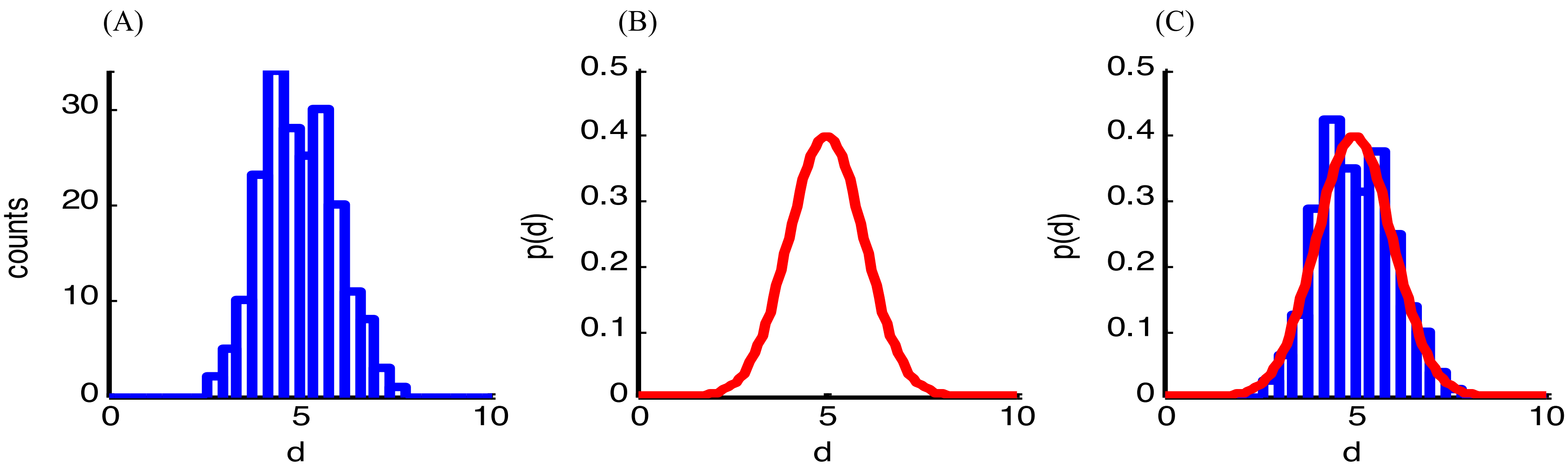
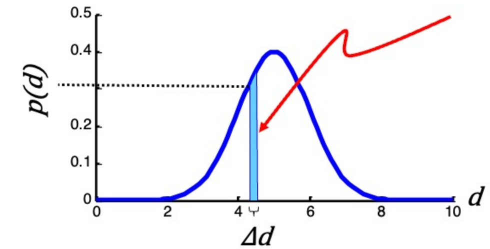


Fig 2.1. (A) Histogram showing data from *200* repetitions of an experiment in which datum, *d*, is measured. Noise causes observations to scatter about their mean value, $\langle d \rangle = 5$. (B) Probability density function (p.d.f), $p(d)$, of the data. (C) Histogram (blue) and p.d.f. (red) superimposed. Note that the histogram has a shape similar to the p.d.f.. *MatLab* script gda02_01.

in general
probability is the integral



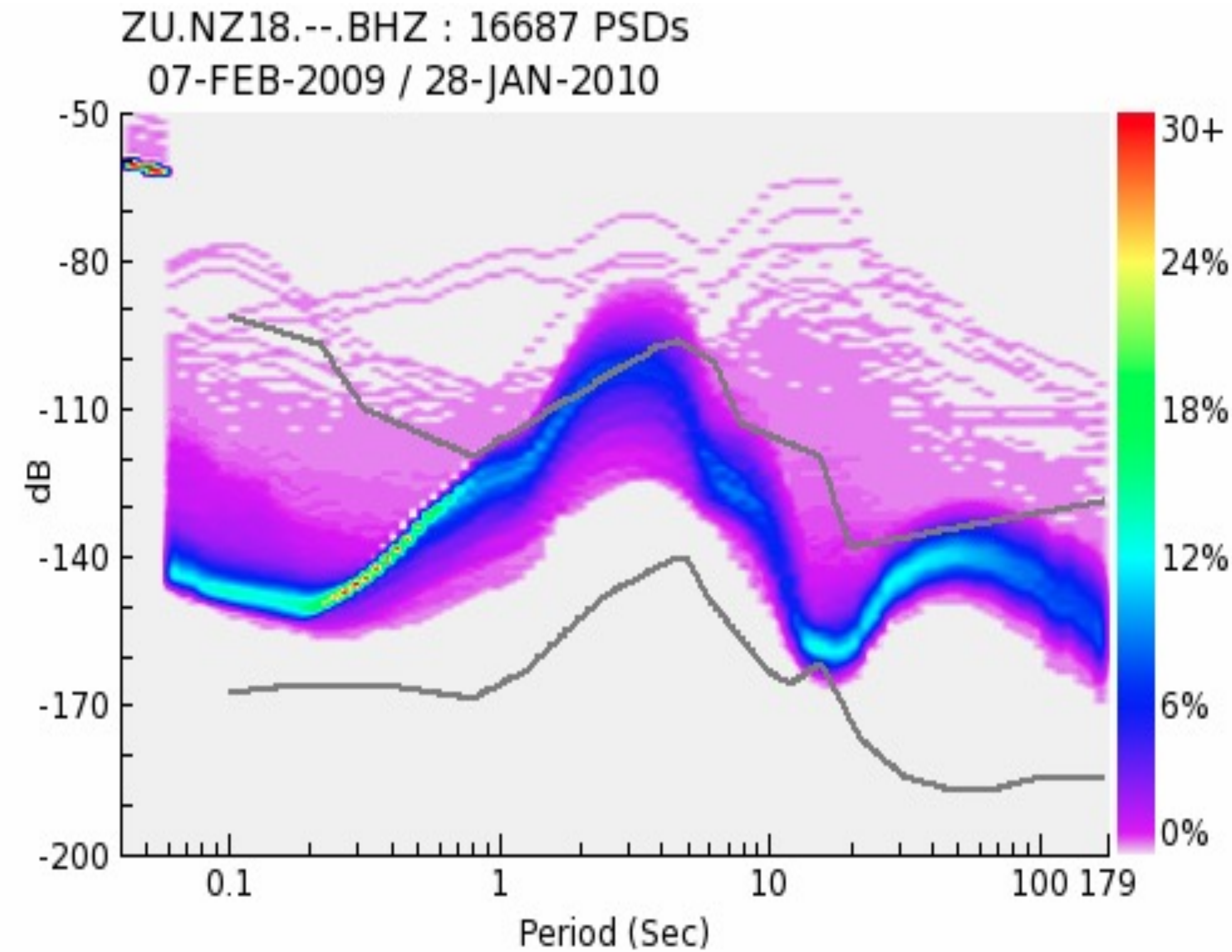
$$P(d_1, d_2) = \int_{d_1}^{d_2} p(d)\, \mathrm{d}d$$

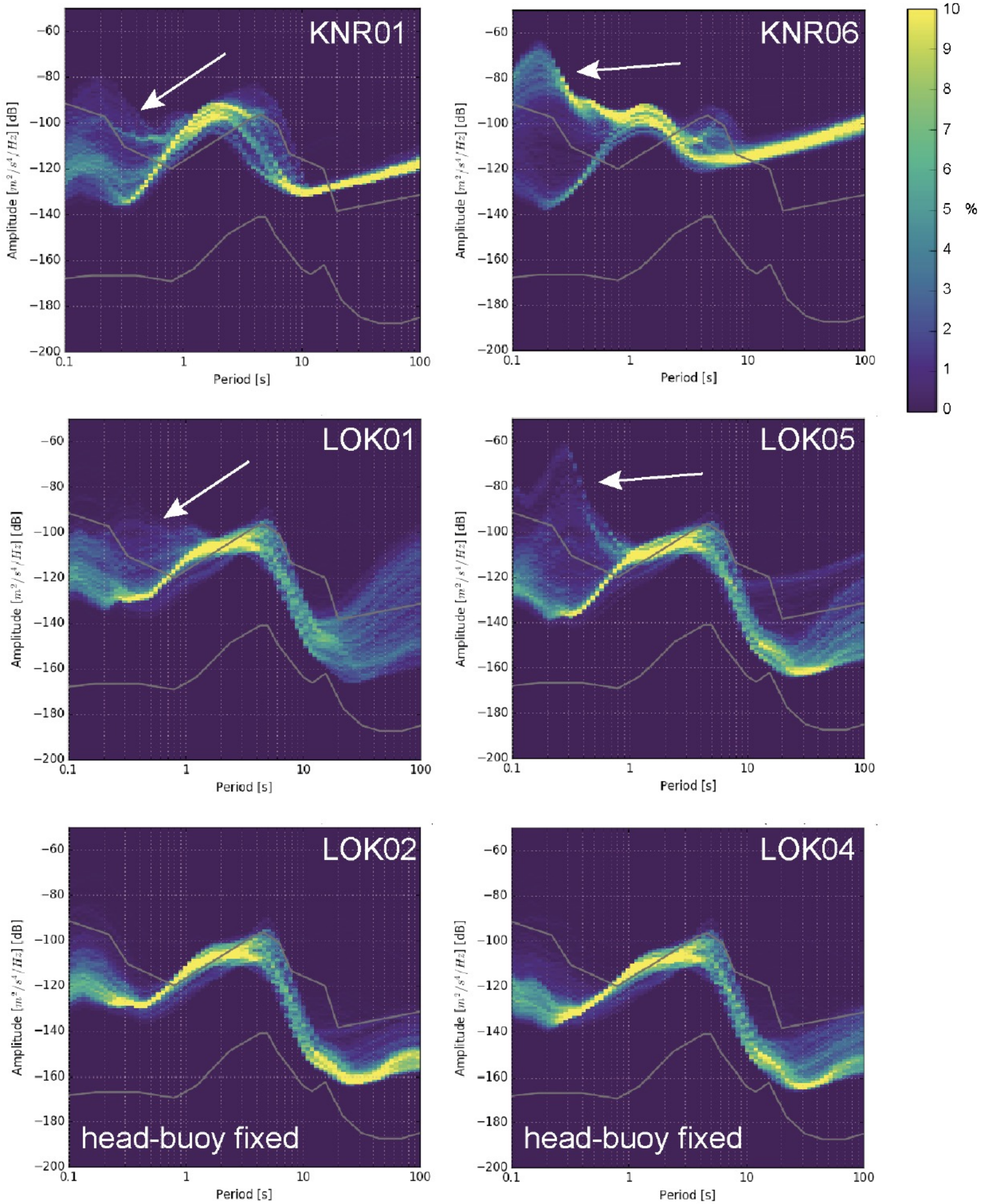
probability that *d* is between
d_1 and d_2

The probability P is a number between 0 and 1 (0% and 100%).
The probability that the random variable will take on a value between d_1 and d_2 is given by the intergral.

