Lecture 2

Probability and Measurement Error, Part 1

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 random variables and their probability density functions

introduce correlation and the multivariate Gaussian distribution

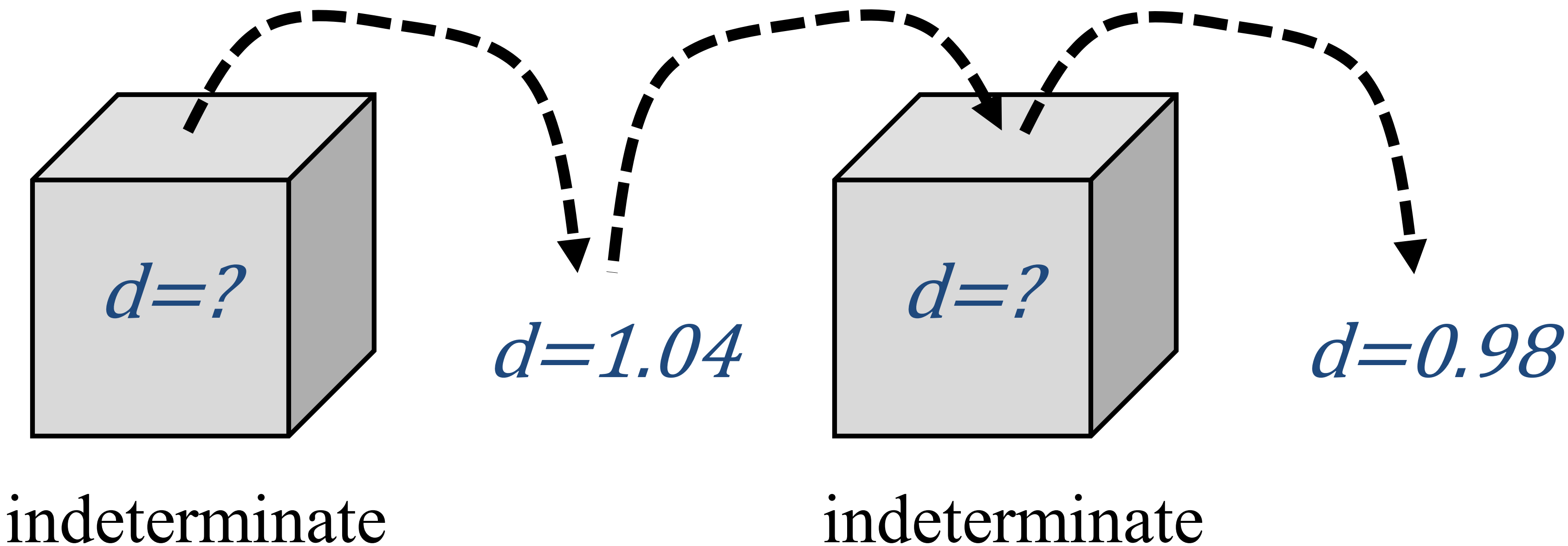
relate error propagation to functions of random variables

Part 1

random variables and their probability density functions

random variable, *d*

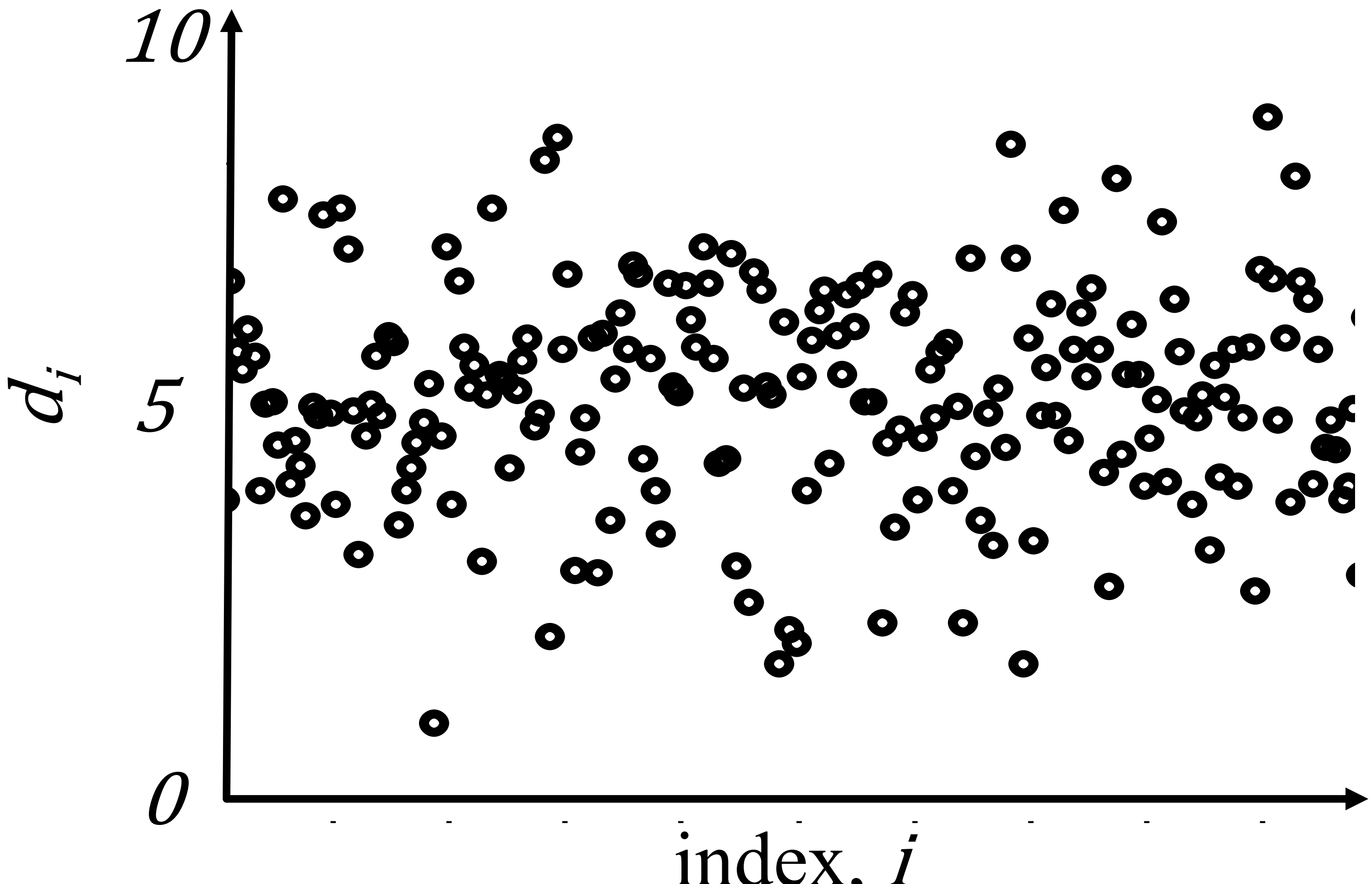
no fixed value until it is *realized*

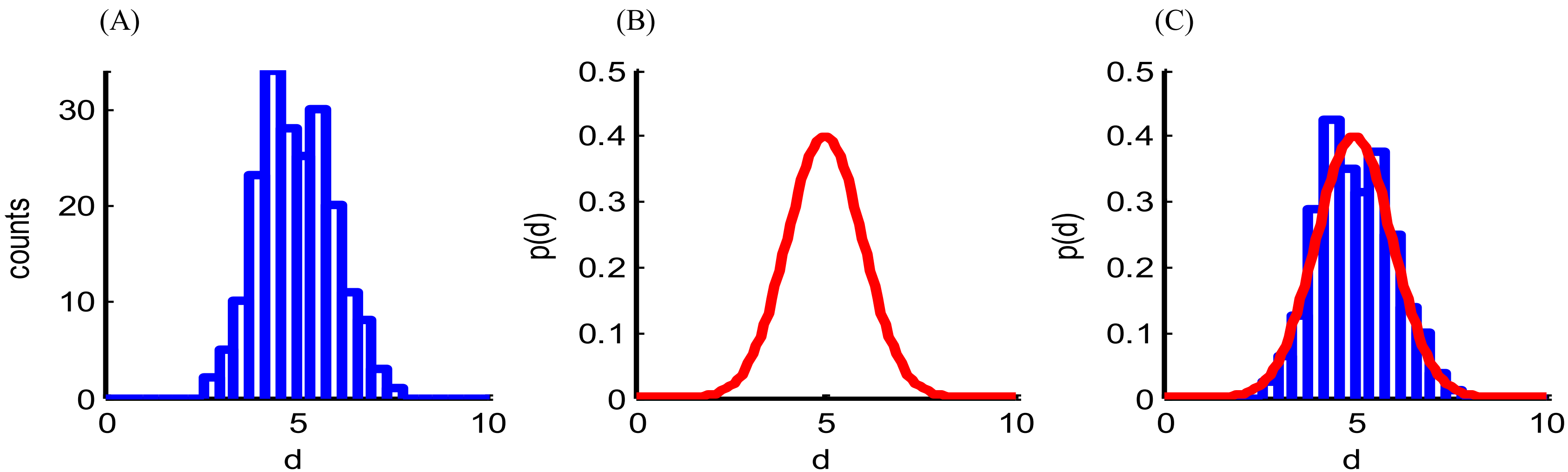


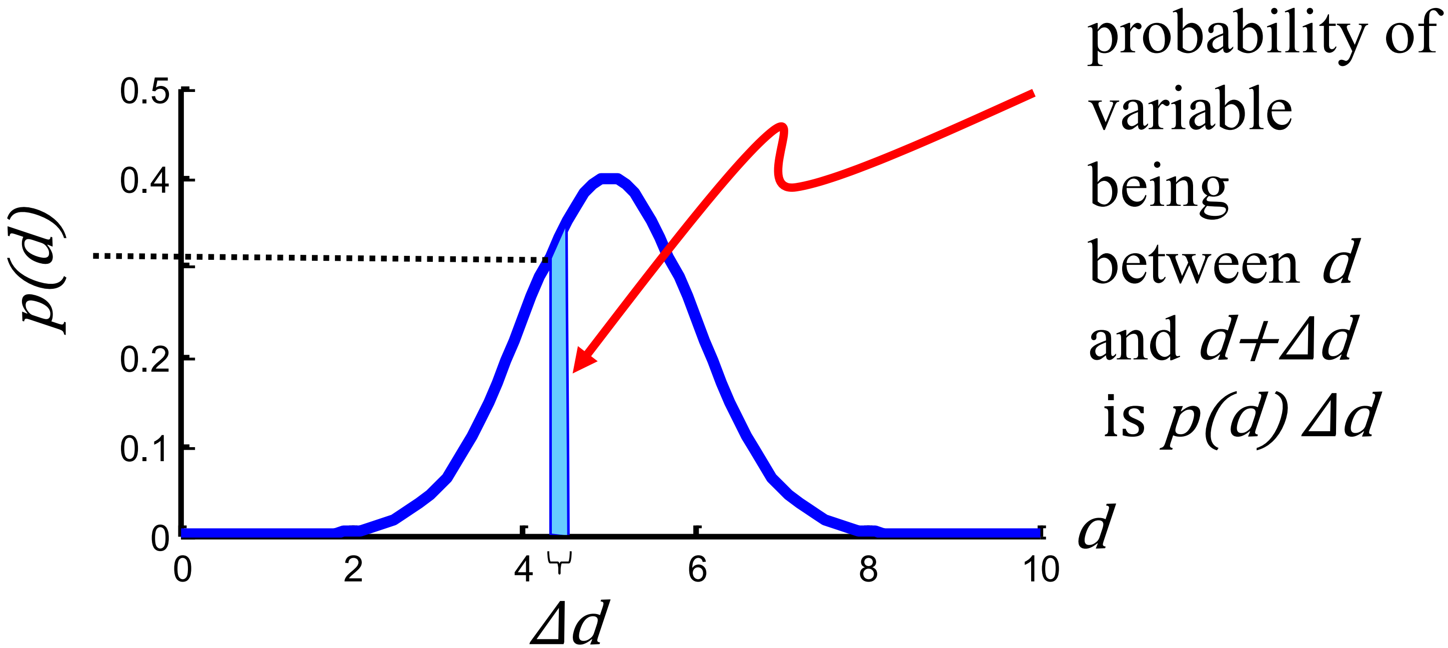
random variables have systematics

tendency to takes on some values more often than others

N realization of data




(A)
(B)
(C)
counts
30
20
10
0
0
5
10
d
p(d)
0.5
0.4
0.3
0.2
0.1
0
0
5
10
d
p(d)
0.5
0.4
0.3
0.2
0.1
0
0
5
10
d



