Lecture 3

Probability and Measurement Error, Part 2

Syllabus

Lecture 01	Describing Inverse Problems
Lecture 02	Probability and Measurement Error, Part 1
Lecture 03	**Probability and Measurement Error, Part 2**
Lecture 04	The L_2 Norm and Simple Least Squares
Lecture 05	A Priori Information and Weighted Least Squared
Lecture 06	Resolution and Generalized Inverses
Lecture 07	Backus-Gilbert Inverse and the Trade Off of Resolution and Variance
Lecture 08	The Principle of Maximum Likelihood
Lecture 09	Inexact Theories
Lecture 10	Nonuniqueness and Localized Averages
Lecture 11	Vector Spaces and Singular Value Decomposition
Lecture 12	Equality and Inequality Constraints
Lecture 13	L_1 , L_∞ Norm Problems and Linear Programming
Lecture 14	Nonlinear Problems: Grid and Monte Carlo Searches
Lecture 15	Nonlinear Problems: Newton's Method
Lecture 16	Nonlinear Problems: Simulated Annealing and Bootstrap Confidence Intervals
Lecture 17	Factor Analysis
Lecture 18	Varimax Factors, Empircal Orthogonal Functions
Lecture 19	Backus-Gilbert Theory for Continuous Problems; Radon's Problem
Lecture 20	Linear Operators and Their Adjoints
Lecture 21	Fréchet Derivatives
Lecture 22	Exemplary Inverse Problems, incl. Filter Design
Lecture 23	Exemplary Inverse Problems, incl. Earthquake Location
Lecture 24	Exemplary Inverse Problems, incl. Vibrational Problems

Purpose of the Lecture

review key points from last lecture

introduce conditional p.d.f.'s and Bayes theorem

discuss confidence intervals

explore ways to compute realizations of random variables

Part 1

review of the last lecture

Joint probability density functions

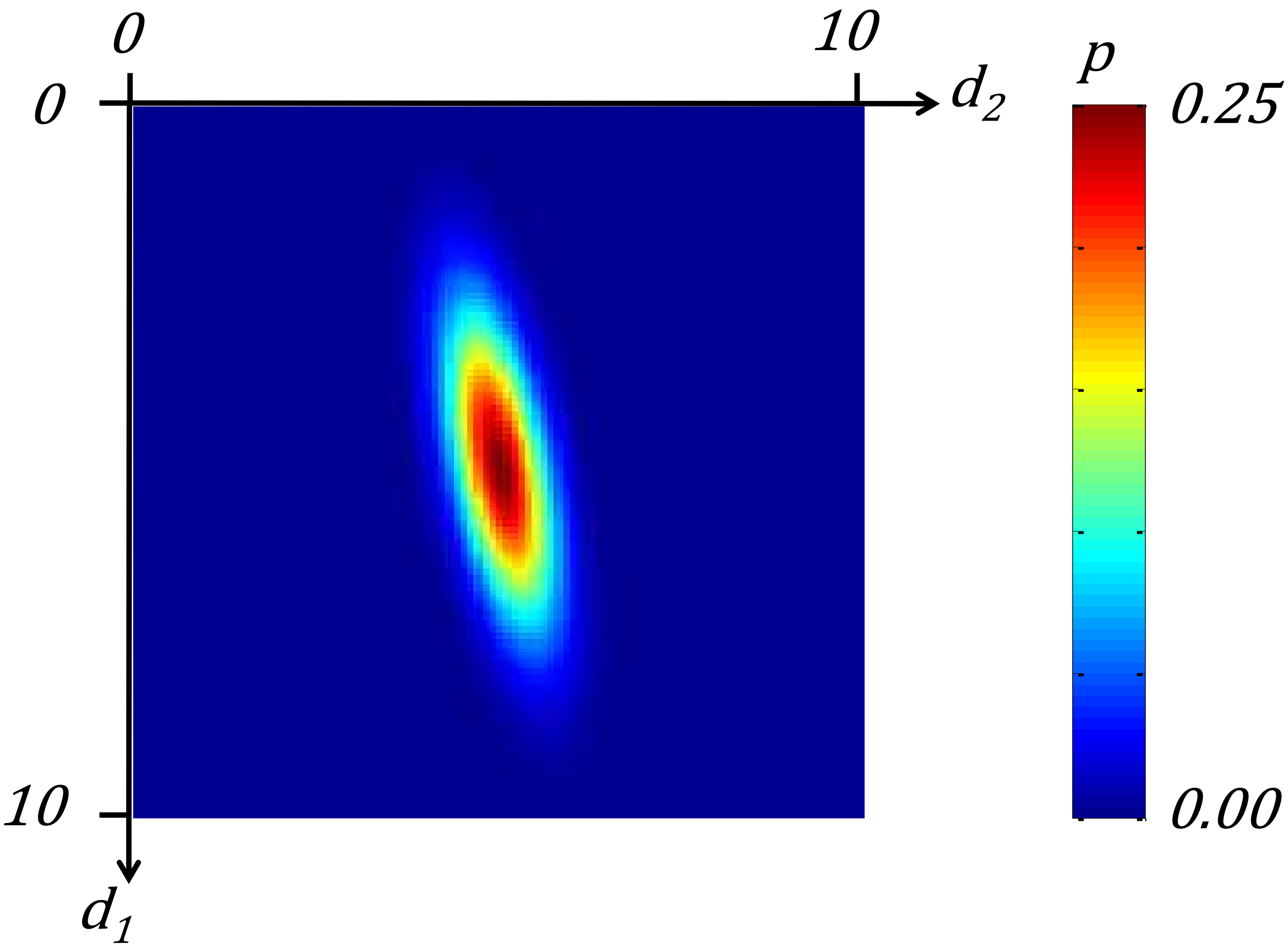
$$p(\mathbf{d}) = p(d_1, d_2, d_3, d_4 \ldots d_N)$$