p.d.f.s of a year of data at an ocean bottom seismometer

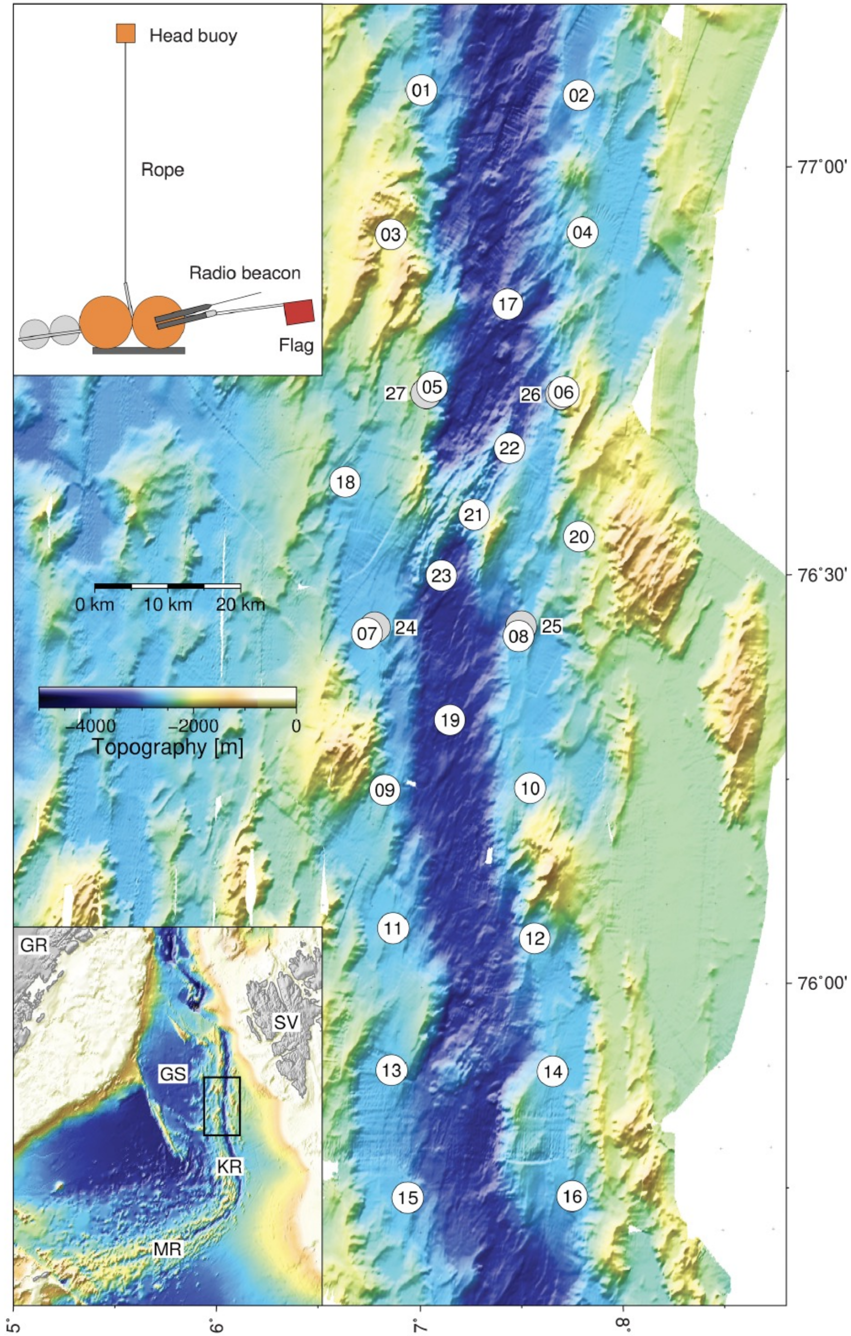
- Power spectra determined for one-hour segments of the entire year of data
- Probability density function used to display the thousands of spectra
- Blues and greens represent frequent values, light purple represents less frequent values



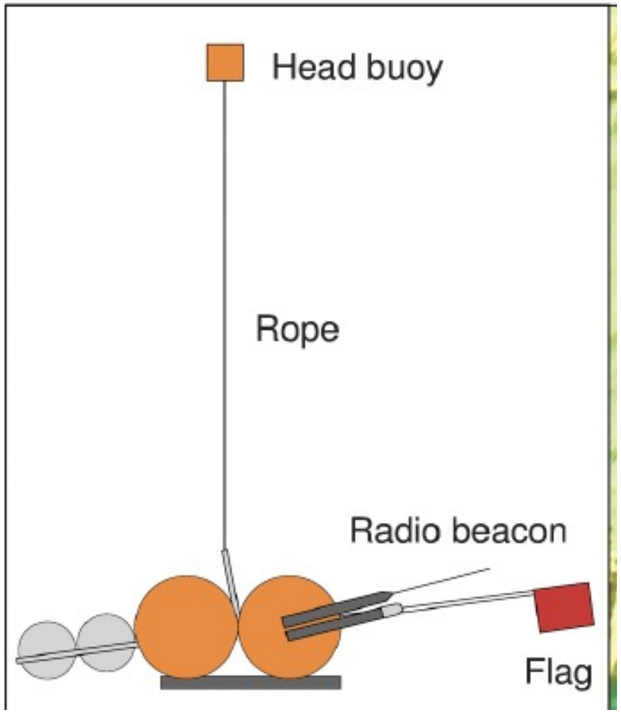


OBS noise characterization

Figure S6: Probabilistic Power Spectral Density plots after McNamara and Buland (2004) calculated from vertical channel seismometer data. **Top row**: Two examples from the



OBS noise characterization



Summarizing a probability density function

Mean, median, mode

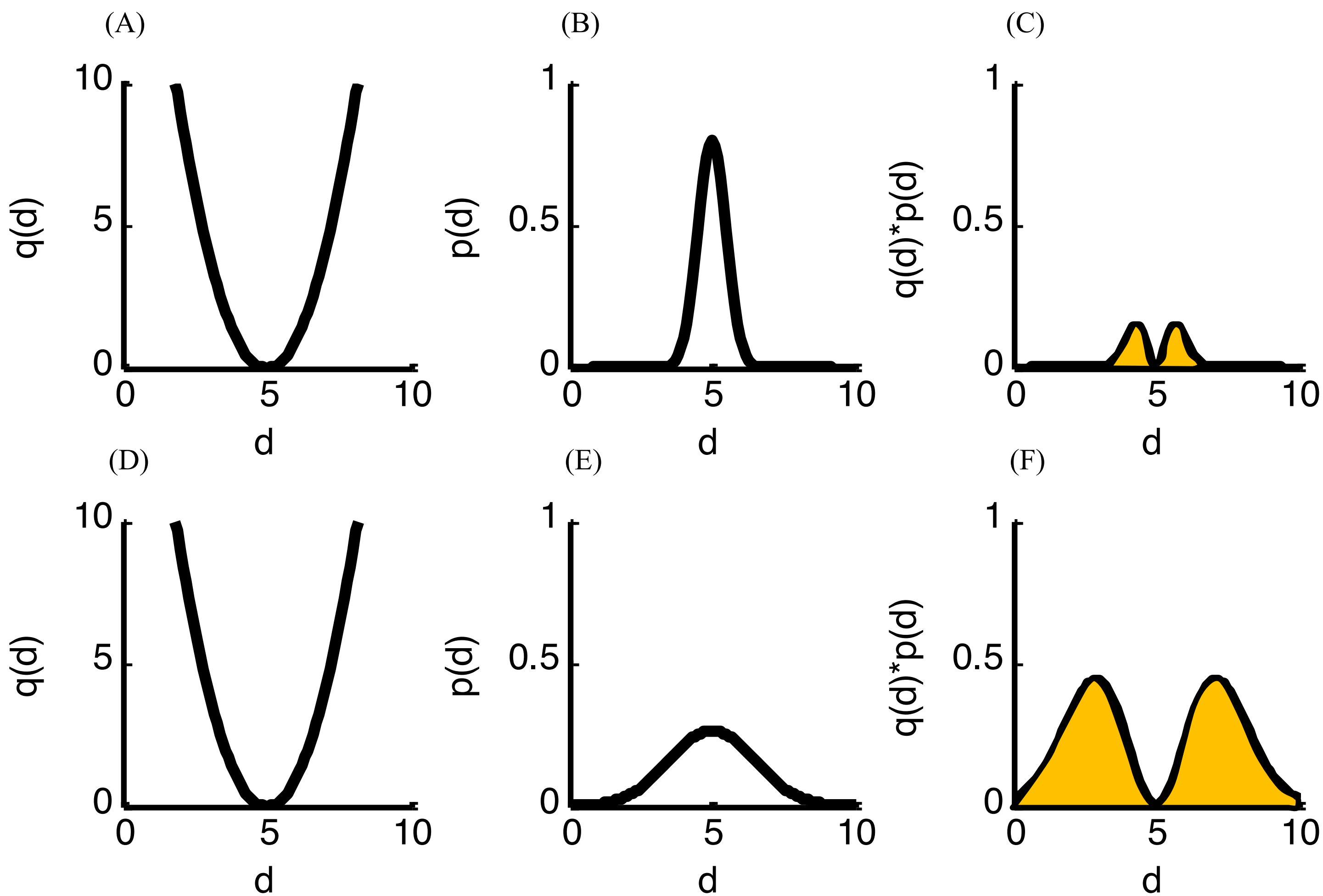
amount of scatter around the typical value

“width of the p.d.f.”

quantifying width

The width of a p.d.f. quantifies scatter.
Small width, small scatter, low observational noise.
Large width, large scatter, high observational noise.