probability of variable being between d and $d+\Delta d$ is $p(d)\,\Delta d$
$p(d)$
0.5
0.4
0.2
0.1
0
0
2
4
6
8
10
d
Δd

in general
probability is the integral

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

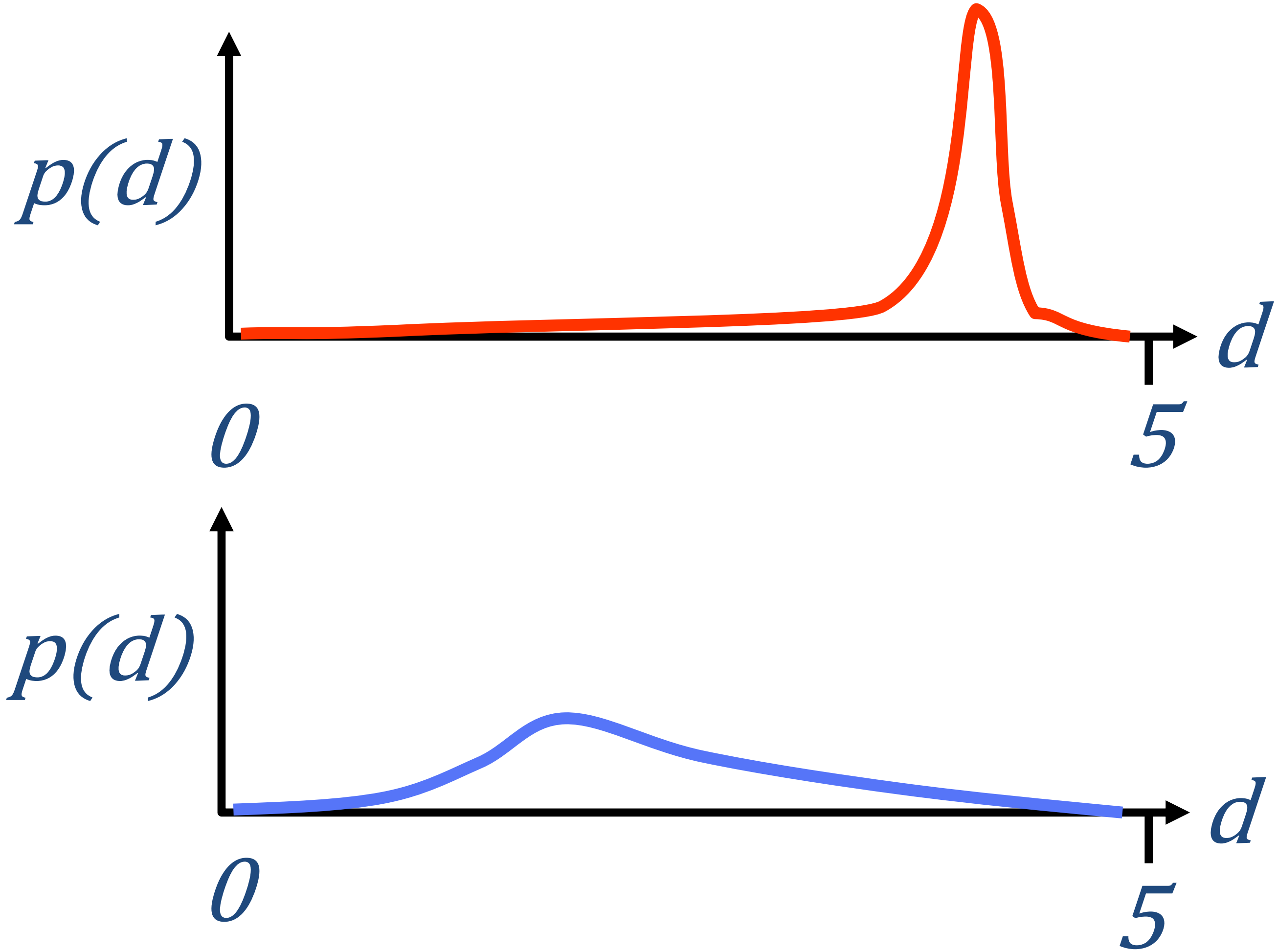
probability that d is between d_1 and d_2

the probability that *d* has some value is 100% or unity

$$P(d_{min}, d_{max}) = \int_{d_{min}}^{d_{max}} p(d)\,\mathrm{d}d = 1$$

probability that *d* is between its minimum and maximum bounds, d_{min} and d_{max}

How do these two p.d.f.'s differ?



Summarizing a probability density function

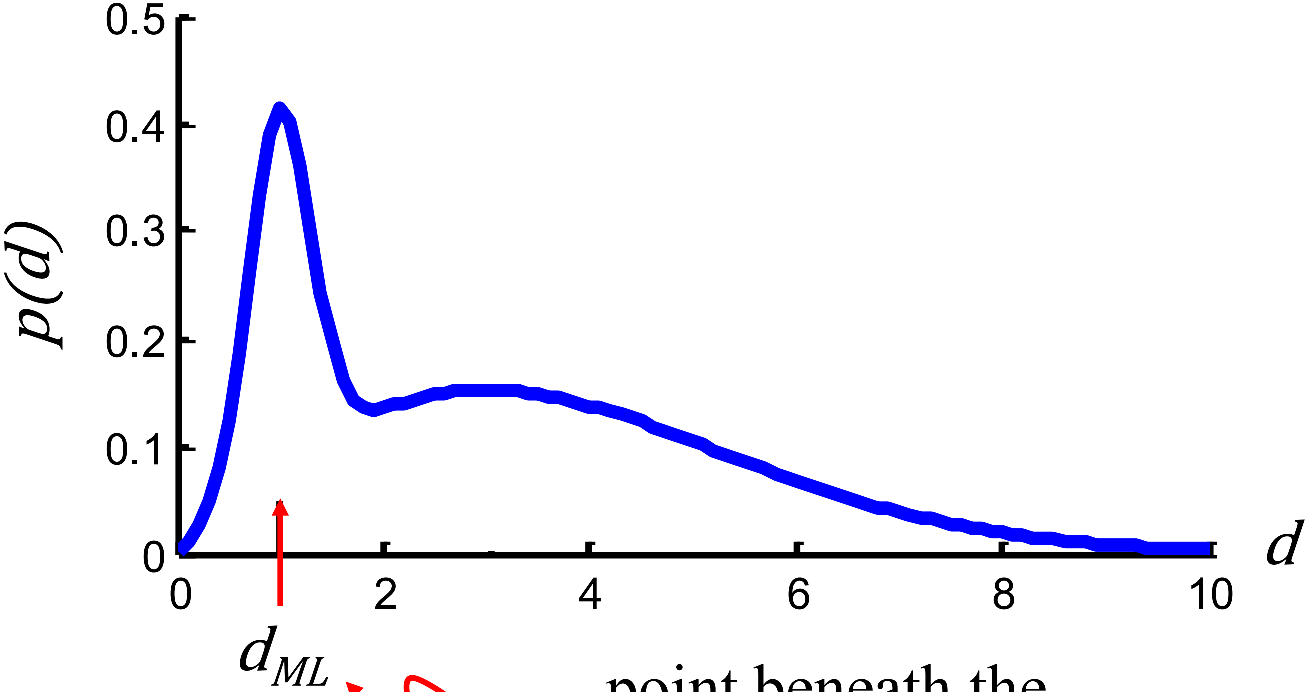
typical value

  “center of the p.d.f.”

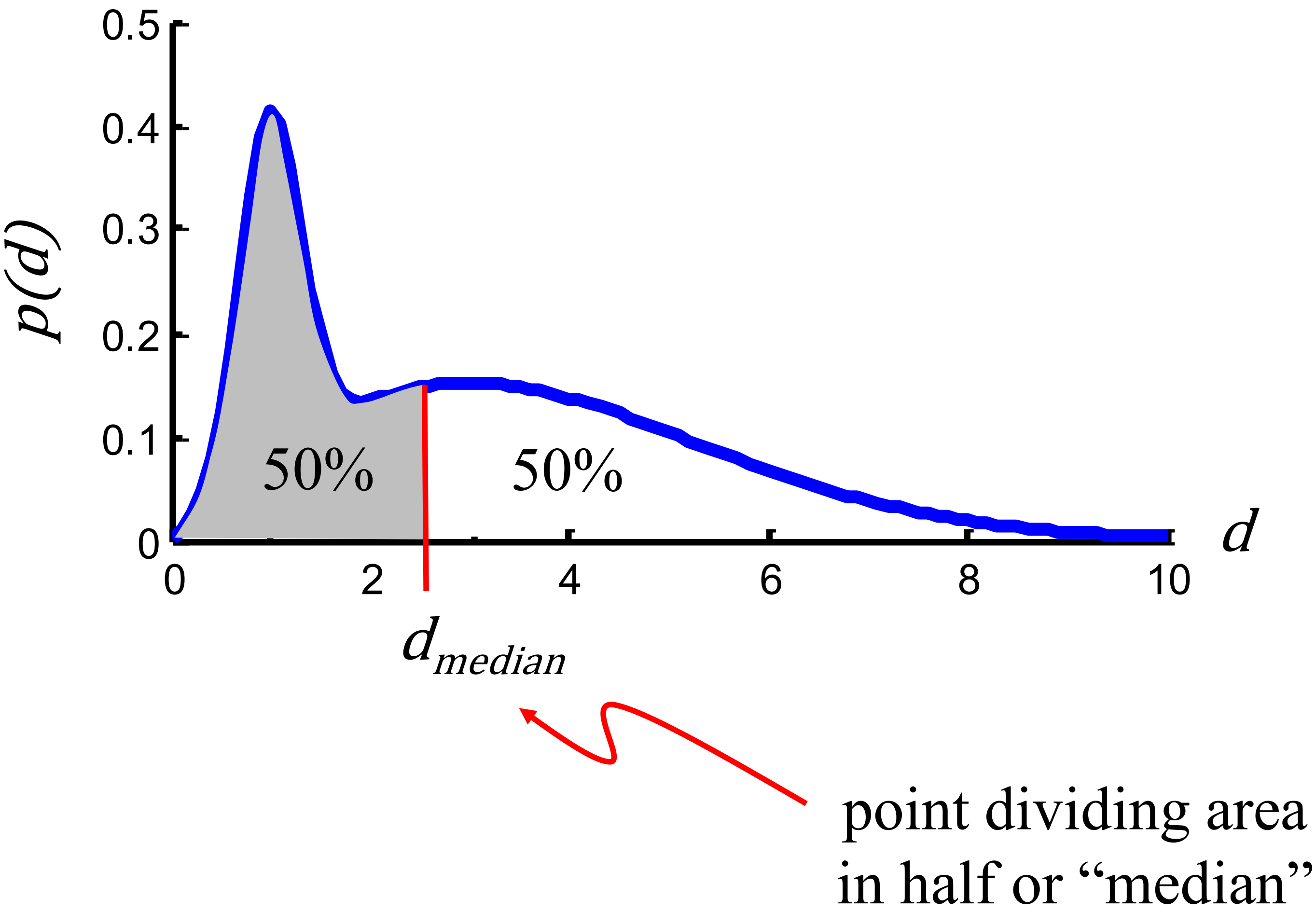
amount of scatter around the typical value

  “width of the p.d.f.”

