

## Correlation in Random Variables

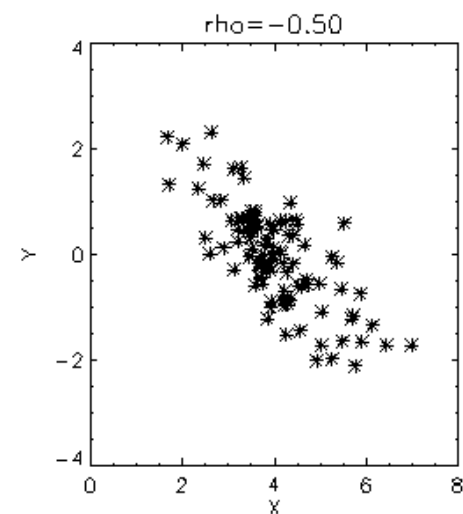
Lecture 11

Spring 2002

## Correlation in Random Variables

Suppose that an experiment produces two random variables,  $X$  and  $Y$ . What can we say about the relationship between them?

One of the best ways to visualize the possible relationship is to plot the  $(X, Y)$  pair that is produced by several trials of the experiment. An example of correlated samples is shown at the right



Lecture 11

1

## Joint Density Function

The joint behavior of  $X$  and  $Y$  is fully captured in the joint probability distribution. For a continuous distribution

$$E[X^m Y^n] = \iint_{-\infty}^{\infty} x^m y^n f_{XY}(x, y) dx dy$$

For discrete distributions

$$E[X^m Y^n] = \sum_{x \in \mathcal{S}_x} \sum_{y \in \mathcal{S}_y} x^m y^n P(x, y)$$

## Covariance Function

The covariance function is a number that measures the common variation of  $X$  and  $Y$ . It is defined as

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

The covariance is determined by the difference in  $E[XY]$  and  $E[X]E[Y]$ .

If  $X$  and  $Y$  were statistically independent then  $E[XY]$  would equal  $E[X]E[Y]$  and the covariance would be zero.

The covariance of a random variable with itself is equal to its variance.  $\text{cov}[X, X] = E[(X - E[X])^2] = \text{var}[X]$

Lecture 11

2

Lecture 11

3

## Correlation Coefficient

The covariance can be normalized to produce what is known as the correlation coefficient,  $\rho$ .

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

The correlation coefficient is bounded by  $-1 \leq \rho \leq 1$ . It will have value  $\rho = 0$  when the covariance is zero and value  $\rho = \pm 1$  when  $X$  and  $Y$  are perfectly correlated or anti-correlated.

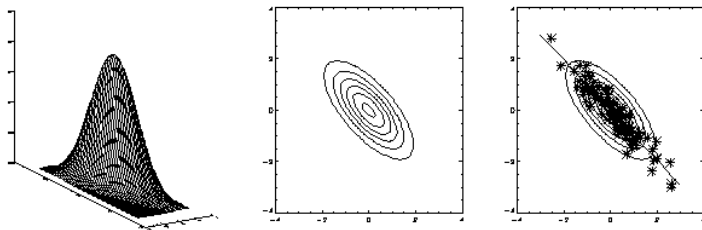
## Autocorrelation Function

The autocorrelation function is very similar to the covariance function. It is defined as

$$R(X, Y) = E[XY] = \text{cov}(X, Y) + E[X]E[Y]$$

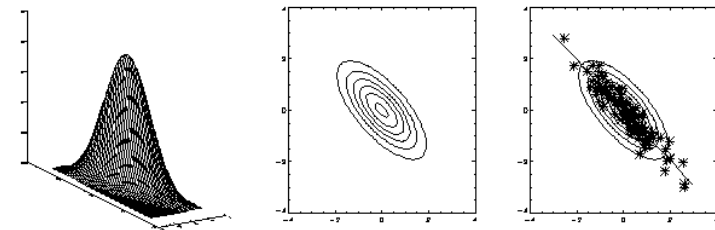
It retains the mean values in the calculation of the value. The random variables are *orthogonal* if  $R(X, Y) = 0$ .

## Normal Distribution



$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2}{2(1-\rho^2)}\right)$$

## Normal Distribution



The orientation of the elliptical contours is along the line  $y = x$  if  $\rho > 0$  and along the line  $y = -x$  if  $\rho < 0$ . The contours are a circle, and the variables are uncorrelated, if  $\rho = 0$ . The center of the ellipse is  $(\mu_x, \mu_y)$ .

