Numerical Methods
Lesson 5

Dr. Jose Feliciano Benitez

Universidad de Sonora

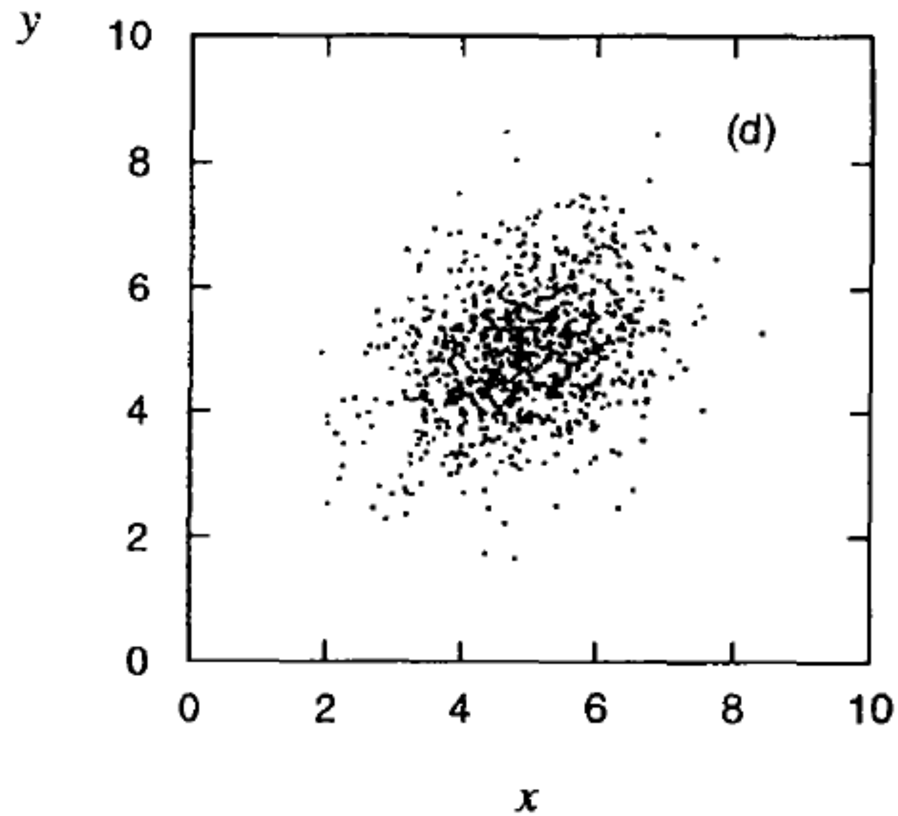
Dr. Benitez Homepage: www.jfbenitez.science
Course page: http://jfbenitez.ddns.net:8080/Courses/MetodosNumericos

Goals for this Lesson

- Joint p.d.f.'s
- Random variable correlations
- Covariance matrix
- Correlation coefficients
- Error propagation

Joint p.d.f.

- Consider two random variables x and y of the same experiment.
- The joint p.d.f. is the function describing the probability distribution in the x-y plane: ***f(x,y)***
- The plot at the right is called a scatter plot of *f(x,y)* in which the **density** of the points is proportional to the value of *f*.

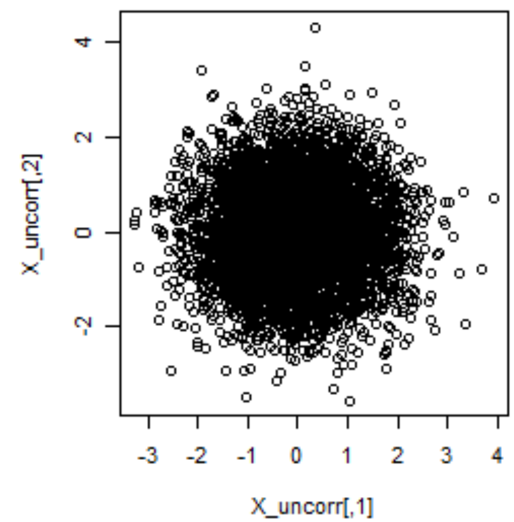


Joint p.d.f. of uncorrelated variables

- Suppose x and y are independent variables. Meaning the knowledge of x does not affect the distribution in y. This means x and y are uncorrelated.
- Assume the distribution in x is described by $f_x(x)$ and the distribution in y by $f_y(y)$, then the joint p.d.f. of x and y is given by :