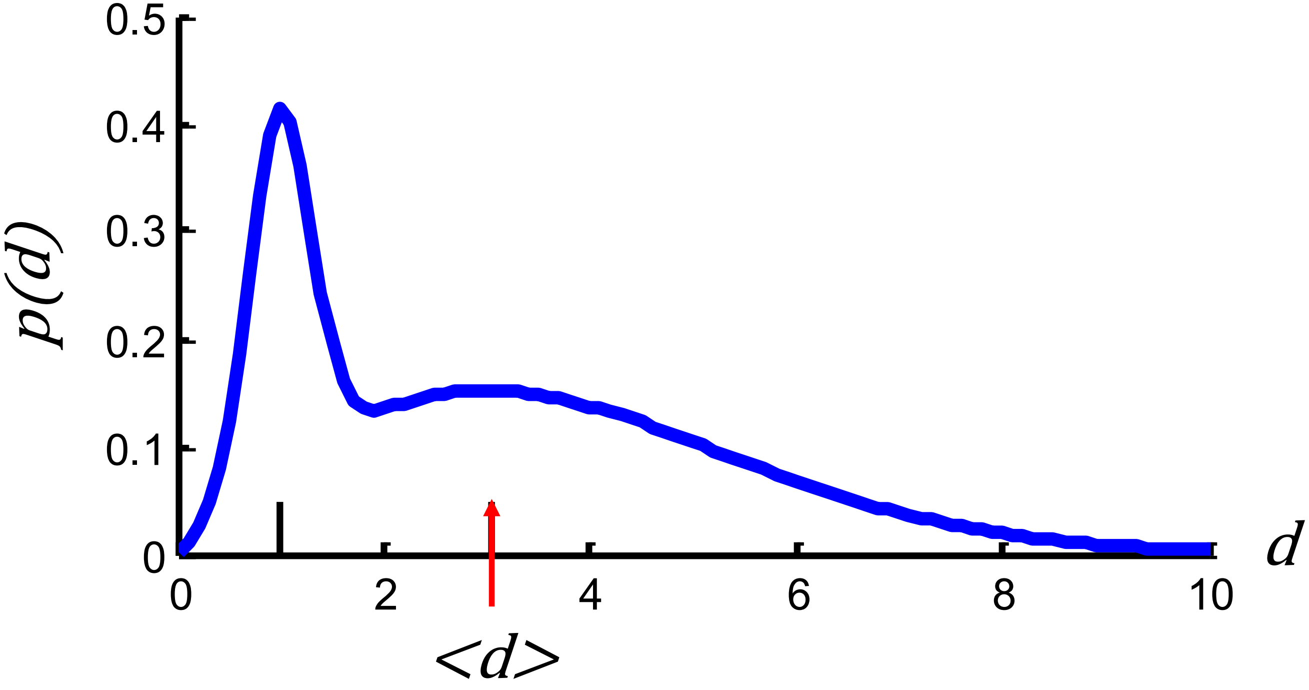
Several possibilities for a typical value



point beneath the peak or "maximum likelihood point" or "mode"



p(d)
0.5
0.4
0.3
0.2
0.1
0
50%
50%
0
2
4
6
8
10
d
d_{median}
point dividing area
in half or "median"



balancing point or
"mean" or "expectation"

can all be different

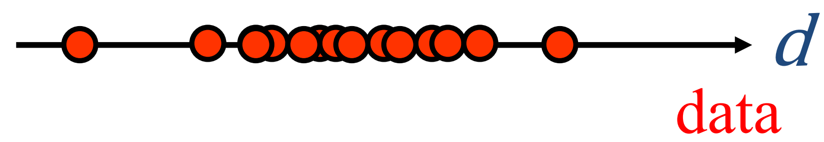
$$d_{ML} \neq d_{median} \neq \langle d \rangle$$

formula for "mean" or "expected value"

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

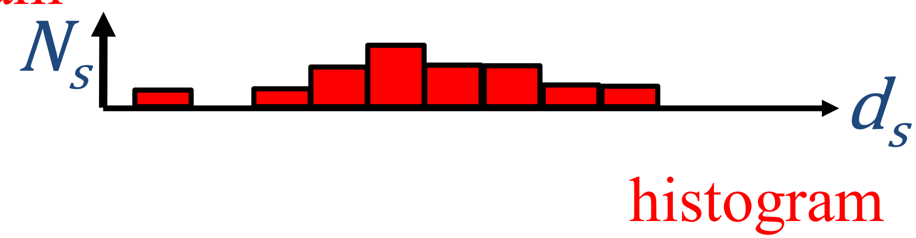
step 1: usual formula for mean

$$<d> = \frac{1}{N}\sum_{i=0}^{N} d_i$$



step 2: replace data with its histogram

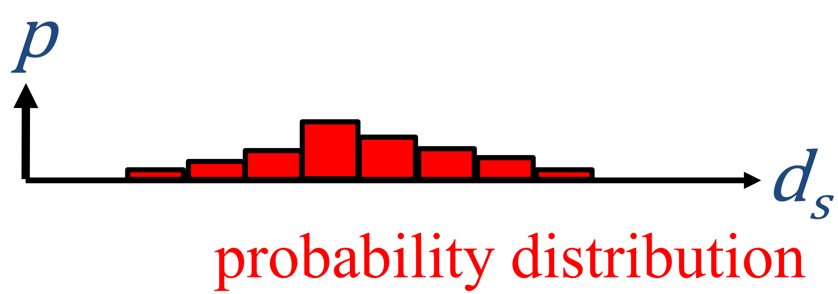
$$<d> \approx \frac{1}{N}\sum_{s=0}^{M} d^{(s)} N_s$$



step 3: replace histogram with probability distribution.

$$<d> \approx \sum_{s=0}^{M} d^{(s)} \frac{N_s}{N}$$

$$\approx \sum_{s=0}^{M} d^{(s)} P(d_s)$$



If the data are continuous, use analogous formula containing an integral:

$$<d> \approx \quad \sum_{s=0}^{M} d^{(s)} \; P(d_s)$$

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

quantifying width

This function grows away from the typical value:

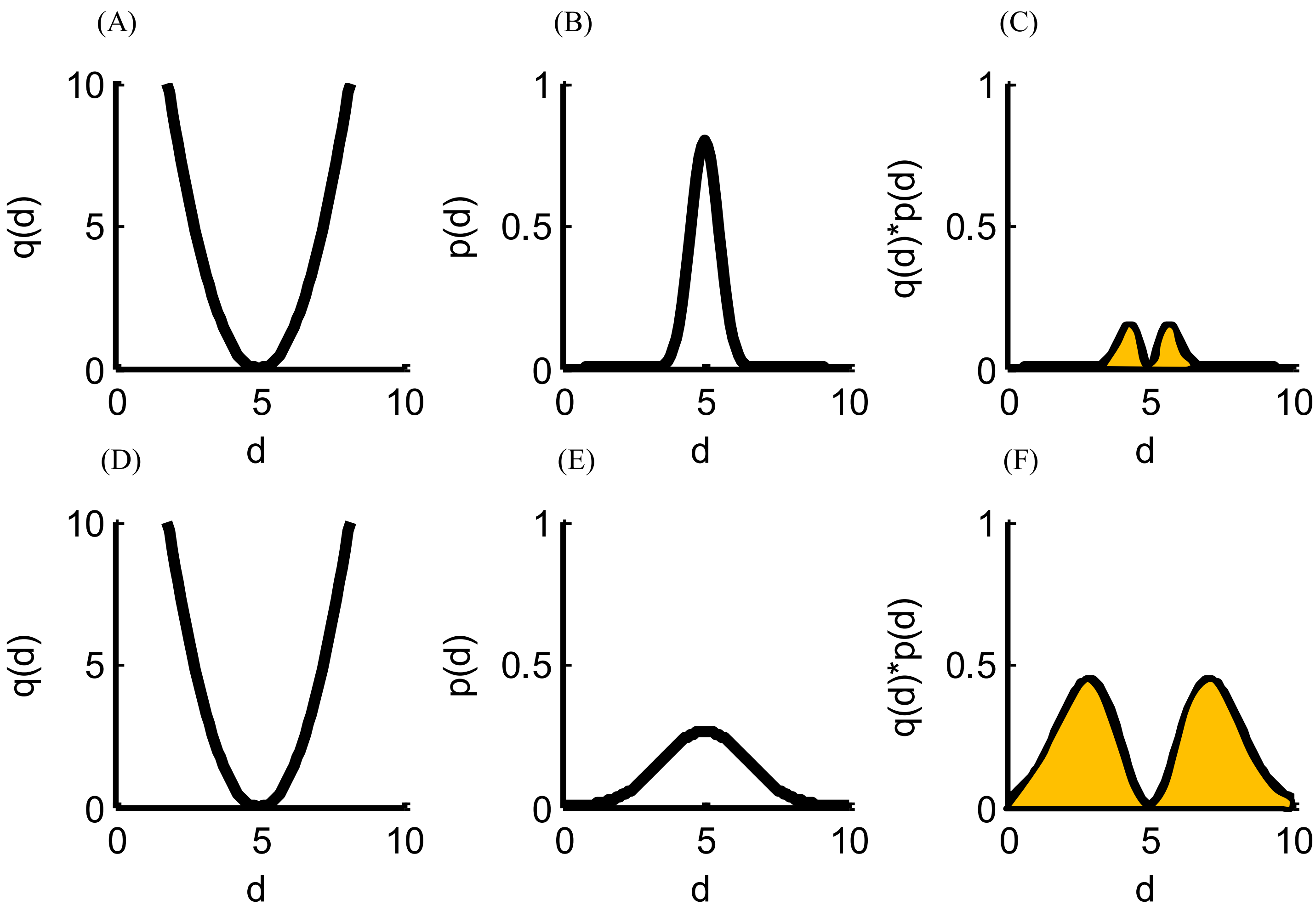
$$q(d) = (d - \langle d \rangle)^2$$

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 $q(d)p(d)$


(A)
(B)
(C)
q(d)
p(d)
q(d)*p(d)
0
5
10
0.5
1
d
(D)
(E)
(F)

variance

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

mean

width is actually square root of variance, that is, σ

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"

MabLab scripts for mean and variance

from tabulated p.d.f. p

```
dbar = Dd*sum(d.*p);
q = (d-dbar).^2;
sigma2 = Dd*sum(q.*p);
sigma = sqrt(sigma2);
```

from realizations of data

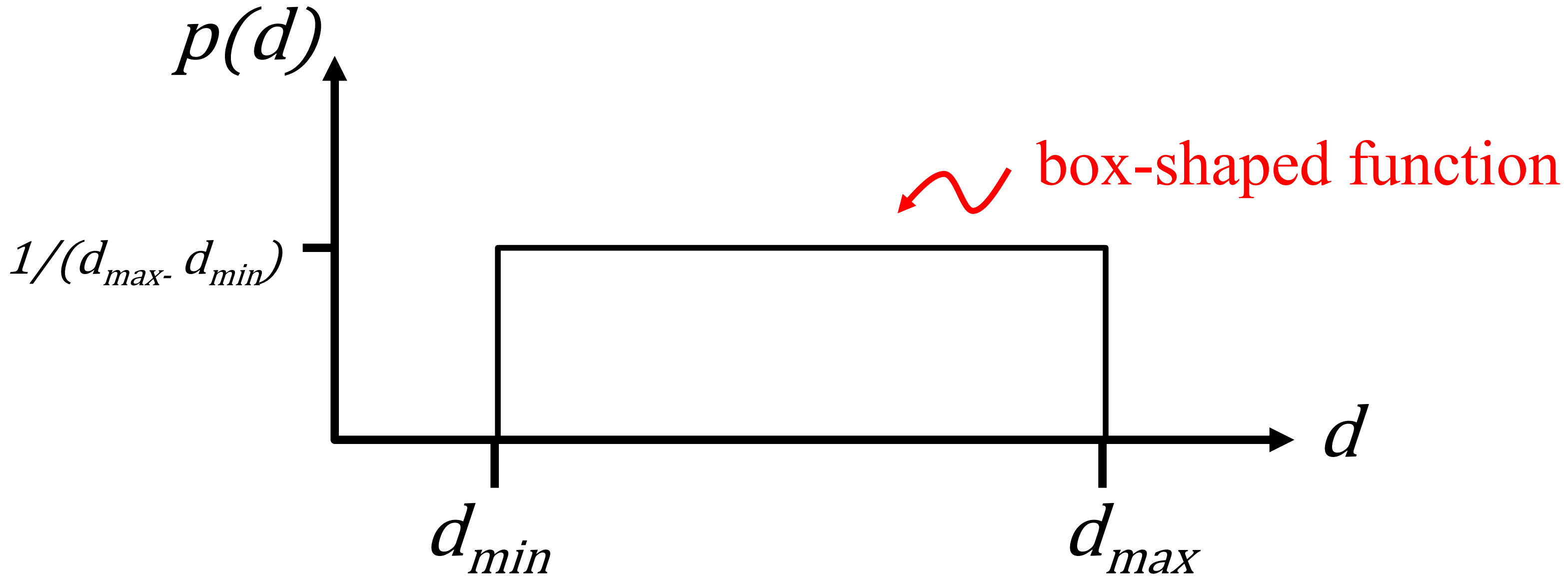
```
dbar = mean(dr);
sigma = std(dr);
sigma2 = sigma^2;
```

two important probability density functions:

uniform

Gaussian (or Normal)

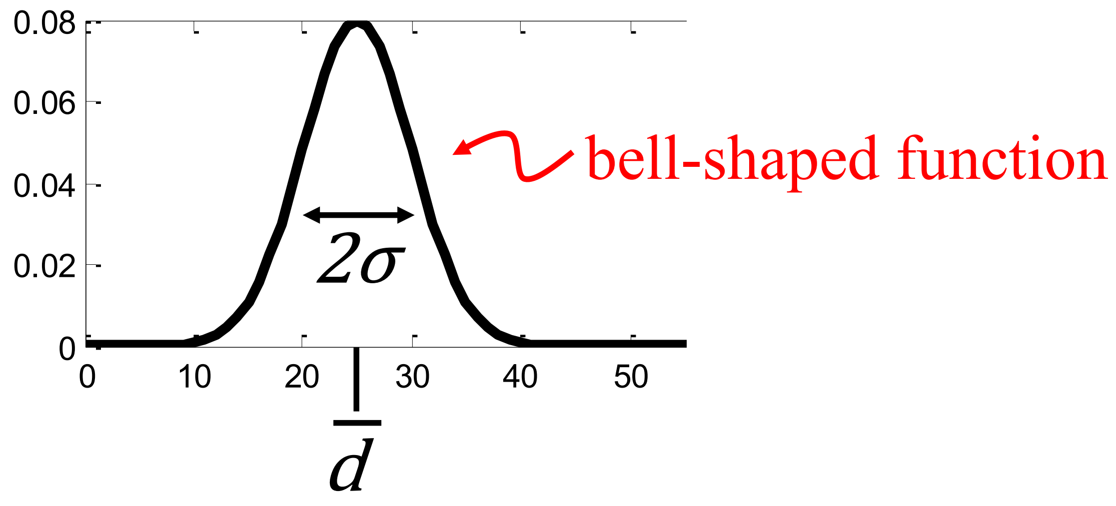
uniform p.d.f.



