Statistical Methods in Particle Physics

Lecture 3

October 24, 2011

Silvia Masciocchi, GSI Darmstadt

s.masciocchi@gsi.de

Winter Semester 2011 / 12

Outline

This lecture:

- Short revision of:
  - Probability density function
  - Cumulative distribution function
- Functions of random variables
- Expectation values
- Covariance, correlation
- Error propagation

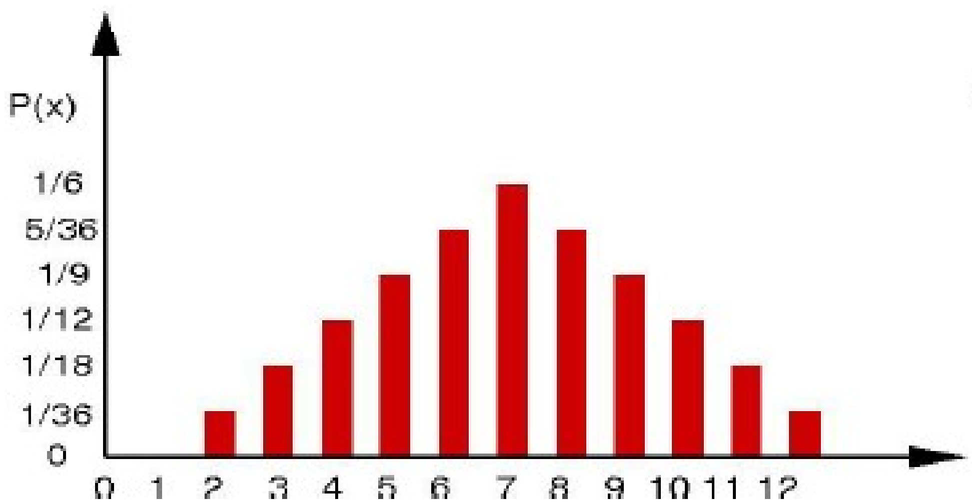
Next time:

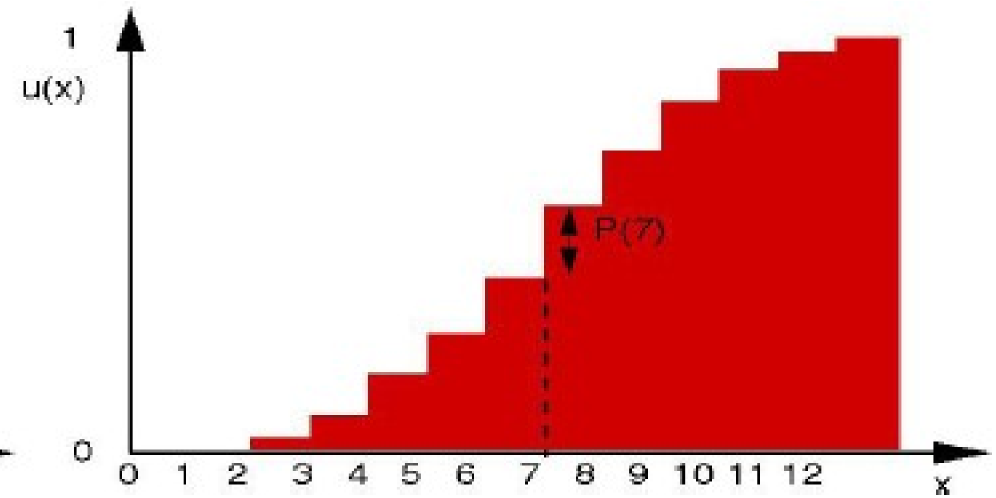
- Catalog of pdf's

Random variables

Typical example: throwing two dices

- Result of each “experiment”:
  {11, 12, 13, 14, 15, 16, 21, 22, … , 63, 64, 65, 66}
- Random variable x = sum of dices
  → possible values (discrete!): x_i = 2, 3, 4, … , 11, 12
- Probability for each value x_i





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

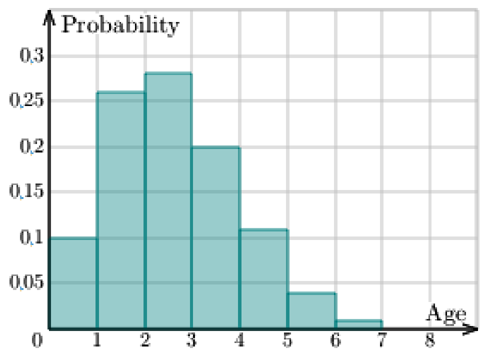
Cumulative distribution u(x):
probability to observe x or smaller value

Discrete and continuous

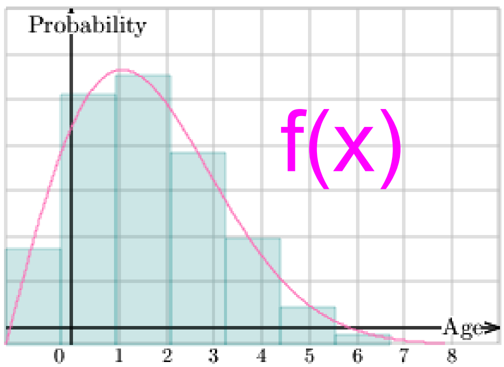
One more example: age of a rented car (x)