$$f(x, y) = f_x(x)\, f_y(y)$$

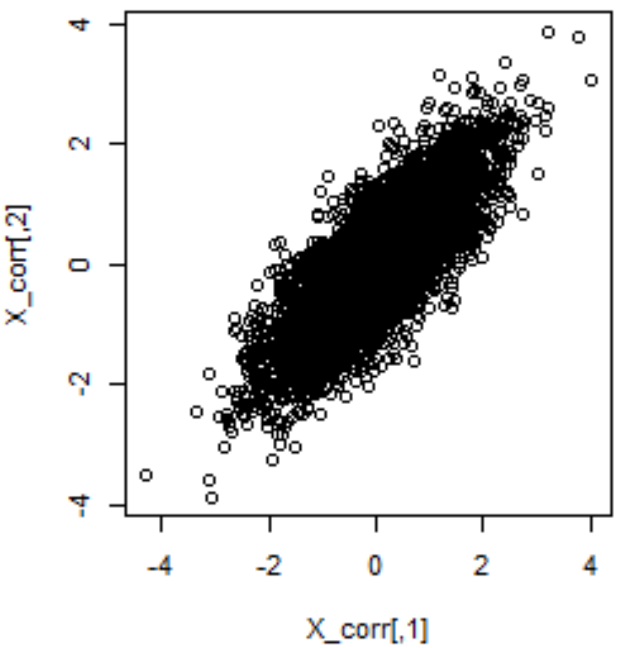
meaning the two parts are *factorized*.



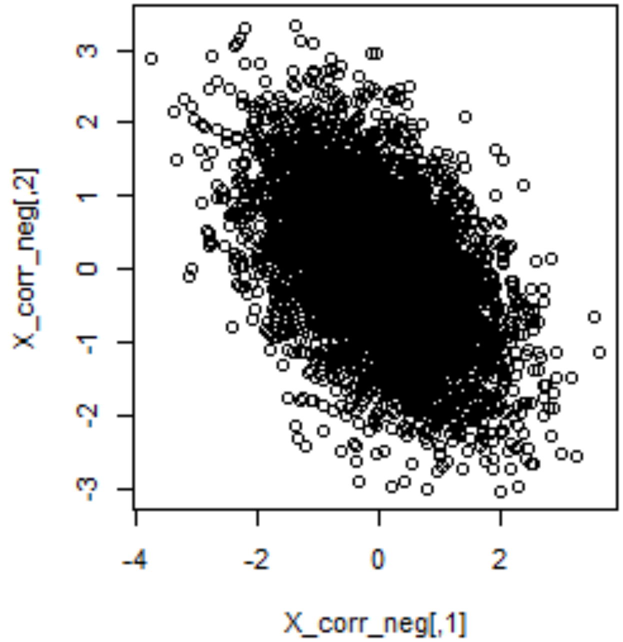
Correlated variables

- In general, there may be some degree of correlation between variables. The degree of correlation can have values between -1 and 1.

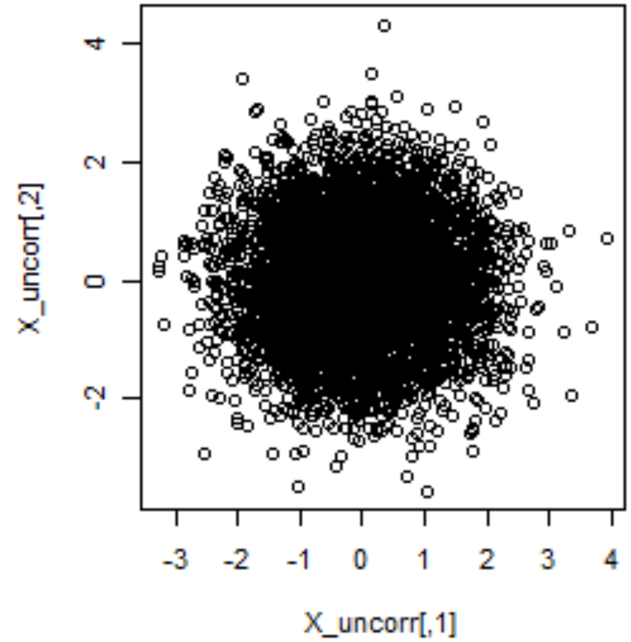
Positive



Negative



Not correlated



Covariance cov[x,y]

- The covariance of two random variables is denoted by Vxy or cov[x,y] is defined as:

$$
\begin{aligned}
V_{xy} &= E[(x-\mu_x)(y-\mu_y)] = E[xy] - \mu_x\mu_y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, y\, f(x,y)\, dx\, dy \; - \; \mu_x\mu_y ,
\end{aligned}
$$

Where *f(x,y)* is the joint p.d.f. , not necessarily factorizable if the variables are correlated.