probability that the data are near **d**

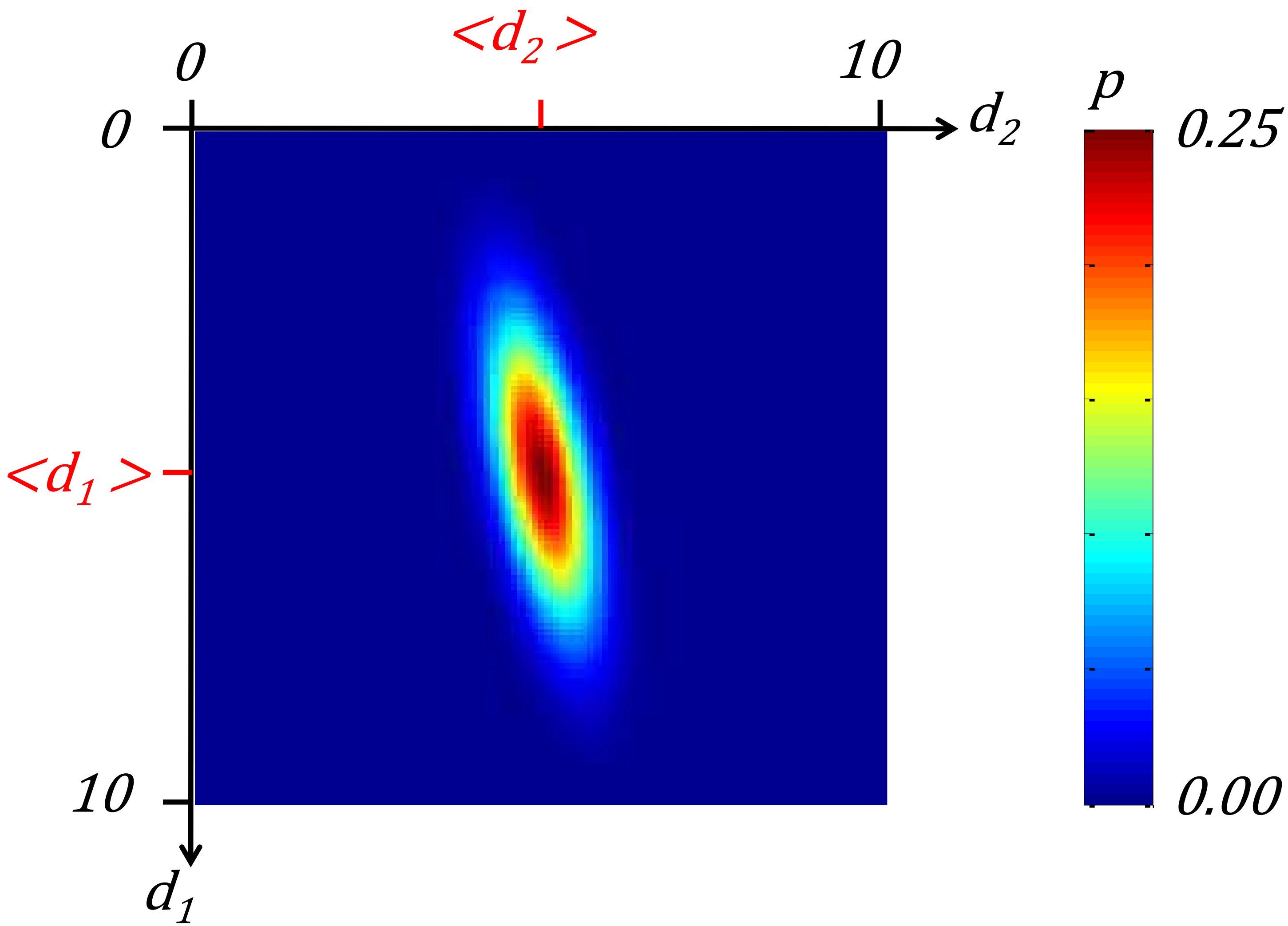
$$p(\mathbf{m}) = p(m_1, m_2, m_3, m_4 \ldots m_M)$$

