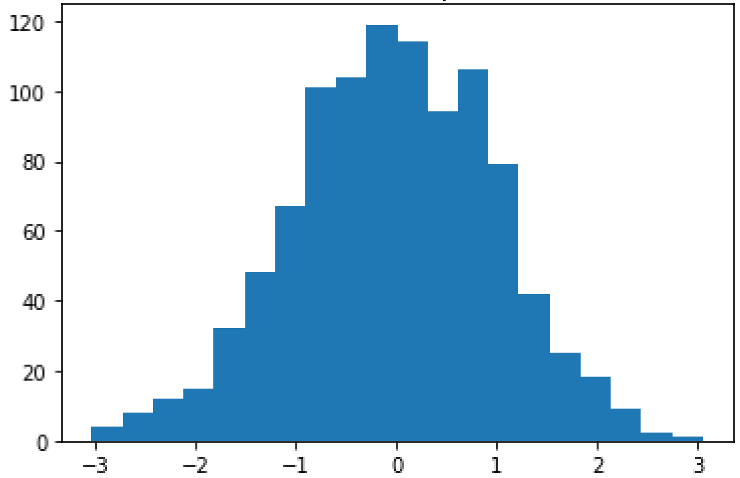
Math Methods and Experimental Physics

Monte Carlo Methods or why I showed the unif. dist.
Central limit theorem or why I showed the normal dist.
Multi-dimensional pdfs
Correlation
Error propagation

Monte Carlo sampling

Problem: We have a pdf: f(x) defined over $(-\infty,\ \infty)$. We want to produce a list of random numbers that are sampled from f(x). In numpy, you can find a long list of common pdfs, to sample from. But let's say f(x) is not on this list.

10,000 normally distributed values, μ=0, σ^2 = 1



A histogram of 1000 random numbers that follow a normal distribution (μ=0,σ=1) using numpy.random.normal()

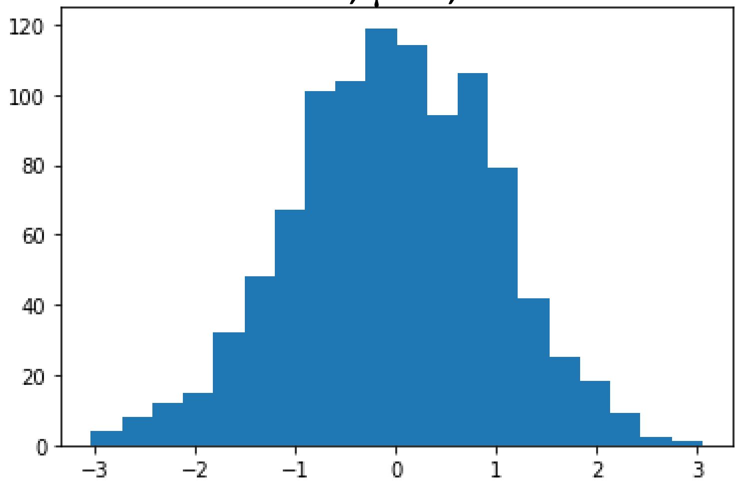
The most basic thing you'll find is a pseudo-random number generator of the uniform distribution.