Correlation coefficient ρ_{xy}

- The correlation coefficient ρ_{xy} is defined by:

$$\rho_{xy} = \frac{V_{xy}}{\sigma_x \sigma_y}$$

- where $\boldsymbol{\sigma}_x$ and $\boldsymbol{\sigma}_y$ are the standard deviations for x and y, defined by their corresponding marginal p.d.f.'s.
- The correlation coefficient has values between **-1 and 1.**

Examples

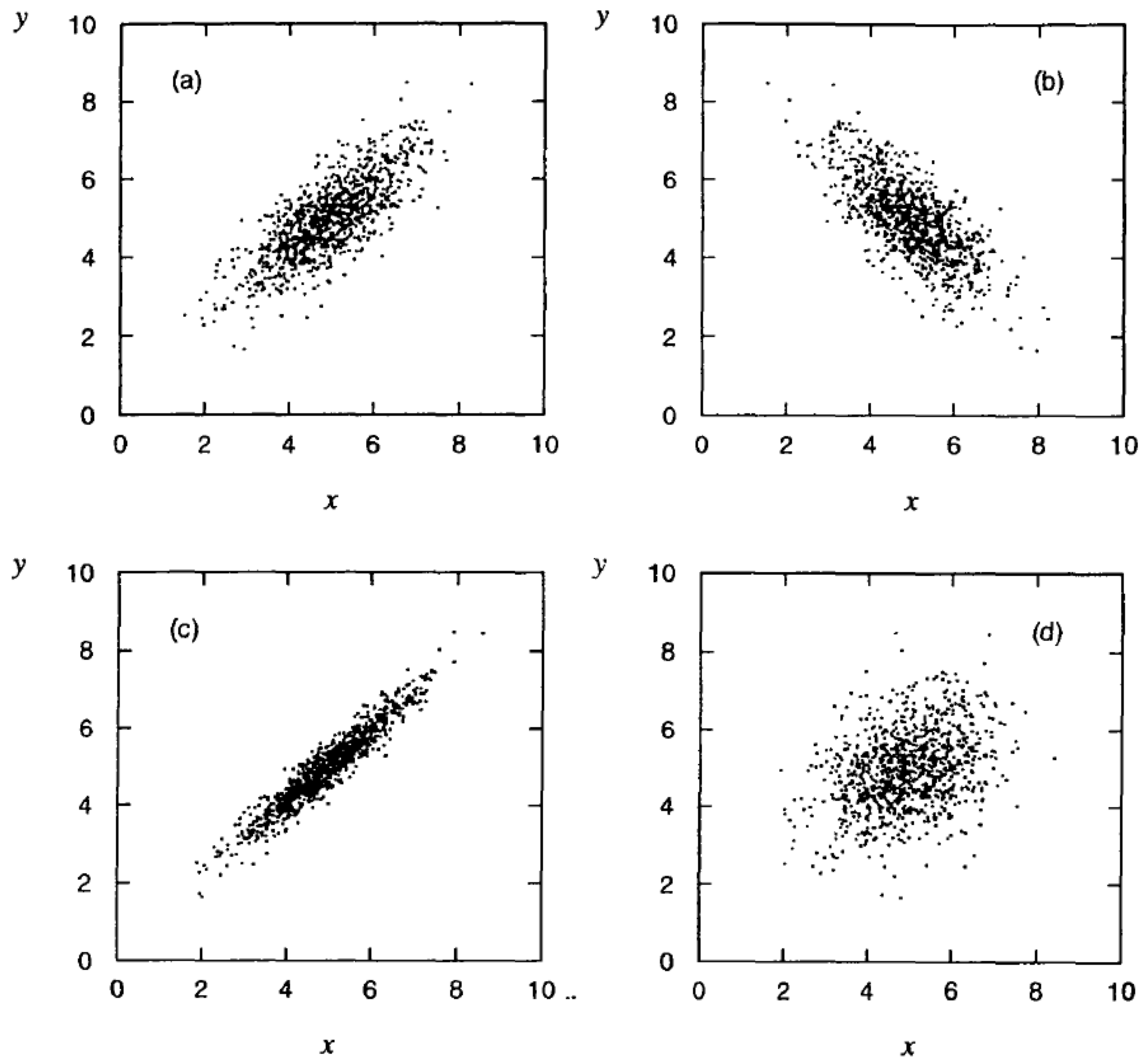


Fig. 1.9 Scatter plots of random variables x and y with (a) a positive correlation, $\rho = 0.75$, (b) a negative correlation, $\rho = -0.75$, (c) $\rho = 0.95$, and (d) $\rho = 0.25$. For all four cases the standard deviations of x and y are $\sigma_x = \sigma_y = 1$.

Coefficient for uncorrelated variables

- The correlation coefficient for uncorrelated variables is 0.
- Proof:

$V_{xy} = E[xy] - \mu_x \mu_y$

$E[xy] = \int\int xyf(x,y)dxdy$

$\quad = \int xf_x(x)dx \; * \int yf_y(y)dy \; = \; E[x] * E[y] = \mu_x \mu_y$

Because f(x,y) is factorizable.

Therefore: $V_{xy} = 0$

Error propagation

Covariance for a function of n random variables

- Suppose we have a function of n random variables
  $y(x_i)$ where i= 1,2,3,…,n
- The joint p.d.f. is $f(x_i)$, but may not be known.
- If we know the mean values μ_i and the covariance matrix V_{ij}, then $E[y]$ and the variance σ_y^2 can be estimated.

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

The expectation value of y is to first order

$$E[y(\mathbf{x})] \approx y(\boldsymbol{\mu}),$$

Important result: Variance of y

$$
\begin{aligned}
E[y^2(\mathbf{x})] &\approx y^2(\boldsymbol{\mu}) + 2y(\boldsymbol{\mu}) \cdot \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\mathbf{x}=\boldsymbol{\mu}} E[x_i - \mu_i] \\