Graphical representation of width



$q(d) = (d-\langle d \rangle)^2$

$p(d)$

$q(d)p(d)$

quantify width as the area under $q(d)p(d)$

Fig 2.4

This function grows away from the typical value:

$$q(d) = (d - \langle d \rangle)^2$$

so the function $q(d)p(d)$ is

small if most of the area is near $\langle d \rangle$, that is, a narrow $p(d)$

large if most of the area is far from $\langle d \rangle$, that is, a wide $p(d)$

so quantify width as the area under $q(d)p(d)$

‘deviation’ of a given measurement is difference between that measurement and the mean

deviation = (d-<d>)

variance is average of the deviation squared

ie, plug *(d-<d>)²* *in to*

$$= \int_{d_{min}}^{d_{max}} d\, p(d)\, \mathrm{d}d$$

recall

$$<d> \; = \int_{d_{min}}^{d_{max}} d\, p(d)\, \mathrm{d}d$$

variance

$$\sigma^2 = \int_{-\infty}^{+\infty} (d - \underbrace{\langle d \rangle}_{\text{mean}})^2 \, p(d) \, \mathrm{d}d$$

Standard deviation is σ and is a measure of the width of the distribution

estimating mean and variance from data

$$\langle d \rangle^{est} = \frac{1}{N}\sum_{i=1}^{N} d_i \quad \text{and} \quad (\sigma^2)^{est} = \frac{1}{N-1}\sum_{i=1}^{N} (d_i - \langle d \rangle^{est})^2$$

estimating mean and variance from data

$$\langle d \rangle^{est} = \frac{1}{N}\sum_{i=1}^{N} d_i \quad \text{and} \quad (\sigma^2)^{est} = \frac{1}{N-1}\sum_{i=1}^{N}(d_i - \langle d \rangle^{est})^2$$

usual formula for "sample mean"

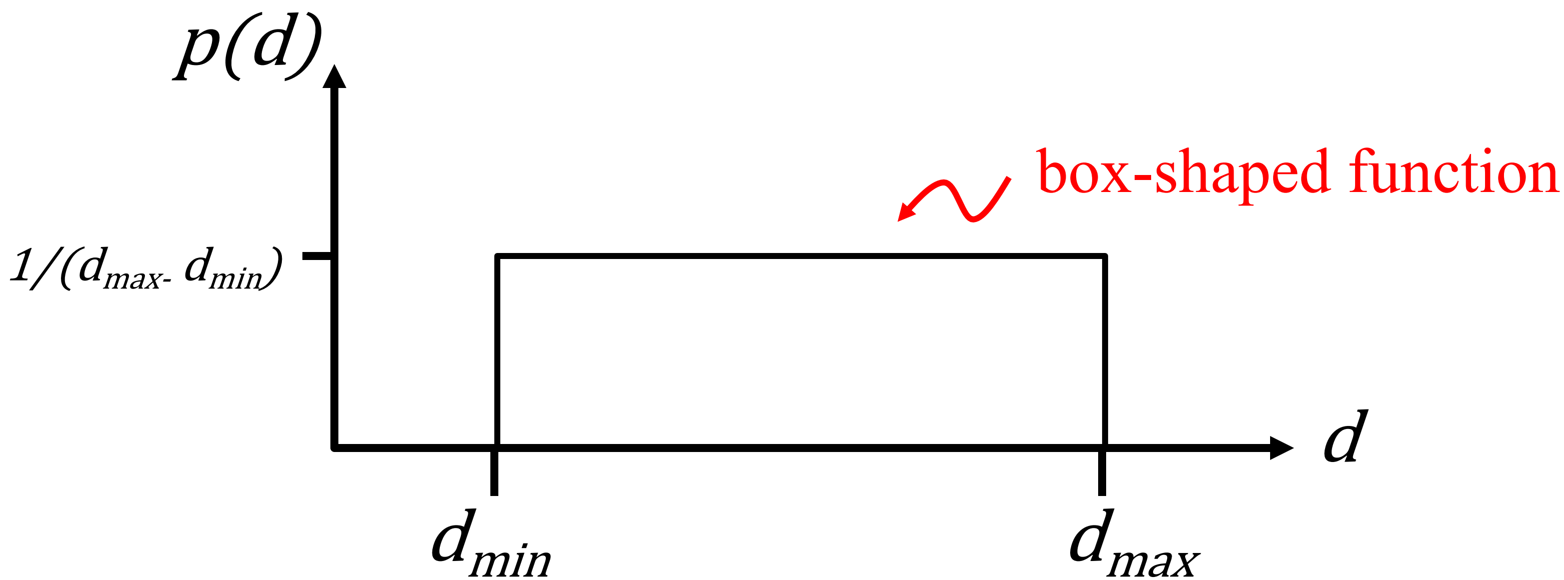
usual formula for square of "sample standard deviation"

(Matlab has functions *mean*(d) and *std*(d)) two important probability density functions:

uniform

Gaussian (or Normal)

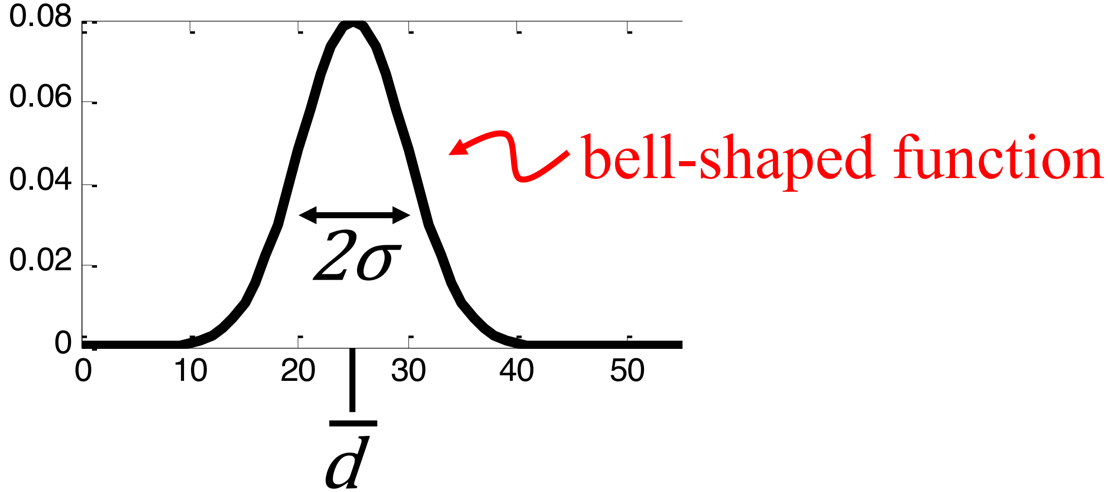
uniform p.d.f.



probability is the same everywhere in the range of possible values

Gaussian (or "Normal") p.d.f.

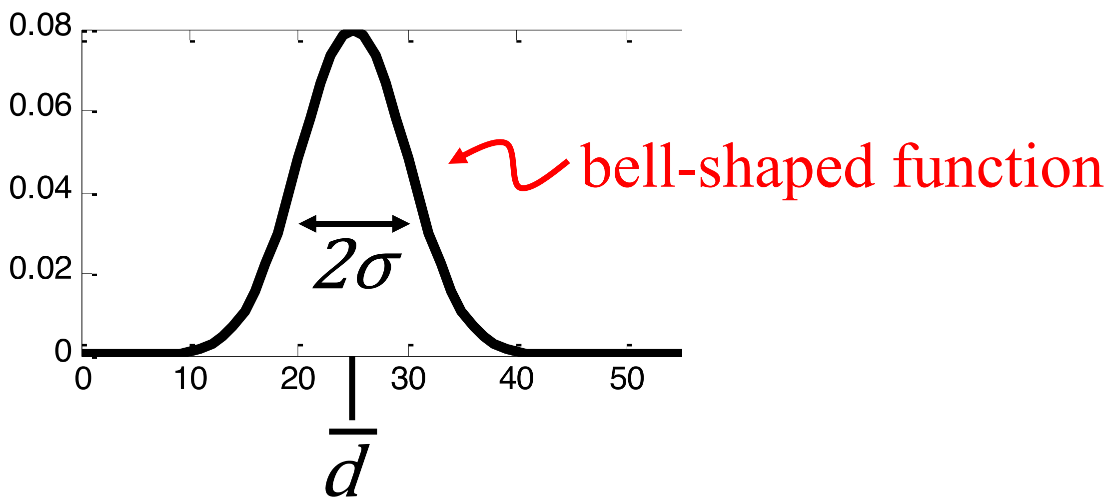
$$p(d) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{(d - \langle d \rangle)^2}{2\sigma^2}\right]$$



Large probability near the mean, $\bar{d}$.
Variance is σ^2.

Gaussian (or "Normal") p.d.f.

$$p(d) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{(d-\langle d\rangle)^2}{2\sigma^2}\right]$$



Symmetric around the mean, so mean=median=maximum likelihood. Relatively short-tailed (little probability far from the mean).

Gaussian p.d.f. probability between $<d> \pm n\sigma$

n	P, %
1	68.27
2	95.45
3	99.73

correlated errors

correlation means that a relationship exists between the noise in two different data types

uncorrelated random variables

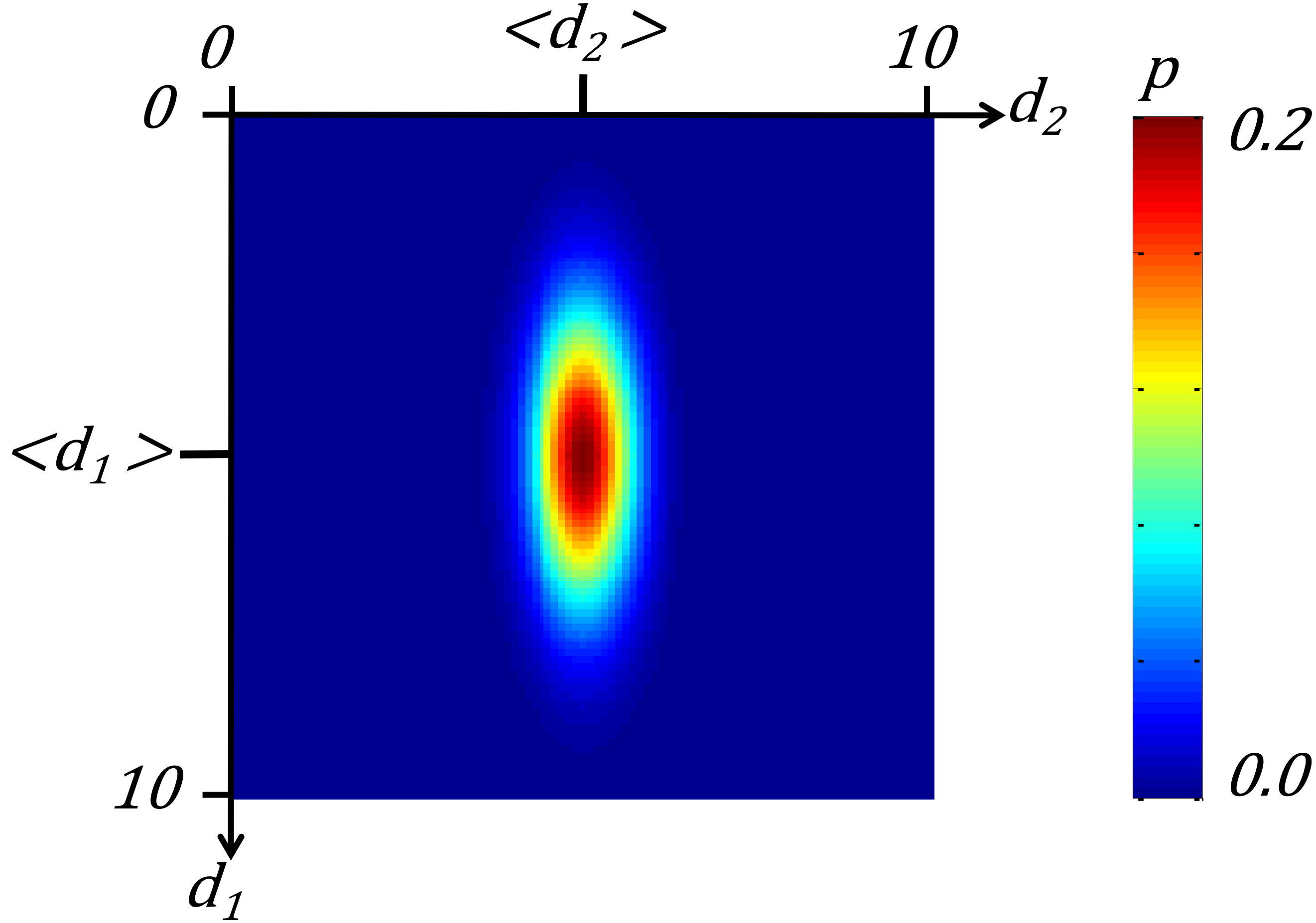
no pattern between values of one variable and values of another

