Properties of Expectations

Correlation [Ross S7.4]

The **correlation** [coefficient] of two random variables X and Y is defined to be

$$\rho(X,Y) = \frac{Cov[X,Y]}{\sqrt{Var[X]\,Var[Y]}}$$

Proposition 32.1 $-1 \leq \rho(X,Y) \leq 1$

Why?

Let $Var[X] = \sigma_X^2$ and $Var[Y] = \sigma_Y^2$.

$$\begin{aligned}
0 &\leq Var\left[\frac{X}{\sigma_X} + \frac{Y}{-\sigma_Y}\right] && (32.1)\\
&= Var\left[\frac{X}{\sigma_X}\right] + Var\left[\frac{Y}{-\sigma_Y}\right] + 2Cov\left[\frac{X}{\sigma_X}, \frac{Y}{-\sigma_Y}\right]\\
&= \frac{Var[X]}{\sigma_X^2} + \frac{Var[Y]}{\sigma_Y^2} - 2\frac{Cov[X,Y]}{\sigma_X\sigma_Y}\\
&= 2 - 2\rho(X,Y)
\end{aligned}$$

$$\Rightarrow \rho(X,Y) \leq 1 \qquad (32.2)$$

$$\begin{aligned}
0 &\leq Var\left[\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right]\\
&= \frac{Var[X]}{\sigma_X^2} + \frac{Var[Y]}{\sigma_Y^2} + 2\frac{Cov[X,Y]}{\sigma_X\sigma_Y}\\
&= 2 - 2\rho(X,Y)
\end{aligned}$$

$$\Rightarrow -1 \leq \rho(X,Y)$$

Now, if $Var[Z] = 0$, then $P[Z = \underbrace{\text{some constant}}_{E[Z]}] = 1$.

If $\rho(X,Y) = 1$, then (32.1) + (32.2) imply

$$Var\left[\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right] = 0$$

hence

$$\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = \frac{\mu_X}{\sigma_X} - \frac{\mu_Y}{\sigma_Y}$$

and therefore

$$Y = \mu_Y + \frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

If $\rho(X,Y) = -1$, then

$$Y = \mu_Y - \frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

The correlation coefficient *measures the degree of linearity* between X and Y.

$\rho(X,Y)$ close to ± 1 indicates high degree of linearity betwen X and Y.

$\rho(X,Y) > 0$ indicates Y tends to increase when X does; we say X and Y are positively correlated.

$\rho(X,Y) < 0$ indicates Y tends to decrease when X does; we say X and Y are negatively correlated.

If $\rho(X,Y) = 0$ then X and Y are called **uncorrelated**.