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

n	P, %
1	68.27
2	95.45
3	99.73

Part 2

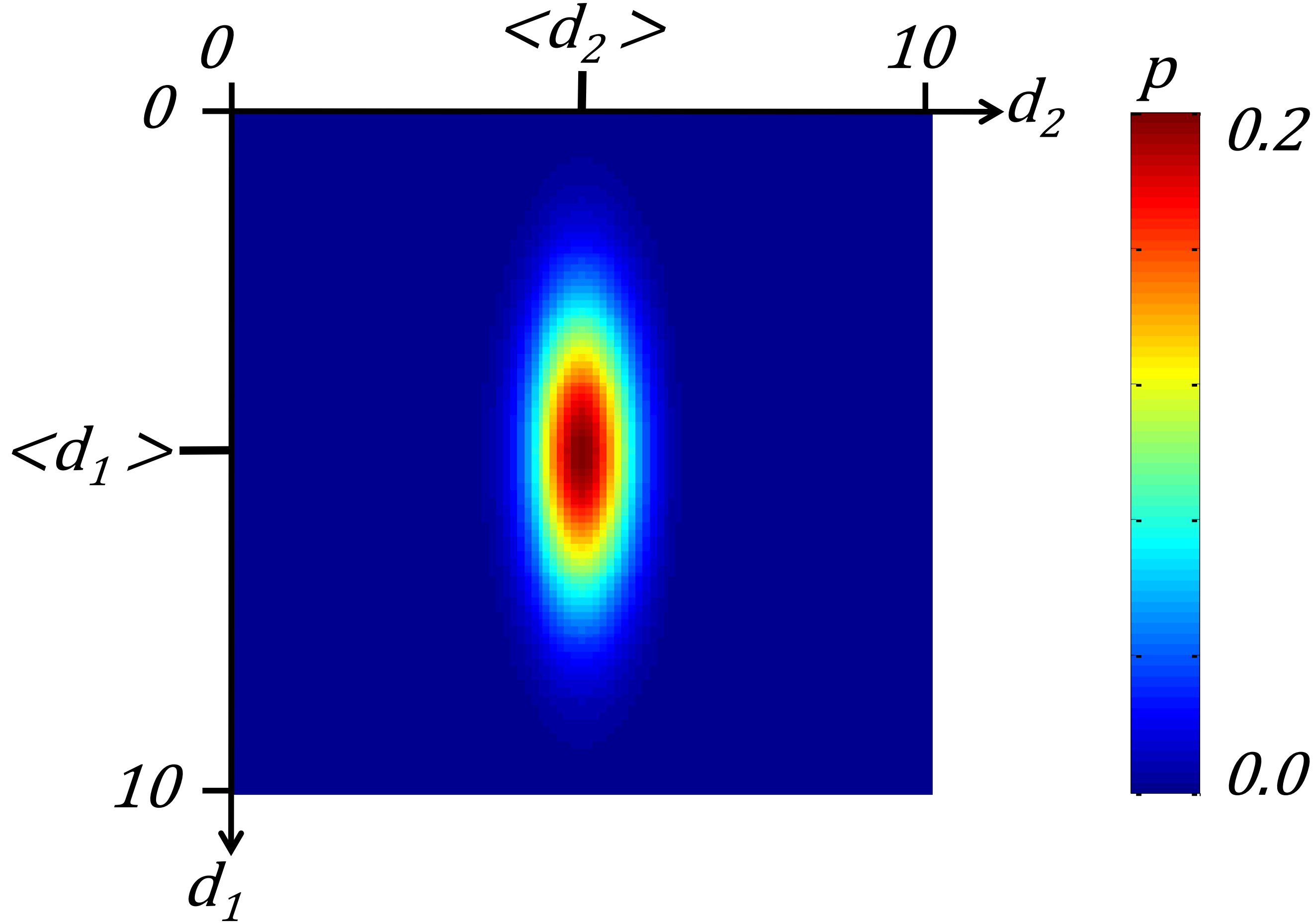
correlated errors

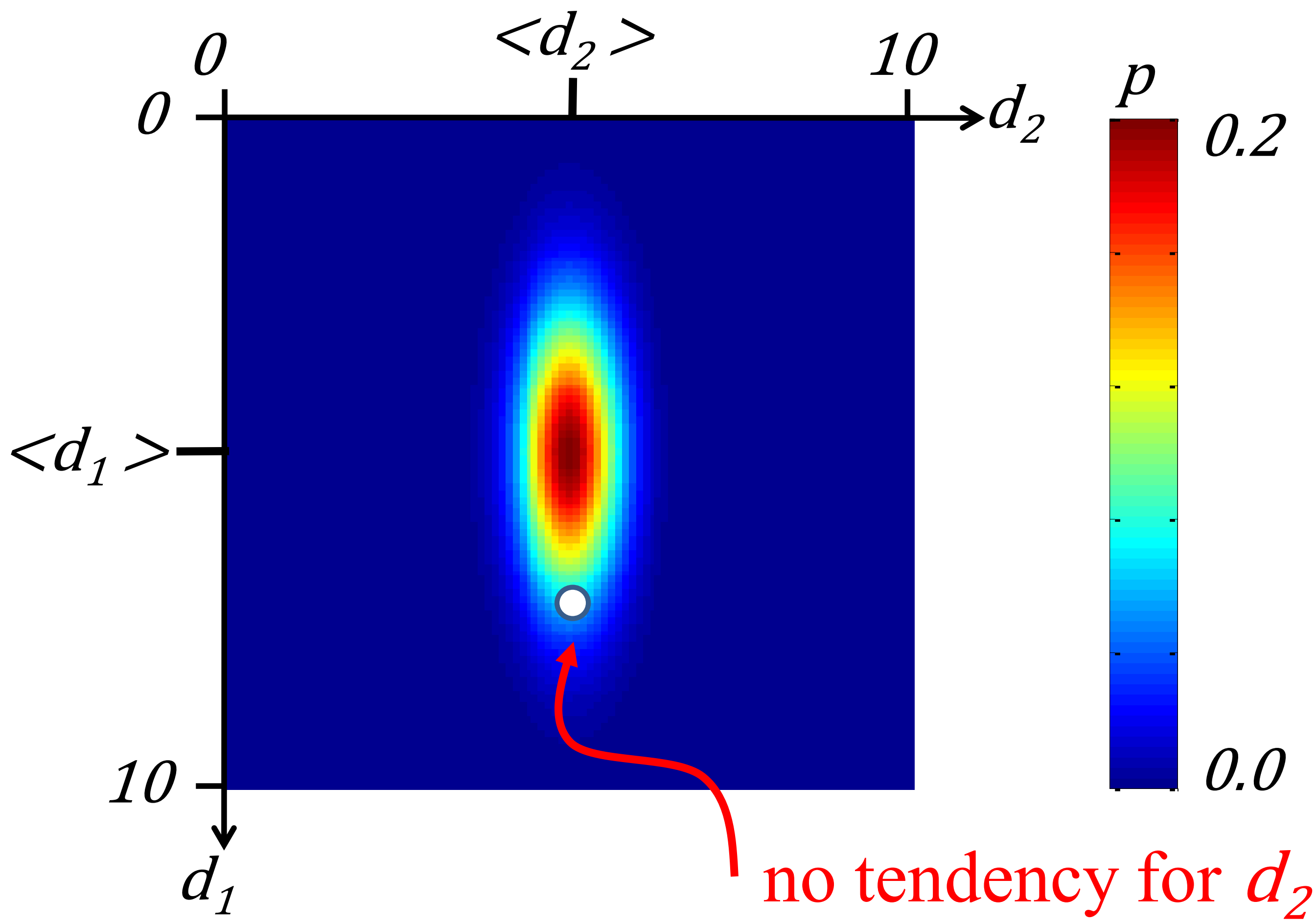
uncorrelated random variables

no pattern of between values of one variable and values of another

when d_1 is higher than its mean d_2 is higher or lower than its mean with equal probability

