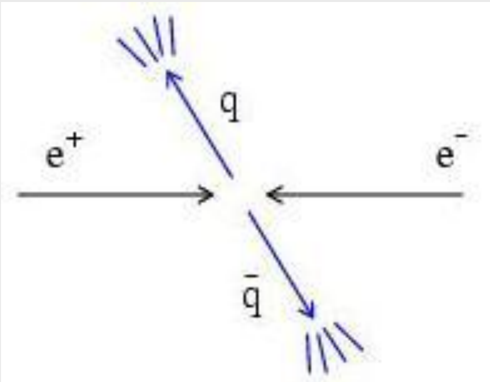
Statistische Methoden der Datenanalyse

Kapitel 1: Fundamentale Konzepte

Professor Markus Schumacher

Freiburg / Sommersemester 2009

Data analysis in particle physics



Observe events of a certain type

Measure characteristics of each event (particle momenta, number of muons, energy of jets,...)

Theories (e.g. SM) predict distributions of these properties up to free parameters, e.g., α, G_F, M_Z, α_s, m_H, ...

Some tasks of data analysis:

Estimate (measure) the parameters;

Quantify the uncertainty of the parameter estimates;

Test the extent to which the predictions of a theory are in agreement with the data.

Dealing with uncertainty

In particle physics there are various elements of uncertainty:

theory is not deterministic

quantum mechanics

random measurement errors

present even without quantum effects

things we could know in principle but don't

e.g. from limitations of cost, time, ...

We can quantify the uncertainty using PROBABILITY

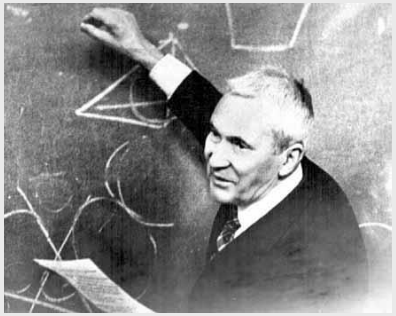
A definition of probability

Consider a set *S* with subsets *A*, *B*, ...

For all $A \subset S, P(A) \geq 0$

$P(S) = 1$

If $A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$



Kolmogorov axioms (1933)

From these axioms we can derive further properties, e.g.

$P(\overline{A}) = 1 - P(A)$

$P(A \cup \overline{A}) = 1$

$P(\emptyset) = 0$

if $A \subset B$, then $P(A) \leq P(B)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional probability, independence

Also define conditional probability of *A* given *B* (with *P(B) ≠ 0*):

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

E.g. rolling dice: $P(n < 3 \,|\, n \text{ even}) = \frac{P((n<3) \cap n \text{ even})}{P(\text{even})} = \frac{1/6}{3/6} = \frac{1}{3}$

Subsets *A, B* independent if: $P(A \cap B) = P(A)P(B)$

If *A, B* independent, $P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$

N.B. do not confuse with disjoint subsets, i.e. $A \cap B = \emptyset$

Interpretation of probability

I. Relative frequency

A, *B*, ... are outcomes of a repeatable experiment

$$P(A) = \lim_{n \to \infty} \frac{\text{times outcome is } A}{n}$$

cf. quantum mechanics, particle scattering, radioactive decay...

II. Subjective probability

A, *B*, ... are hypotheses (statements that are true or false)

$$P(A) = \text{degree of belief that } A \text{ is true}$$

- Both interpretations consistent with Kolmogorov axioms.
- In particle physics frequency interpretation often most useful, but subjective probability can provide more natural treatment of non-repeatable phenomena:

systematic uncertainties, probability that Higgs boson exists,...

Bayes' theorem

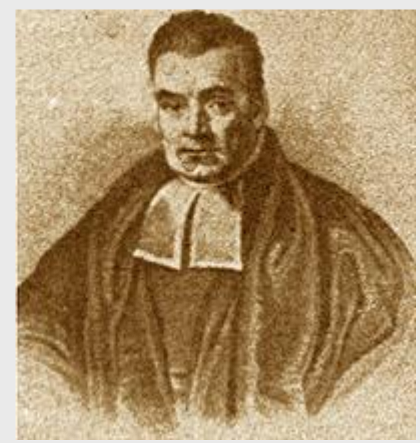
From the definition of conditional probability we have,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{and} \quad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

but $P(A \cap B) = P(B \cap A)$, so

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' theorem



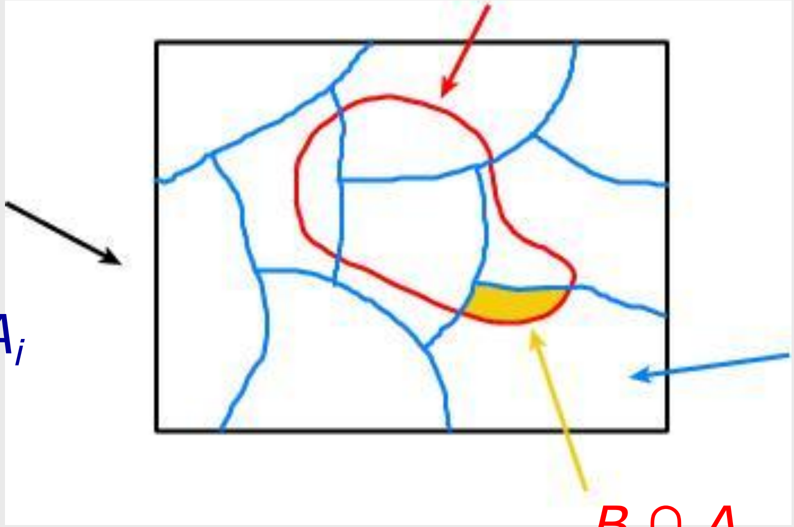
First published (posthumously) by the Reverend Thomas Bayes (1702−1761)

An essay towards solving a problem in the doctrine of chances, Philos. Trans. R. Soc. **53** (1763) 370; reprinted in Biometrika, **45** (1958) 293.

The law of total probability

Consider a subset B of the sample space S,

divided into disjoint subsets A_i such that $\cup_i A_i = S$,



$\rightarrow$ $$B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i),$$

$\rightarrow$ $$P(B) = P(\cup_i (B \cap A_i)) = \sum_i P(B \cap A_i)$$

$\rightarrow$ $$P(B) = \sum_i P(B|A_i) P(A_i)$$ law of total probability

Bayes' theorem becomes

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$$

An example using Bayes' theorem

Suppose the probability (for anyone) to have AIDS is:

$$P(\text{AIDS}) = 0.001$$
$$P(\text{no AIDS}) = 0.999$$

← prior probabilities, i.e., before any test carried out

Consider an AIDS test: result is $+$ or $-$

$$P(+|\text{AIDS}) = 0.98$$
$$P(-|\text{AIDS}) = 0.02$$

← probabilities to (in)correctly identify an infected person

$$P(+|\text{no AIDS}) = 0.03$$
$$P(-|\text{no AIDS}) = 0.97$$

← probabilities to (in)correctly identify an uninfected person

Suppose your result is $+$. How worried should you be?

Bayes' theorem example (cont.)

The probability to have AIDS given a + result is

$$P(\text{AIDS}|+) = \frac{P(+|\text{AIDS})P(\text{AIDS})}{P(+|\text{AIDS})P(\text{AIDS}) + P(+|\text{no AIDS})P(\text{no AIDS})}$$

$$= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999}$$

$= 0.032$ ← posterior probability

i.e. you're probably OK!

Your viewpoint: my degree of belief that I have AIDS is 3.2%

Your doctor's viewpoint: 3.2% of people like this will have AIDS

Frequentist Statistics − general philosophy

In frequentist statistics, probabilities are associated only with the data, i.e., outcomes of repeatable observations (shorthand: $\vec{x}$).