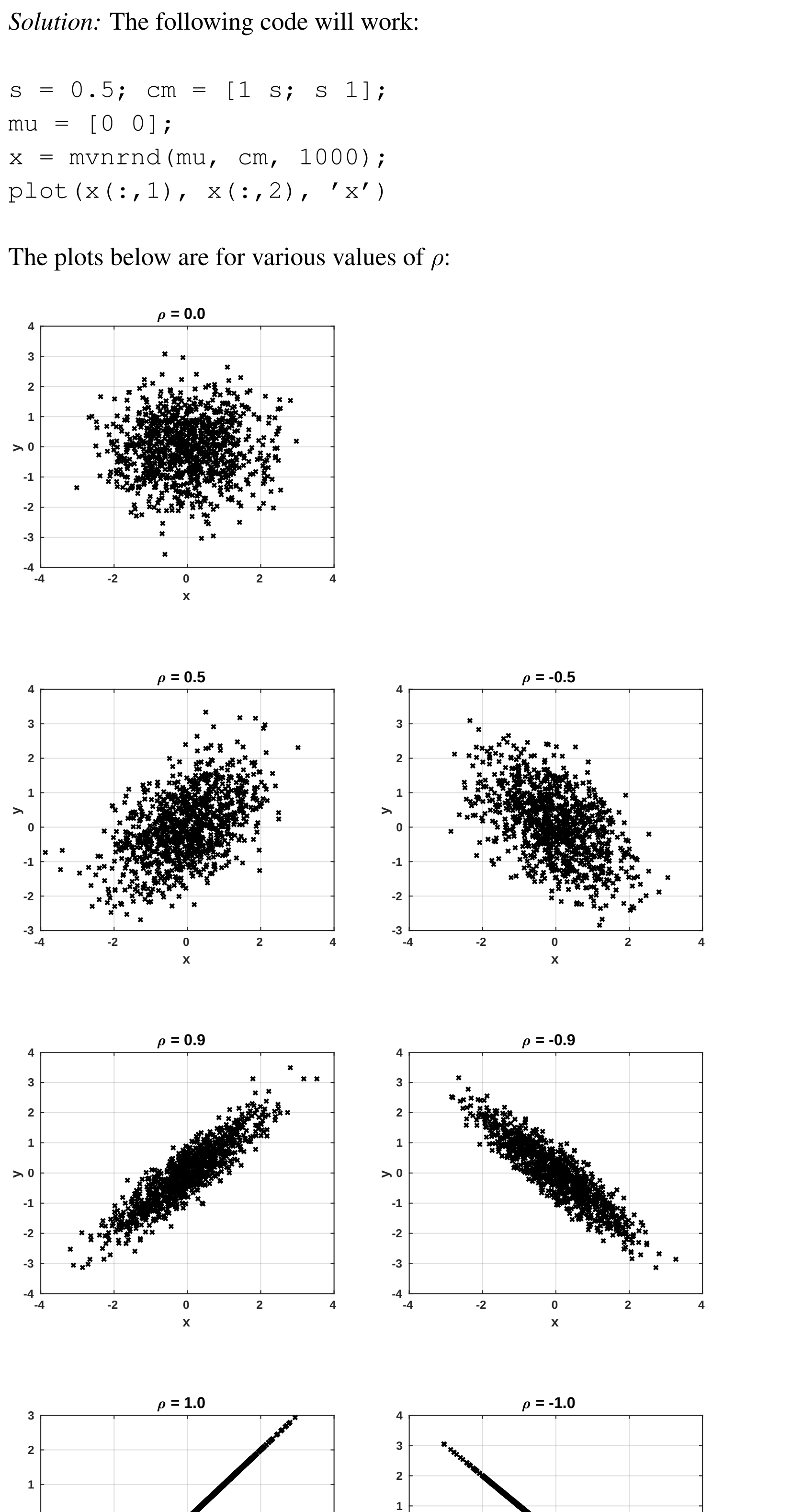
Example 32.1: [Matlab] For a bivariate Gaussian with parameters μ_X, μ_Y, σ_X, σ_Y and ρ, it turns out that ρ is the correlation coefficient of the two Gaussians (see Notes #34).

Use Matlab to generate 1000 realizations of a bivariate Gaussian pair (X,Y) with means 0, variances 1, and correlation coefficient 0.5. Plot the 1000 pairs. Repeat for correlation coefficient 0.9. What do you observe?

Solution: The following code will work:

```
s = 0.5; cm = [1 s; s 1];
mu = [0 0];
x = mvnrnd(mu, cm, 1000);
plot(x(:,1), x(:,2), 'x')
```

The plots below are for various values of ρ:



Example 32.2: Let $X_1, \ldots, X_n$ be iid with variance σ^2, and recall that $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is the sample mean. $X_i - \bar{X}$ is called the i**th deviation**.

Show that

$$Cov[X_i - \bar{X}, \bar{X}] = 0$$

for each $i = 1, \ldots, n$.

Solution:

$$\begin{aligned}
Cov[X_i - \bar{X}, \bar{X}] &= Cov[X_i, \bar{X}] - Cov[\bar{X}, \bar{X}]\\
&= Cov[X_i, \frac{1}{n}\sum_{j=1}^n X_j] - Var[\bar{X}]\\
&= \frac{1}{n}Cov[X_i, \sum_{j=1}^n X_j] - \frac{\sigma^2}{n}\\
&= \frac{1}{n}\sum_{j=1}^n Cov[X_i, X_j] - \frac{\sigma^2}{n}\\
&= \frac{1}{n}Cov[X_i, X_i] - \frac{\sigma^2}{n}\\
&= \frac{1}{n}\sigma^2 - \frac{\sigma^2}{n}\\
&= 0
\end{aligned}$$

where $Var[\bar{X}] = \sigma^2/n$ from Example 31.2.

Note: While we use the terms **correlation coefficient** and **correlation** to both denote $\rho(X,Y)$, some books/authors use the term **correlation coefficient** as we do, and the term **correlation** to mean $E[XY]$.