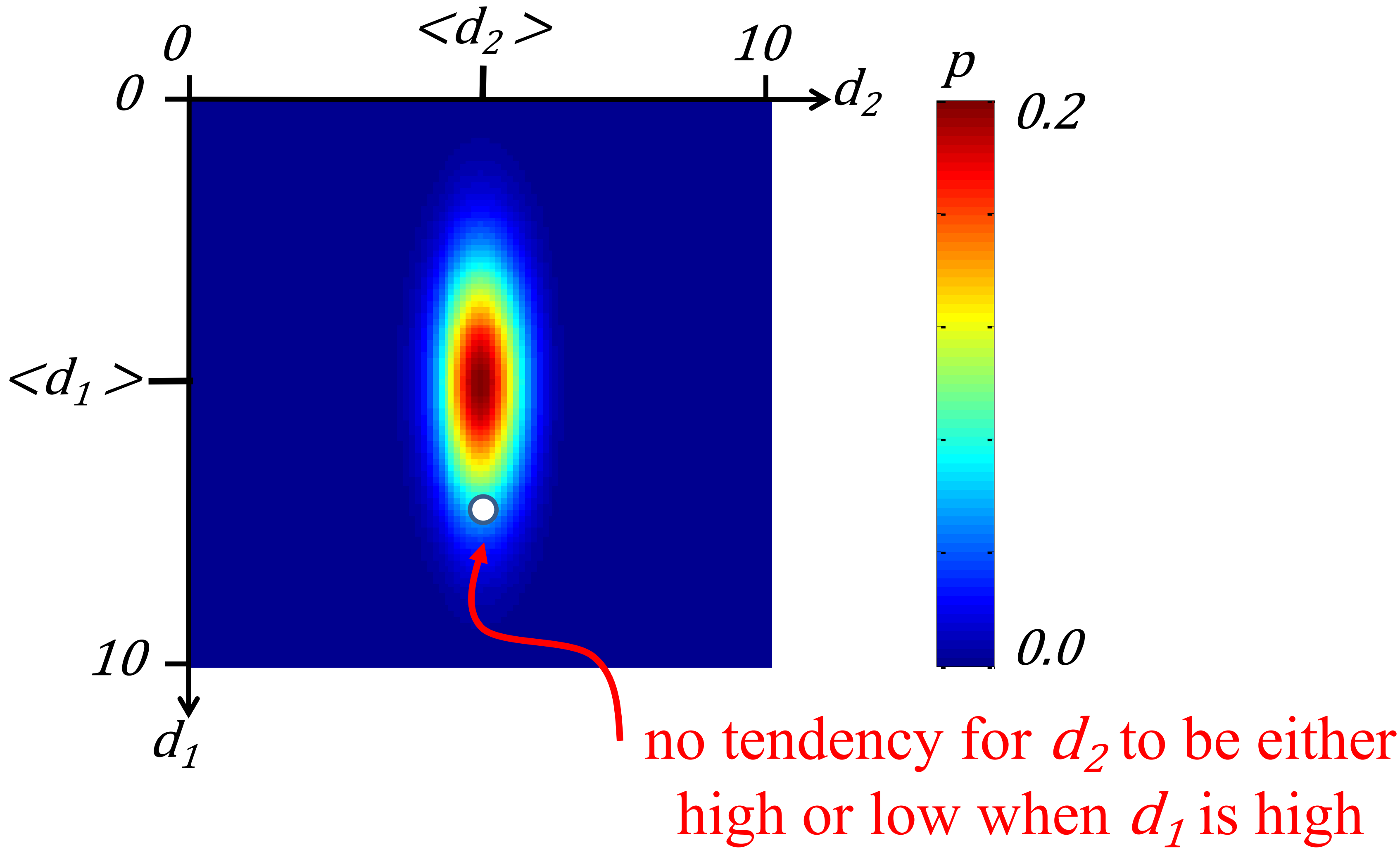
Examples

1. Length of a metal rod as a function of temperature

2. Length of a rod vs. time

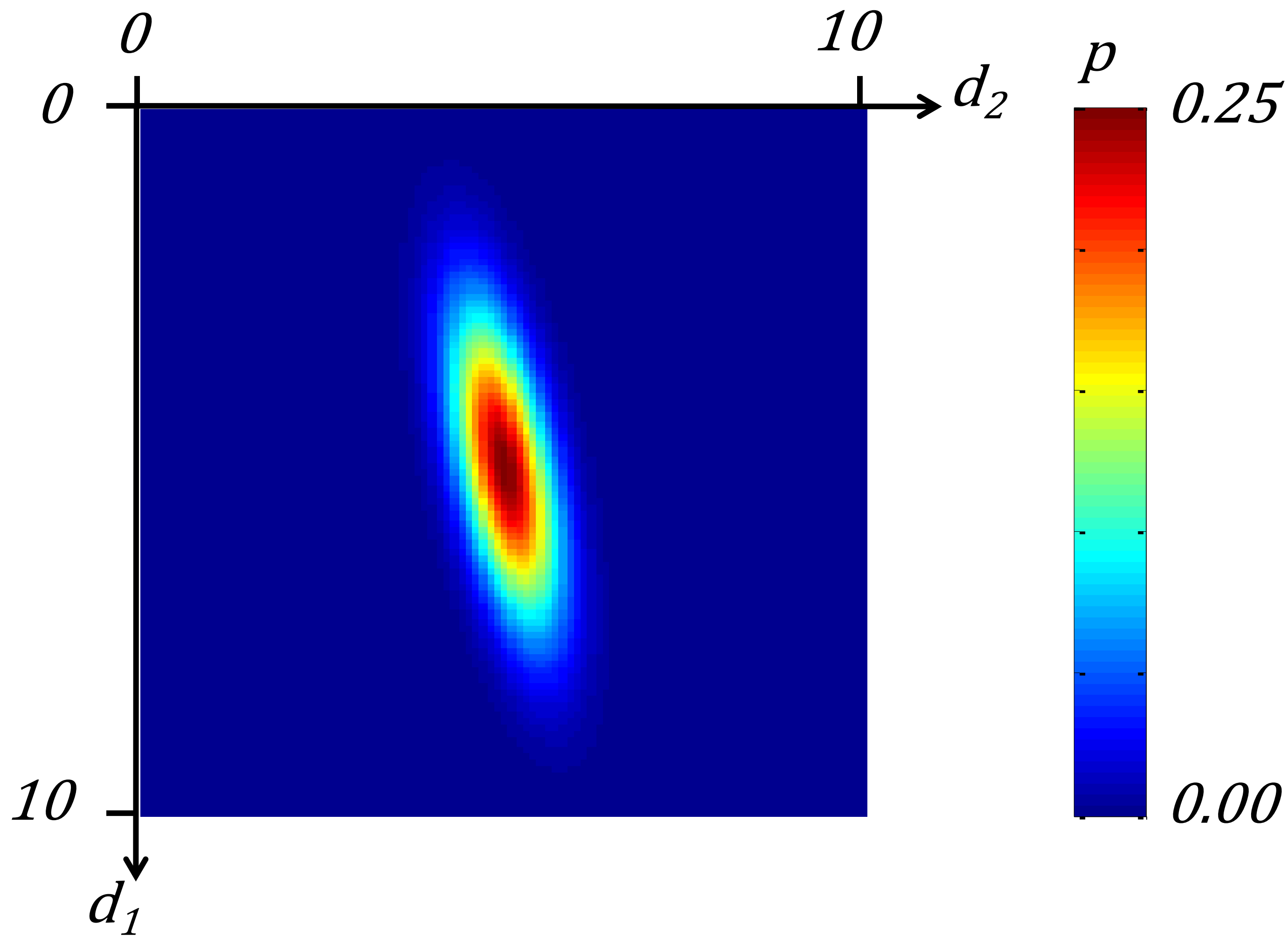
joint probability density function uncorrelated case