Age (Years)	0-1	1-2	2-3	3-4	4-5	5-6	6-7
Probability	.10	.26	.28	.20	.11	.04	.01

The histogram of probabilities can be described by a function f



x



x

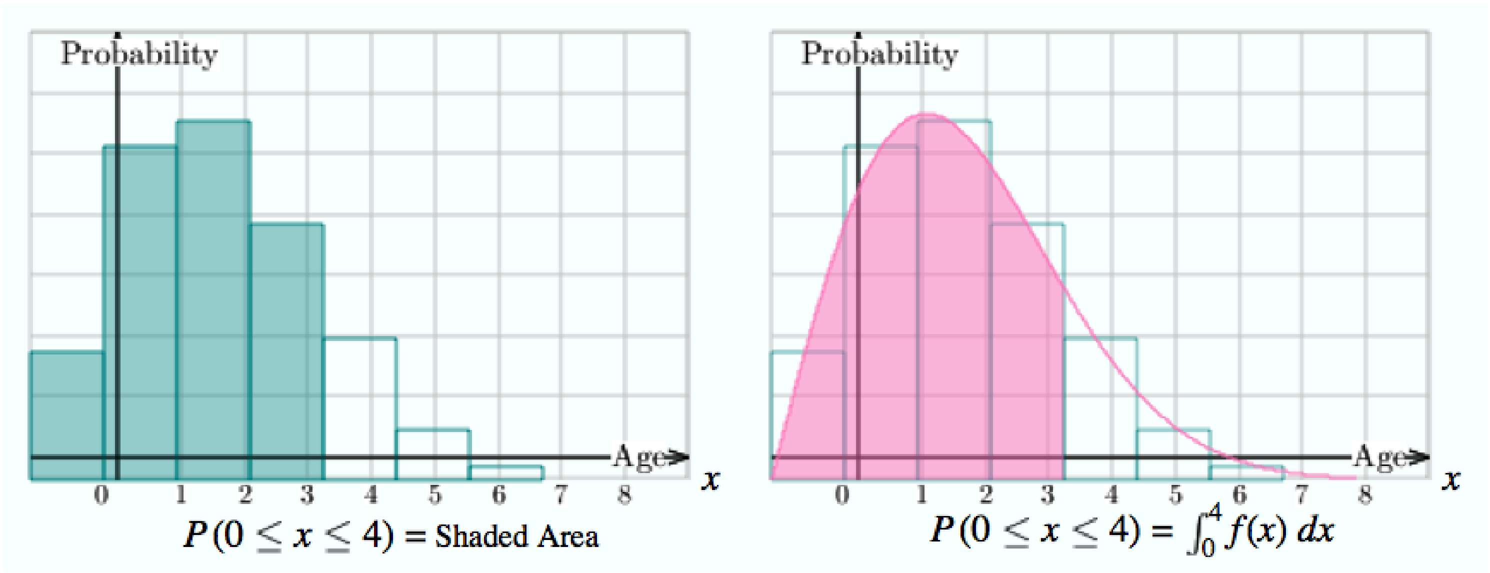
f(x) is called **probability density function**.

The domain of f is the whole range of values which x can take.

We use it to calculate the probability of the given variable x to be in an interval [a,b]

Discrete and continuous

Probability for the rental car to have age between 0 and 4 years:



Probability density functions

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)**

With:

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

Normalization
(x must be somewhere)

Note:

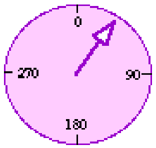
- $f(x) \geq 0$
- f(x) is NOT a probability ! It has dimension 1/x !

Exercises

1. For what constant k is $f(x)=ke^{-x}$ a probability density function on [0,1]?

    If f is any non-negative function with domain some interval (a,b), then the process of choosing a suitable constant k to make $\int_a^b k\, f(x)dx = 1$ is called **normalizing** the function f

2. Suppose that you spin the dial shown in the figure so that it comes to rest at a random position. Model this with a suitable probability density function, and use it to find the probability that the dial will land somewhere between 5° and 300°.



The **uniform density function** on the interval [a,b] is the constant function defined by