&+ E\left[\left(\sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\mathbf{x}=\boldsymbol{\mu}} (x_i - \mu_i)\right)\left(\sum_{j=1}^{n} \left[\frac{\partial y}{\partial x_j}\right]_{\mathbf{x}=\boldsymbol{\mu}} (x_j - \mu_j)\right)\right] \\
&= y^2(\boldsymbol{\mu}) + \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\mathbf{x}=\boldsymbol{\mu}} V_{ij}, \qquad (1.52)
\end{aligned}
$$

so that the variance $\sigma_y^2 = E[y^2] - (E[y])^2$ is given by

$$
\sigma_y^2 \approx \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\mathbf{x}=\boldsymbol{\mu}} V_{ij}. \qquad (1.53)
$$

For uncorrelated variables

For the case where the x_i are not correlated, i.e. $V_{ii} = \sigma_i^2$ and $V_{ij} = 0$ for $i \neq j$, equations (1.53) and (1.54) become

$$\sigma_y^2 \approx \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]^2_{\mathbf{x}=\boldsymbol{\mu}} \sigma_i^2 \tag{1.57}$$

Examples with two variables and possible correlation ($V_{12} \neq 0$)

- Example: $y = x_1 + x_2$

$$\sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2V_{12}.$$

- Example: $y = x_1 * x_2$

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

In this last formula y, x_1, x_2 are the mean values.

Exercise

- Analyze the data:
  /home/DATA/MetodosNumericos/Lesson5.dat

Contains pairs of values: x y

- Graph the data as scatter plot. You can use a TGraph.
- Calculate:
  - mean values: mu_x and mu_y
  - standard deviations: sigma_x and sigma_y
  - Covariance: V_{xy}
  - correlation coefficient: rho_xy
- For z=x + y calculate mean value variance
- For z=x*y calculate mean and variance

For all above calculations ***since you do not know the joint p.d.f.*** you can approximate it as follows:

f(x,y) = 1/N for x,y = the values in the data file,

Here N is the total number of data pairs in the file.

The integrals in the formulas are evaluated by a discrete loop over the data pairs in the file.