Monte Carlo sampling: Inverse transform

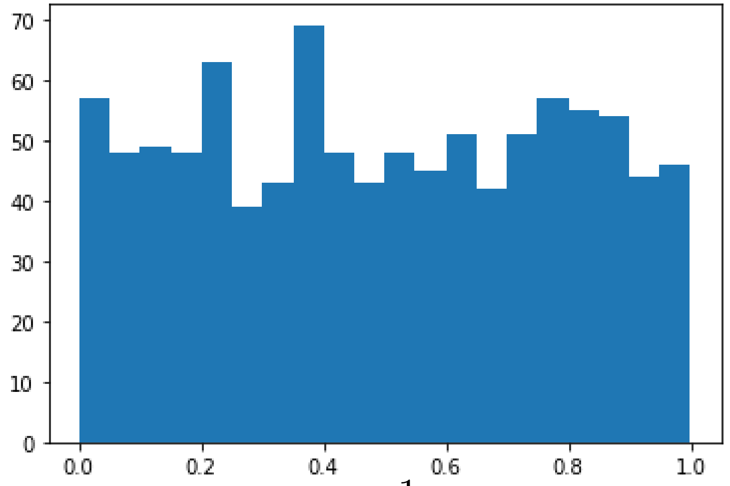
Let F(x) be the cumulative distribution function (cdf) of f(x).

If a random variable z is distributed according to f(x), then F(z) (= F(f(x))) will be uniformly distributed in [0,1].

10,000 normally distributed values, μ=0, σ^2 = 1



z



$$\frac{1}{2}\left(1 + erf(z/\sqrt{2})\right)$$

Monte Carlo sampling: Inverse transform

Let's invert this process!

a) Generate a uniformly distributed list y in [0,1].
b) Find $x = F^{-1}(y)$

Then x, will be distributed following f(x)

Central limit theorem

In the limit of n large, the sum of n independent continuous random variables x_i with means μ_i and variances σ^2_i becomes a Gaussian random variable with mean

$$\mu = \sum_i \mu_i$$

and with variance

$$\sigma^2 = \sum_i \sigma_i^2$$

The importance of this theorem is that the form of pdf(x_i) is irrelevant.

Central limit theorem – simple example

You have N measurements of a voltage, V_i. The voltage variances, are δV^2_i

The average voltage <V> is normally distributed with mean:

$$\bar{V} = \frac{1}{N} \sum_i V_i$$

And variance (unbiased):

$$\delta \bar{V}^2 = \frac{1}{N-1} \sum_i \delta V_i^2$$

And you don't care whether V_i are normally distributed or not.

(*) I distinguish <V> a distribution and $\bar{V}$, a point estimate.

Multi-dimensional pdfs

A single observation can result in multiple values.

Example: A neutrino detected in IceCube results in measurement of direction in the sky (δ,RA), energy, time, etc.