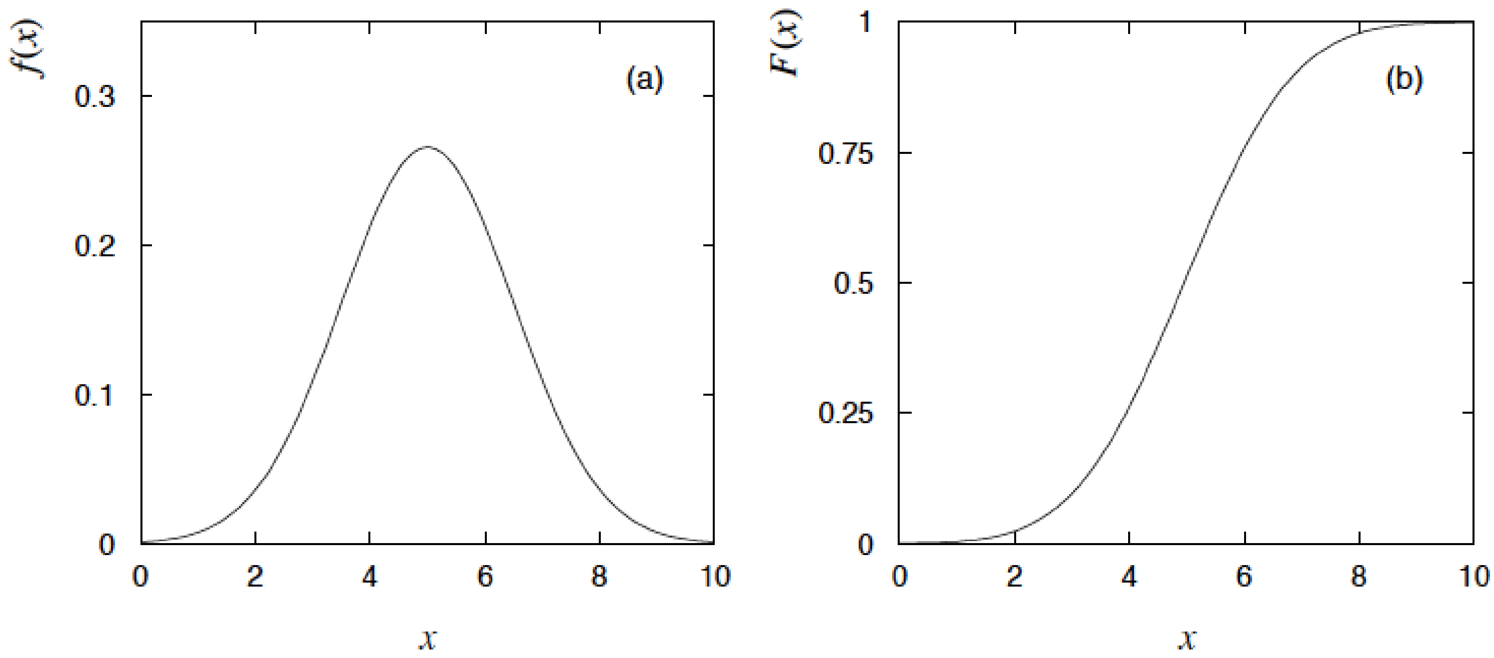
$$f(x) = \frac{1}{b-a}$$

Cumulative distribution function (cdf)

Given a pdf f(x'), probability to have outcome less then or equal to x, is:

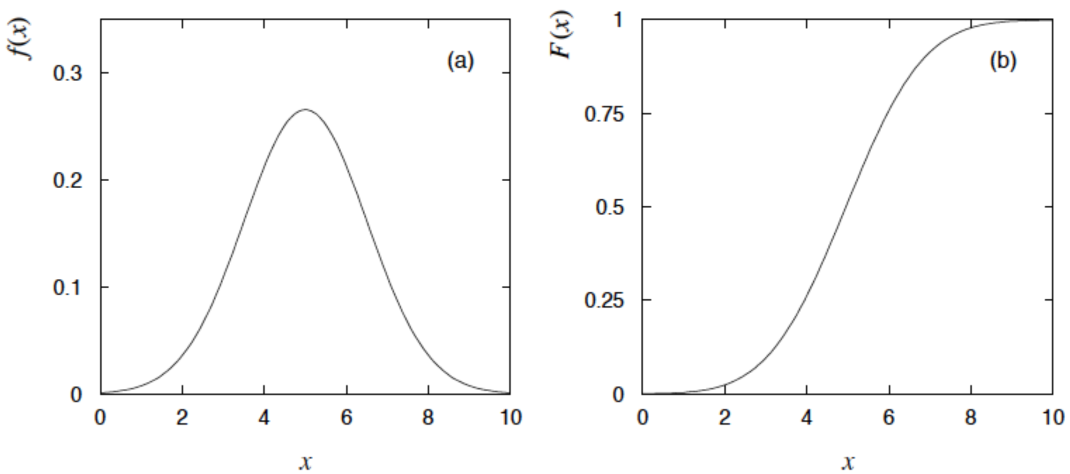
$$\int_{-\infty}^{x} f(x')\,dx' = F(x)$$

Cumulative distribution function



Cumulative distribution function (cdf)

$$\int_{-\infty}^{x} f(x')dx' = F(x)$$ **Cumulative distribution function**



- F(x) is a continuously non-decreasing function
- $F(-\infty)= 0$, $F(\infty)=1$
- For well behaved distributions:

$$\text{pdf}: \quad f(x) = \frac{\partial F(x)}{\partial x}$$

Exercise

1. Given the probability density function:

$$f(x) = \begin{cases} |1-x| & \text{for } x \text{ in } [0,2] \\ 0 & \text{elsewhere} \end{cases}$$

- compute the cdf F(x)
- what is the probability to find x > 1.5 ?
- what is the probability to find x in [0.5,1] ?

DONE

Exercises

- **T-Shirts** The age (in years) of randomly chosen T-shirts in your wardrobe from last summer is distributed according to the density function $f(x)=10/9x^2$ with $1 \leq x \leq 10$. Find the probability that a randomly chosen T-shirt is between 2 and 8 years old.

- **The Doomsday Meteor** The probability that a "doomsday meteor" will hit the earth in any given year and release a billion megatons or more of energy is on the order of 0.000 000 01. If X is the year in which a doomsday meteor hits the earth, then it may be modeled with an associated probability density function given by $f(x)=ae^{-ax}$ with a=0000 000 01.

  (a) What is the probability that the earth will be hit by a doomsday meteor at least once during the next 100 years? (Give the answer correct to 2 significant digits.)

  (b) What is the probability that the earth has been hit by a doomsday meteor at least once since the appearance of life (about 4 billion years ago)?