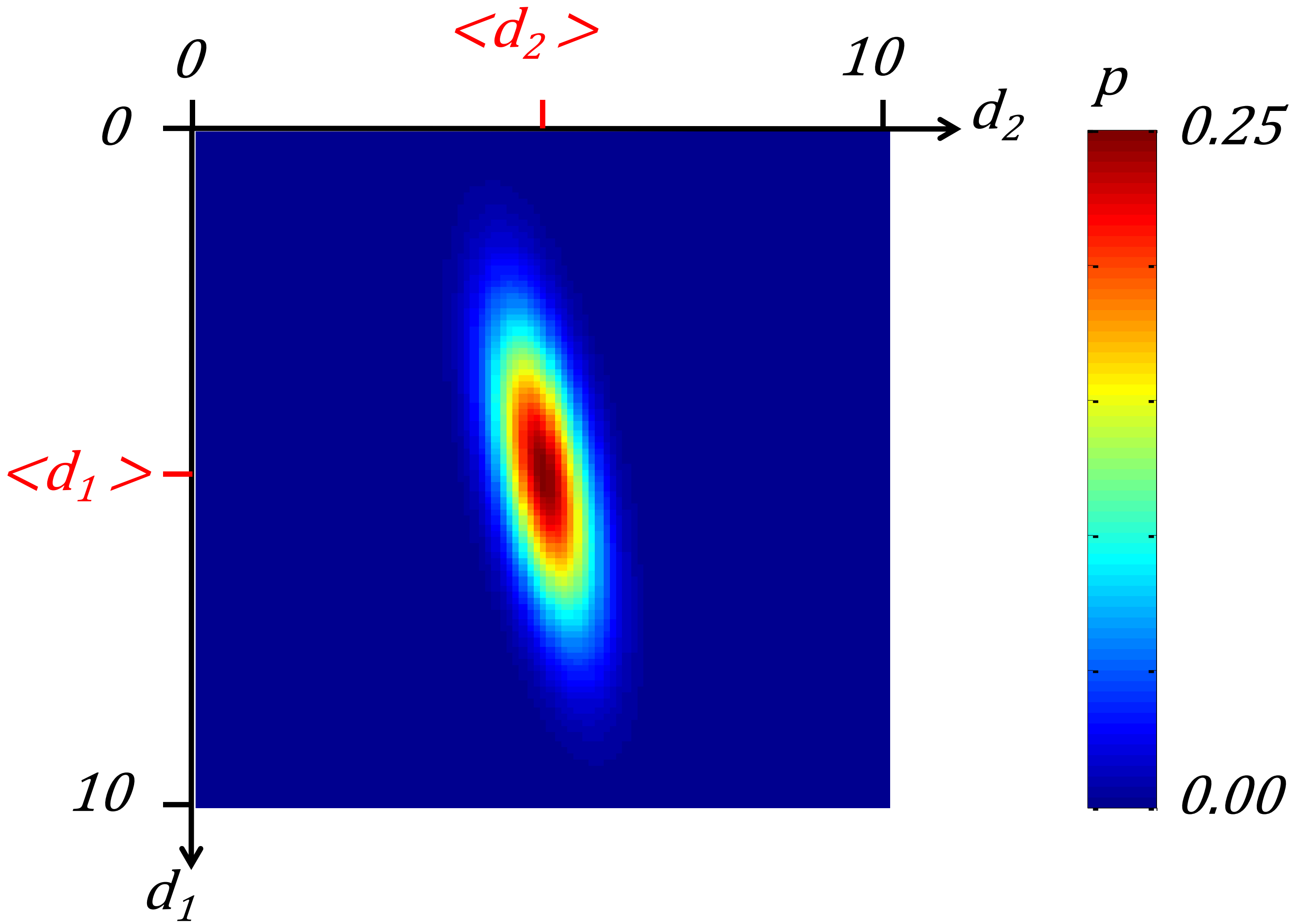


0
$\langle d_2 \rangle$
10
d_2
0
$\langle d_1 \rangle$
10
d_1
p
0.2
0.0
no tendency for d_2 to be either high or low when d_1 is high

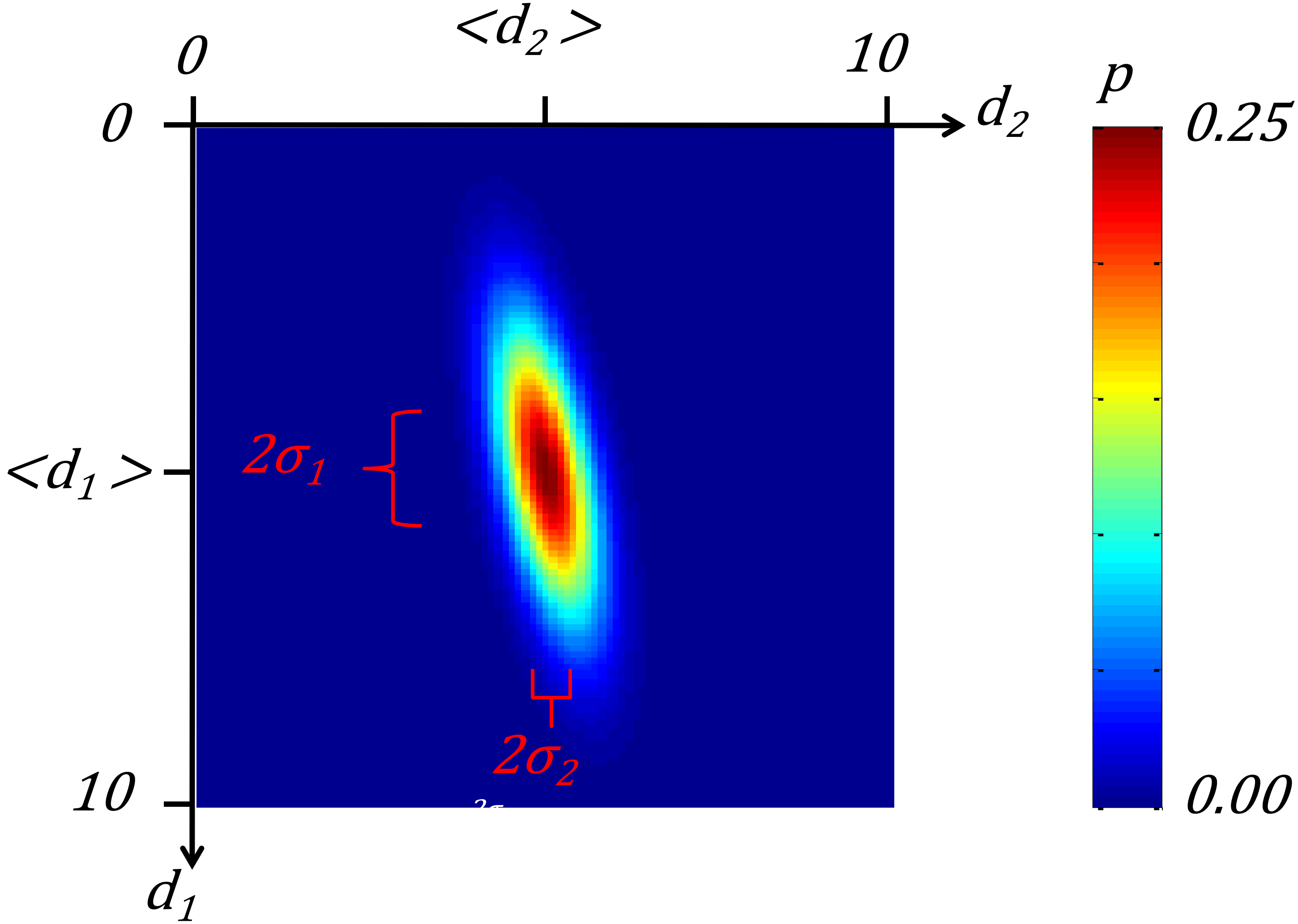
Joint p.d.f. of two data, $p(d_1,d_2)$

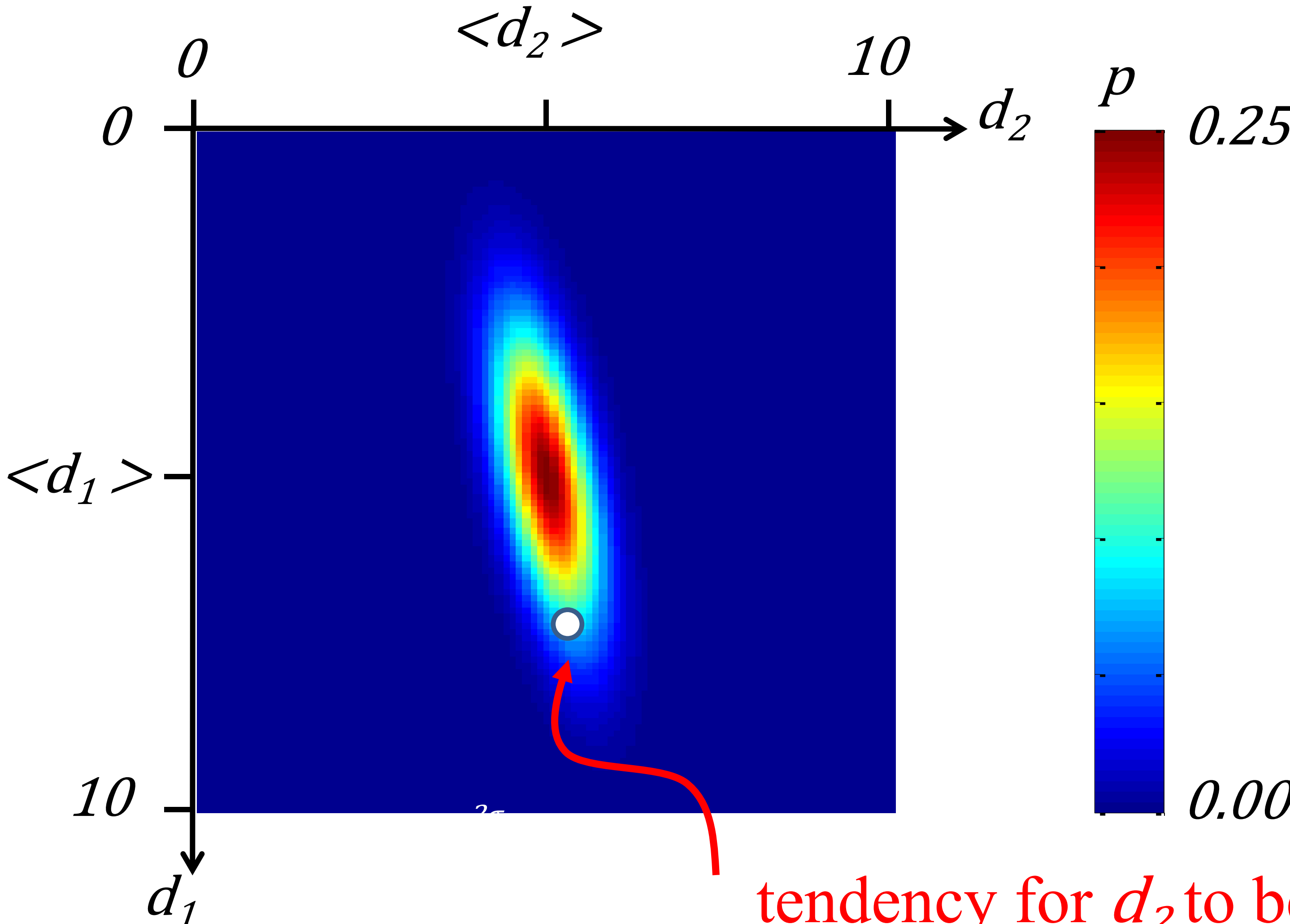


mean values $<d_1>$ and $<d_2>$



variances σ_1^2 and σ_2^2





tendency for d_2 to be high when d_1 is high (positive correlation)

formula for covariance

(covariance is the degree of correlation)

$$\mathrm{cov}(d_1, d_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [d_1 - \langle d_1 \rangle]\,[d_2 - \langle d_2 \rangle]\, p(\mathbf{d})\, \mathrm{d}d_1\, \mathrm{d}d_2$$

Similar to equation for variance

formula for covariance

$$\mathrm{cov}(d_1, d_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [d_1 - \langle d_1 \rangle]\, [d_2 - \langle d_2 \rangle]\, p(\mathbf{d})\, \mathrm{d}d_1\, \mathrm{d}d_2$$

+ positive correlation high d_1 high d_2

- negative correlation high d_1 low d_2

formula for covariance

$$\mathrm{cov}(d_1, d_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [d_1 - \langle d_1 \rangle]\, [d_2 - \langle d_2 \rangle]\, p(\mathbf{d})\, \mathrm{d}d_1\, \mathrm{d}d_2$$