Probability = limiting frequency

Probabilities such as

P (Higgs boson exists),
P ($0.117 < \alpha_s < 0.121$),

etc. are either 0 or 1, but we don't know which.

The tools of frequentist statistics tell us what to expect, under the assumption of certain probabilities, about hypothetical repeated observations.

The preferred theories (models, hypotheses, ...) are those for which our observations would be considered 'usual'.

Bayesian Statistics – general philosophy

In Bayesian statistics, use subjective probability for hypotheses:

probability of the data assuming hypothesis H (the likelihood)

prior probability, i.e., before seeing the data

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H)\,dH}$$

posterior probability, i.e., after seeing the data

normalization involves sum over all possible hypotheses

Bayes' theorem has an "if-then" character: If your prior probabilities were $\pi(H)$, then it says how these probabilities should change in the light of the data.

No general prescription for priors (subjective!)

Random variables and probability density functions

A random variable is a numerical characteristic assigned to an element of the sample space; can be discrete or continuous.

Suppose outcome of experiment is continuous value *x*

$$P(x \text{ found in } [x, x+dx]) = f(x)\,dx$$

→ *f(x)* = probability density function (pdf)

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

x must be somewhere

Or for discrete outcome x_i with e.g. i = 1, 2, ... we have

$$P(x_i) = p_i$$

probability mass function

$$\sum_i P(x_i) = 1$$

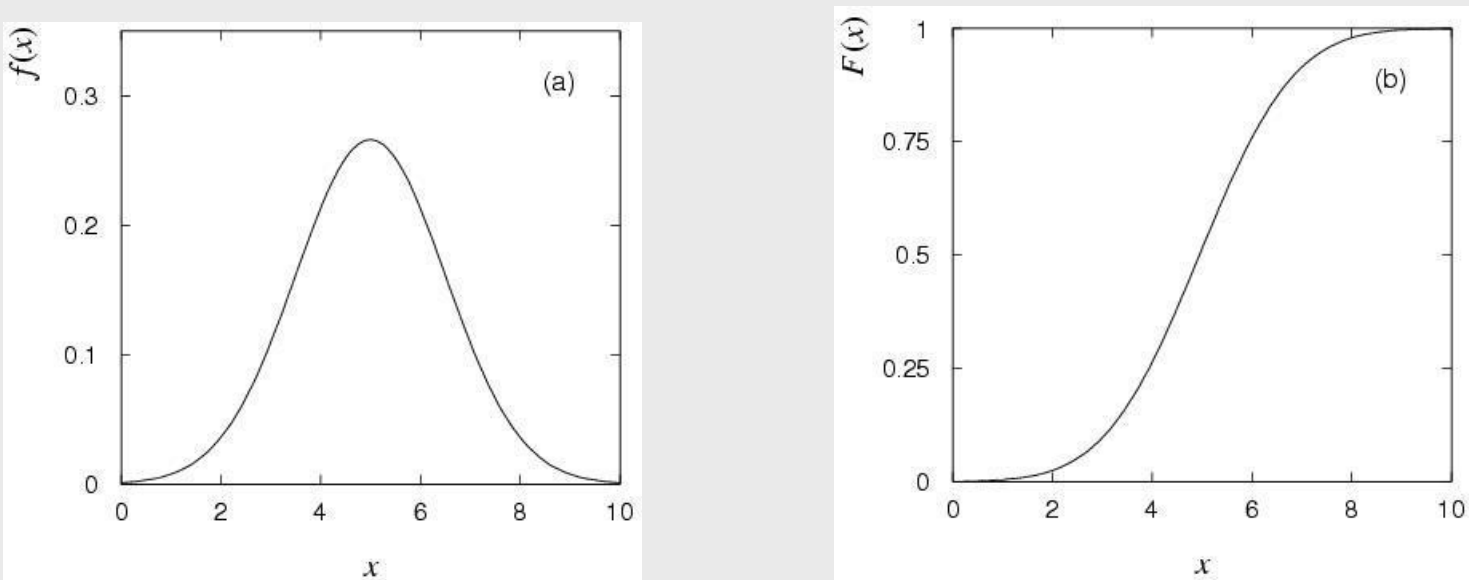
x must take on one of its possible values