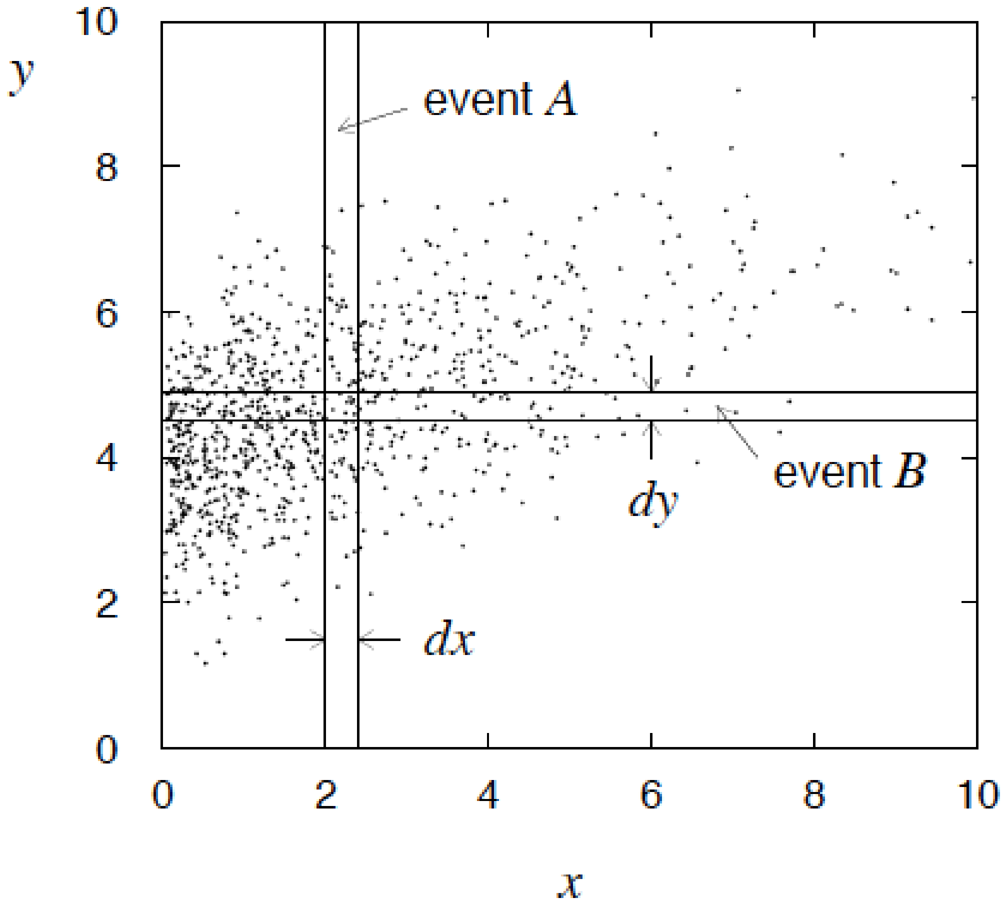
Multivariate distributions

$$f(x,y)$$

The outcome of the experiment is characterized by more than 1 quantity, e.g. by x and y

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

Joint pdf



Normalization:

$$\iint f(x,y)\, dx\, dy = 1$$

$$\iint \ldots \int f(x_1, x_2, \ldots x_n)\, dx_1\, dx_2 \ldots dx_n = 1$$

Exercise

Let X and Y have the joint probability density function

$$f(x,y) = \frac{3}{2}x^2(1-y) \qquad \text{for } -1<x<1,\ -1<y<1$$

Let A={(x,y): 0<x<1, 0<y<x}

Find the probability that (X,Y) falls in A.

Marginal pdf's

From a multivariate distribution

$$f(x,y)\,dx\,dy$$

(e.g. scatter plot)

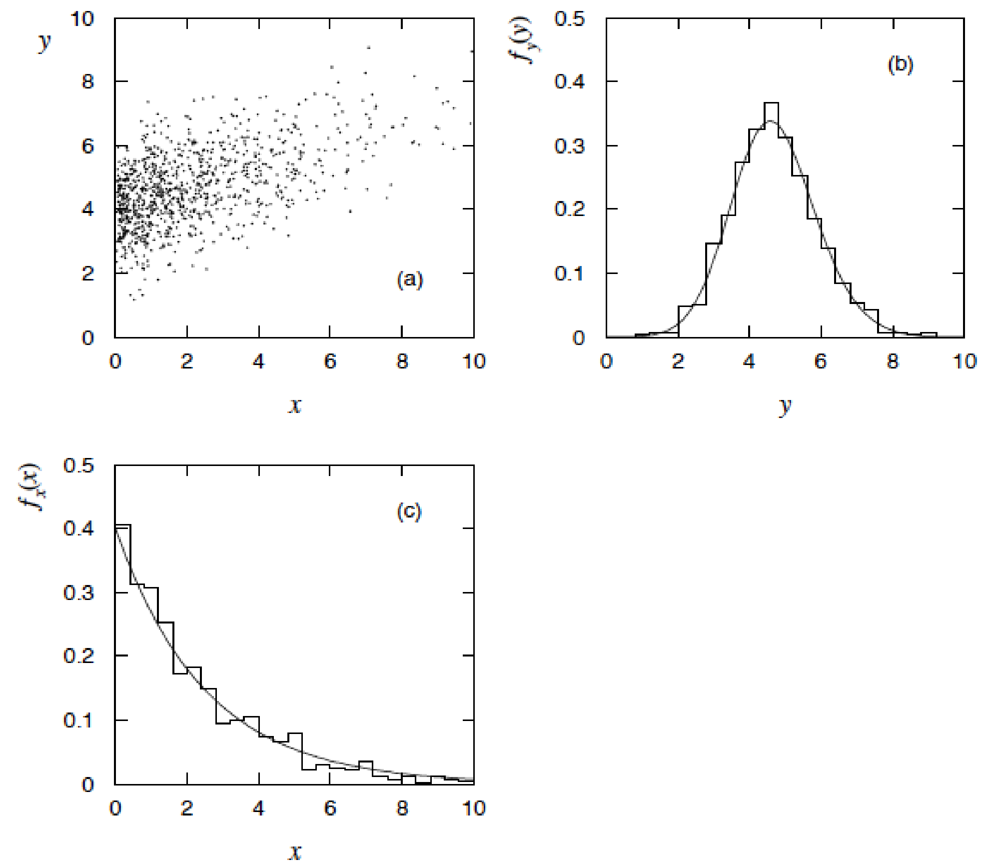
we might be in interested only in the pdf of ONE of the components (x or y, here)

→ projection of joint pdf onto individual axes

Marginal pdf

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

$$f_y(y) = \int f(x,y)\,dx$$



Distribution of a single variable which is part of a multivariate distribution

Conditional pdf

Recall the conditional probability:

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

We define:

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

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

Conditional probability density functions

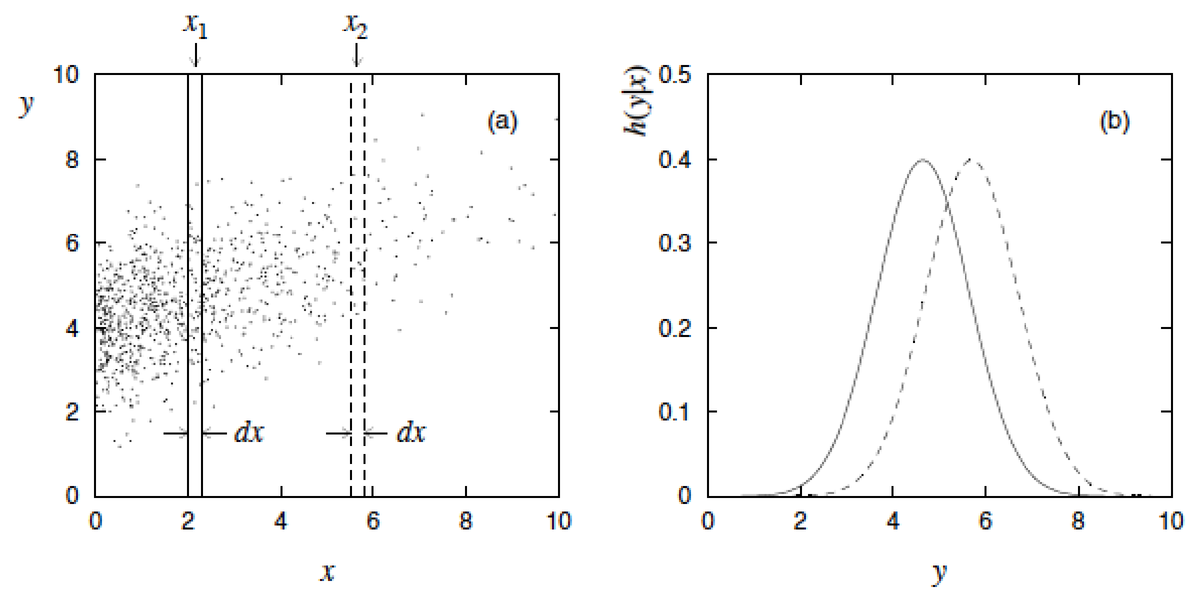
Bayes' theorem becomes: $$g(x|y) = \frac{h(y|x)\,f_x(x)}{f_y(y)}$$

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

Then: x, y independent if $f(x,y) = f_x(x)\,f_y(y)$

Conditional pdf (2)

Example: joint pdf f(x,y) is used to find the conditional pdf's $h(y|x_1)$ and $h(y|x_2)$



Basically treat some of the random variables as constant, then divide the joint pdf by the marginal pdf of those variables being held constant

→ so that what is left has the correct normalization $\int h(y|x)\,dy = 1$

Exercise

A soda machine has a random amount Y_2 gallons of soda at the beginning of the day and dispenses Y_1 gallons over the course of the day (which must be less than or equal to Y_2). The two variables have the following joint density:

$$f(y_1, y_2) = \begin{cases} \frac{1}{2}, 0 \le y_1 \le y_2 \le 2 \\ 0 \quad \text{elsewhere} \end{cases}$$

Find the conditional density of Y_1 given $Y_2 = y_2$ and the probability that less than ½ gallon will be sold if the machine has 1.5 gallon at the start of the day.

Functions of a random variable

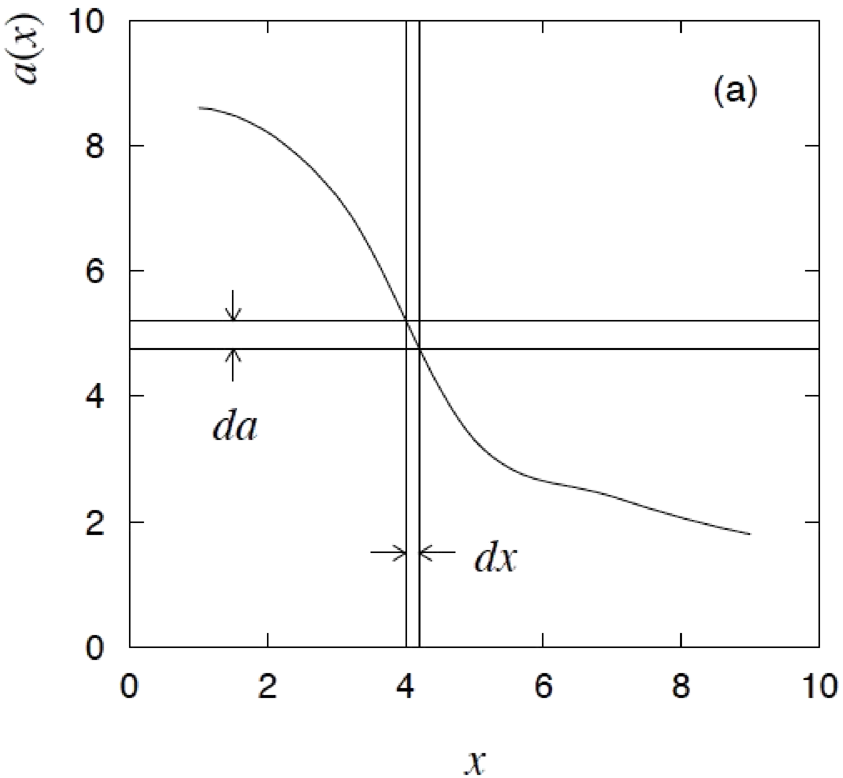
A function of a random variable is itself a random variable.

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].