Covariance of data with itself is the variance

joint p.d.f.
mean is a vector
covariance is a symmetric matrix

$$\langle d \rangle_i = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} d_i \, p(\mathbf{d}) \, \mathrm{d}d_1 \cdots \mathrm{d}d_N$$

$$[\mathrm{cov}\,\mathbf{d}]_{ij} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} [d_i - \langle d_i \rangle]\,[d_j - \langle d_j \rangle]\, p(\mathbf{d}) \, \mathrm{d}d_1 \cdots \mathrm{d}d_N$$

diagonal elements: variances
off-diagonal elements: covariances

joint p.d.f.
mean is a vector
covariance is a symmetric matrix

$$\langle d \rangle_i = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} d_i \, p(\mathbf{d}) \, \mathrm{d}d_1 \cdots \mathrm{d}d_N$$

$$[\mathrm{cov}\,\mathbf{d}]_{ij} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} [d_i - \langle d_i \rangle] \, [d_j - \langle d_j \rangle] \, p(\mathbf{d}) \, \mathrm{d}d_1 \cdots \mathrm{d}d_N$$

diagonal elements: variances *(width of data set)*
off-diagonal elements: covariances *(how much data are correlated)*

For uncorrelated data of equal variance
$[\mathrm{cov}\,\mathbf{d}] = \sigma_d^2 \mathbf{I}$

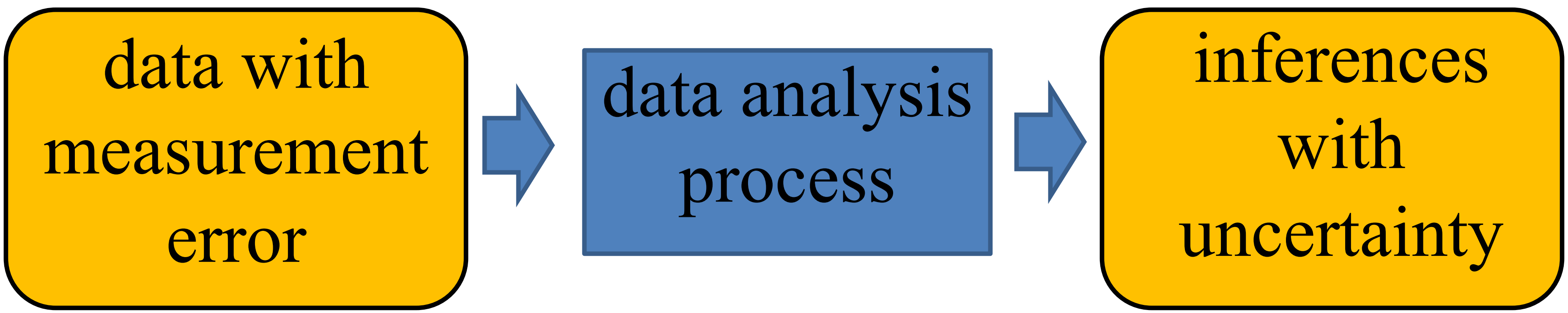
estimating covariance from a table D of data

$$[\text{cov}\ \mathbf{d}]_{ij}^{est} = \frac{1}{K}\sum_{k=1}^{K}\left(D_{ki} - \langle D_i \rangle^{est}\right)\left(D_{kj} - \langle D_j \rangle^{est}\right)$$

D_{ki}: realization k of data-type i

Later we'll show how to use covariance matrix as a weighting matrix in inversion

error in measurement implies uncertainty in inferences



In other words, uncertainty in data leads to uncertainty in model parameters

functions of random variables

given $p(\mathbf{d})$

with $\mathbf{m}=\mathbf{f}(\mathbf{d})$

what is $p(\mathbf{m})$?

rule for error propagation

Given m = Md

$$[\text{cov}\,\mathbf{m}] = \mathbf{M}\,[\text{cov}\,\mathbf{d}]\,\mathbf{M}^{\mathrm{T}}$$

Equation 2.17b

Covariance Matrix

Covariance matrix numerical example

The table below displays scores on math, English, and art tests for 5 students. Note that data from the table is represented in matrix $\mathbf{D}$, where each column in the matrix shows scores on a test and each row shows scores for a student.

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \underset{\mathbf{D}}{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}$$

Given the data represented in matrix $\mathbf{D}$, compute the variance of each test and the covariance between the tests.

http://stattrek.com/matrix-algebra/covariance-matrix.aspx

Covariance Matrix

Covariance matrix numerical example

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}$$

$\mathbf{D}$

mean

variance

$$\langle d \rangle^{est} = \frac{1}{N} \sum_{i=1}^{N} d_i \quad \text{and} \quad (\sigma^2)^{est} = \frac{1}{N-1} \sum_{i=1}^{N} (d_i - \langle d \rangle^{est})^2$$

http://stattrek.com/matrix-algebra/covariance-matrix.aspx

Covariance Matrix

Covariance matrix numerical example

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}$$

D

K = total number of students
i, j = which test
(Cov d)ij = covariance between the ith and jth tests

covariance

$$[\mathrm{cov}\ \mathbf{d}]_{ij}^{est} = \frac{1}{K}\sum_{k=1}^{K}\left(D_{ki} - \langle D_i \rangle^{est}\right)\left(D_{kj} - \langle D_j \rangle^{est}\right)$$

http://stattrek.com/matrix-algebra/covariance-matrix.aspx

Sample covariance from sample of data

$$[\text{cov}\,\mathbf{d}]_{ij}^{est} = \frac{1}{K}\sum_{k=1}^{K}\left(D_{ki} - \langle D_i\rangle^{est}\right)\left(D_{kj} - \langle D_j\rangle^{est}\right)$$

D_{ki}: realization k of data-type i

k is a student (row of the matrix D),
i is a test (column of the matrix D)

in MatLab, `C=cov(D)`

Covariance

For two random variable vectors A and B, the covariance is defined as

$$\operatorname{cov}(A, B) = \frac{1}{N-1} \sum_{i=1}^{N} (A_i - \mu_A)^* (B_i - \mu_B)$$

where μ_A is the mean of A, μ_B is the mean of B, and $*$ denotes the complex conjugate.

The *covariance matrix* of two random variables is the matrix of pair-wise covariance calculations between each variable,

$$C = \begin{pmatrix} \operatorname{cov}(A, A) & \operatorname{cov}(A, B) \\ \operatorname{cov}(B, A) & \operatorname{cov}(B, B) \end{pmatrix}.$$

http://www.mathworks.com/help/matlab/ref/cov.html

Covariance

For two random variable vectors A and B, the covariance is defined as

For our example,
i is the counter for students (N total students)
A, B are the first two tests

$$\mathrm{cov}(A,B) = \frac{1}{N-1}\sum_{i=1}^{N}(A_i - \mu_A)^*(B_i - \mu_B)$$

where μ_A is the mean of A, μ_B is the mean of B, and $*$ denotes the complex conjugate.

The *covariance matrix* of two random variables is the matrix of pair-wise covariance calculations between each variable,

$$C = \begin{pmatrix} \mathrm{cov}(A,A) & \mathrm{cov}(A,B) \\ \mathrm{cov}(B,A) & \mathrm{cov}(B,B) \end{pmatrix}.$$

If we had just 2 tests, test A and test B, the covariance matrix would be 2 x 2. Since we have 3 tests in our example, the covariance matrix will be 3x3.

http://www.mathworks.com/help/matlab/ref/cov.html

The *covariance matrix* of two random variables is the matrix of pair-wise covariance calculations between each variable,

$$C = \begin{pmatrix} \operatorname{cov}(A, A) & \operatorname{cov}(A, B) \\ \operatorname{cov}(B, A) & \operatorname{cov}(B, B) \end{pmatrix}.$$

For a matrix A whose columns are each a random variable made up of observations, the covariance matrix is the pair-wise covariance calculation between each column combination. In other words, $C(i, j) = \operatorname{cov}(A(:, i), A(:, j))$.

http://www.mathworks.com/help/matlab/ref/cov.html

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \underset{\mathbf{D}}{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}$$

$$[\mathrm{cov}\ \mathbf{d}]_{ij}^{est} = \frac{1}{K}\sum_{k=1}^{K}\left(D_{ki} - \langle D_i\rangle^{est}\right)\left(D_{kj} - \langle D_j\rangle^{est}\right)$$

Or if you prefer,

$$\mathrm{cov}(A, B) = \frac{1}{N-1}\sum_{i=1}^{N}(A_i - \mu_A)^*(B_i - \mu_B)$$

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \underset{\mathbf{D}}{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}$$

Let’s calculate the covariance between the English and the Art tests

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \underbrace{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}_{\mathbf{D}}$$

Let’s calculate the covariance between the English and the Art tests

Mean of the English test = (60+90+60+60+30)/5 = 60
Mean of the Art test = (90+30+60+90+30)/5 = 60

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \underbrace{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}_{\mathbf{D}}$$

Let's calculate the covariance between the English and the Art tests

Mean of the English test = (60+90+60+60+30)/5 = 60
Mean of the Art test = (90+30+60+90+30)/5 = 60

Cov(English, Art) = (60-60)(90-60)+(90-60)(30-60)+(60-60)(60-60)+(60-60)(90-60)+(30-60)(30-60)

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \underset{\mathbf{D}}{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}$$

Let's calculate the covariance between the English and the Art tests

Mean of the English test = (60+90+60+60+30)/5 = 60
Mean of the Art test = (90+30+60+90+30)/5 = 60

Cov(English, Art) = (60-60)(90-60)+(90-60)(30-60)+(60-60)(60-60)+(60-60)(90-60)+(30-60)(30-60)
=0+(30)(-30)+0+0+(30)(30) = 0

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}$$

D

$$[\mathrm{cov}\ \mathbf{d}]_{ij}^{est} = \frac{1}{K}\sum_{k=1}^{K}(D_{ki} - \langle D_i\rangle^{est})\ (D_{kj} - \langle D_j\rangle^{est})$$

$$\mathrm{Cov(D)} = \begin{bmatrix} 2520/5 & 1800/5 & 900/5 \\ 1800/5 & 1800/5 & 0/5 \\ 900/5 & 0/5 & 3600/5 \end{bmatrix} = \begin{bmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{bmatrix}$$

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \underbrace{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}_{\mathbf{D}}$$

$$\mathrm{Cov}(D) = \begin{bmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{bmatrix}$$

Which test has the biggest variance?

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \underset{\mathbf{D}}{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}$$

$$\mathrm{Cov(D)} = \begin{bmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{bmatrix}$$

Shown in red along the diagonal, we see the variance of scores for each test. The art test has the biggest variance (720); and the English test, the smallest (360). So we can say that art test scores are more variable than English test scores.

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \underset{\mathbf{D}}{\begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}}$$

$$\text{Cov(D)} = \begin{bmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{bmatrix}$$

The covariance between math and English is positive (360), and the covariance between math and art is positive (180). This means the scores tend to covary in a positive way. As scores on math go up, scores on art and English also tend to go up; and vice versa.

The covariance between English and art, however, is zero. This means there tends to be no predictable relationship between the movement of English and art scores.

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\Rightarrow \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}$$

$$\mathbf{D}$$

$$\text{Cov(D)} = \begin{bmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{bmatrix}$$

If the covariance between any tests had been negative, it would have meant that the test scores on those tests tend to move in opposite directions. That is, students with relatively high scores on the first test would tend to have relatively low scores on the second test.