Cumulative distribution function

Probability to have outcome less than or equal to *x* is

$$\int_{-\infty}^{x} f(x')\,dx' \equiv F(x)$$

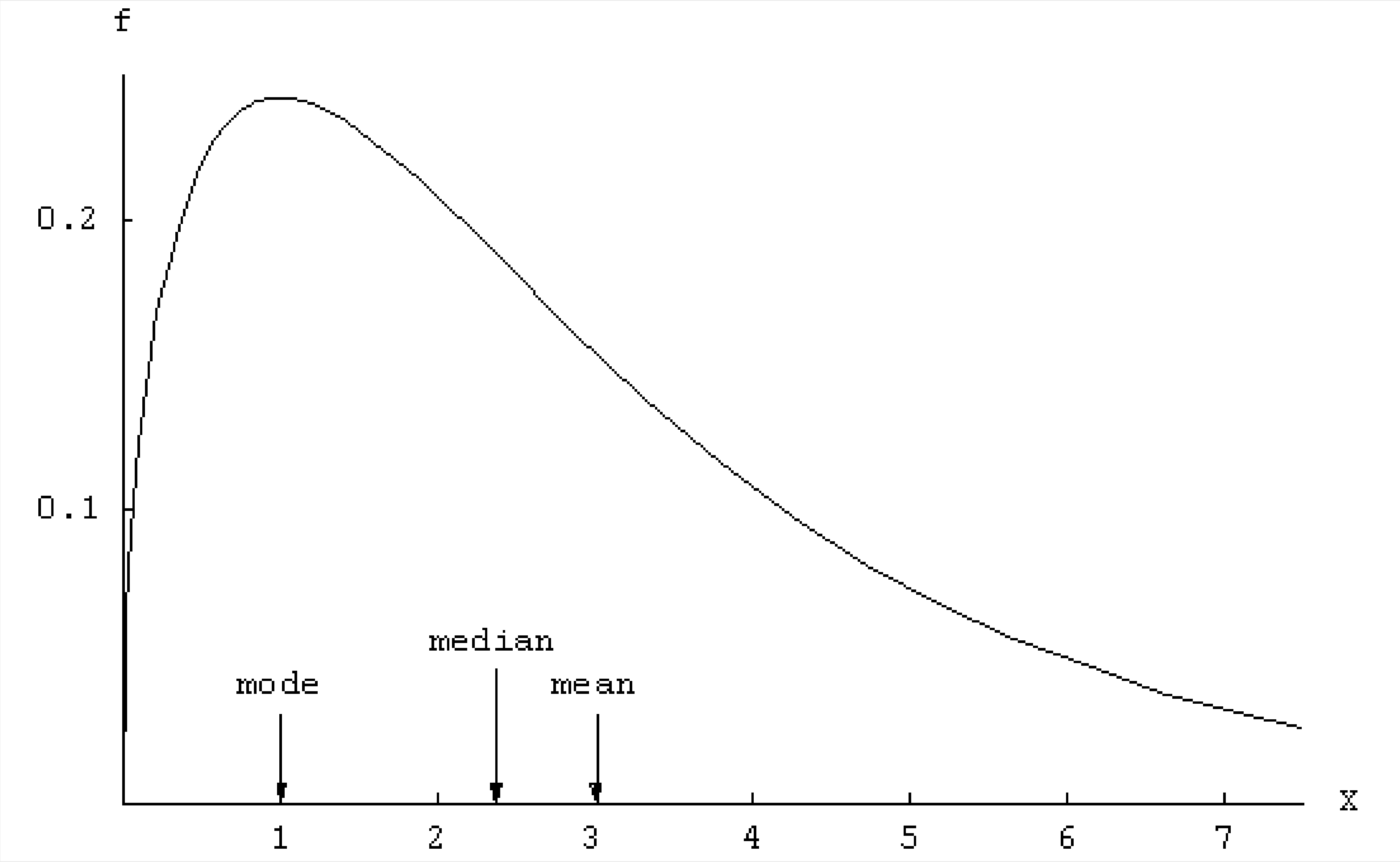
cumulative distribution function



Alternatively define pdf with $f(x) = \dfrac{\partial F(x)}{\partial x}$

Mean, median and mode

e.g.: Maxwells's velocity distribution



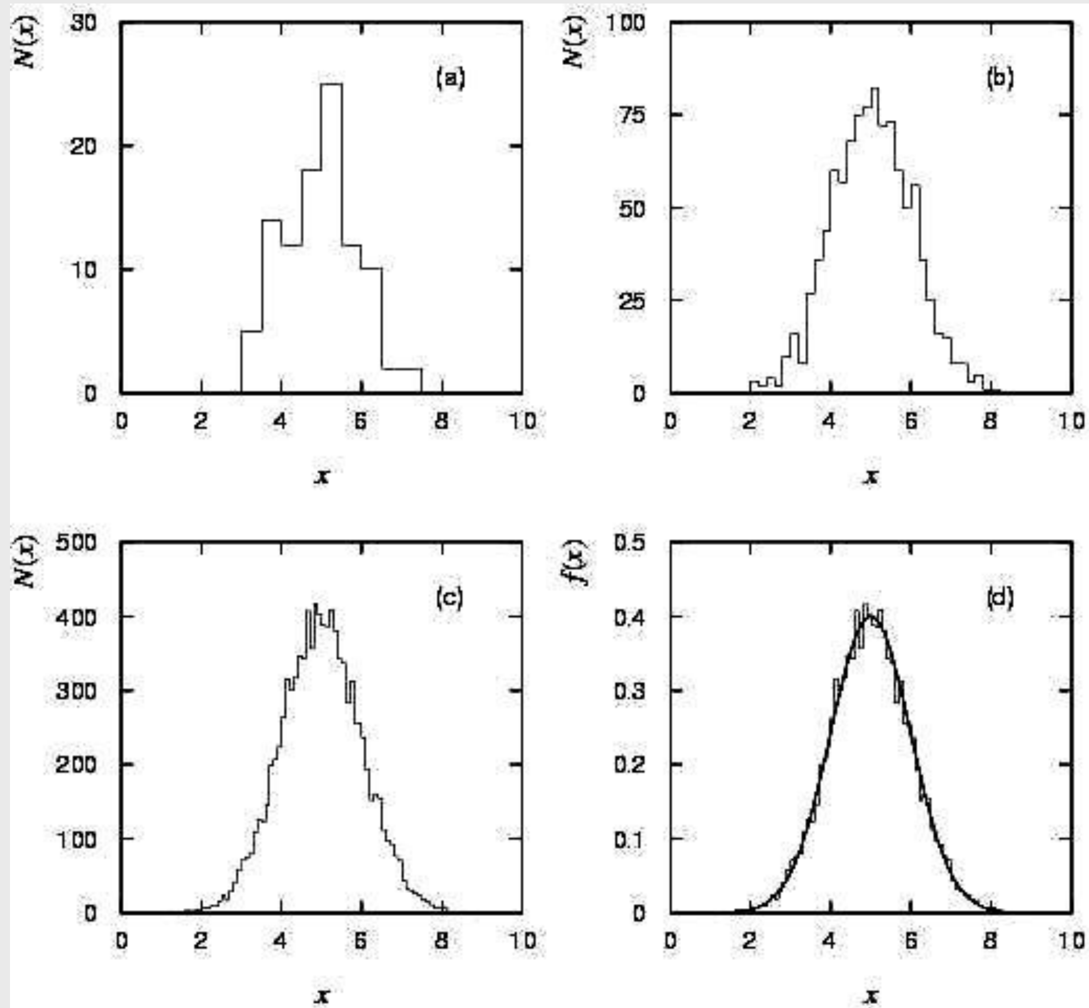
pdf = histogram with

- infinite data sample,
- zero bin width,
- normalized to unit area

$$f(x) = \frac{N(x)}{n\Delta x}$$

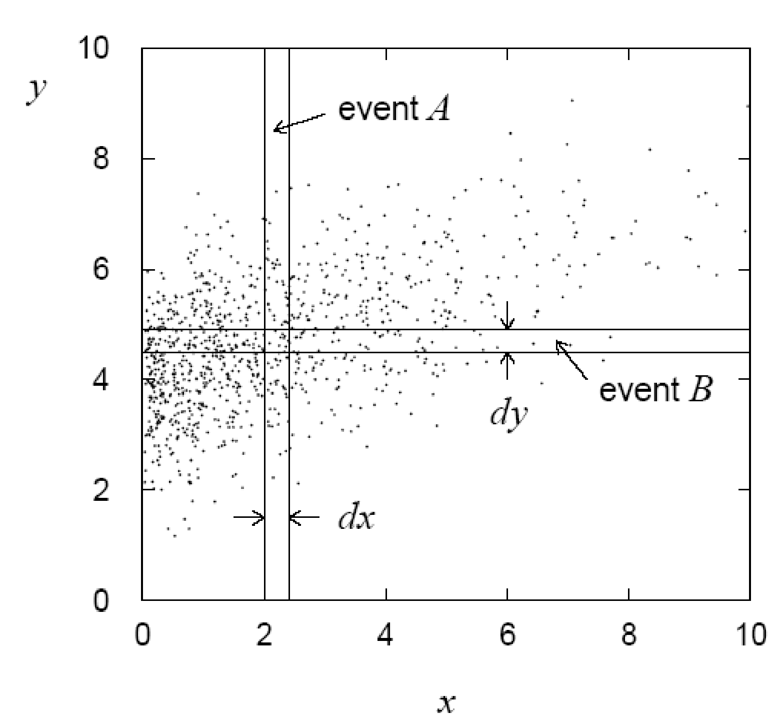
$n =$ number of entries

$\Delta x =$ bin width



Multivariate distributions

Outcome of experiment charac-terized by several values, e.g. an n-component vector, $(x_1, \ldots x_n)$