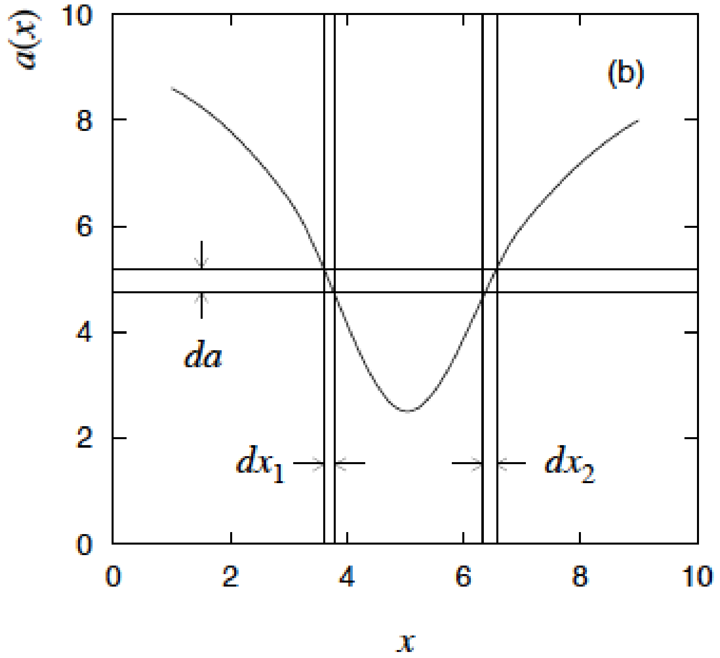


For one-variable case with unique inverse this is simply:

$$g(a)\,da = \left|\int_{x(a)}^{x(a+da)} f(x')\,dx'\right| = \int_{x(a)}^{x(a)+\left|\frac{dx}{da}\right|da} f(x')\,dx' \qquad g(a) = f(x(a))\left|\frac{dx}{da}\right|$$

Functions without unique inverse

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



Example: $a = x^2, \quad x = \pm\sqrt{a}, \quad dx = \pm\frac{da}{2\sqrt{a}}$

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

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

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

Functions of more than one random variable

Consider the random variables $\vec{x} = (x_1, x_2, ..., x_n)$

And the function $a(\vec{x})$

Its probability density function is:

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

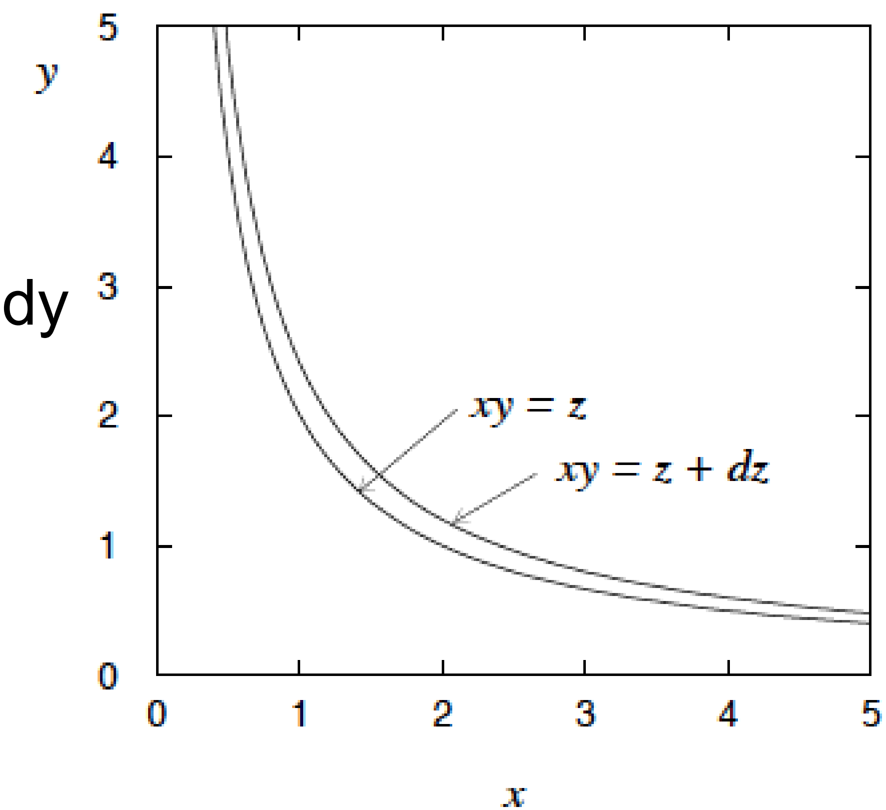
dS = region of $\vec{x}$ space between (hyper)surfaces defined by:

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

Example

Consider the random variables x, y>0, which follow the joint pdf f(x,y).
Consider the function z=xy. What is its pdf g(z)?

$$g(z)\,dz = \int \ldots \int_{dS} f(x,y)\,dx\,dy$$
$$= \int_0^{\infty} dx \int_{z/x}^{(z+dz)/x} f(x,y)\,dy$$
$$g(z) = \int_0^{\infty} f\left(x,\frac{z}{x}\right)\frac{dx}{x}$$
$$= \int_0^{\infty} f\left(\frac{z}{y},y\right)\frac{dy}{y}$$



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.

The joint pdf of the vector of functions is $g(\vec{y})$ Is

$$g(\vec{y}) = |J| f(\vec{x})$$

where J is the Jacobian determinant:

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

Expectation value

Consider a continuous random variable x with pdf f (x).