probability that the model parameters are near **m**

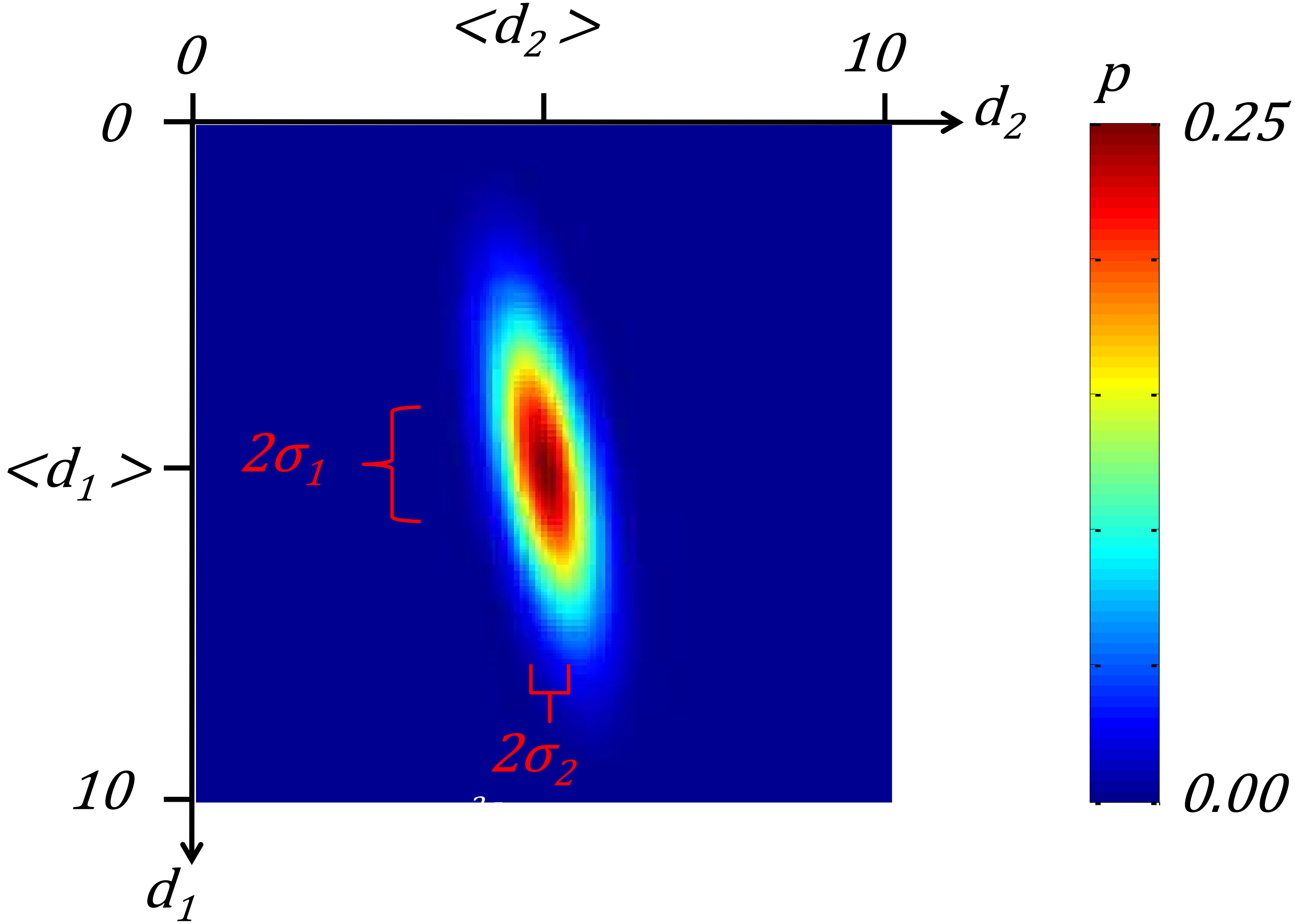
Joint p.d.f. or two data, $p(d_1,d_2)$



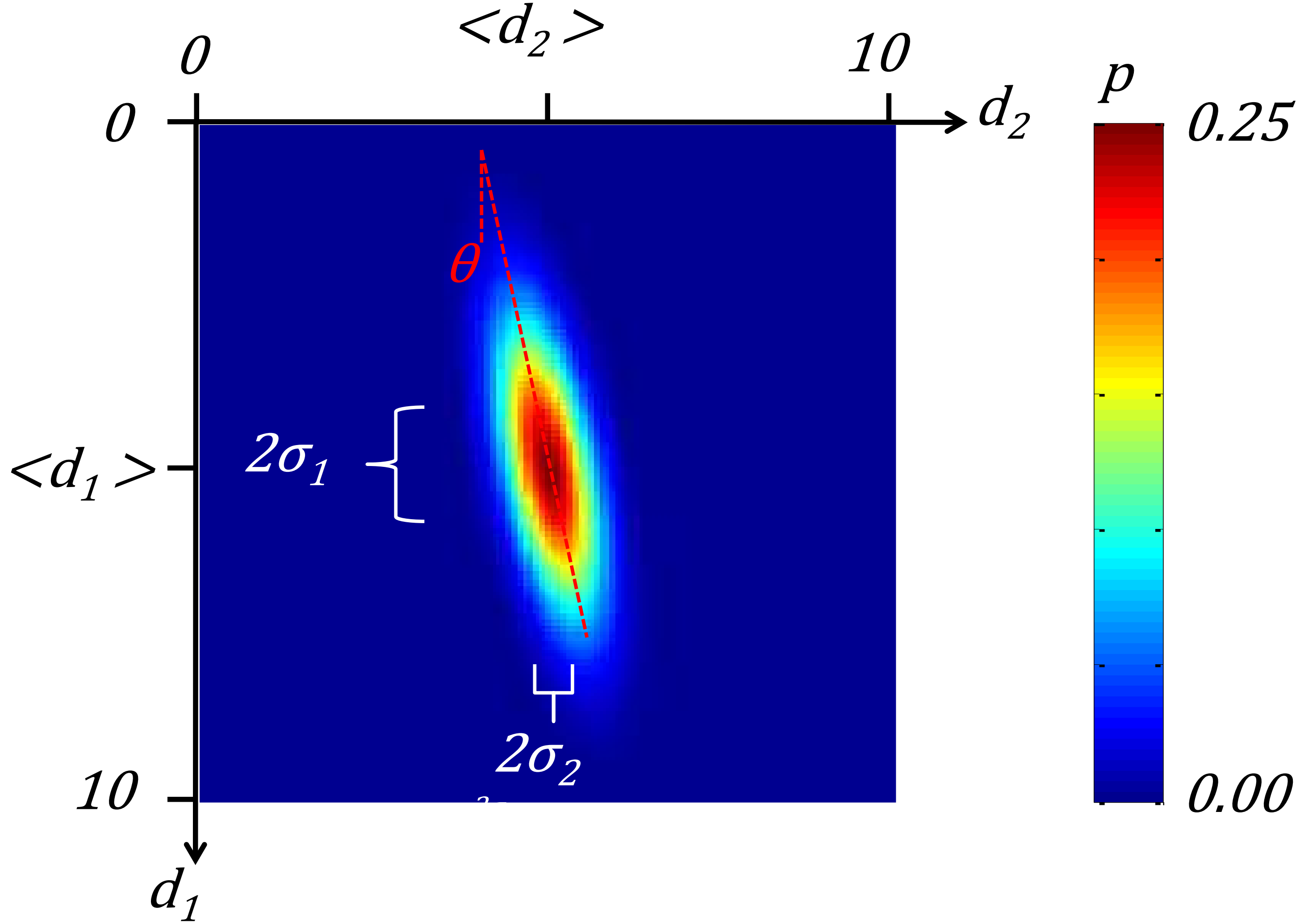
means $<d_1>$ and $<d_2>$



variances σ_1^2 and σ_2^2



covariance – degree of correlation



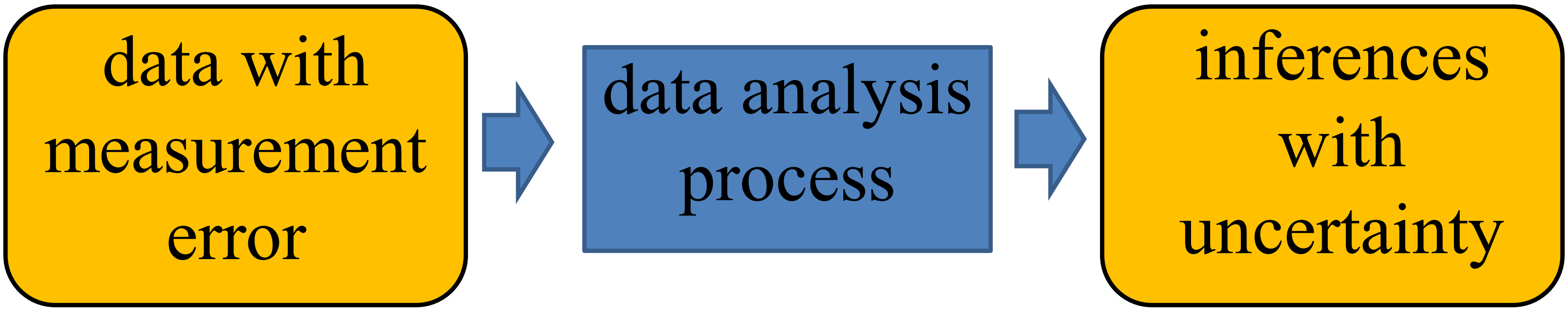
summarizing a joint p.d.f.
mean is a vector
covariance is a symmetric matrix