$$P(A \cap B) = f(x, y)\, dx\, dy$$

joint pdf

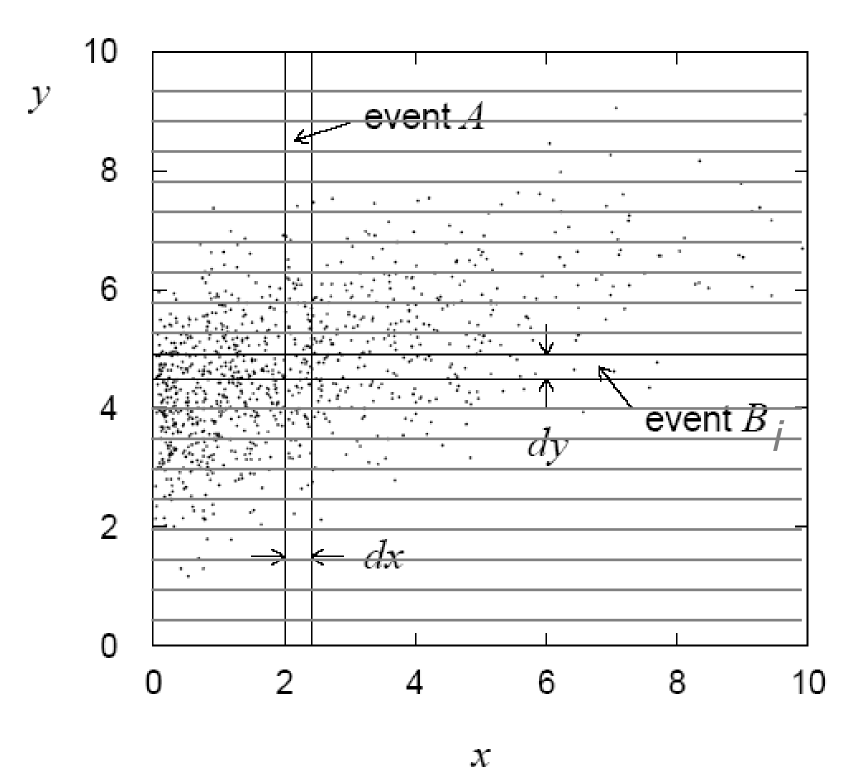
Normalization: $$\int \cdots \int f(x_1, \ldots, x_n) dx_1 \cdots dx_n = 1$$

Marginal pdf

Sometimes we want only pdf of some (or one) of the components:

$$
\begin{aligned}
P(A) &= \sum_i P(A \cap B_i) \\
&= \sum_i f(x, y_i)\, dy\, dx \\
&\to \int f(x,y)\, dy\, dx
\end{aligned}
$$

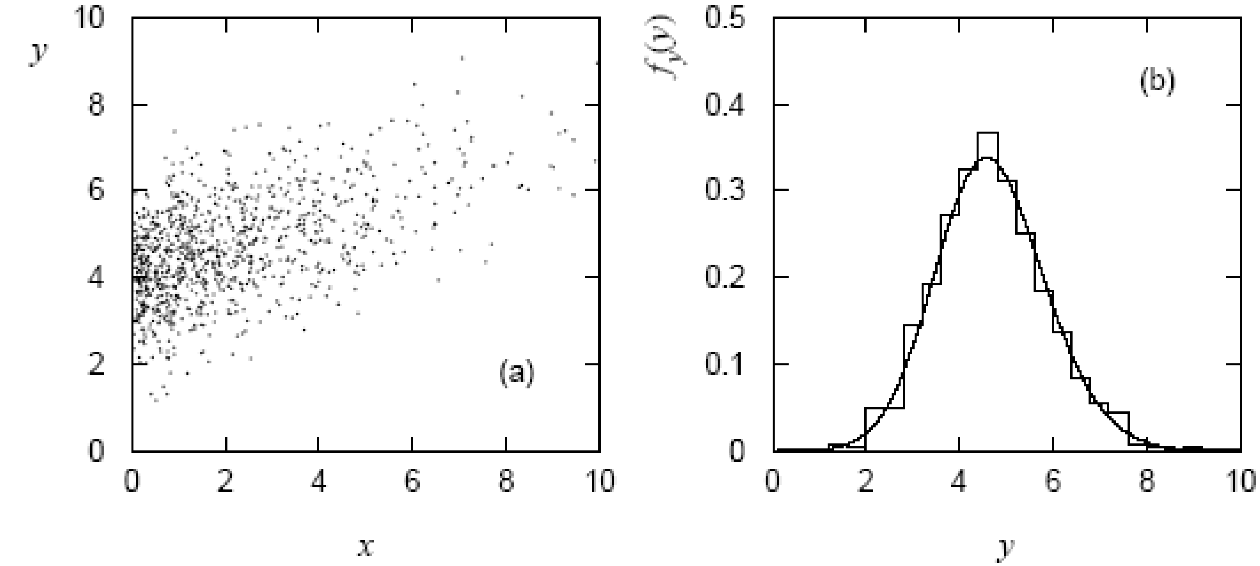
$$f_x(x) = \int f(x,y)\, dy$$

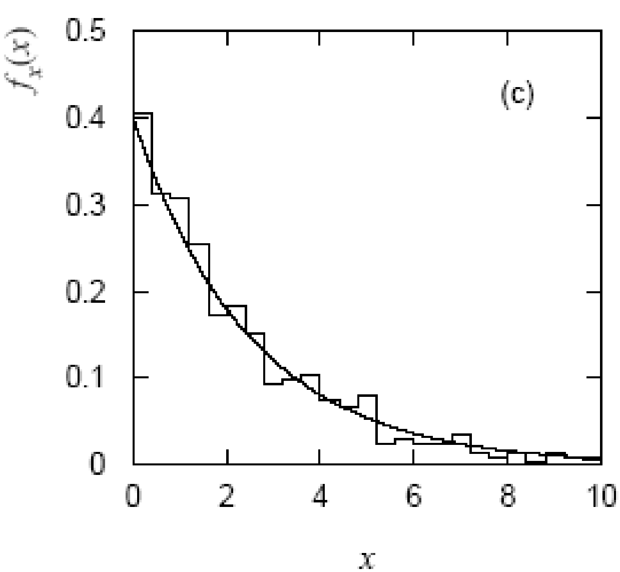


→ marginal pdf $f_1(x_1) = \int \cdots \int f(x_1, \ldots, x_n)\, dx_2 \ldots dx_n$

x_1, x_2 independent if $f(x_1, x_2) = f_1(x_1) f_2(x_2)$

Marginal pdf (2)





Marginal pdf ~ projection of joint pdf onto individual axes.

Conditional pdf

Sometimes we want to consider some components of joint pdf as constant. Recall conditional probability:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{f(x,y)\,dx\,dy}{f_x(x)\,dx}$$

→ conditional pdfs:

$$h(y|x) = \frac{f(x,y)}{f_x(x)}\,, \qquad g(x|y) = \frac{f(x,y)}{f_y(y)}$$

Bayes' theorem becomes:

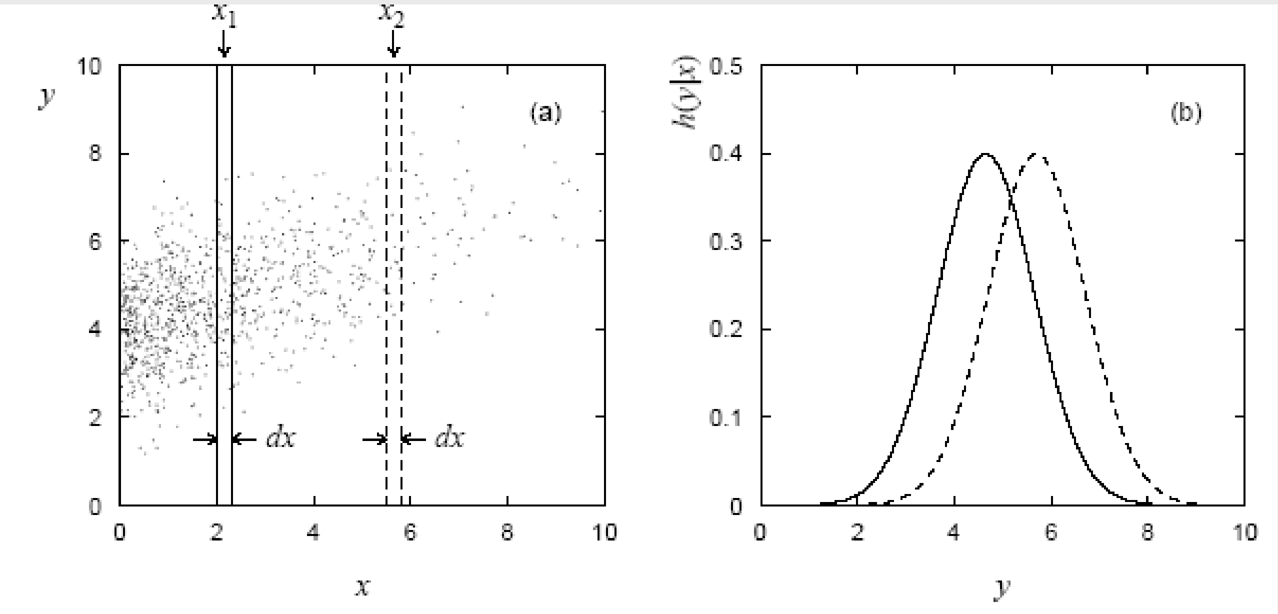
$$g(x|y) = \frac{h(y|x)f_x(x)}{f_y(y)}\,.$$