Define **expectation (mean) value** as

$$E[x] = \int x f(x) dx$$

E[x] is NOT a function of x, it is rather a parameter of f(x)

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 and standard deviation

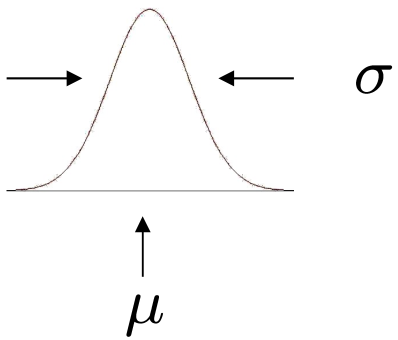
Variance:

μ

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

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

Standard deviation: $\sigma = \sqrt{\sigma^2}$ Same dimension as x



Exercises

- Find the mean of the random variable X that has probability density function f given by:

  $f(x) = x^2 / 3$ for $-1<x<2$

- Suppose that X has the power distribution with parameter a > 1, which has density:

  $f(x) = (a - 1)x^{-a}$ for $x > 1$

  Show that:

$$E[x] = \begin{cases} \infty, \text{if } 1 < a \leq 2 \\ \dfrac{a-1}{a-2}, \text{if } a > 2 \end{cases}$$

- Let the random variable x have the probability density function

$$f(x) = \begin{cases} 3x^2, \text{if } 0 \leq x \leq 1 \\ 0, \text{elsewhere} \end{cases}$$

Calculate its variance.

Covariance and correlation

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

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

Can be written as:

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

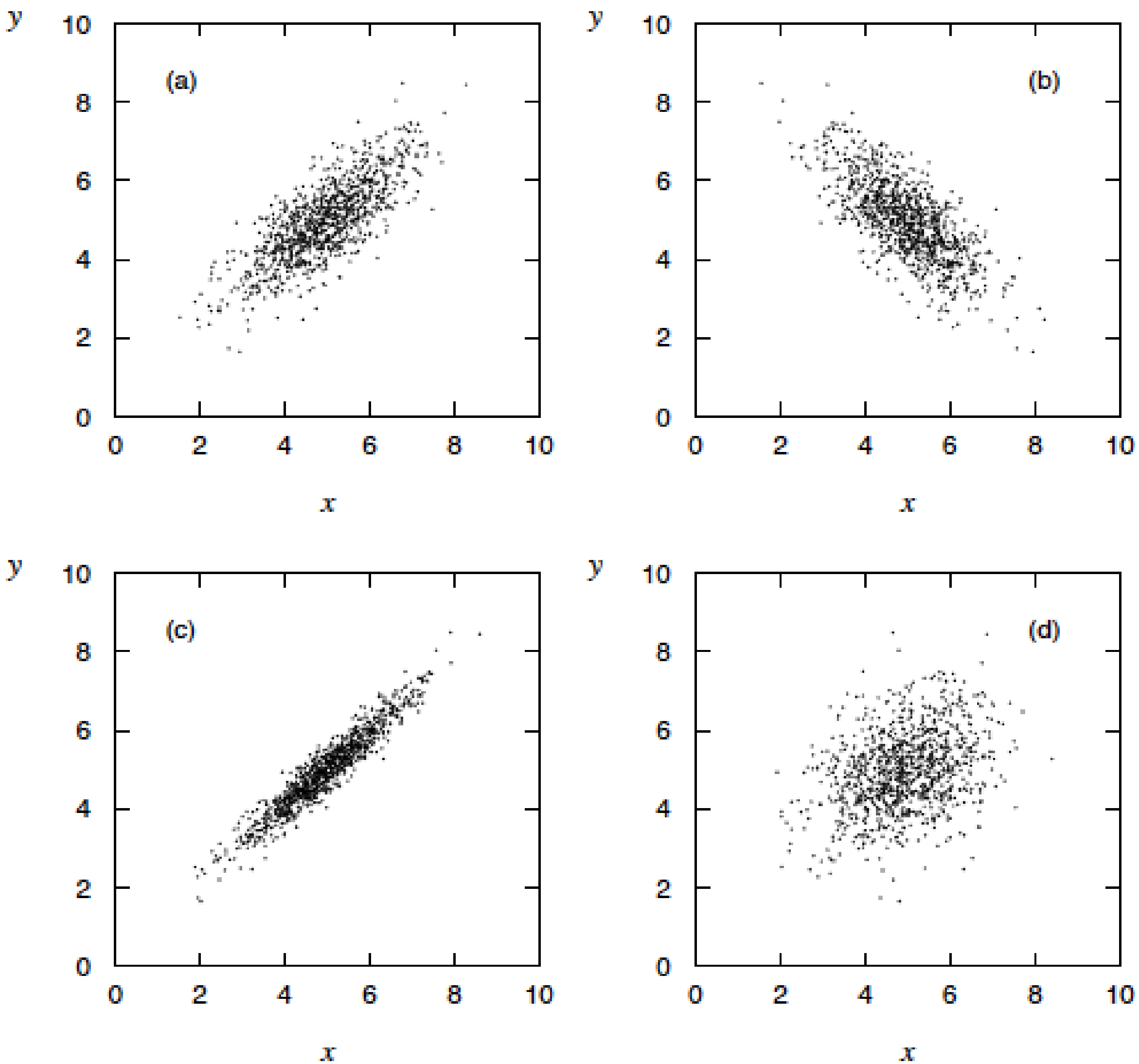
Correlation coefficient (dimensionless) defined as:

$$\rho_{xy} = \frac{\mathrm{cov}[x,y]}{\sigma_x\sigma_y}, \qquad -1\leq\rho_{xy}\leq+1$$

Correlation coefficient

$\rho = 0.75$

$\rho = -0.75$



$\rho = 0.95$

$\rho = 0.25$

Independent variables

If x and y are independent, i.e. $f(x,y) = f_x(x)\, f_y(y)$, then:

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

Therefore: $\mathrm{cov}[x,y] = 0$

x and y are 'uncorrelated'

Note!! The converse is NOT always true!!!

Exercise

Let [x y] be an absolutely continuous random vector with domain:

$$R_{XY} = \{(x, y) : 0 \leq x \leq y \leq 2\}$$

i.e. R_{XY} is the set of all couples (x,y) such that $0 \leq y \leq 2$ and $0 \leq x \leq y$.

Let the joint probability density function of [x y] be:

$$f(x, y) = \begin{cases} \frac{3}{8} y, & \text{if } (x, y) \in R_{XY} \\ 0 & \text{, otherwise} \end{cases}$$

Compute the covariance between X and Y.