Multi-dimensional distributions

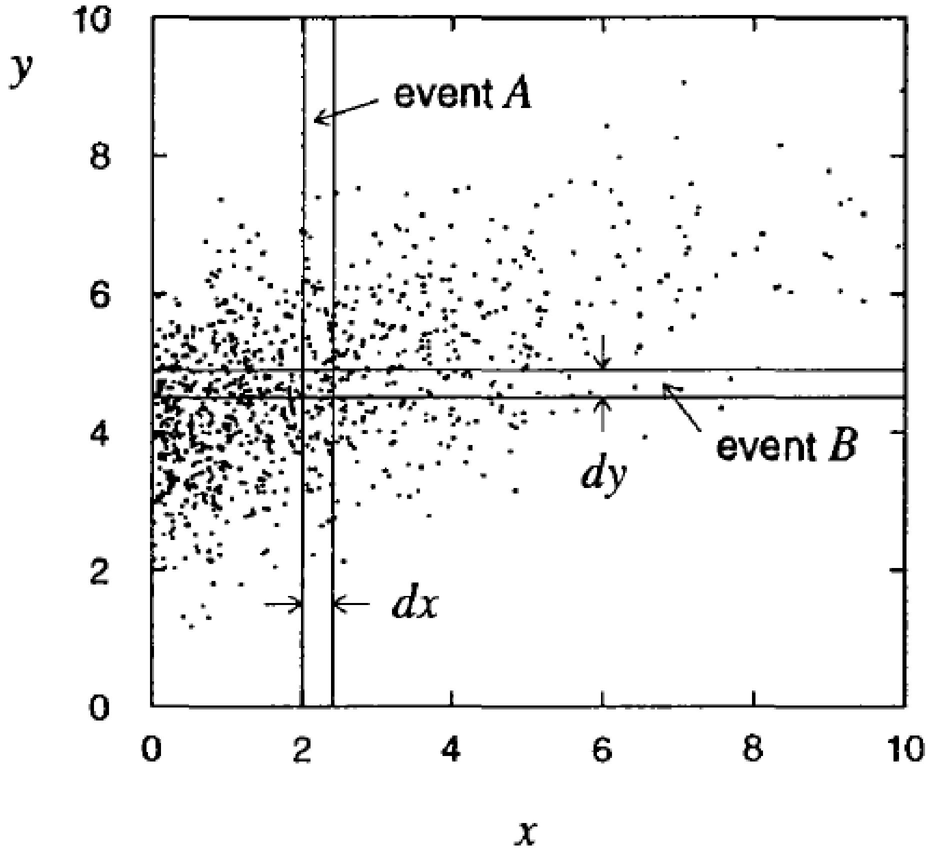
We can have a joint pdf

f(x,y) such that:

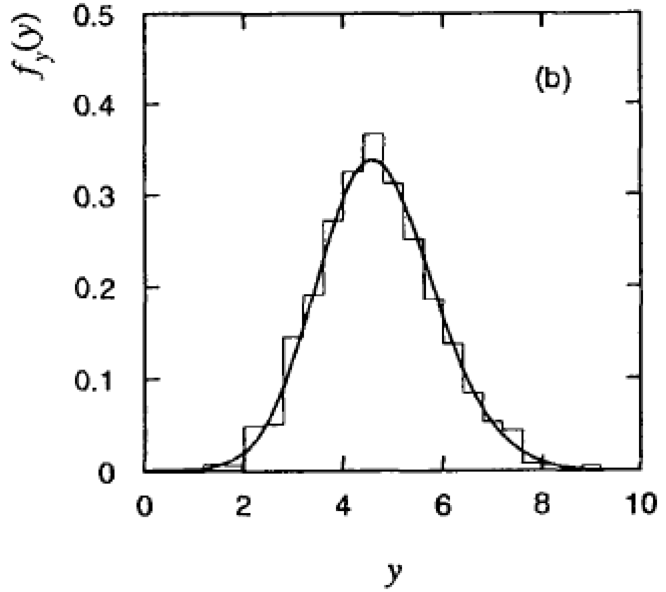
$$\int_S f(x,y)dxdy = 1$$

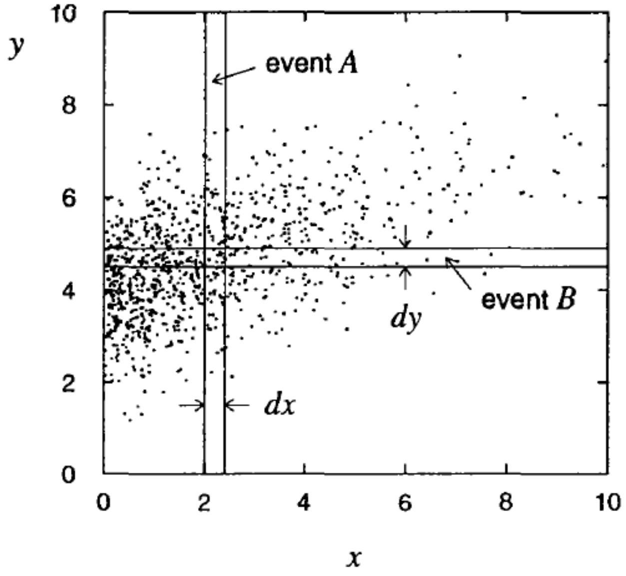
f(x,y)dxdy is the probability in the squared defined by (x,y) and (x+dx,y+dy)

This is the book example
Histograms are useful too.