joint probability density function
uncorrelated case

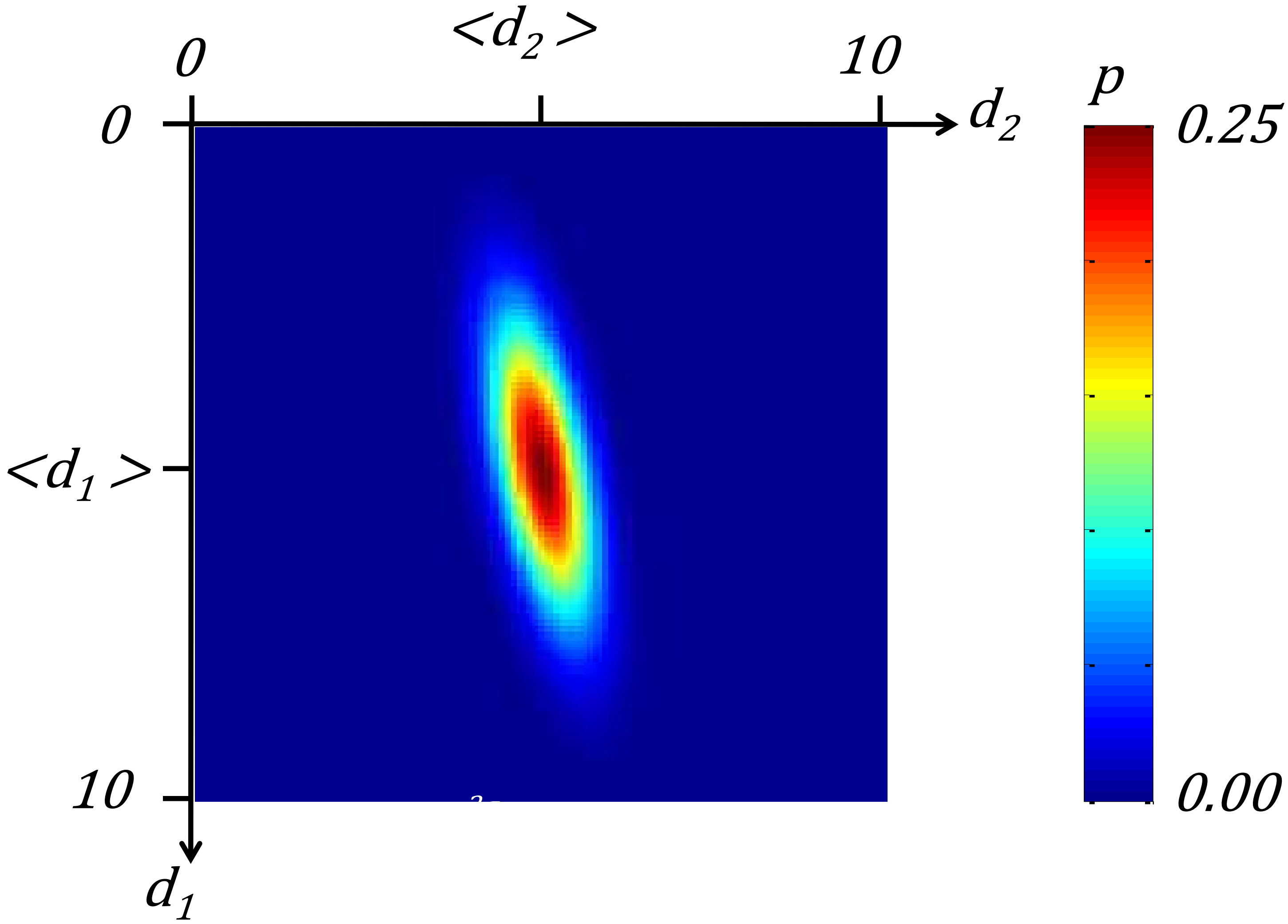




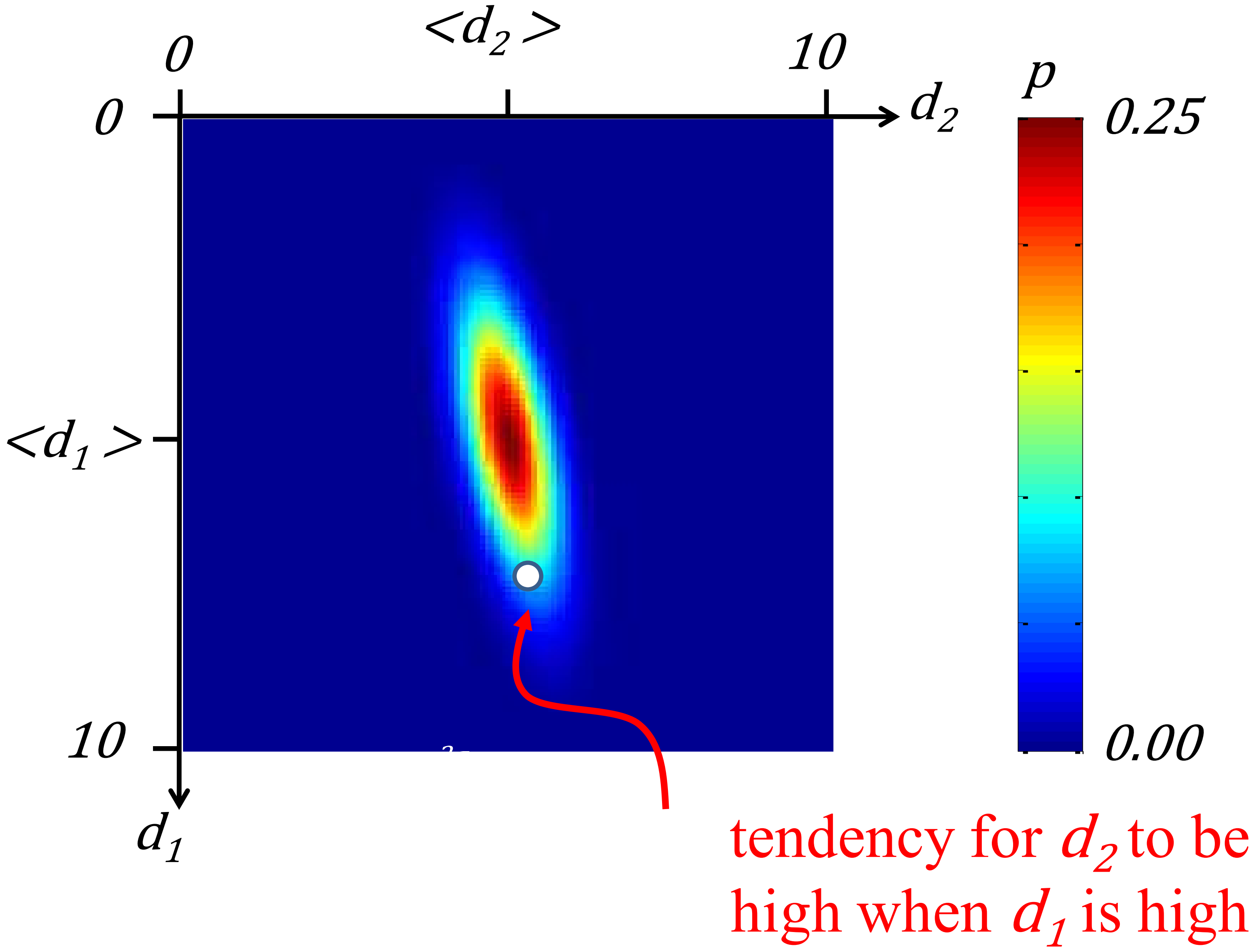
no tendency for d_2 to be either high or low when d_1 is high in uncorrelated case

joint p.d.f.
is just the product of individual p.d.f.'s

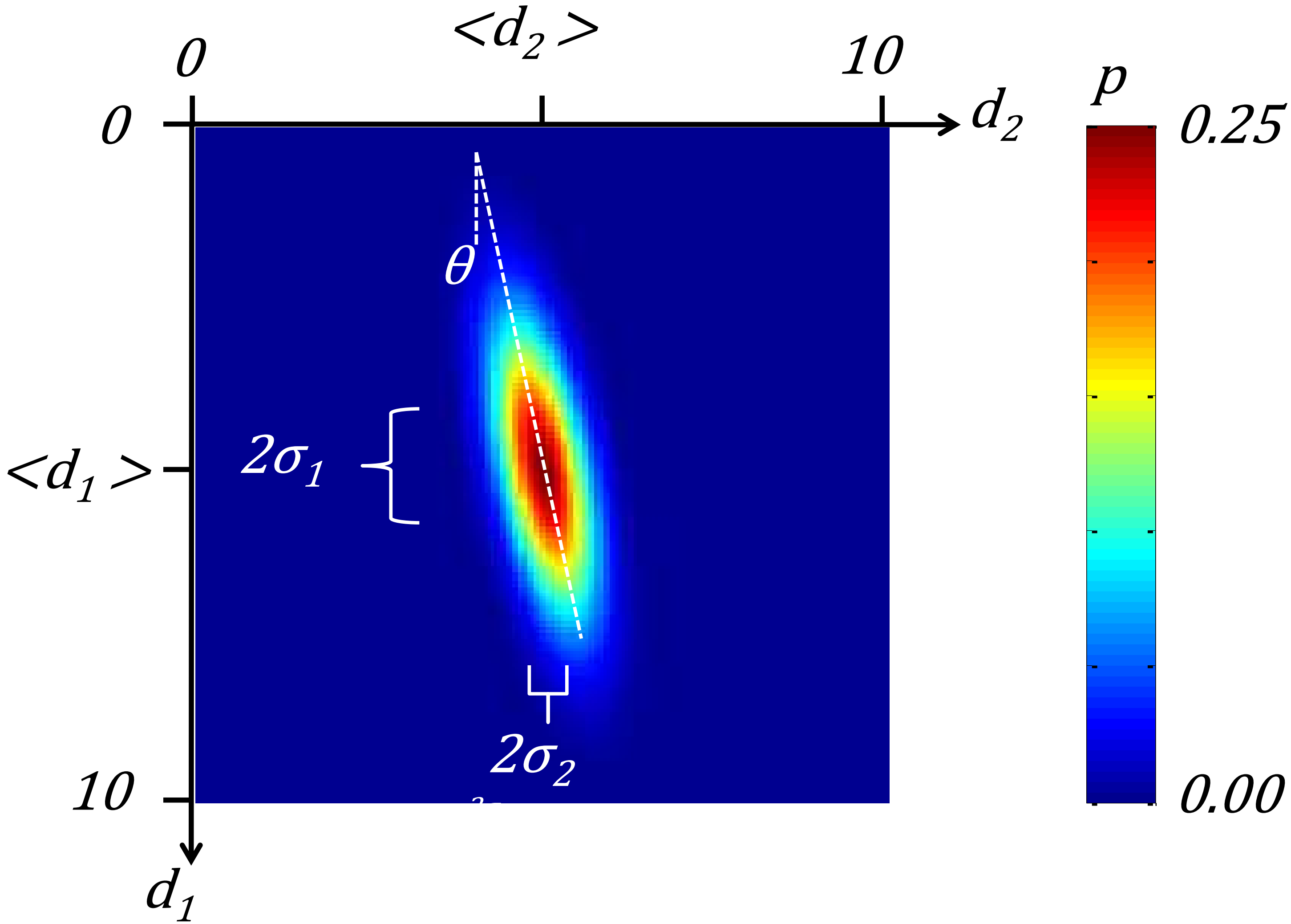
$$p(\mathbf{d}) = p(d_1)\, p(d_2)\, p(d_3)\, \cdots\, p(d_N)$$