Recall *A*, *B* independent if $P(A \cap B) = P(A)P(B)\,.$

→ *x*, *y* independent if $f(x,y) = f_x(x)f_y(y)\,.$

Conditional pdfs (2)

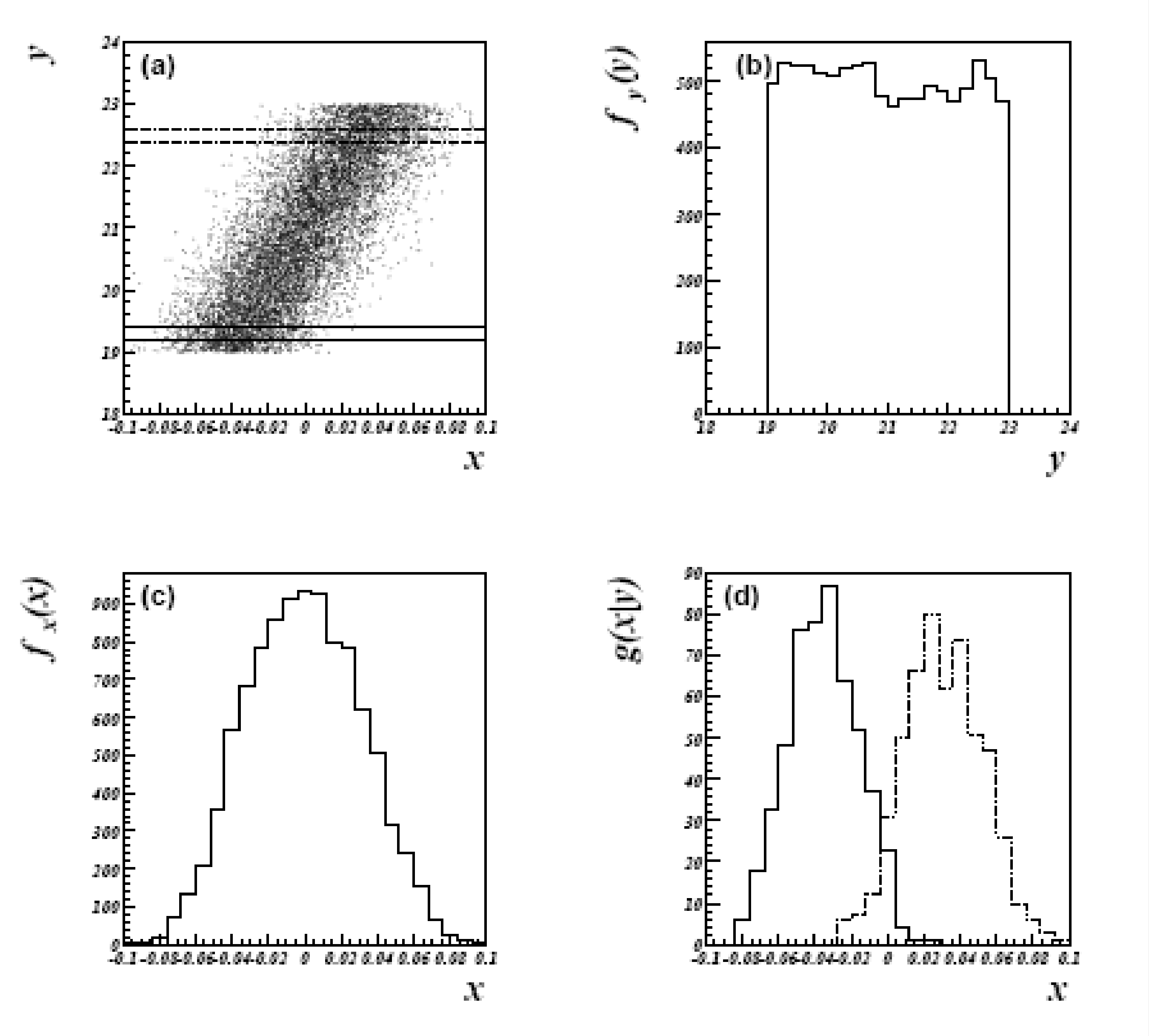
E.g. joint pdf $f(x,y)$ used to find conditional pdfs $h(y|x_1)$, $h(y|x_2)$:



Basically treat some of the r.v.s as constant, then divide the joint pdf by the marginal pdf of those variables being held constant so that what is left has correct normalization, e.g.,

$$\int h(y|x)\,dy = 1 .$$

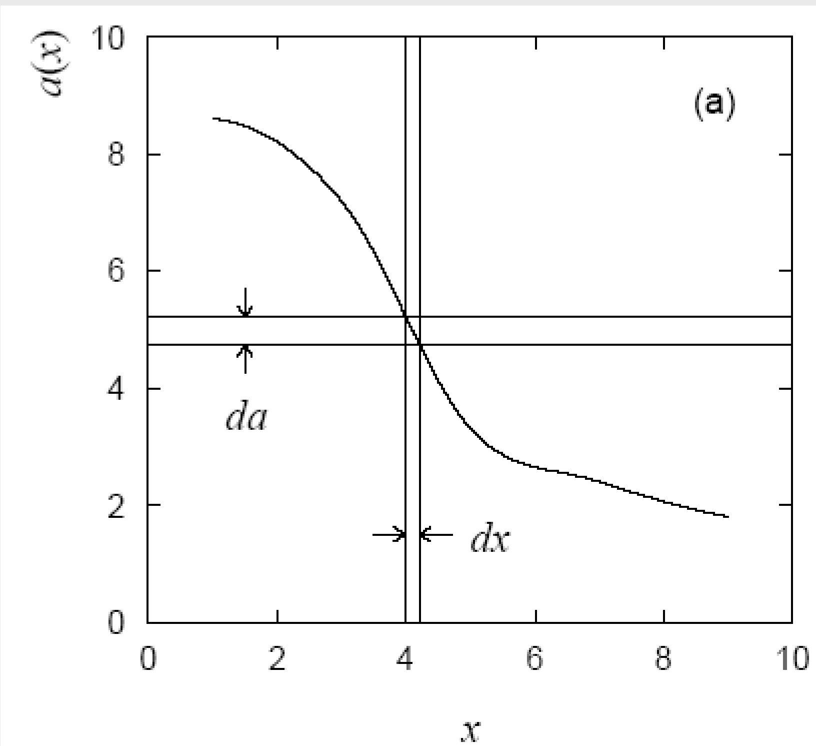
WDF für zwei Variablen mit Abhängigkeit



A function of a random variable is itself a random variable.

Suppose *x* follows a pdf *f*(*x*), consider a function *a*(*x*).

What is the pdf *g*(*a*)?