Multi-dimensional distributions

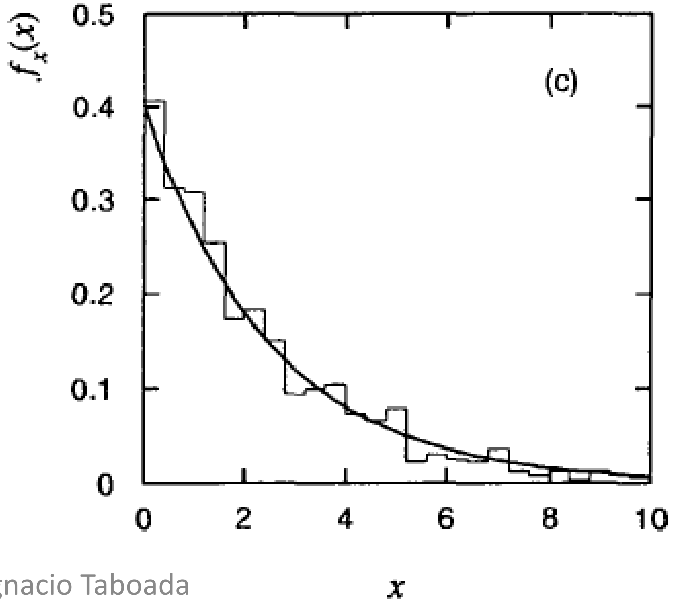




Marginal pdfs

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

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

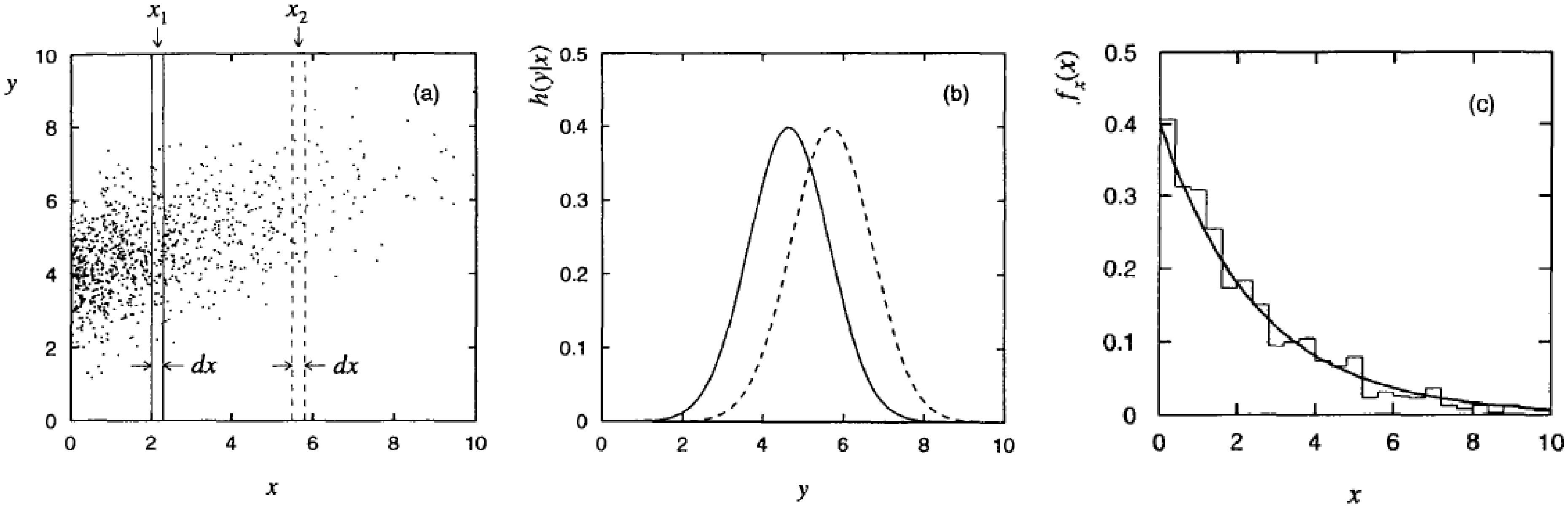


Independence of multi-dimensional data

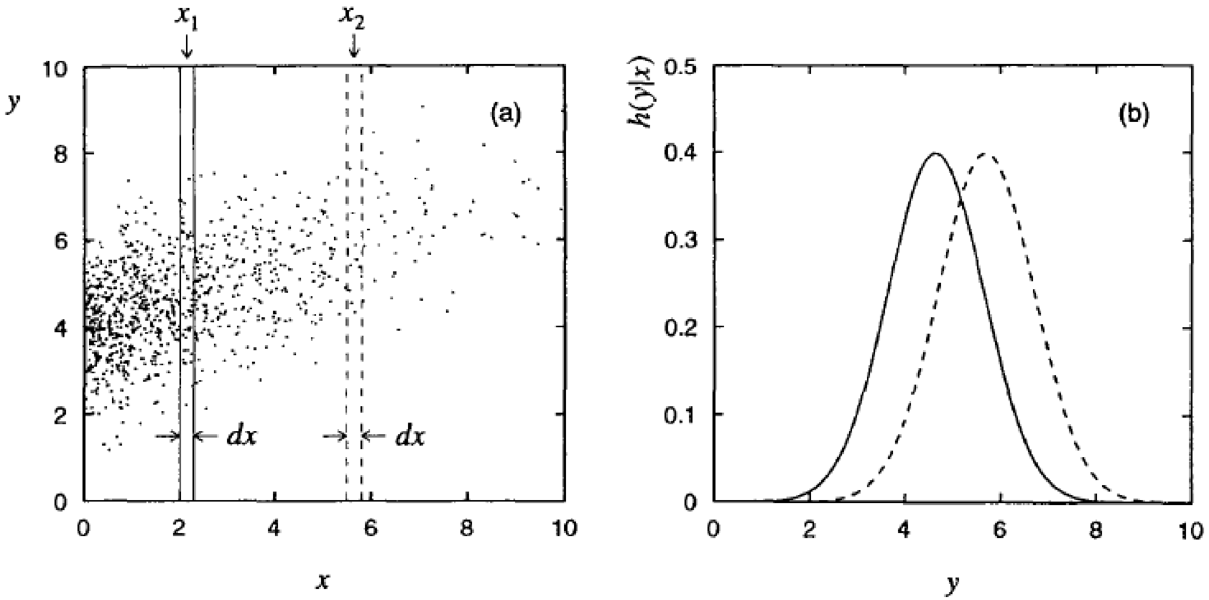
We can define conditional pdf:

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

Note that h(y|x) is a function of y given the *parameter* x.



Independent multi-dimensional data



If x,y are independent, then h(y|x) and g(x|y) are independent of the parameter x and y respectively. In the plot above, x and y are not independent.