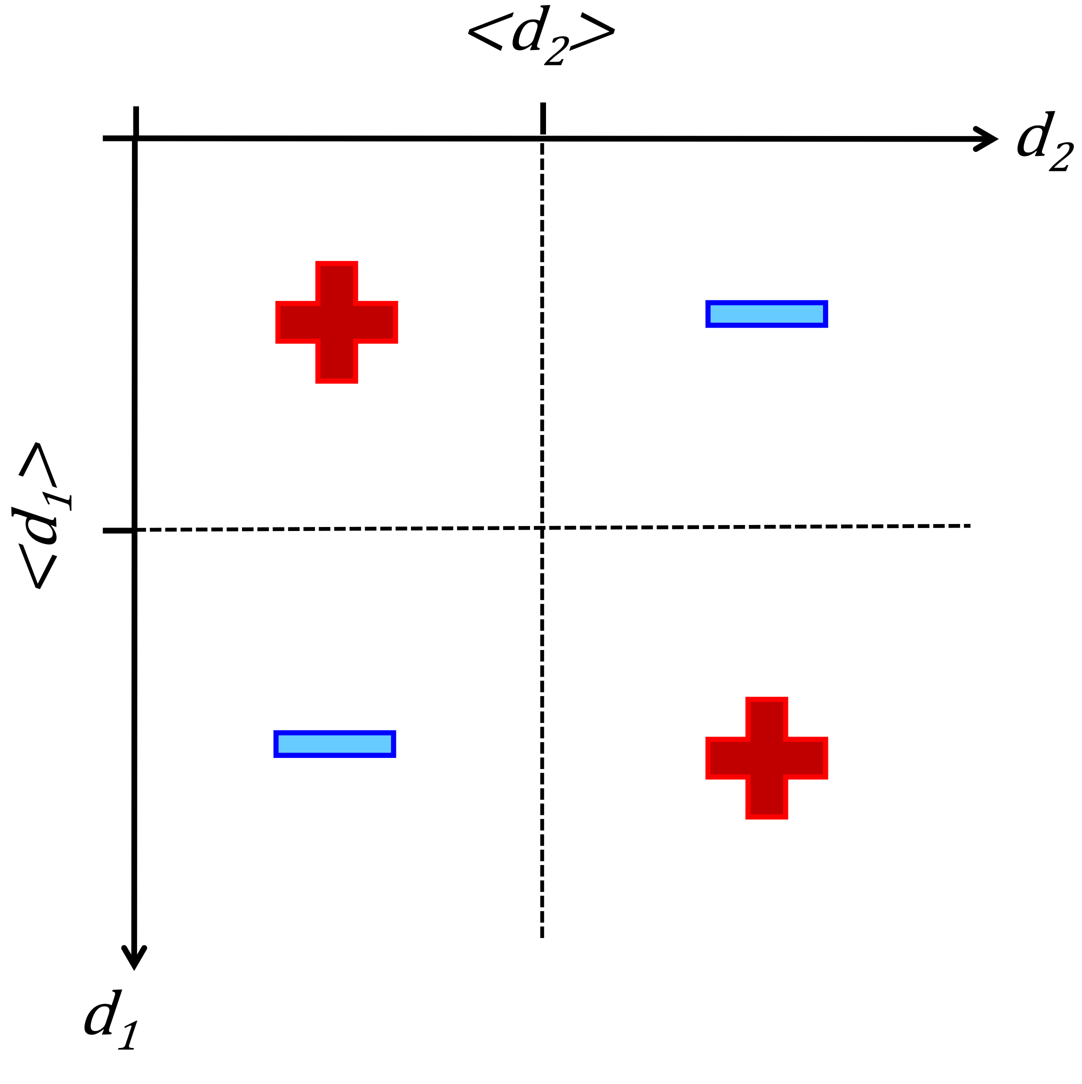
<d2>
0
10
0
d2
p
0.25
<d1>
10
d1
0.00

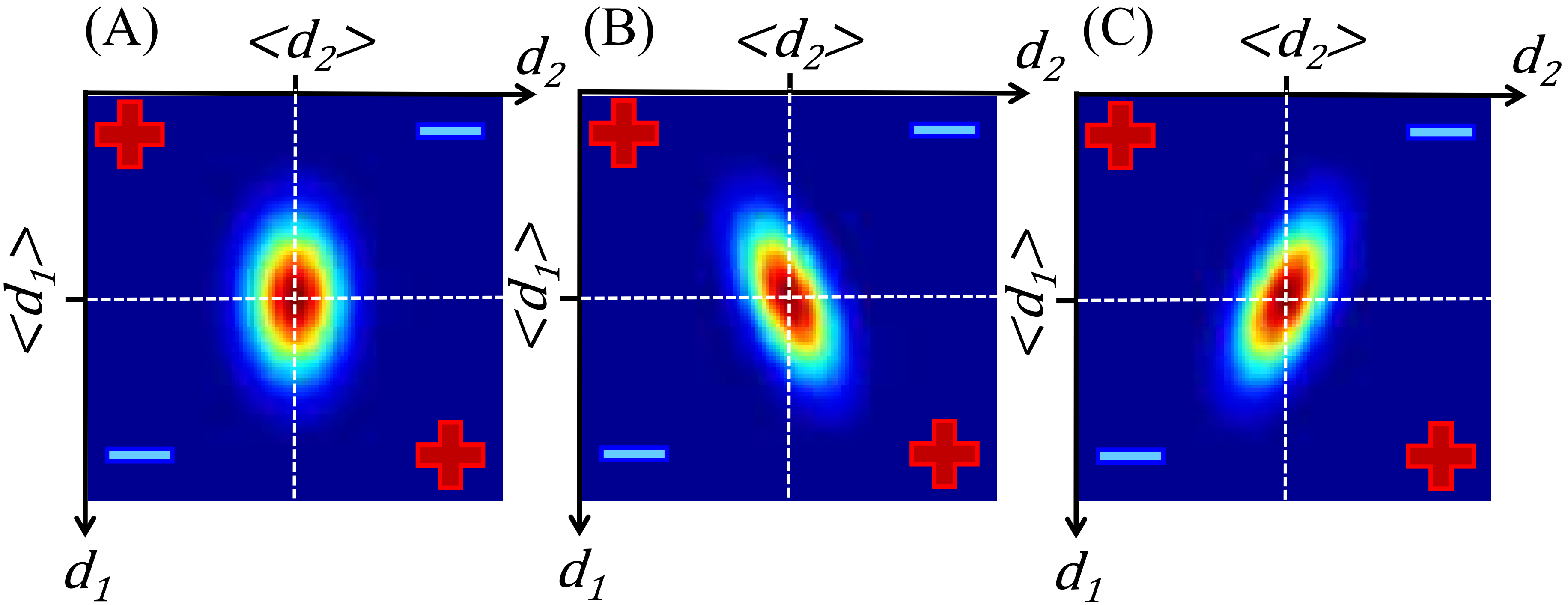


$<d_2>$
0
10
d_2
0
$<d_1>$
10
d_1
p
0.25
0.00
tendency for d_2 to be high when d_1 is high



<d2>
0
10
d2
0
θ
<d1>
2σ1
2σ2
10
d1
p
0.25
0.00


$<d_2>$
d_2
$<d_1>$
d_1



(A)
(B)
(C)
<d2>
d2
<d1>
d1

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

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

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

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

diagonal elements: variances
off-diagonal elements: covariances

estimating covariance from a table D of data

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

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

in MatLab, `C=cov(D)`

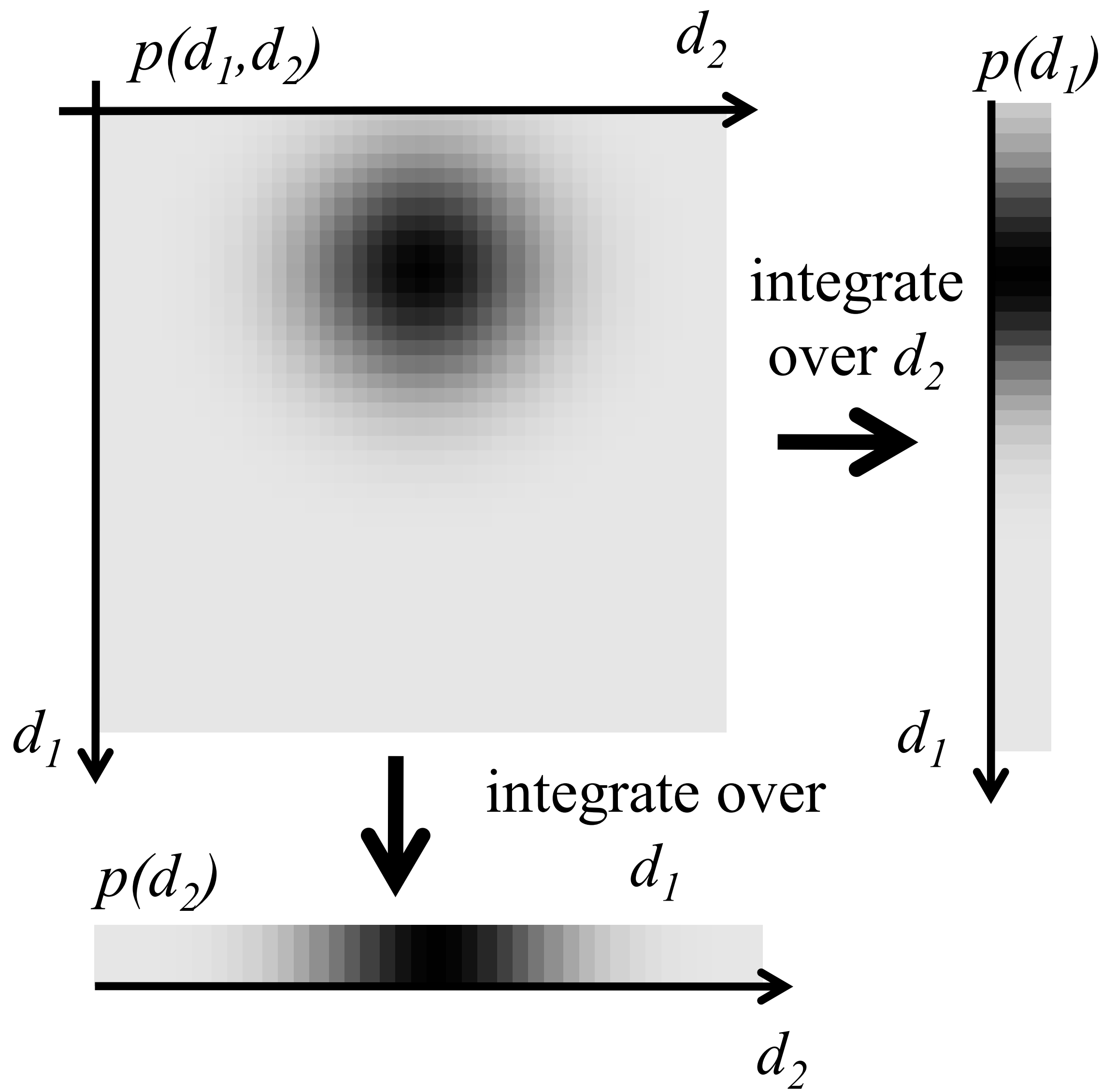
univariate p.d.f. formed from joint p.d.f.

$$p(\boldsymbol{d}) \rightarrow p(d_i)$$

behavior of d_i irrespective of the other ds

$$p(d_i) = \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{(N-1 \; integrals)} p(\mathbf{d}) \, \mathrm{d}d_j \, \mathrm{d}d_k \cdots \mathrm{d}d_l$$

integrate over everything but d_i



p(d1,d2)
d2
d1
p(d1)
integrate
over d2
d1
integrate over
d1
p(d2)
d2

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

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

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

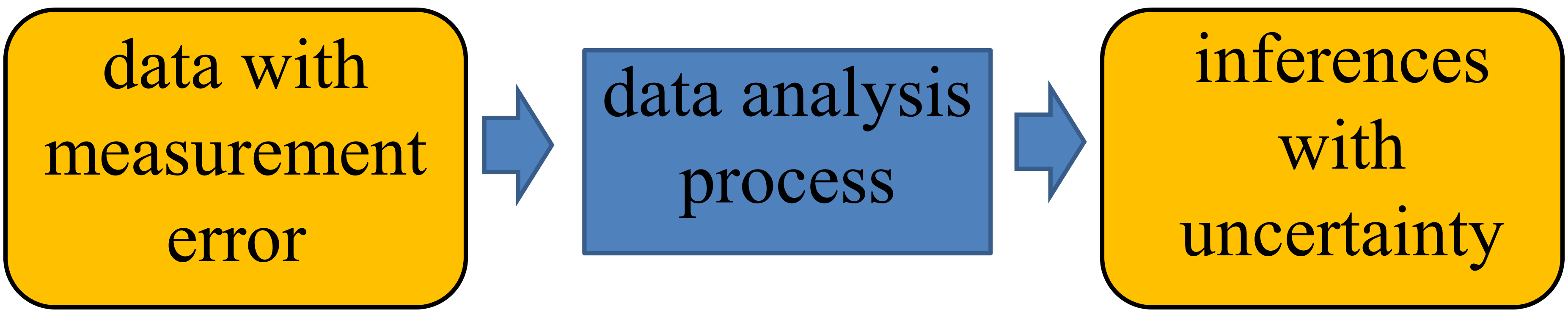
mean

covariance matrix

Part 3

functions of random variables

functions of random variables



functions of random variables

given $p(\mathbf{d})$

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

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

univariate p.d.f.
use the chair rule

$$1 = \int_{d_{\min}}^{d_{max}} p(d)\,\mathrm{d}d \;= \int_{d(m_{\min})}^{d(m_{max})} p[d(m)]\frac{\mathrm{d}d}{\mathrm{d}m}\mathrm{d}m = \int_{m_{min}}^{m_{max}} p(m)\,\mathrm{d}m$$

univariate p.d.f.
use the chair rule

$$1 = \int_{d_{min}}^{d_{max}} p(d)\,\mathrm{d}d = \int_{d(m_{min})}^{d(m_{max})} p[d(m)]\,\frac{\mathrm{d}d}{\mathrm{d}m}\,\mathrm{d}m = \int_{m_{min}}^{m_{max}} p(m)\,\mathrm{d}m$$

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

rule for
transforming a
univariate p.d.f.

multivariate p.d.f.

$$1 = \int p(\mathbf{d})\, \mathrm{d}^N d = \int p[\mathbf{d}(\mathbf{m})] \left|\frac{\partial \mathbf{d}}{\partial \mathbf{m}}\right| \mathrm{d}^N m = \int p[\mathbf{d}(\mathbf{m})]\, J(\mathbf{m})\, \mathrm{d}^N m = \int p(\mathbf{m})\, \mathrm{d}^N m$$

Jacobian determinant

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

determinant of matrix with elements $\partial d_i/\partial m_j$

multivariate p.d.f.

$$1 = \int p(\mathbf{d})\, \mathrm{d}^{\mathrm{N}} d = \int p[\mathbf{d}(\mathbf{m})] \left|\frac{\partial \mathbf{d}}{\partial \mathbf{m}}\right| \mathrm{d}^{\mathrm{N}} m = \int p[\mathbf{d}(\mathbf{m})]\, J(\mathbf{m})\, \mathrm{d}^{\mathrm{N}} m = \int p(\mathbf{m})\, \mathrm{d}^{\mathrm{N}} m$$

Jacobian determinant

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

rule for transforming a multivariate p.d.f.