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

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

diagonal elements: variances
off-diagonal elements: covariances

error in measurement
implies
uncertainty in inferences



functions of random variables

given $p(\mathbf{d})$

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

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

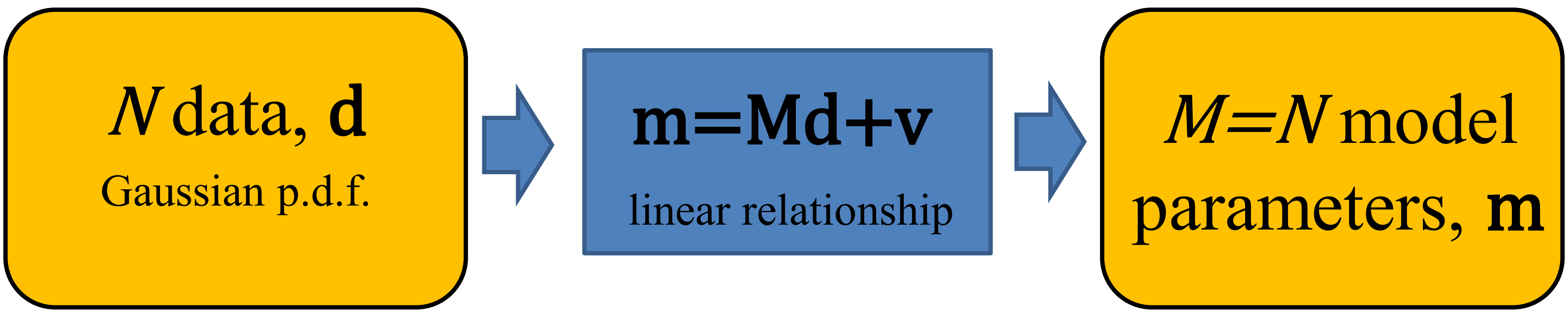
given $p(\mathbf{d})$ and $\mathbf{m}(\mathbf{d})$
then

$$p(\mathbf{m}) = p[\mathbf{d}(\mathbf{m})] \left|\frac{\partial \mathbf{d}}{\partial \mathbf{m}}\right| = p[\mathbf{d}(\mathbf{m})]\, J(\mathbf{m})$$

Jacobian determinant

$$\det \begin{pmatrix} \partial d_1/\partial m_1 & \partial d_1/\partial m_2 & \ldots \\ \partial d_2/\partial m_1 & \partial d_2/\partial m_2 & \ldots \\ \ldots & \ldots & \ldots \end{pmatrix}$$

multivariate Gaussian example



given

$$p(\mathbf{d}) = \frac{1}{(2\pi)^{\frac{N}{2}} \ |[\mathrm{cov}\ \mathbf{d}]|^{½}} \ \exp\left(-½\ [\mathbf{d} - \langle\mathbf{d}\rangle]^{\mathrm{T}}\ [\mathrm{cov}\ \mathbf{d}]^{-1}\ [\mathbf{d} - \langle\mathbf{d}\rangle]\right)$$

and the linear relation

$$\mathbf{m}=\mathbf{M}\mathbf{d}+\mathbf{v}$$

what's $p(\mathbf{m})$?

answer

$$p(\mathbf{m}) = \frac{1}{(2\pi)^{\frac{M}{2}}\,|[\mathrm{cov}\,\mathbf{m}]|^{½}} \exp\!\left(-½[\mathbf{m}-\langle\mathbf{m}\rangle]^{\mathrm{T}}[\mathrm{cov}\,\mathbf{m}]^{-1}[\mathbf{m}-\langle\mathbf{m}\rangle]\right)$$

with

$$\langle\mathbf{m}\rangle = \mathbf{M}\,\langle\mathbf{d}\rangle + \mathbf{v} \qquad \text{and} \qquad [\mathrm{cov}\,\mathbf{m}] = \mathbf{M}\,[\mathrm{cov}\,\mathbf{d}]\,\mathbf{M}^{\mathrm{T}}$$

answer

$$p(\mathbf{m}) = \frac{1}{(2\pi)^{\frac{M}{2}}\,|[\mathrm{cov}\,\mathbf{m}]|^{½}} \exp\left(-½[\mathbf{m} - \langle\mathbf{m}\rangle]^{\mathrm{T}}[\mathrm{cov}\,\mathbf{m}]^{-1}[\mathbf{m} - \langle\mathbf{m}\rangle]\right)$$

also Gaussian

with

$$\langle\mathbf{m}\rangle = \mathbf{M}\,\langle\mathbf{d}\rangle + \mathbf{v} \quad \text{and} \quad [\mathrm{cov}\,\mathbf{m}] = \mathbf{M}\,[\mathrm{cov}\,\mathbf{d}]\,\mathbf{M}^{\mathrm{T}}$$

rule for error propagation

rule for error propagation

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

holds even when $M \neq N$
and for non-Gaussian distributions

rule for error propagation

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

memorize

holds even when $M \neq N$
and for non-Gaussian distributions

example

given

given N uncorrelated Gaussian data with uniform variance $\sigma_d^{\,2}$

and formula for sample mean

$$m_1 = \frac{1}{N}\sum_{i=1}^{N} d_i = \frac{1}{N}\,[1, 1, 1, \cdots, 1]\,\mathbf{d}$$

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

$$[\text{cov}\ \mathbf{d}\ ] = \sigma_d^{\,2}\,\mathbf{I} \quad \text{and}$$

$$\mathbf{M} = [1,\ 1,\ 1,\ \ldots,\ 1]/N$$

$$[\text{cov}\ \mathbf{m}\ ] = \sigma_d^{\,2}\,\mathbf{M}\mathbf{M}^{\mathrm{T}} = \sigma_d^{\,2}\,\mathrm{N}/\mathrm{N}^2 = (\sigma_d^{\,2}/\mathrm{N})\mathbf{I} = \sigma_m^{\,2}\,\mathbf{I}$$