CHAPTER

3

Solution of the Linear, Gaussian Inverse Problem, Viewpoint 1: The Length Method

OUTLINE

Menke Chapter 3

Chapter 3 Learning Goals

Introduce the concept of prediction error and the norms that quantify it

Develop the Least Squares Solution

Develop the Minimum Length Solution

The Linear Inverse Problem

$$\mathbf{Gm} = \mathbf{d}$$

model parameters

data

data kernel

an estimate of the model parameters can be used to predict the data

$$\mathbf{G}\mathbf{m}^{\text{est}} = \mathbf{d}^{\text{pre}}$$

but the prediction may not match the observed data
(e.g. due to observational error)

$$\mathbf{d}^{\text{pre}} \neq \mathbf{d}^{\text{obs}}$$

this mismatch leads us to define the prediction error (also called misfit)

$$\mathbf{e} = \mathbf{d}^{\text{obs}} - \mathbf{d}^{\text{pre}}$$

when the model parameters exactly predict the data, $\mathbf{e} = 0$

example of prediction error (misfit) for line fit to data

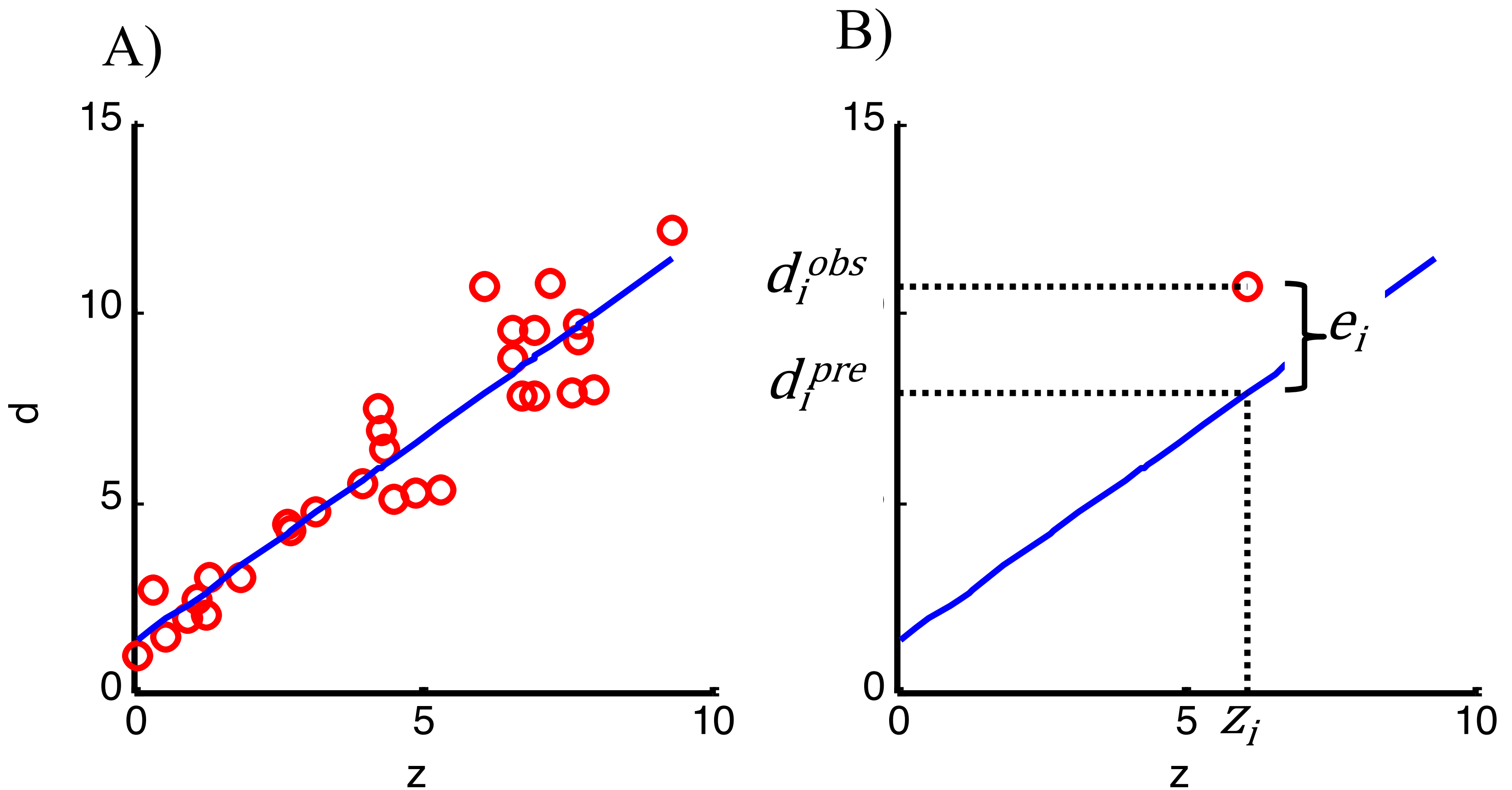


Fig 3.1. A) Least squares fitting of a straight line to *(z, d)* data. B) The error, e_i, for each observation is the difference between the observed and predicted datum: $e_i = d_i^{obs} - d_i^{pre}$.

"norm"

rule for quantifying the overall size of the error vector **e**

lots of possible ways to do it

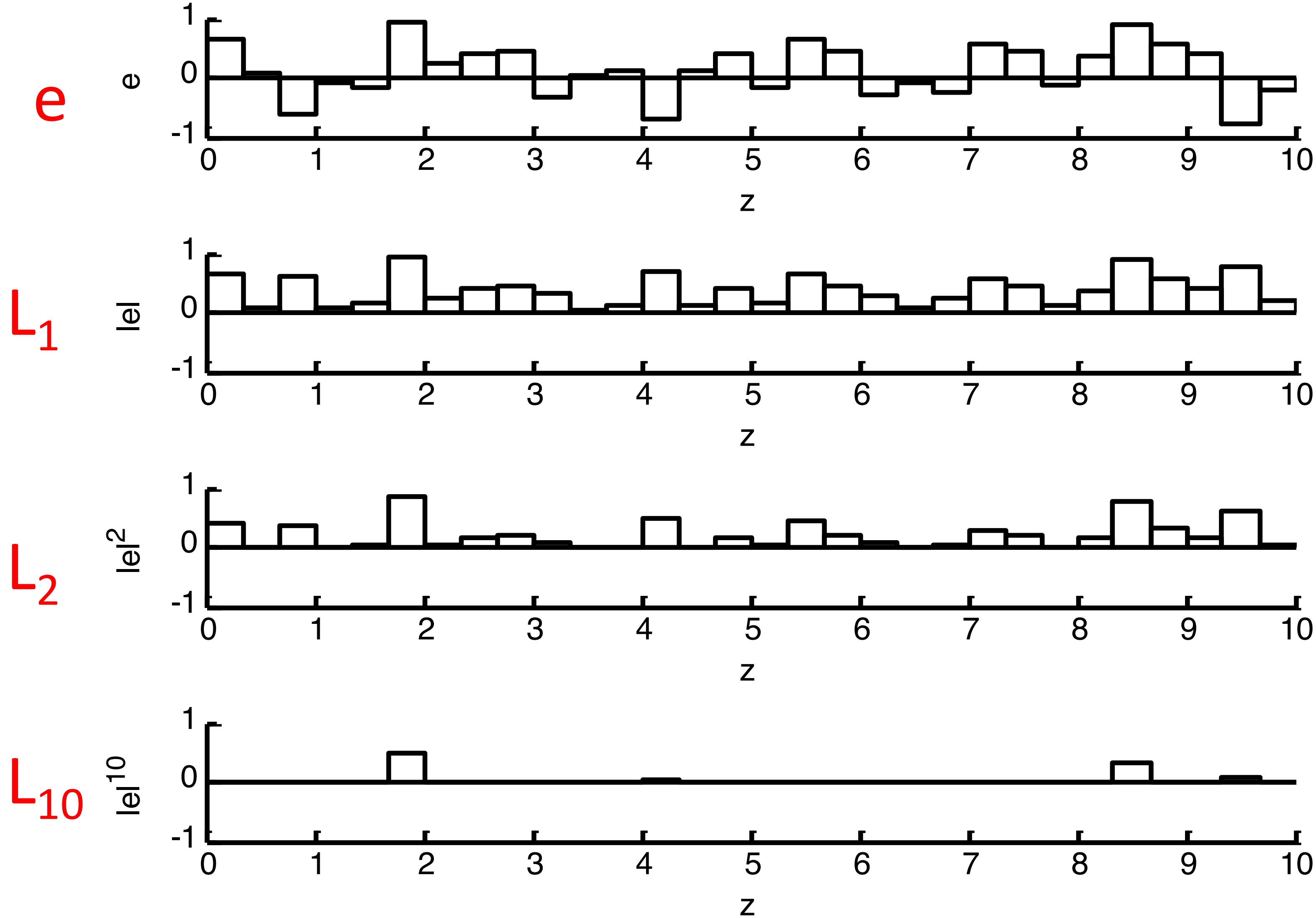
L_n family of norms

$$L_1 \text{ norm:} \quad \|\mathbf{e}\|_1 = \left[\sum_i |e_i|^1\right]$$

$$L_2 \text{ norm:} \quad \|\mathbf{e}\|_2 = \left[\sum_i |e_i|^2\right]^{½}$$

$$L_n \text{ norm:} \quad \|\mathbf{e}\|_n = \left[\sum_i |e_i|^n\right]^{1/n}$$

higher norms give increasing weight to largest element of **e**



guiding principle for solving an inverse problem

find the $\mathbf{m}^{\text{est}}$

that minimizes $E = ||\mathbf{e}||$

Called misfit, error, or cost function

with

$$\mathbf{e} = \mathbf{d}^{\text{obs}} - \mathbf{d}^{\text{pre}}$$

and

$$\mathbf{d}^{\text{pre}} = \mathbf{G}\mathbf{m}^{\text{est}}$$

but which norm to use?

it makes a difference!