Covariance

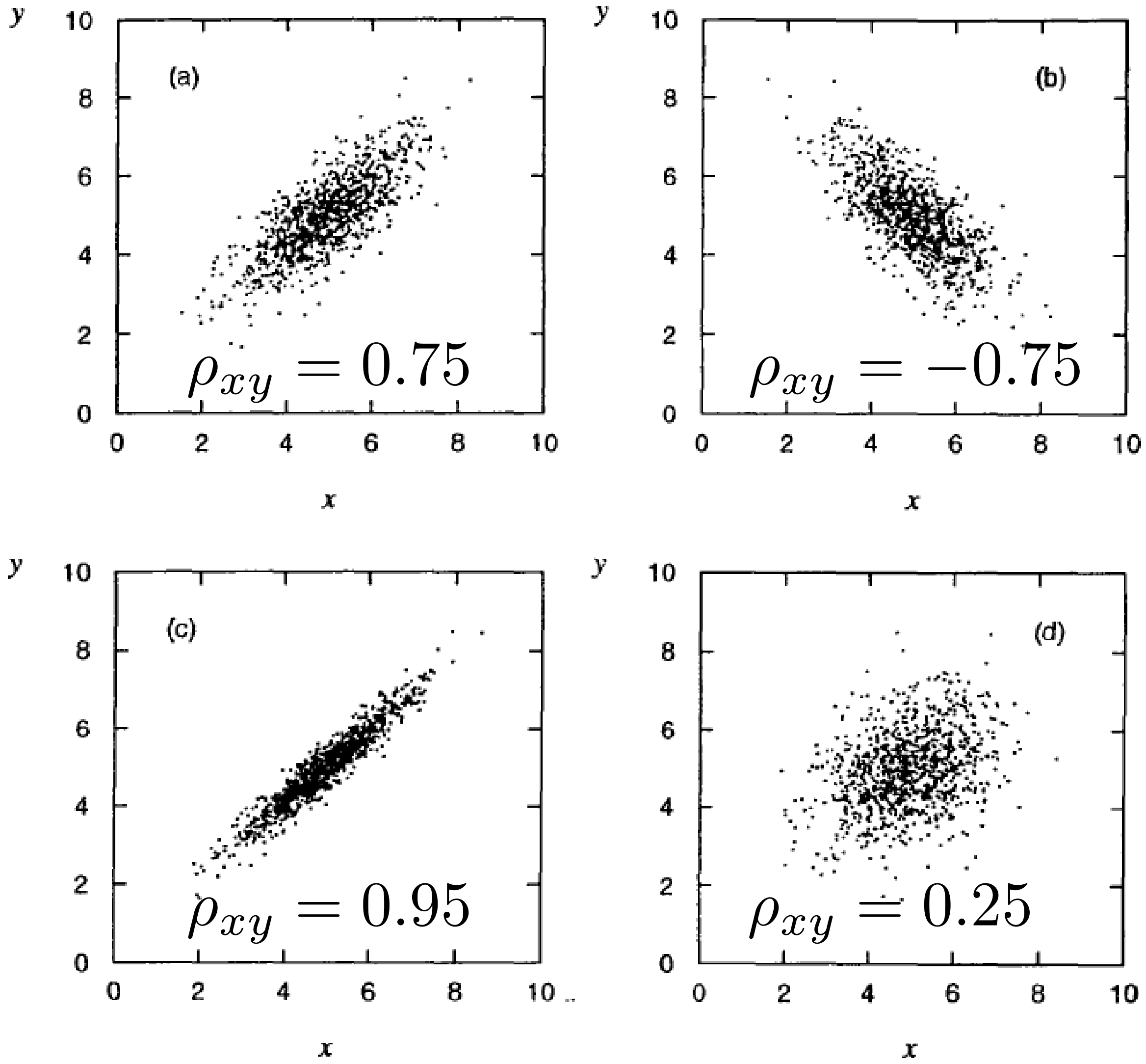
$$V_{xy} = E((x-\bar{x})(y-\bar{y})) = E(xy) - \bar{x}\bar{y} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy - \bar{x}\bar{y}$$

We can define also V_{xx}, V_{yy}, V_{yx} to form a 2x2 symmetric matrix. This can easily be extended to more dimensions. Since this matrix is symmetric it can always be diagonalized. The eigenvectors define new <u>decorrelated variables</u> x', y' (Whether the decorrelated variables have physical meaning, it's another issue).