## Linear Estimation

The task is to construct a rule for the prediction  $\hat{Y}$  of  $Y$  based on an observation of  $X$ .

If the random variables are correlated then this should yield a better result, on the average, than just guessing. We are encouraged to select a linear rule when we note that the sample points tend to fall about a sloping line.

$$\hat{Y} = aX + b$$

where  $a$  and  $b$  are parameters to be chosen to provide the best results. We would expect  $a$  to correspond to the slope and  $b$  to the intercept.

## Minimize Prediction Error

To find a means of calculating the coefficients from a set of sample points, construct the predictor error

$$\varepsilon = E[(Y - \hat{Y})^2]$$

We want to choose  $a$  and  $b$  to minimize  $\varepsilon$ . Therefore, compute the appropriate derivatives and set them to zero.

$$\begin{aligned}\frac{\partial \varepsilon}{\partial a} &= -2E[(Y - \hat{Y}) \frac{\partial \hat{Y}}{\partial a}] = 0 \\ \frac{\partial \varepsilon}{\partial b} &= -2E[(Y - \hat{Y}) \frac{\partial \hat{Y}}{\partial b}] = 0\end{aligned}$$

These can be solved for  $a$  and  $b$  in terms of the expected values. The expected values can be themselves estimated from the sample set.

## Prediction Error Equations

The above conditions on  $a$  and  $b$  are equivalent to

$$\begin{aligned}E[(Y - \hat{Y})X] &= 0 \\ E[Y - \hat{Y}] &= 0\end{aligned}$$

The prediction error  $Y - \hat{Y}$  must be orthogonal to  $X$  and the expected prediction error must be zero.

Substituting  $\hat{Y} = aX + b$  leads to a pair of equations to be solved for  $a$  and  $b$ .

$$\begin{aligned}E[(Y - aX - b)X] &= E[XY] - aE[X^2] - bE[X] = 0 \\ E[Y - aX - b] &= E[Y] - aE[X] - b = 0\end{aligned}$$

$$\begin{bmatrix} E[X^2] & E[X] \\ E[X] & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} E[XY] \\ E[Y] \end{bmatrix}$$

## Prediction Error

$$\begin{aligned}a &= \frac{\text{cov}(X, Y)}{\text{var}(X)} \\ b &= E[Y] - \frac{\text{cov}(X, Y)}{\text{var}(X)} E[X]\end{aligned}$$

The prediction error with these parameter values is

$$\varepsilon = (1 - \rho^2)\text{var}(Y)$$

When the correlation coefficient  $\rho = \pm 1$  the error is zero, meaning that perfect prediction can be made.

When  $\rho = 0$  the variance in the prediction is as large as the variation in  $Y$ , and the predictor is of no help at all.

For intermediate values of  $\rho$ , whether positive or negative, the predictor reduces the error.

## Linear Predictor Program

---

The program `lp.pro` computes the coefficients  $[a, b]$  as well as the covariance matrix  $C$  and the correlation coefficient,  $\rho$ . The covariance matrix is

$$C = \begin{bmatrix} \text{var}[X] & \text{cov}[X, Y] \\ \text{cov}[X, Y] & \text{var}[Y] \end{bmatrix}$$

Usage example:

```
N=100
X=Randomn(seed,N)
Z=Randomn(seed,N)
Y=2*X-1+0.2*Z
p=lp(X,Y,c,rho)
print,'Predictor Coefficients=',p
print,'Covariance matrix'
print,c
print,'Correlation Coefficient=',rho
```

## Program lp

---

```
function lp,X,Y,c,rho,mux,muy
;Compute the linear predictor coefficients such that
;Yhat=aX+b is the minimum mse estimate of Y based on X.

;Shorten the X and Y vectors to the length of the shorter.
n=n_elements(X) < n_elements(Y)
X=(X[0:n-1])*
Y=(Y[0:n-1])*

;Compute the mean value of each.
mux=total(X)/n
muy=total(Y)/n
```

Continued on next page

## Program lp (continued)

---

```
;Compute the covariance matrix.
V=[[X-mux],[Y-muy]]
C=V##transpose(V)/(n-1)

;Compute the predictor coefficient and constant.
a=c[0,1]/c[0,0]
b=muy-a*mux

;Compute the correlation coefficient
rho=c[0,1]/sqrt(c[0,0]*c[1,1])

Return,[a,b]
END
```

## Example

---

```
Predictor Coefficients= 1.99598 -1.09500
Correlation Coefficient=0.812836
Covariance Matrix
      0.950762    1.89770
      1.89770    5.73295
```

## IDL Regress Function

IDL provides a number of routines for the analysis of data. The function REGRESS does multiple linear regression.