or

$$\sigma_m^{\,2} = (\sigma_d^{\,2}/\mathrm{N})$$

so

error of sample mean
decreases with number of data

$$\sigma_m = \sigma_d / \sqrt{N}$$

decrease is rather slow , though,
because of the square root

Part 2

conditional p.d.f.'s and Bayes theorem

joint p.d.f. $p(d_1,d_2)$

probability that d_1 is near a given value

and

probability that d_2 is near a given value

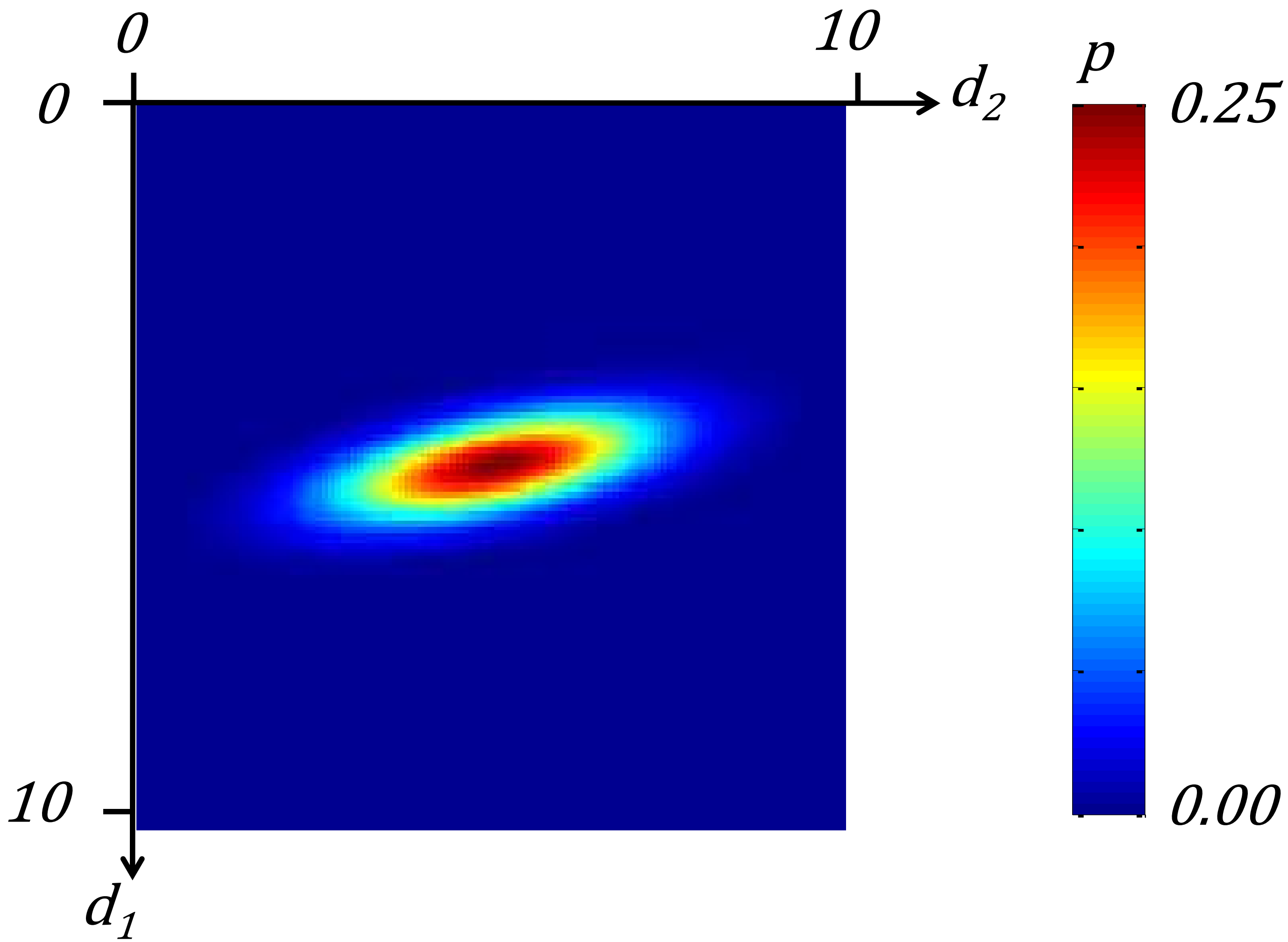
conditional p.d.f. $p(d_1|d_2)$

probability that d_1 is near a given value

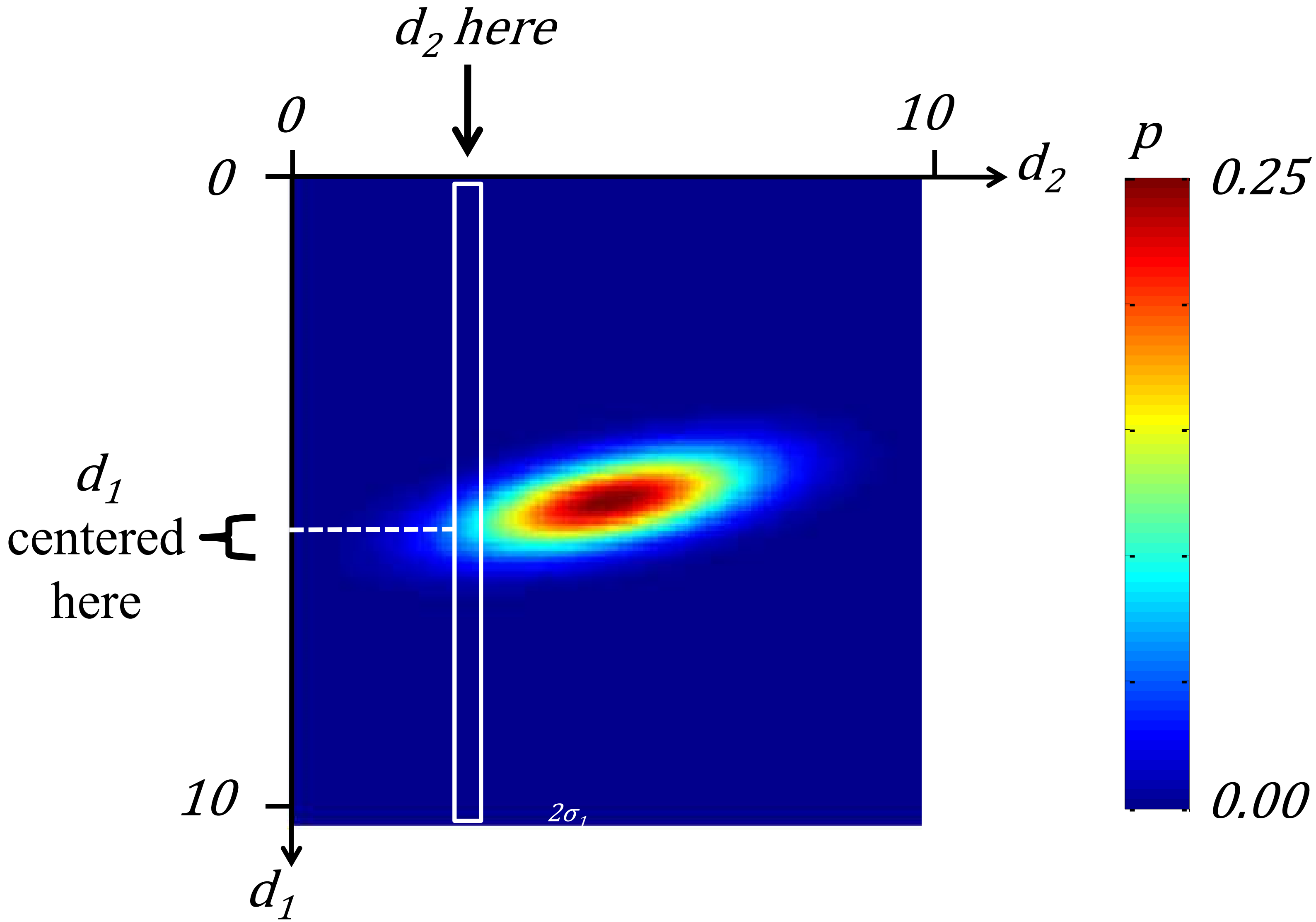
given that we know that

d_2 is near a given value

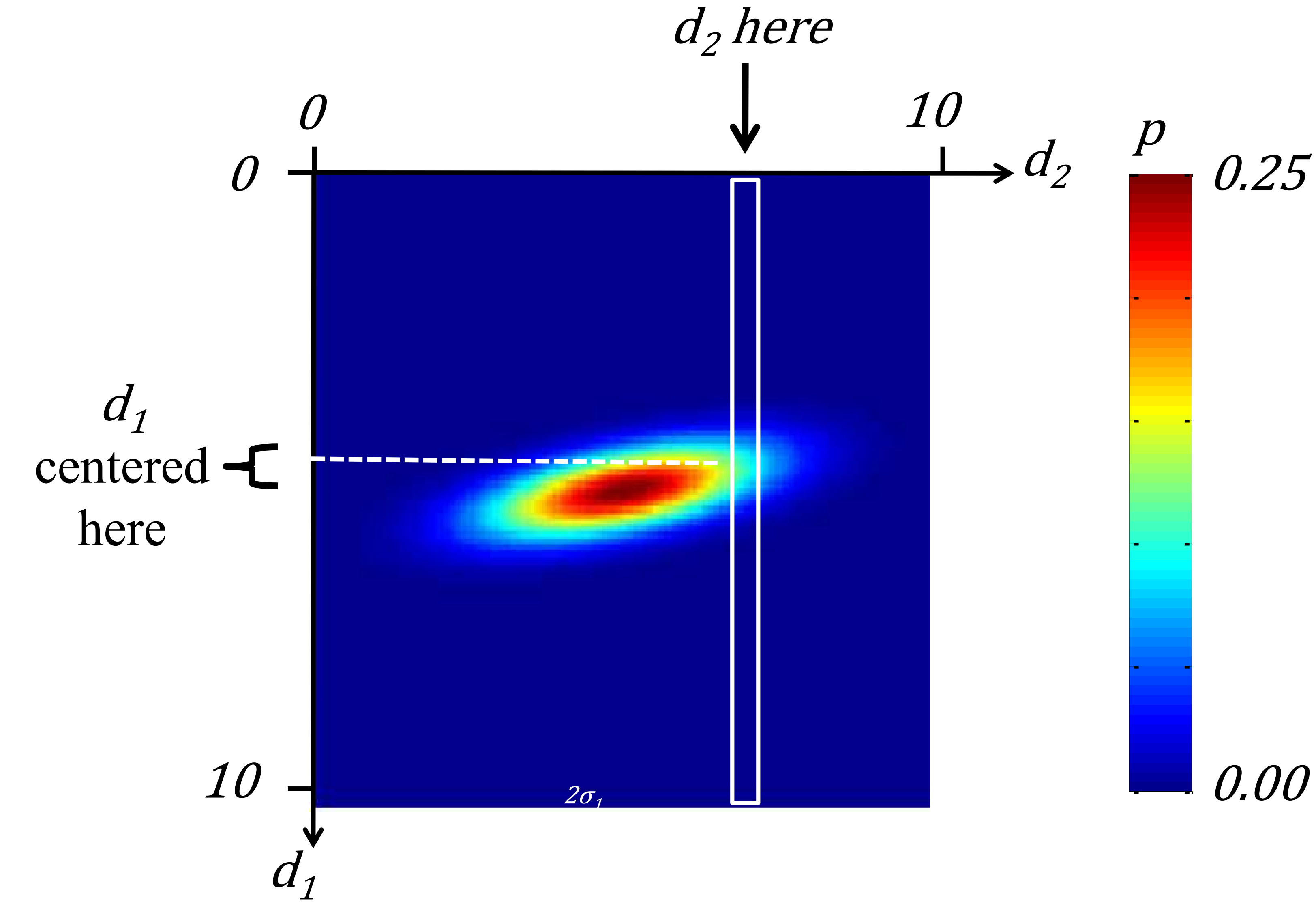
Joint p.d.f.



Joint p.d.f.



Joint p.d.f.



so, to convert a

joint p.d.f. $p(d_1,d_2)$

to a

conditional p.d.f.'s $p(d_1|d_2)$

evaluate the joint p.d.f. at d_2

and

normalize the result to unit area

$$p(d_1|d_2) = \frac{p(d_1, d_2)}{\int p(d_1, d_2)\,\mathrm{d}d_1}$$

$$p(d_1|d_2) = \frac{p(d_1, d_2)}{\int p(d_1, d_2)\, \mathrm{d}d_1}$$

area under
p.d.f. for
fixed d_2

$$p(d_1|d_2) = \frac{p(d_1, d_2)}{\int p(d_1, d_2)\, \mathrm{d}d_1} = \frac{p(d_1, d_2)}{p(d_2)}$$

similarly

conditional p.d.f. $p(d_2|d_1)$

probability that d_2 is near a given value

given that we know that

d_1 is near a given value

$$p(d_2|d_1) = \frac{p(d_1, d_2)}{\int p(d_1, d_2)\, \mathrm{d}d_2} = \frac{p(d_1, d_2)}{p(d_1)}$$