Error propagation

Suppose we measure a set of values $\vec{x} = (x_1, ..., x_n)$

which follow some joint pdf $f(\vec{x})$.

$f(\vec{x})$ might be not fully known. But we have the covariances:

V_{ij} = cov[x_i,x_j], and the means $\vec{\mu} = E[\vec{x}]$ (in practice only estimates)

Now consider a function $y(\vec{x})$.

What is the variance of $y(\vec{x})$?

Hard way: use joint pdf $f(\vec{x})$ to find the pdf g(y) ,

Then from g(y) find

$$V[y] = E[y^2] - (E[y])^2$$

Often NOT practical. $f(\vec{x})$ may not even be fully known ...

Error propagation - 2

Expand $y(\vec{x})$ to the first order in a Taylor series about $\vec{\mu}$

$$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)$$

To find the variance V[y] we need $E[y^2]$ and E[y]:

$$E[y(\vec{x})] \approx y(\vec{\mu}) \quad \text{since} \quad E[x_i - \mu_i] = 0$$

Error propagation - 3

$$E[y^2(\vec{x})] \approx y^2(\vec{\mu}) + 2y(\vec{\mu}) \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} E[x_i - \mu_i]$$

$$+ E\left[\left(\sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)\right)\left(\sum_{j=1}^{n} \left[\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} (x_j - \mu_j)\right)\right]$$

$$= y^2(\vec{\mu}) + \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Putting the ingredients together gives the variance of $y(\vec{x})$

$$\sigma_y^2 \approx \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Error propagation - 4

If the x_i are uncorrelated, i.e. $V_{ij} = \sigma^2_i \delta_{ij}$, then this becomes:

$$\sigma_y^2 \approx \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]^2_{\vec{x}=\vec{\mu}} \sigma_i^2$$

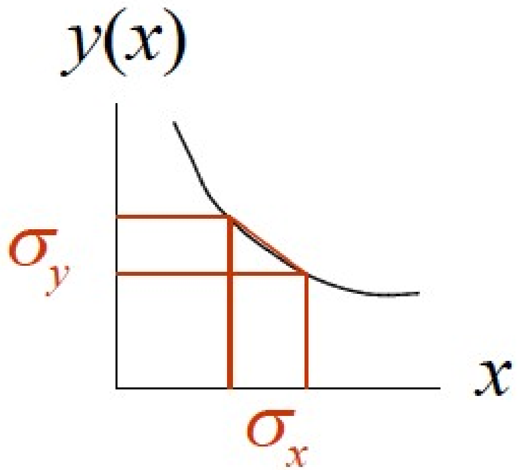
Similar for a set of m functions $\vec{y}(\vec{x}) = (y_1(\vec{x}), \ldots, y_m(\vec{x}))$

$$U_{kl} = \mathrm{cov}[y_k, y_l] \approx \sum_{i,j=1}^{n} \left[\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Or in matrix notation $U = A V A^T$, where $A_{ij} = \left[\frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}}$

Error propagation - 5

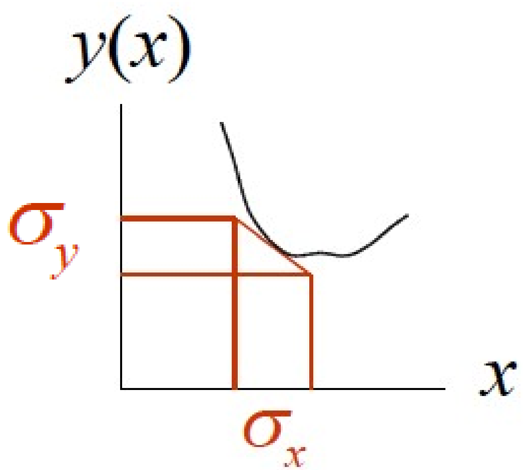
These are the error propagation formulae: the covariances which summarize the "errors" in measurements of $\vec{x}$, are propagated to the new quantities $\vec{y}(\vec{x})$



LIMITATION:

Exact only if $\vec{y}(\vec{x})$ linear.

Approximation breaks down if function is nonlinear over a region comparable in size to the σ_i



N.B. We said nothing about the pdf of the x_i , e.g. it does not have to be Gaussian

Error propagation: SPECIAL CASES

$$y = x_1 + x_2 \quad \rightarrow \quad \sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2\mathrm{cov}[x_1, x_2]$$

$$y = x_1 x_2 \quad \rightarrow \quad \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + 2\frac{\mathrm{cov}[x_1, x_2]}{x_1 x_2}$$

That is, **if the x_i are uncorrelated:**

Add errors quadratically for the sum (or difference),

Add relative errors quadratically for product (or ratio)

correlations can change this completely...

Error propagation – MORE SPECIAL

Consider $y = x_1 - x_2$ with:

$$\mu_1 = \mu_2 = 10, \quad \sigma_1 = \sigma_2 = 1, \quad \rho = \frac{\mathrm{cov}[x_1, x_2]}{\sigma_1 \sigma_2} = 0$$

$$V[y] = 1^2 + 1^2 = 2 \quad \rightarrow \quad \sigma_y = 1.4$$

Now suppose $\rho=1$ (full correlation). Then:

$$V[y] = 1^2 + 1^2 - 2 = 0 \quad \rightarrow \quad \sigma_y = 0$$

i.e. for 100% correlation, the error in the difference goes to 0 !!

Wrapping up lecture 3

- Probability density functions. Described by:
  - Expectation values (mean, variance)
  - Covariance
  - Correlation

- Given a function of a random variable, we know how to find the variance of the function using error propagation

NEXT TIME:

- Examples of probability functions:

  binomial, multinomial, Poisson, uniform, exponential, Gaussian
  - Central limit theorem

  Chi-square, Cauchy, Landau