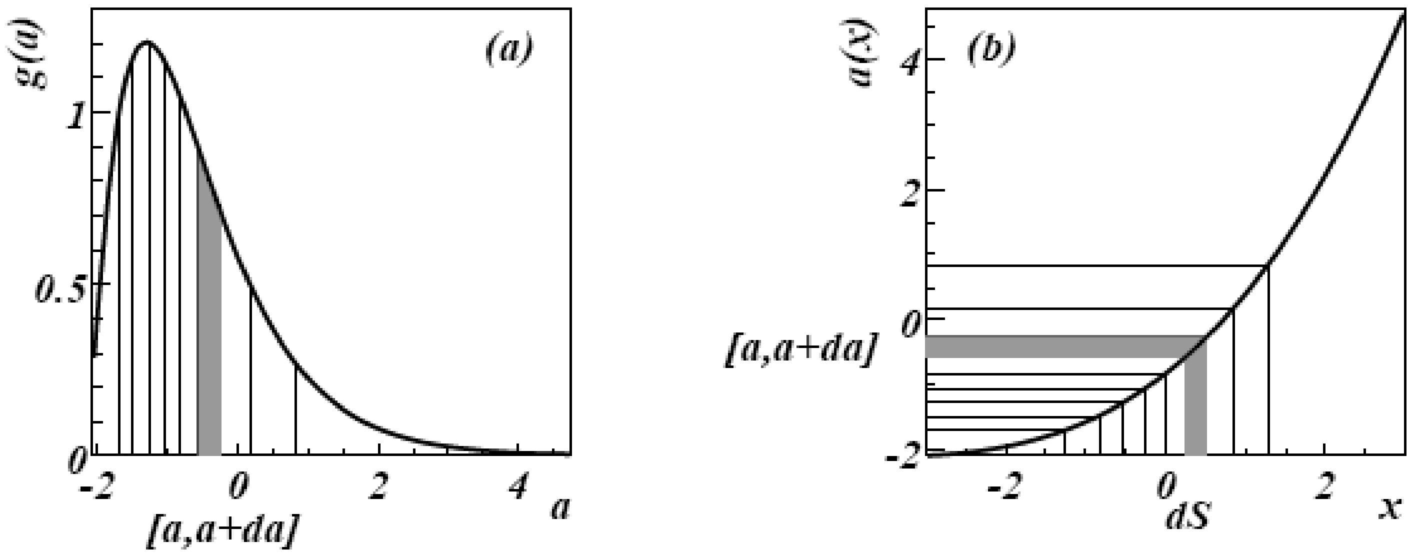
$$g(a)\,da = \int_{dS} f(x)\,dx$$

dS = region of *x* space for which *a* is in [*a*, *a*+*da*].

For one-variable case with unique inverse this is simply

$$g(a)\,da = f(x)\,dx$$

$$\rightarrow \quad g(a) = f(x(a)) \left|\frac{dx}{da}\right|$$



probability in a-space g(a)da

equals one in x-space f(x)dS

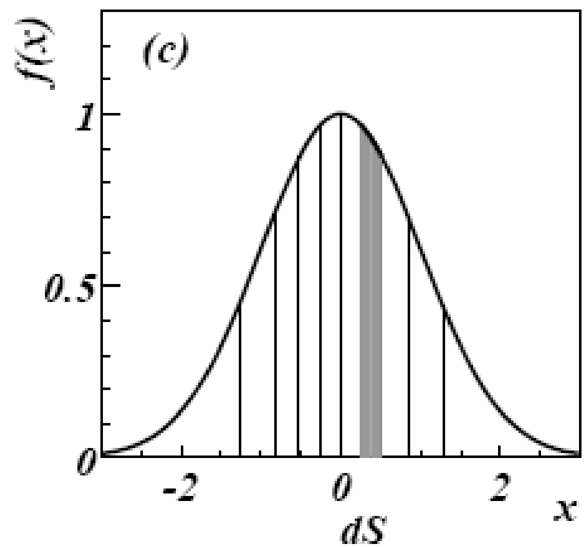
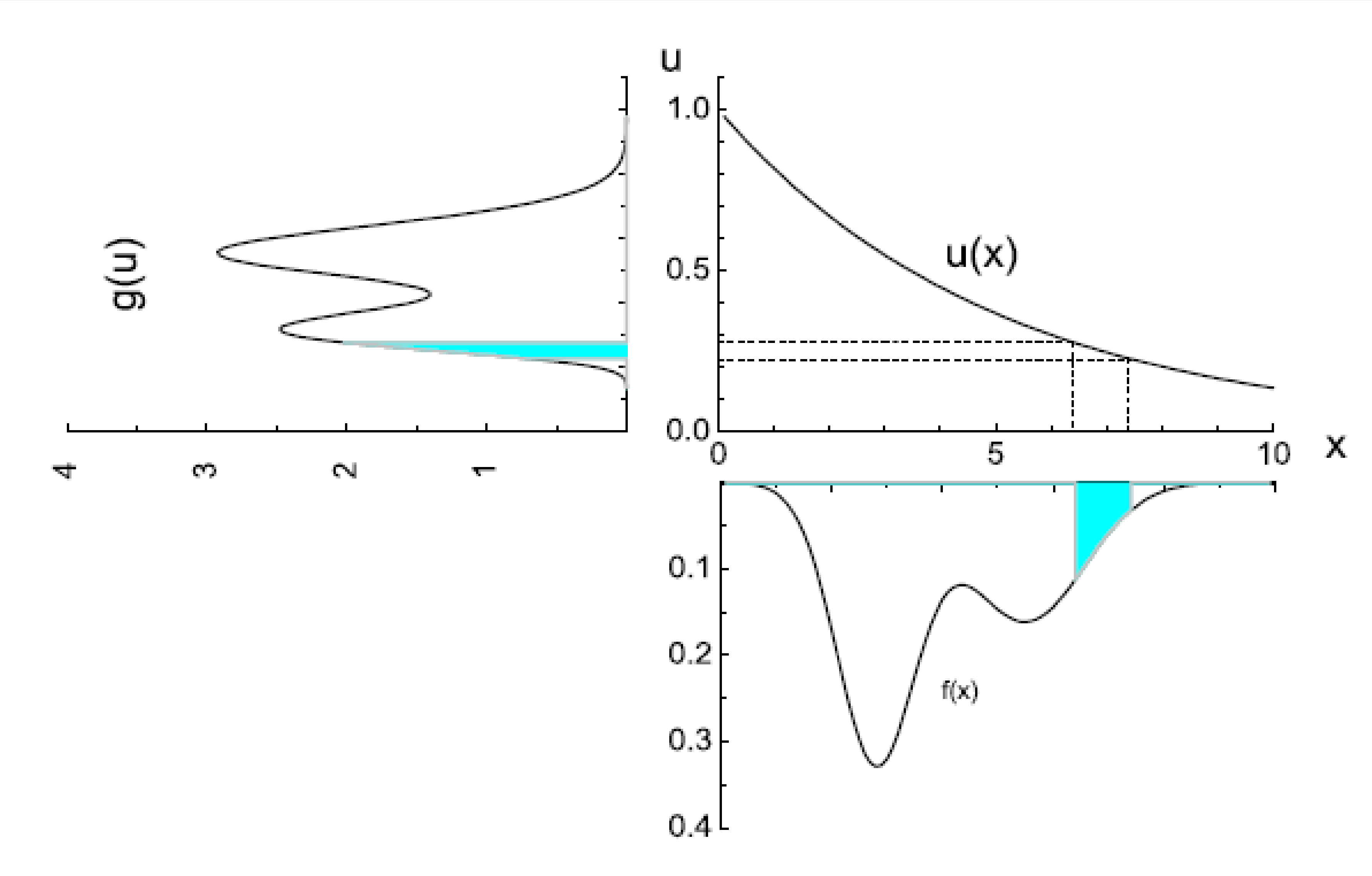
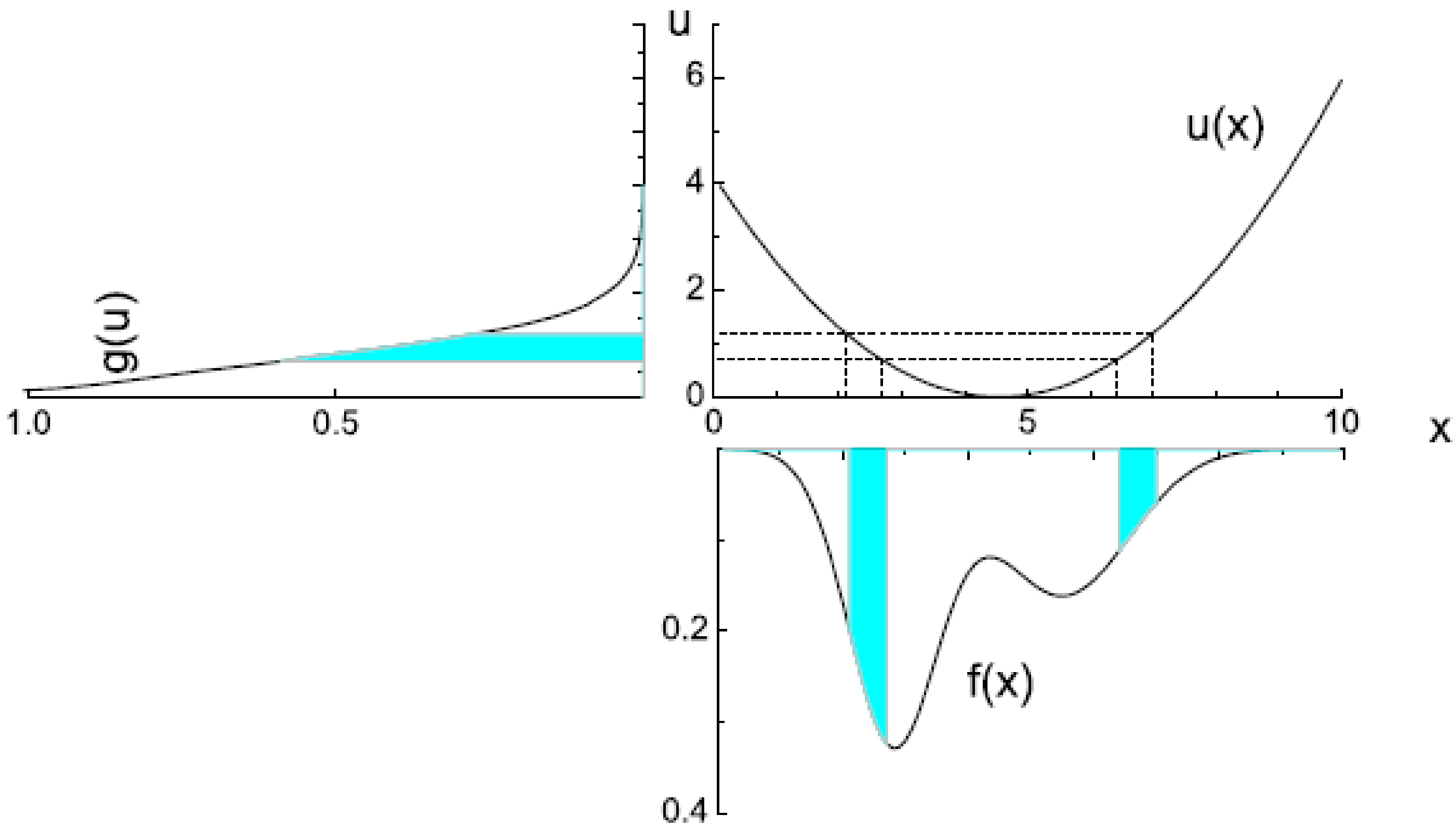