putting both together

$$p(d_1, d_2) = p(d_1|d_2)p(d_2) = p(d_2|d_1)\,p(d_1)$$

rearranging to achieve a result called Bayes theorem

$$p(d_1|d_2) = \frac{p(d_2|d_1)\,p(d_1)}{p(d_2)} = \frac{p(d_2|d_1)\,p(d_1)}{\int p(d_1,d_2)\,\mathrm{d}d_1} = \frac{p(d_2|d_1)\,p(d_1)}{\int p(d_2|d_1)\,p(d_1)\,\mathrm{d}d_1}$$

$$p(d_2|d_1) = \frac{p(d_1|d_2)\,p(d_2)}{p(d_1)} = \frac{p(d_1|d_2)\,p(d_2)}{\int p(d_1,d_2)\,\mathrm{d}d_2} = \frac{p(d_1|d_2)\,p(d_2)}{\int p(d_1|d_2)\,p(d_2)\,\mathrm{d}d_2}$$

rearranging to achieve a result called Bayes theorem

$$p(d_1|d_2) = \frac{p(d_2|d_1)\,p(d_1)}{p(d_2)} = \frac{p(d_2|d_1)\,p(d_1)}{\int p(d_1,d_2)\,\mathrm{d}d_1} = \frac{p(d_2|d_1)\,p(d_1)}{\int p(d_2|d_1)\,p(d_1)\,\mathrm{d}d_1}$$

three alternate ways to write *p(d₂)*

$$p(d_2|d_1) = \frac{p(d_1|d_2)\,p(d_2)}{p(d_1)} = \frac{p(d_1|d_2)\,p(d_2)}{\int p(d_1,d_2)\,\mathrm{d}d_2} = \frac{p(d_1|d_2)\,p(d_2)}{\int p(d_1|d_2)\,p(d_2)\,\mathrm{d}d_2}$$

three alternate ways to write *p(d₁)*

Important

$$p(d_1/d_2) \neq p(d_2/d_1)$$

example
probability that you will die given that you have pancreatic cancer is 90%
(fatality rate of pancreatic cancer is very high)

but
probability that a dead person died of pancreatic cancer is 1.3%
(most people die of something else)

Example using Sand

discrete values
d_1: grain size S=small B=Big
d_2: weight L=Light H=heavy

joint p.d.f.

$$P(d_1, d_2) = \begin{bmatrix} d_1 \backslash d_2 & L & H \\ S & 0.8000 & 0.0010 \\ B & 0.1000 & 0.0990 \end{bmatrix}$$

joint p.d.f.

$$P(d_1, d_2) = \begin{array}{c} d_1 \backslash d_2 \\ S \\ B \end{array} \begin{array}{cc} L & H \\ 0.8000 & 0.0010 \\ 0.1000 & 0.0990 \end{array}$$

univariate p.d.f.'s

$$P(d_1) = \begin{array}{c} d_1 \\ S \\ B \end{array} \begin{array}{c} \\ 0.8010 \\ 0.1990 \end{array} \quad \text{and} \quad P(d_2) = \begin{array}{ccc} d_2 & L & H \\ & 0.9000 & 0.1000 \end{array}$$

joint p.d.f.

$$P(d_1, d_2) = \begin{array}{ccc} d_1\backslash d_2 & L & H \\ S & 0.8000 & 0.0010 \\ B & 0.1000 & 0.0990 \end{array}$$

most grains are small and light

univariate p.d.f.'s

$$P(d_1) = \begin{array}{cc} d_1 & \\ S & 0.8010 \\ B & 0.1990 \end{array} \text{ and } \quad P(d_2) = \begin{array}{ccc} d_2 & L & H \\ & 0.9000 & 0.1000 \end{array}$$

most grains are small

most grains are light

conditional p.d.f.'s

$$P(d_1|d_2) = \begin{bmatrix} d_1 \backslash d_2 & L & H \\ S & 0.8888 & 0.0100 \\ B & 0.1111 & 0.9900 \end{bmatrix} \text{ and } P(d_2|d_1) = \begin{bmatrix} d_1 \backslash d_2 & L & H \\ S & 0.9986 & 0.0012 \\ B & 0.5025 & 0.4974 \end{bmatrix}$$

if a grain is
light it's
probably
small

conditional p.d.f.'s

$$P(d_1|d_2) = \begin{array}{c|cc} d_1\backslash d_2 & L & H \\ \hline S & 0.8888 & 0.0100 \\ B & 0.1111 & 0.9900 \end{array} \quad \text{and} \quad P(d_2|d_1) = \begin{array}{c|cc} d_1\backslash d_2 & L & H \\ \hline S & 0.9986 & 0.0012 \\ B & 0.5025 & 0.4974 \end{array}$$

if a grain is
heavy it's
probably big

conditional p.d.f.'s

$$P(d_1|d_2) = \begin{bmatrix} d_1\backslash d_2 & L & H \\ S & 0.8888 & 0.0100 \\ B & 0.1111 & 0.9900 \end{bmatrix} \text{ and } P(d_2|d_1) = \begin{bmatrix} d_1\backslash d_2 & L & H \\ S & 0.9986 & 0.0012 \\ B & 0.5025 & 0.4974 \end{bmatrix}$$

if a grain is small it's probabilty light

conditional p.d.f.'s

$$P(d_1|d_2) = \begin{bmatrix} d_1\backslash d_2 & L & H \\ S & 0.8888 & 0.0100 \\ B & 0.1111 & 0.9900 \end{bmatrix} \text{ and } P(d_2|d_1) = \begin{bmatrix} d_1\backslash d_2 & L & H \\ S & 0.9986 & 0.0012 \\ B & 0.5025 & 0.4974 \end{bmatrix}$$

if a grain is big
the chance is
about even that
its light or heavy

If a grain is big the chance is about even that its light or heavy
?

What’s going on?