Defining: $\rho_{xy} = \dfrac{V_{xy}}{\sigma_x \sigma_y}$

The (unitless) correlation coefficient matrix is: $Corr(x,y) = \begin{pmatrix} 1, \rho_{xy} \\ \rho_{xy}, 1 \end{pmatrix}$

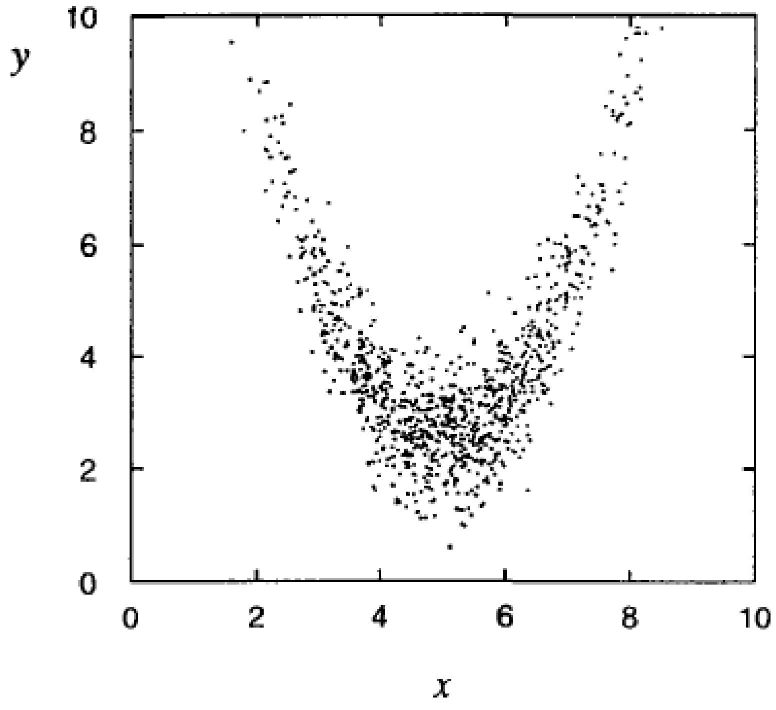
Covariance



Covariance

Uncorrelated data (x and y) have $\rho_{xy} = 0$

But, it's possible to have $\rho_{xy} = 0$ even when there is a correlation.
(You can do more things to data, than a linear transformation)



Error propagation

Let's have a measurement of $\vec{x}$. The joint pdf $f(\vec{x})$ is unknown, but the expectation values $\vec{\mu}$ and co-variance matrix V_{ij} are known
(Given the data you can always get point estimates).

Let's have $y(\vec{x})$ be a function of our measurements. What is σ_y?

$$y(\vec{x}) \sim y(\vec{\mu}) + \sum_i \left.\frac{\partial y}{\partial x_i}\right|_{\vec{\mu}} (x_i - \mu_i)$$

Clearly: $E[y(\vec{x})] \sim y(\vec{\mu})$

So:

$$E[y^2(\vec{x})] = y^2(\vec{\mu}) + 2y(\vec{\mu}) \sum_i \frac{\partial y}{\partial x_i}|_\mu \cancel{E[x_i - \mu_i]}$$

$$+E\left[\left((\sum_i \left.\frac{\partial y}{\partial x_i}\right|_{\vec{\mu}} (x_i - \mu_i))(\sum_j \left.\frac{\partial y}{\partial x_j}\right|_{\vec{\mu}} (x_j - \mu_j))\right)\right]$$

Error propagation

So:

$$E[y^2] \sim y^2(\vec{\mu}) + \sum_{i,j} \frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\bigg|_{\vec{\mu}} V_{ij}$$

And we can the estimate of the variance of y:

$$\boxed{\sigma_y^2 \sim E[y^2(\vec{x}] - (E[y(\vec{x}])^2 = \sum_{i,j} \frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\bigg|_{\vec{\mu}} V_{ij}}$$

For the simple case: $y = x_1 + x_2$

$$\sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2\frac{\partial y}{\partial x_1}\frac{\partial y}{\partial x_2}V_{12}$$

And for uncorrelated x_1, x_2 ("Errors add in quadrature")

$$\sigma_y^2 = \sigma_1^2 + \sigma_2^2$$