Compute the predictor coefficient  $a$  and constant  $b$  by

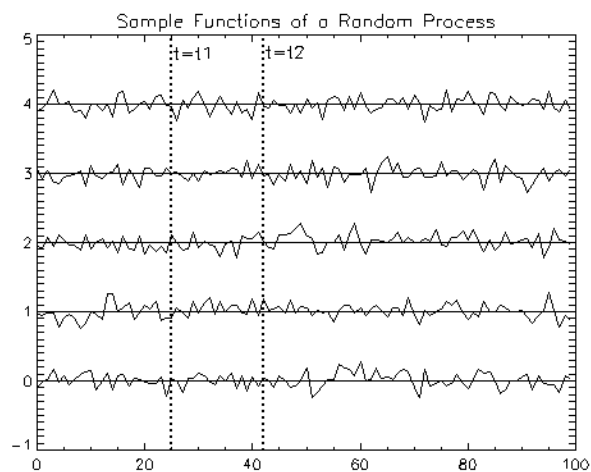
```
a=regress(X,Y,Const=b)
print,[a,b]
1.99598 -1.09500
```

## New Concept

# Introduction to Random Processes

Today we will just introduce the basic ideas.

## Random Processes



## Random Process

- A random variable is a function  $X(e)$  that maps the set of experiment outcomes to the set of numbers.
- A *random process* is a rule that maps every outcome  $e$  of an experiment to a *function*  $X(t, e)$ .
- A random process is usually conceived of as a function of time, but there is no reason to not consider random processes that are functions of other independent variables, such as spatial coordinates.
- The function  $X(u, v, e)$  would be a function whose value depended on the location  $(u, v)$  and the outcome  $e$ , and could be used in representing random variations in an image.

## Random Process

---

- The domain of  $e$  is the set of outcomes of the experiment. We assume that a probability distribution is known for this set.
- The domain of  $t$  is a set,  $\mathcal{T}$ , of real numbers.
- If  $\mathcal{T}$  is the real axis then  $X(t, e)$  is a *continuous-time* random process
- If  $\mathcal{T}$  is the set of integers then  $X(t, e)$  is a *discrete-time* random process
- We will often suppress the display of the variable  $e$  and write  $X(t)$  for a continuous-time RP and  $X[n]$  or  $X_n$  for a discrete-time RP.

## Random Process

---

- A RP is a family of functions,  $X(t, e)$ . Imagine a giant strip chart recording in which each pen is identified with a different  $e$ . This family of functions is traditionally called an *ensemble*.
- A single function  $X(t, e_k)$  is selected by the outcome  $e_k$ . This is just a time function that we could call  $X_k(t)$ . Different outcomes give us different time functions.
- If  $t$  is fixed, say  $t = t_1$ , then  $X(t_1, e)$  is a random variable. Its value depends on the outcome  $e$ .
- If both  $t$  and  $e$  are given then  $X(t, e)$  is just a number.

## Moments and Averages

---

$X(t_1, e)$  is a random variable that represents the set of samples across the ensemble at time  $t_1$

If it has a probability density function  $f_X(x; t_1)$  then the moments are

$$m_n(t_1) = E[X^n(t_1)] = \int_{-\infty}^{\infty} x^n f_X(x; t_1) dx$$

The notation  $f_X(x; t_1)$  may be necessary because the probability density may depend upon the time the samples are taken.

The mean value is  $\mu_X = m_1$ , which may be a function of time.

The central moments are

$$E[(X(t_1) - \mu_X(t_1))^n] = \int_{-\infty}^{\infty} (x - \mu_X(t_1))^n f_X(x; t_1) dx$$

## Pairs of Samples

---

The numbers  $X(t_1, e)$  and  $X(t_2, e)$  are samples from the same time function at different times.

They are a pair of random variables  $(X_1, X_2)$ .

They have a joint probability density function  $f(x_1, x_2; t_1, t_2)$ .

From the joint density function one can compute the marginal densities, conditional probabilities and other quantities that may be of interest.