figure from Lutz Feld

Mapping the x and a spaces

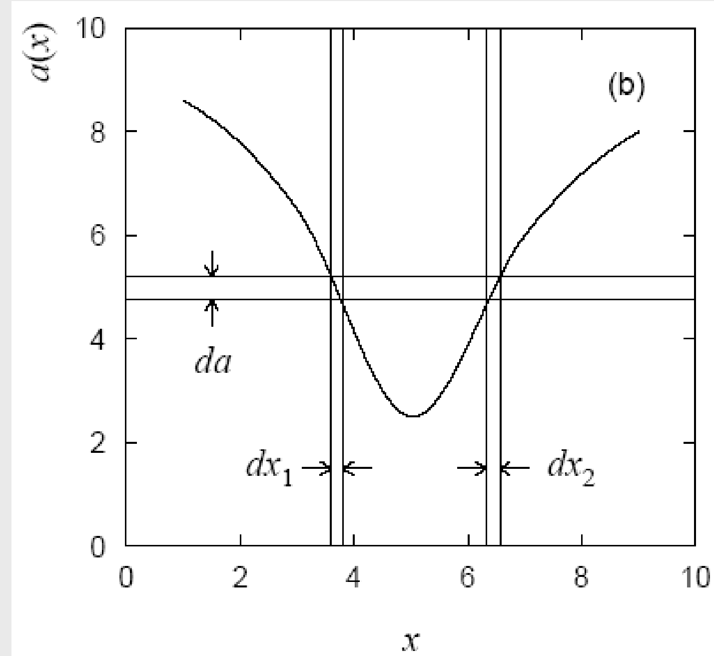


Mapping the x and a spaces: two „branches"



Functions without unique inverse

If inverse of *a*(*x*) not unique, include all *dx* intervals in *dS* which correspond to *da*:



Example:

$$a = x^2, \;\; x = \pm\sqrt{a}, \;\; dx = \pm\frac{da}{2\sqrt{a}}\,.$$

$$dS = \left[\sqrt{a}, \sqrt{a} + \frac{da}{2\sqrt{a}}\right] \cup \left[-\sqrt{a} - \frac{da}{2\sqrt{a}}, -\sqrt{a}\right]$$

$$g(a) = \frac{f(\sqrt{a})}{2\sqrt{a}} + \frac{f(-\sqrt{a})}{2\sqrt{a}}$$

Consider r.v.s $\vec{x} = (x_1, \ldots, x_n)$ and a function $a(\vec{x})$.

$$g(a')da' = \int \ldots \int_{dS} f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