Bayes theorem provides the answer

$$P(H|B) = \frac{P(B|H)\,P(H)}{P(B|L)\,P(L) + P(B|H)\,P(H)} =$$

$$= \frac{0.9900 \times 0.1000}{0.1111 \times 0.9000 + 0.9900 \times 0.1000} = \frac{0.0990}{0.1000 + 0.0990} = \frac{0.0990}{0.1990} = 0.4974$$

Bayes theorem provides the answer

$$P(H|B) = \frac{P(B|H)\,P(H)}{P(B|L)\,P(L) + P(B|H)\,P(H)}$$

probability of a big grain given it's heavy

the probability of a big grain
=
probability of a big grain given it's light
+
probability of a big grain given its heavy

Bayes theorem provides the answer

$$P(H|B) = \frac{P(B|H)\,P(H)}{P(B|L)\,P(L) + P(B|H)\,P(H)} =$$

$$= \frac{0.9900 \times 0.1000}{0.1111 \times 0.9000 + 0.9900 \times 0.1000} = \frac{0.0990}{0.1000 + 0.0990} = \frac{0.0990}{0.1990} = 0.4974$$

only a few percent of light grains are big
but
there are a lot of light grains

this term dominates the result

Bayesian Inference
use observations to *update* probabilities

before the observation: probability that its heavy is 10%, because heavy grains make up 10% of the total.

observation: the grain is big

after the observation: probability that the grain is heavy has risen to 49.74%

Part 2

Confidence Intervals

suppose that we encounter in the literature the result

$m_1 = 50 \pm 2$ (95%) and $m_2 = 30 \pm 1$ (95%)

what does it mean?

joint p.d.f.

$p(m_1,m_2)$

univariate p.d.f.

$p(m_1)$

univariate p.d.f.

$p(m_2)$

compute mean $<m_1>$ and variance σ_1^2

compute mean $<m_2>$ and variance σ_2^2

$m_1 = 50 \pm 2$ (95%) and $m_2 = 30 \pm 1$ (95%)

$<m_1>$ $2\sigma_1$ $<m_2>$ $2\sigma_2$

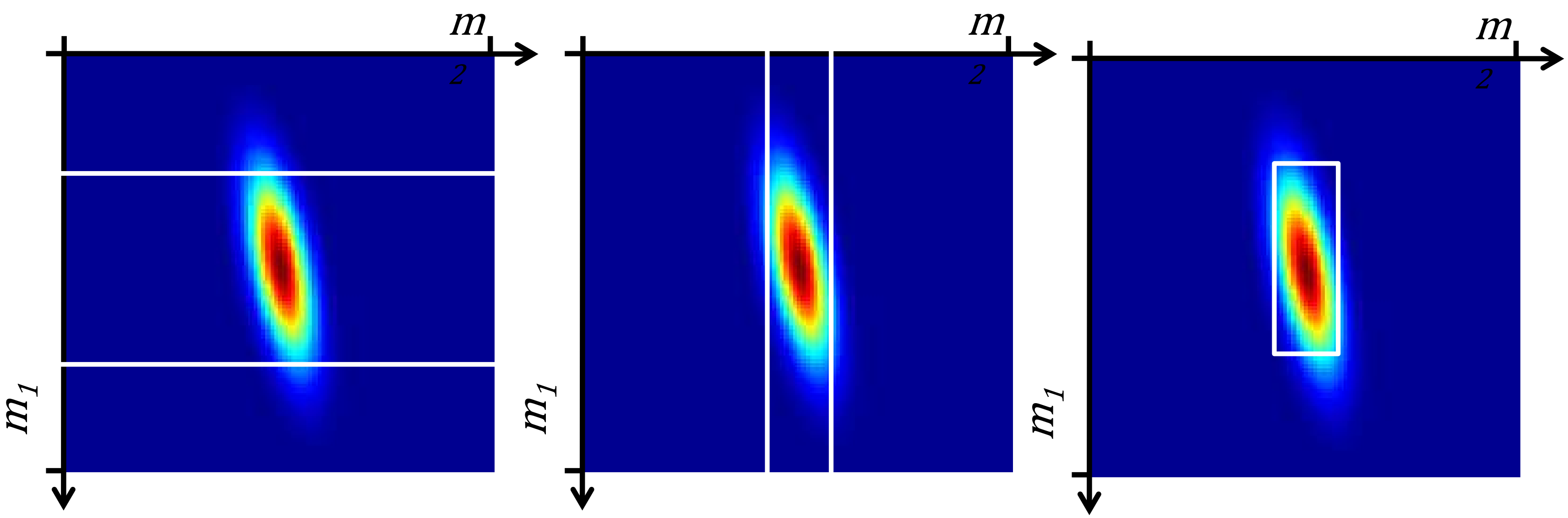
irrespective of the value of m_2, there is a 95% chance that m_1 is between 48 and 52,

$$m_1 = 50 \pm 2 \ (95\%) \quad \text{and} \quad m_2 = 30 \pm 1 \ (95\%)$$

irrespective of the value of m_1, there is a 95% chance that m_1 is between 29 and 31, So what's the probability that both m_1 and m_2 are within 2σ of their means?

That will depend upon the degree of correlation

For uncorrelated model parameters, it's $(0.95)^2 = 0.90$



$$m_1 = \langle m_1 \rangle \pm 2\sigma_1$$

$$m_2 = \langle m_2 \rangle \pm 2\sigma_2$$

$$m_1 = \langle m_1 \rangle \pm 2\sigma_1$$

and

$$m_2 = \langle m_2 \rangle \pm 2\sigma_2$$

Suppose that you read a paper which states values and confidence limits for 100 model parameters

What's the probability that they all fall within their 2σ bounds?

Part 4

computing realizations of random variables

Why?

create noisy "synthetic" or "test" data

generate a suite of hypothetical models, all different from one another

MatLab function **`random()`**

```
m = random('Normal',mbar,sigma,N,1);
```

can do many different p.d.f's

But what do you do if *MatLab* doesn't have the one you need?

One possibility is to use the Metropolis-Hasting algorithm

It requires that you:

1) evaluate the formula for $p(d)$

2) already have a way to generate realizations of Gaussian and Uniform p.d.f.'s

goal: generate a length N vector $\mathbf{d}$ that contains realizations of $p(d)$

steps:

set d_i with $i=1$ to some reasonable value

now for subsequent d_{i+1}

generate a proposed successor d'

from a conditional p.d.f. $q(d'|d_i)$

that returns a value near d_i

generate a number α from a uniform

p.d.f. on the interval (0,1)

accept d' as d_{i+1} if $\alpha < \dfrac{p(d')\, q(d_i|d')}{p(d_i)\, q(d'|d_i)}$

else set $d_{i+1} = d_i$

repeat

A commonly used choice for the conditional p.d.f. is

$$q(d'|d) = \frac{1}{(2\pi)^{½}\,\sigma}\exp\left\{-\frac{(d'-d_i)^2}{2\sigma^2}\right\}$$

here σ is chosen to represent the sixe of the neighborhood, the typical distance of d_{i+1} from d_i

example
exponential p.d.f.

$$p(d) = \tfrac{1}{2}c\,\exp(-|d|/c)$$

Histogram of 5000 realizations