simple example

data with measurement error → data analysis process → inferences with uncertainty

one datum, d
uniform p.d.f.
$0<d<1$
→ $m = 2d$ → one model parameter, m

$m = 2d$ and $d = \frac{1}{2}m$

$d = 0 \rightarrow m = 0$

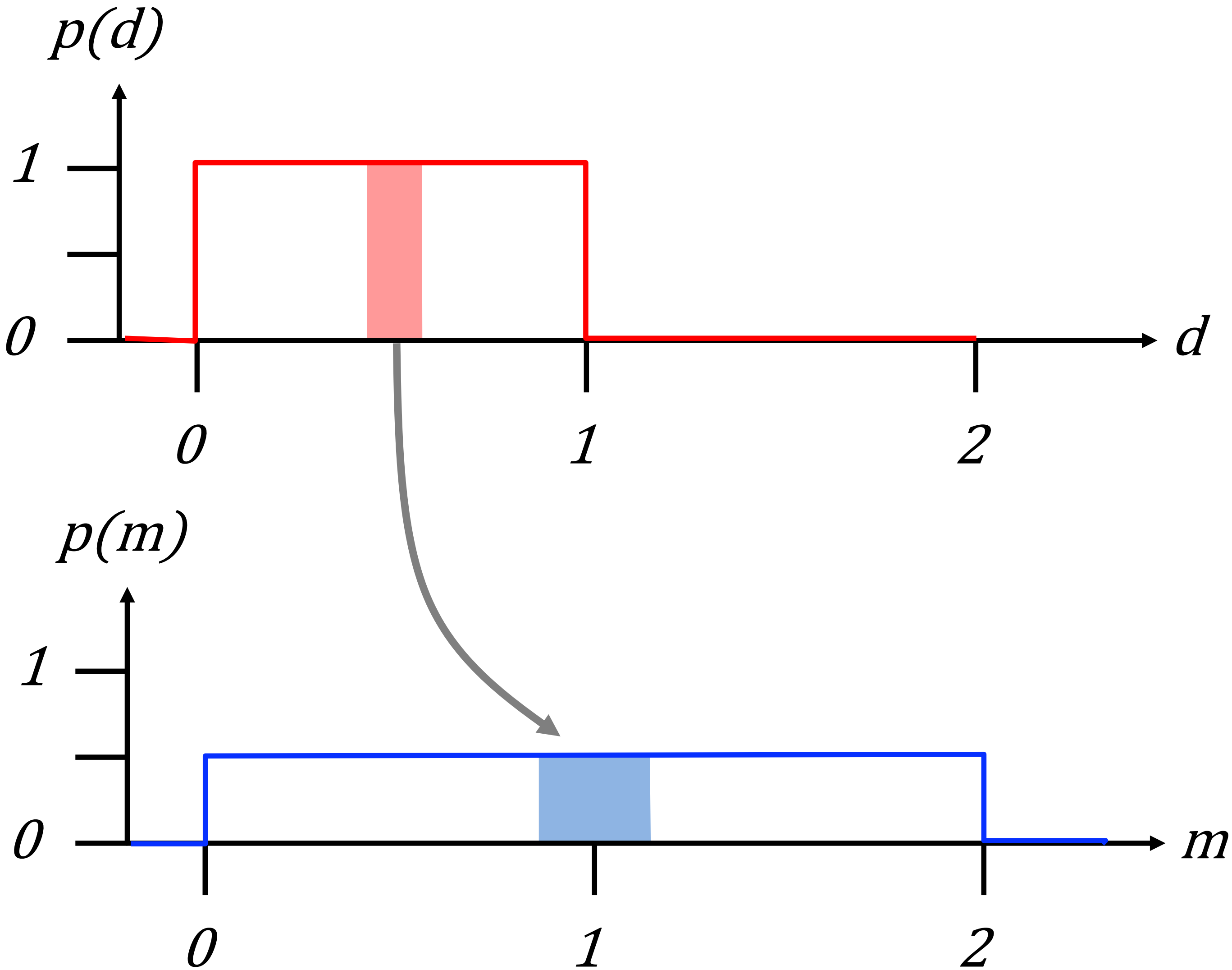
$d = 1 \rightarrow m = 2$

$dd/dm = \frac{1}{2}$

$p(d) = 1$

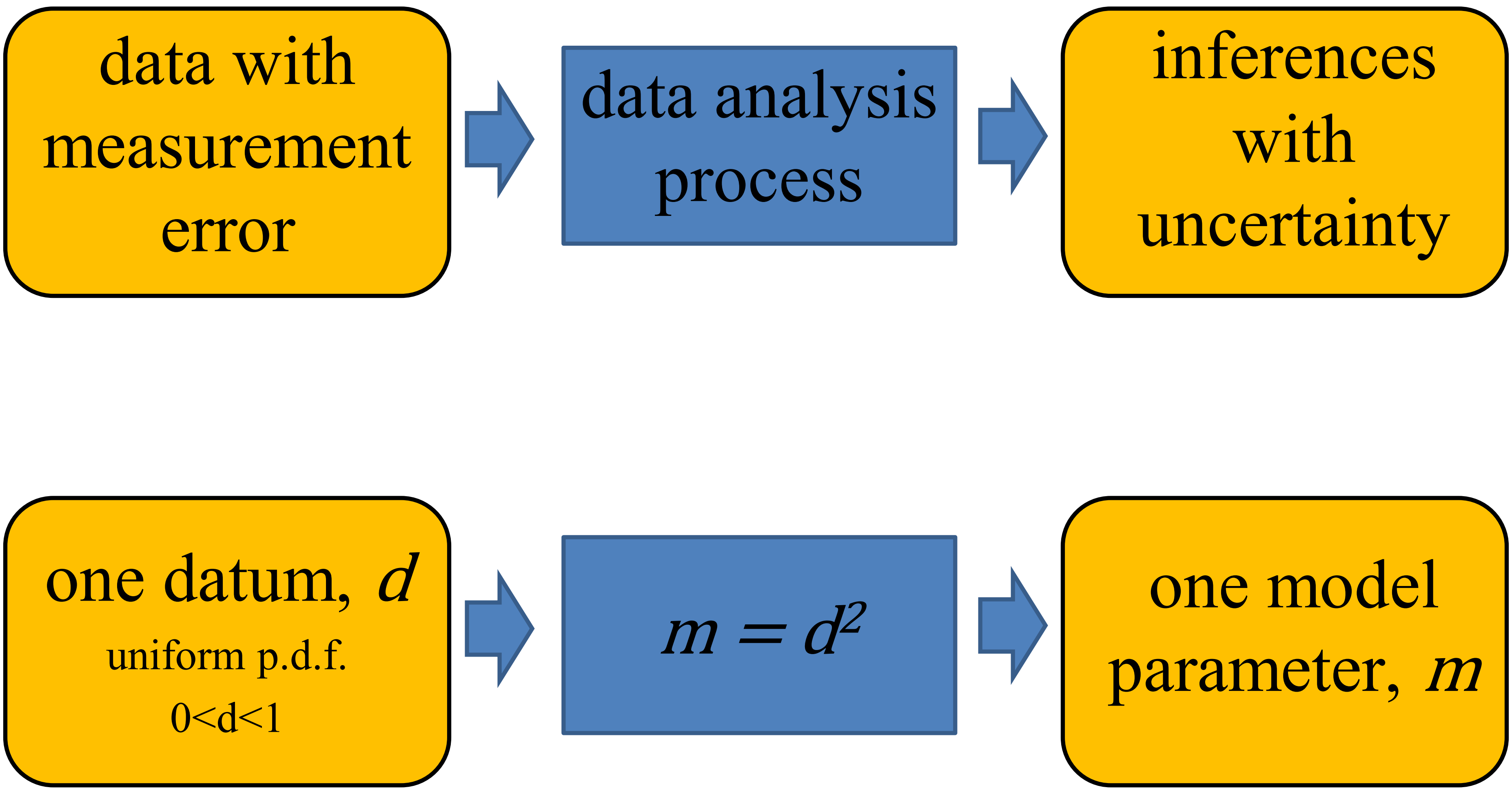
$$p(m) = p[d(m)] \left|\frac{\mathrm{d}d}{\mathrm{d}m}\right|$$

$p(m) = \frac{1}{2}$



p(d)
1
0
d
0
1
2
p(m)
1
0
m
0
1
2

another simple example



$m = d^2$ and $d = m^{½}$

$d = 0 \rightarrow m = 0$

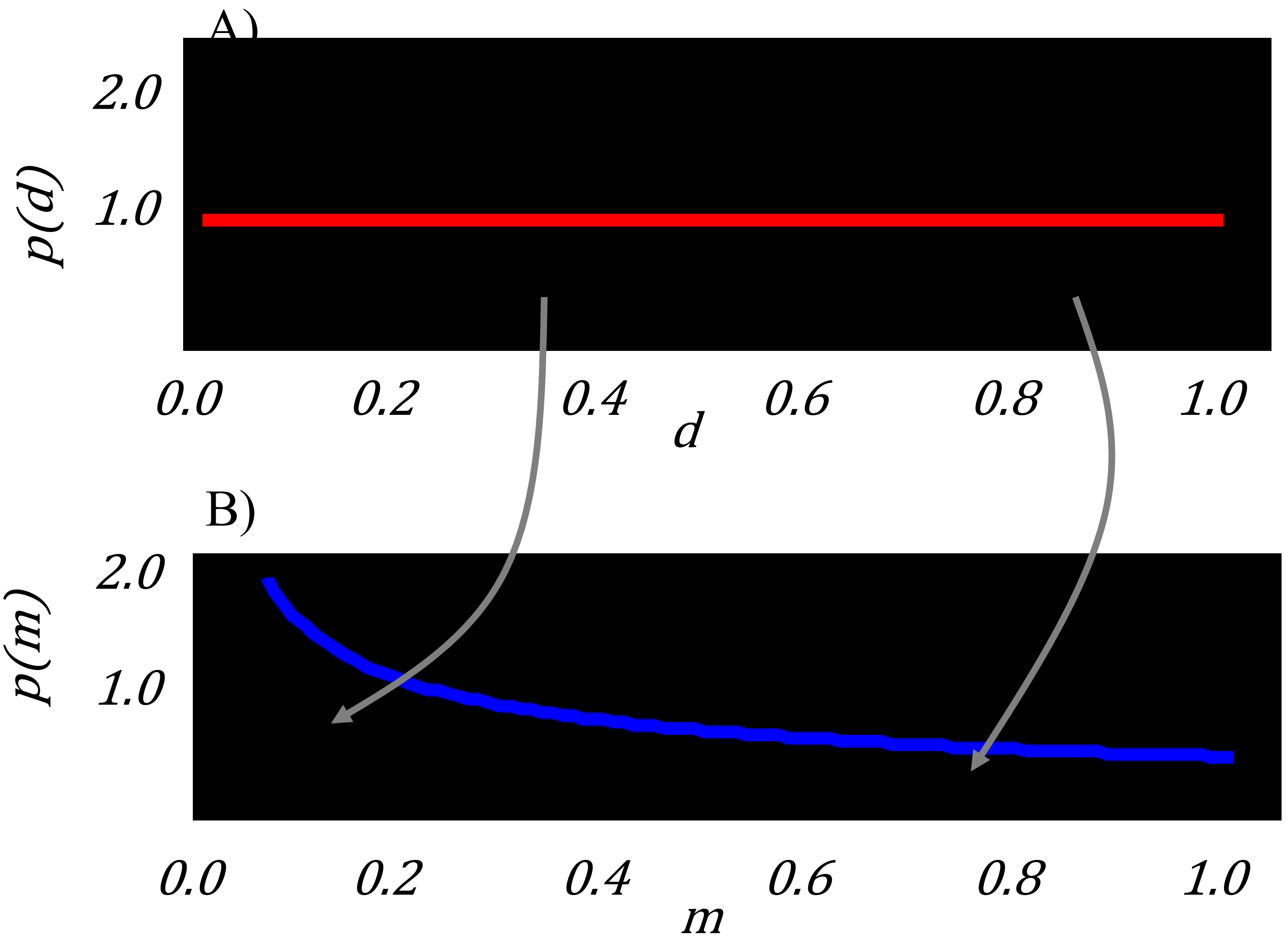
$d = 1 \rightarrow m = 1$

$dd/dm = ½ m^{-½}$

$p(d) = 1$

$$p(m) = p[d(m)] \left| \frac{\mathrm{d}d}{\mathrm{d}m} \right|$$

$p(m) = ½\, m^{-½}$


A)
2.0
1.0
p(d)
0.0
0.2
0.4
0.6
0.8
1.0
d
B)
2.0
1.0
p(m)
0.0
0.2
0.4
0.6
0.8
1.0
m

multivariate example

data with measurement error → data analysis process → inferences with uncertainty