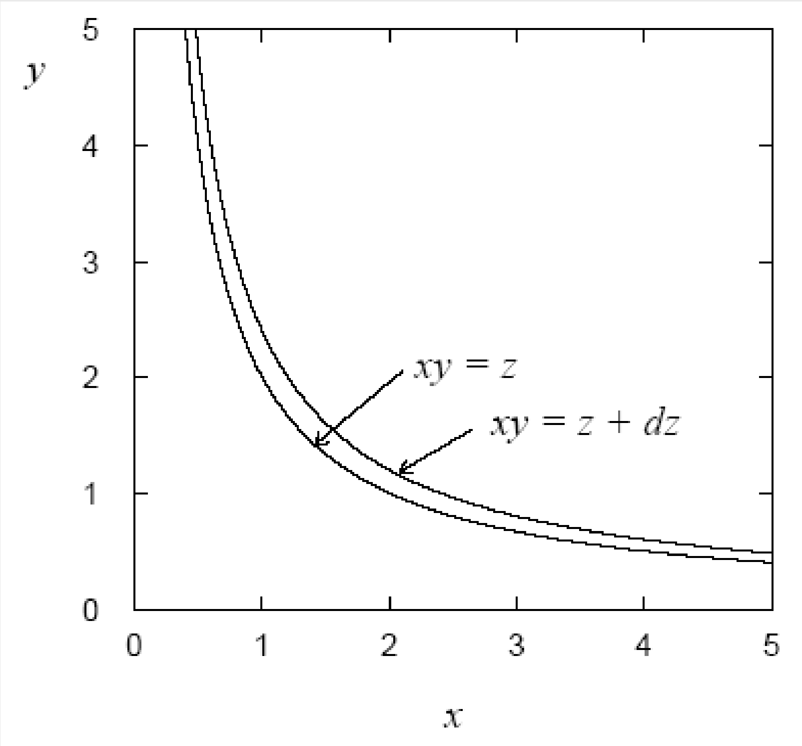
dS = region of *x*-space between (hyper)surfaces defined by

$$a(\vec{x}) = a', \; a(\vec{x}) = a' + da'$$

Prof. Markus Schumacher Chapter 2: Fundamental

Example: r.v.s *x*, *y* > 0 follow joint pdf *f*(*x*,*y*),

consider the function *z* = *xy*. What is *g*(*z*)?



$$g(z)\,dz = \int \ldots \int_{dS} f(x,y)\,dx\,dy$$

$$= \int_0^\infty dx \int_{z/x}^{(z+dz)/x} f(x,y)\,dy$$

$$\rightarrow \quad g(z) = \int_0^\infty f(x, \frac{z}{x})\,\frac{dx}{x}$$

$$= \int_0^\infty f(\frac{z}{y}, y)\,\frac{dy}{y}$$

(Mellin convolution)

More on transformation of variables

Consider a random vector $\vec{x} = (x_1, \ldots, x_n)$ with joint pdf $f(\vec{x})$.

Form *n* linearly independent functions $\vec{y}(\vec{x}) = (y_1(\vec{x}), \ldots, y_n(\vec{x}))$

for which the inverse functions $x_1(\vec{y}), \ldots, x_n(\vec{y})$ exist.

Then the joint pdf of the vector of functions is $g(\vec{y}) = |J| f(\vec{x})$

where *J* is the Jacobian determinant:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & & & \vdots \\ & & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

For e.g. $g_1(y_1)$ integrate $g(\vec{y})$ over the unwanted components.

Expectation values

Consider continuous r.v. x with pdf $f(x)$.

Define expectation (mean $E[x] = \int x\, f(x)\, dx$

Notation (often): $E[x] = \mu$ ~ "centre of gravity" of pdf.

For a function $y(x)$ with pdf $g(y)$,

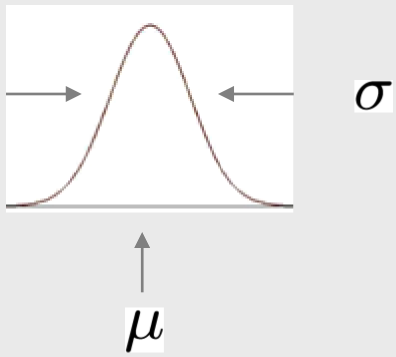
$$E[y] = \int y\, g(y)\, dy = \int y(x) f(x)\, dx$$ (equivalent)

Variance: $V[x] = E[x^2] - \mu^2 = E[(x-\mu)^2]$

Notation: $V[x] = \sigma^2$

Standard deviation: $\sigma = \sqrt{\sigma^2}$

σ ~ width of pdf, same units as x.



Covariance and correlation

Define covariance cov[*x*,*y*] (also use matrix notation V_{xy}) as

$$\mathrm{cov}[x,y] = E[xy] - \mu_x \mu_y = E[(x-\mu_x)(y-\mu_y)]$$

Correlation coefficient (dimensionless) defined as

$$\rho_{xy} = \frac{\mathrm{cov}[x,y]}{\sigma_x \sigma_y}$$

If *x*, *y*, independent, i.e., $f(x,y) = f_x(x) f_y(y)$, then

$$E[xy] = \int\int xy\, f(x,y)\, dx dy = \mu_x \mu_y$$

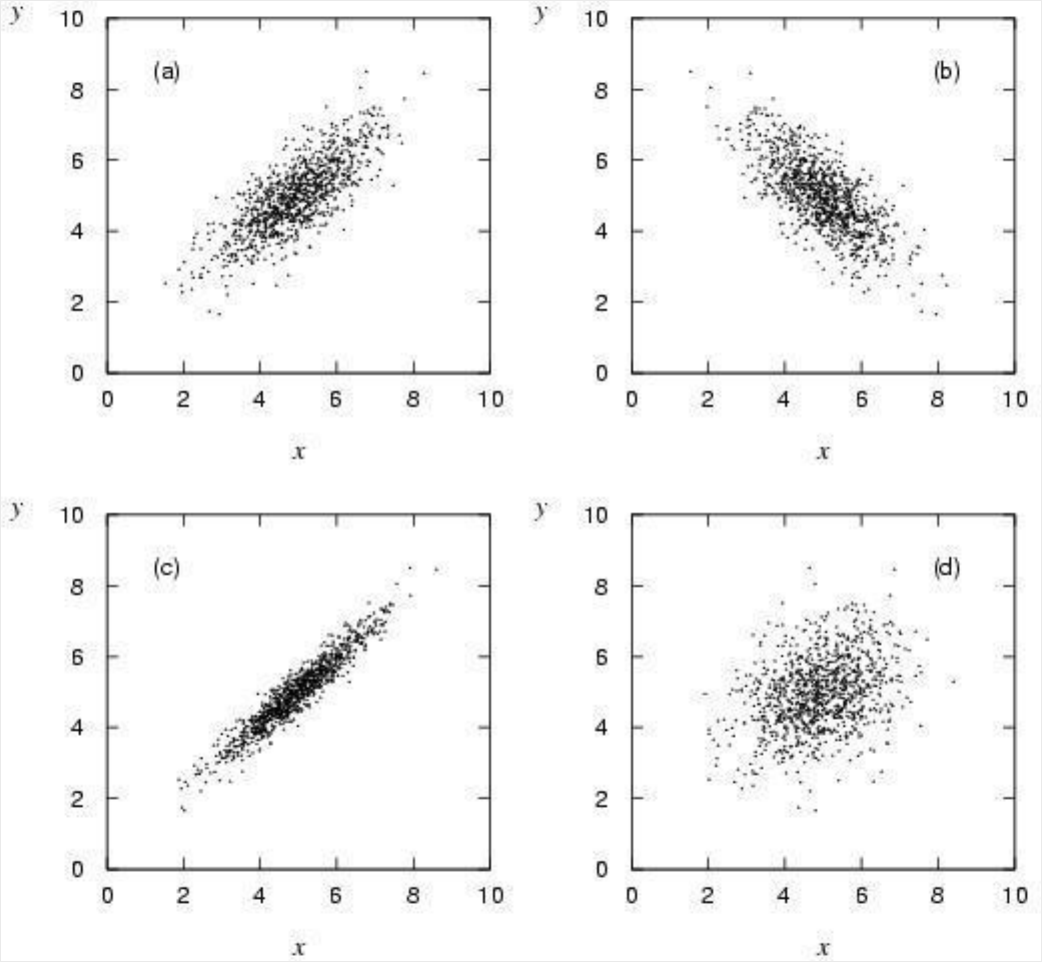
→ $\mathrm{cov}[x,y] = 0$ *x* and *y*, 'uncorrelated'

N.B. converse not always true.

Correlations

$\rho = 0.75$

$\rho = -0.75$



$\rho = 0.95$

$\rho = 0.25$

Korrelationen