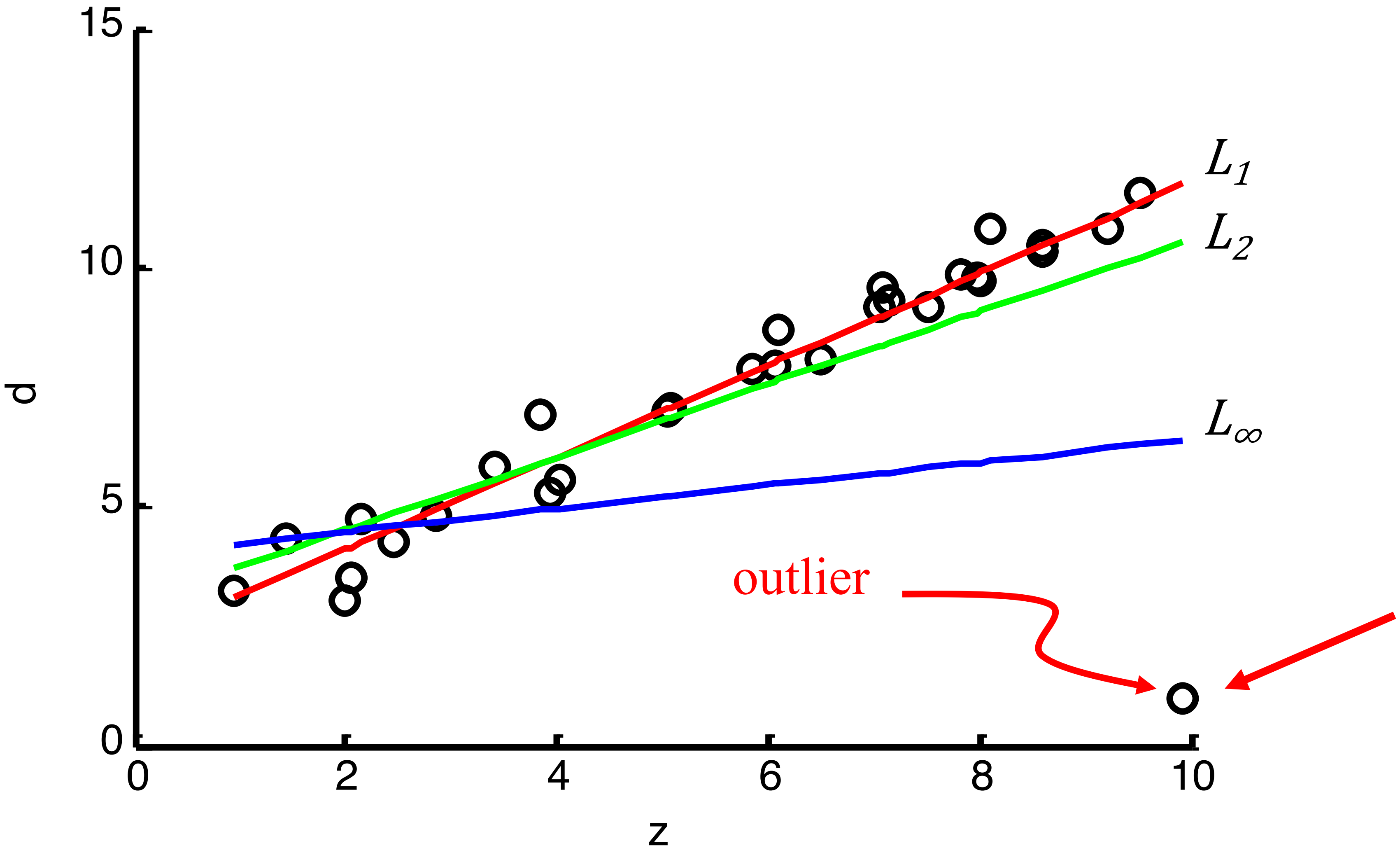


Fig 3.3. Straight line fits to *(z,d)* pairs where the error is measured under the L_1, L_2 and L_∞ norms. The L_1 norm gives the least weight to the one outlier. *MatLab* script gda03_03.

Least Squares Solution to $\mathbf{Gm=d}$

Least Squares = L2 norm

(typically use least squares when data has Gaussian-distributed error)

L_2 norm of error

$$E = \sum_{i=1}^{N} e_i^2 = \mathbf{e}^{\mathrm{T}}\mathbf{e}$$

minimize E

Principle of Least Squares

$$E = \mathbf{e}^T\mathbf{e} = (\mathbf{d} - \mathbf{Gm})^T(\mathbf{d} - \mathbf{Gm}) = \sum_{i=1}^{N}\left[d_i - \sum_{j=1}^{M} G_{ij}m_j\right]\left[d_i - \sum_{k=1}^{M} G_{ik}m_k\right]$$

minimize E with respect to m_q

$$\partial E/\partial m_q = 0$$

Recall $d^{pre}=Gm^{est}$

$e = d^{obs} - d^{pre}$

thus, $e=d^{obs} - Gm^{est}$

$E = e^Te = (d\text{-}Gm)^T(d\text{-}Gm)$

matrix identities:

$E = (d^T\text{-}(Gm)^T)(d\text{-}Gm)$ use $(A+B)^T = A^T+B^T$

$E = (d^T\text{-}m^TG^T)(d\text{-}Gm)$ use $(AB)^T = B^TA^T$

Multiply out

Calculate derivative of E with respect to m and set = 0

(see Menke text p. 45 for details of derivation) eventually get

$$\frac{\partial E}{\partial m_q} = 0 = 2\sum_{k=1}^{M} m_k \sum_{i=1}^{N} G_{iq}G_{ik} - \; -2\sum_{i=1}^{N} G_{iq}d_i$$

or

$$\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{m} - \mathbf{G}^{\mathrm{T}}\mathbf{d} = 0$$

presuming $[\mathbf{G}^{\mathrm{T}}\mathbf{G}]$ has an inverse

Least Square Solution

$$\mathbf{m}^{\mathrm{est}} = [\mathbf{G}^{\mathrm{T}}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{d}$$

example
straight line problem

$$\mathbf{Gm} = \mathbf{d}$$

$$\begin{bmatrix} 1 & z_1 \\ 1 & z_2 \\ \vdots & \vdots \\ 1 & z_N \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} d \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

$$\mathbf{G}^{\mathrm{T}}\mathbf{G} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_N \end{bmatrix} \begin{bmatrix} 1 & z_1 \\ 1 & z_2 \\ \vdots & \vdots \\ 1 & z_N \end{bmatrix} = \begin{bmatrix} N & \sum_{i=1}^{N} z_i \\ \sum_{i=1}^{N} z_i & \sum_{i=1}^{N} z_i^2 \end{bmatrix}$$

$$\mathbf{G}^{\mathrm{T}}\mathbf{d} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_N \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} d_i \\ \sum_{i=1}^{N} d_i z_i \end{bmatrix}$$

$$\mathbf{m}^{\text{est}} = [\mathbf{G}^{\text{T}}\mathbf{G}]^{-1}\mathbf{G}^{\text{T}}\mathbf{d} = \begin{bmatrix} N & \sum_{i=1}^{N} z_i \\ \sum_{i=1}^{N} z_i & \sum_{i=1}^{N} z_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{N} d_i \\ \sum_{i=1}^{N} d_i z_i \end{bmatrix}$$

another example
fitting a plane surface

$$d_i = m_1 + m_2 x_i + m_3 y_i$$

$$\mathbf{Gm} = \mathbf{d}$$

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & y_N \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

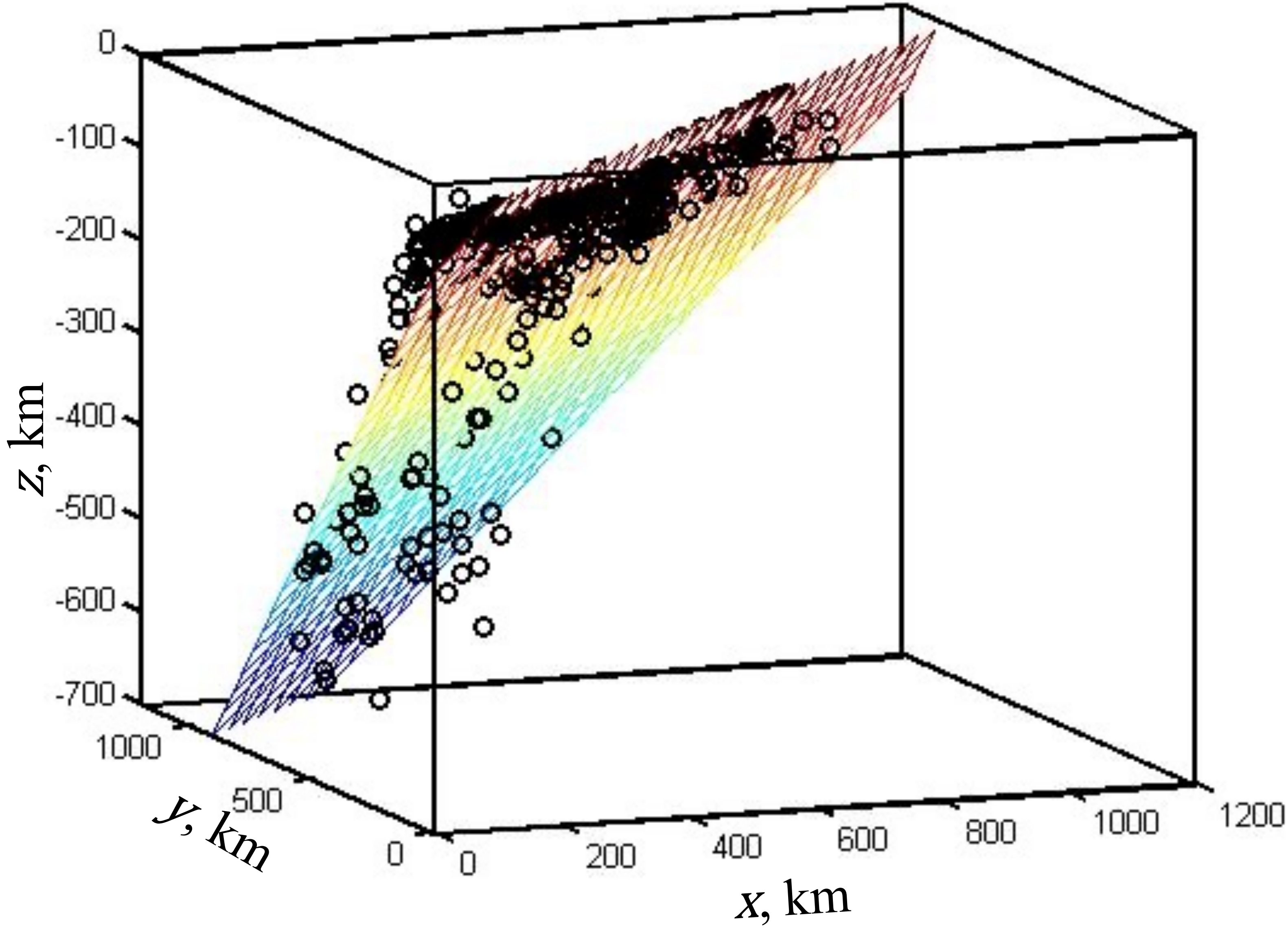


Fig 3.6 (circles) Earthquakes in the Kurile subduction zone, northwest Pacific ocean. The *x*-axis points north and the *y*-axis east. The earthquakes scatter about a dipping planar surface (colored grid), determined using least squares. Data courtesy of the US Geological Survey. *MatLab* script gda03_08.

Next time, what to do

when $[\mathbf{G}^T\mathbf{G}]$ has no inverse

(continue Chapter 3)