2 data, d_1 , d_2
uniform p.d.f.
$0<d<1$
for both

→

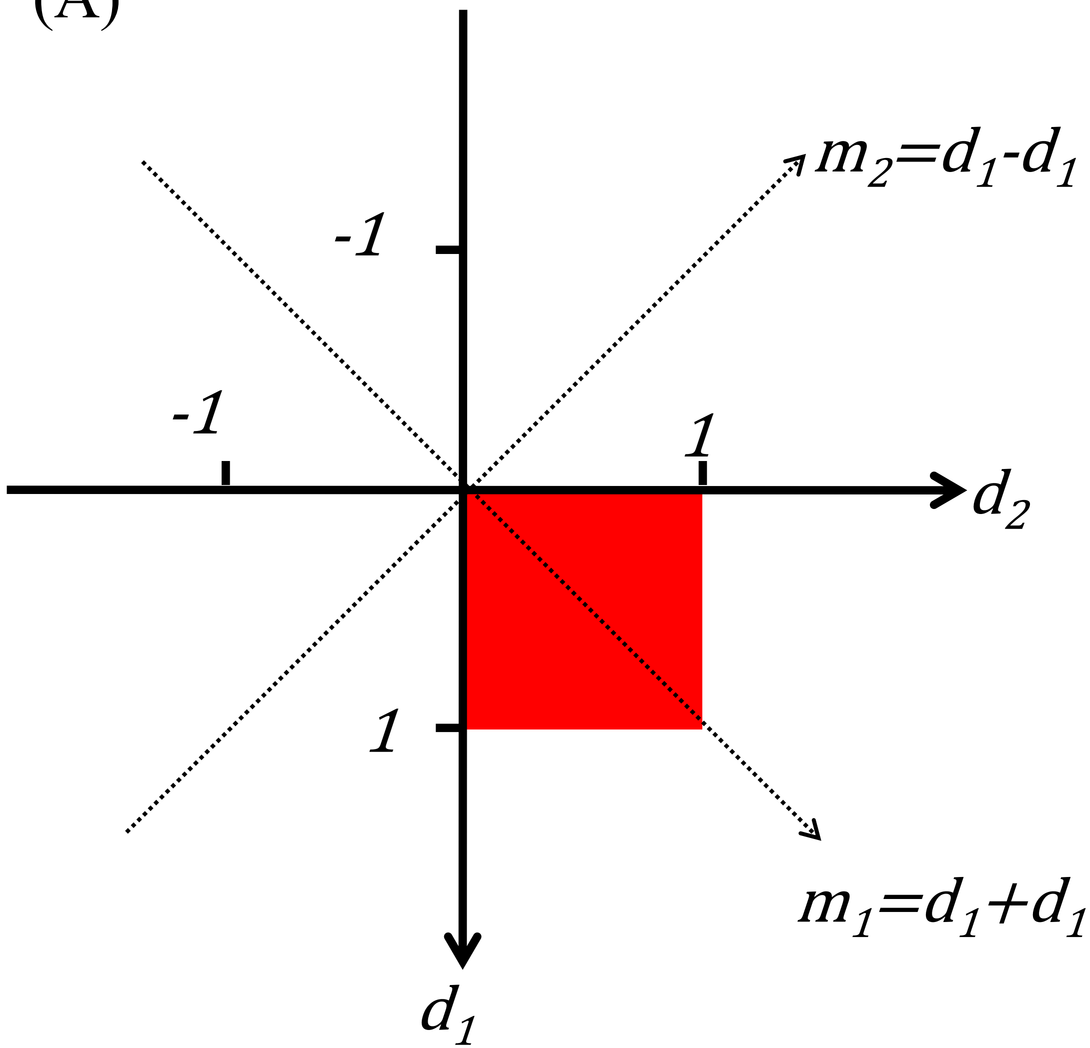
$m_1 = d_1 + d_2$

$m_2 = d_2 - d_2$

→

2 model parameters, m_1 and m_2

(A)



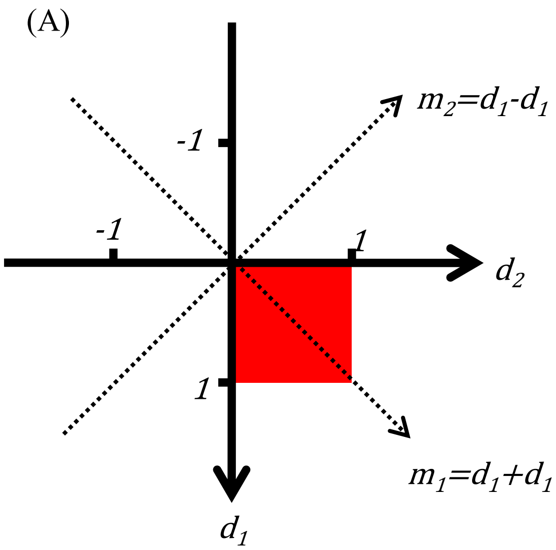
$\mathbf{m}=\mathbf{Md}$ with $\mathbf{M} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $|\det(\mathbf{M})|=2$

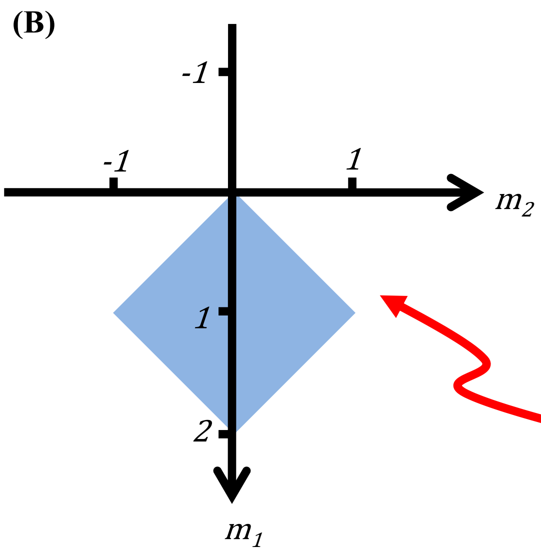
$\mathbf{d}=\mathbf{M}^{-1}\mathbf{m}$ so $\partial \mathbf{d}/ \partial \mathbf{m} = \mathbf{M}^{-1}$ and $\mathbf{J} = |\det(\mathbf{M}^{-1})| = ½$

$$p(\mathbf{d}) = 1$$

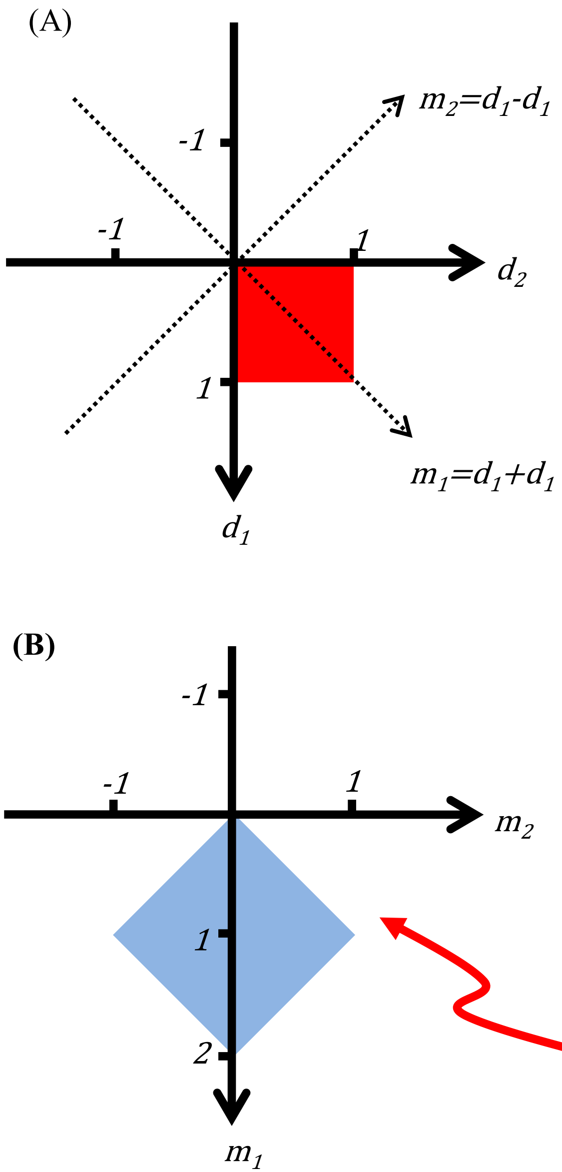
$$p(\mathbf{m}) = p[\mathbf{d}(\mathbf{m})] \left|\frac{\partial \mathbf{d}}{\partial \mathbf{m}}\right| = p[\mathbf{d}(\mathbf{m})]\, J(\mathbf{m})$$

$$p(\mathbf{m}) = ½$$

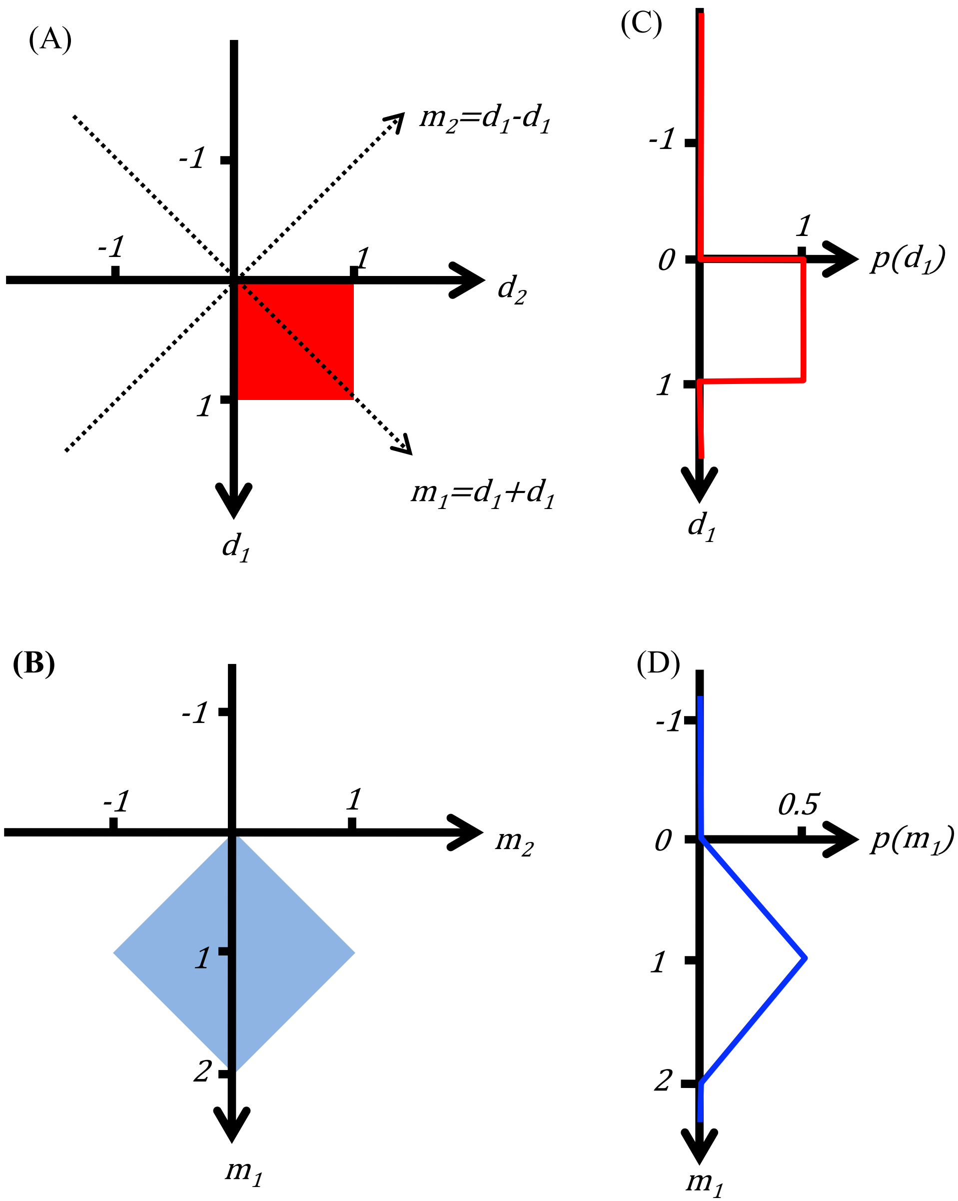




Note that the shape in **m** is different than the shape in **d**



The shape is a square with sides of length √2. The amplitude is ½. So the area is ½×√2×√2 = 1



(A)
$m_2=d_1-d_1$
$m_1=d_1+d_1$
d_2
d_1
-1
1
(C)
$p(d_1)$
d_1
-1
0
1
(B)
m_2
m_1
-1
1
2
(D)
0.5
$p(m_1)$
m_1
-1
0
1
2

moral

$p(\mathbf{m})$ can behavior quite differently from $p(\mathbf{d})$