Random Variables with SCILAB

By

Gilberto E. Urroz, Ph.D., P.E.

Distributed by

infoClearinghouse.com

A "zip" file containing all of the programs in this document (and other SCILAB documents at InfoClearinghouse.com) can be downloaded at the following site:

http://www.engineering.usu.edu/cee/faculty/gurro/Software_Calculators/Scilab_Docs/ScilabBookFunctions.zip

The author's SCILAB web page can be accessed at:

http://www.engineering.usu.edu/cee/faculty/gurro/Scilab.html

Please report any errors in this document to: gurro@cc.usu.edu

Random Variables

A random variable is a function that assigns a numerical value to a particular event within a sample space. For example, let X be a random variable representing the number of rainy days in a particular week in a specific location. Obviously, X can take values of *0, 1, 2, 3, 4, 5, 6*, and *7*. The event associated with the value $X = 3$, for example, can be described as A_3 = *{it rains three days in the week under consideration}.*

The example of the number of rainy days presented above is an example of a *discrete random variable*, i.e., a variable that can only take a finite number of values. Another example of a discrete variable is the number of cars driving through a particular tollbooth in a given highway. This is a discrete variable because it can only take the value zero or a positive integer. Some discrete variables can take non-integer values. For example, to keep track of daily parking times in an airport parking lot, the times may be recorded every half hour. Therefore, if X is a random variable representing the recorded parking times, X can take the values of *0, 0.5, 1.0, … , 23.5, 24.0*.

If we are interested, for example, in analyzing the concentration of a particular contaminant in a water sample, the concentration measured (say, in *ppm* or *mg/l*) is a *continuous random variable*. The compressive strength of concrete samples selected randomly is also a continuous variable since any value of the strength can be measured in a particular test. A continuous random variable, therefore, can take any value in a continuous range. For probability applications, rather than concentrating on the probability of a specific value of a continuous random variable, we deal with the probabilities of ranges.

Notation

When dealing with random variables it is customary to use upper-case letters, e.g., *X, Y, Z*, etc., to refer to the *name* of the random variable, and to use lower-case letters, e.g., x, y, z, etc., to refer to the actual *value* taken by the variable. This notation allows the writing of statements such as $P(X = x)$, read as "*the probability of the random variable X taking the value x.*" We can also write statements such as $P(x_1<X<x_2)$, interpreted as "the probability that the random variable X takes values between x_1 and x_2, not including the extremes of the interval." An expression such as $P(X\geq x)$ represents "*the probability that X is at least equal to x*."

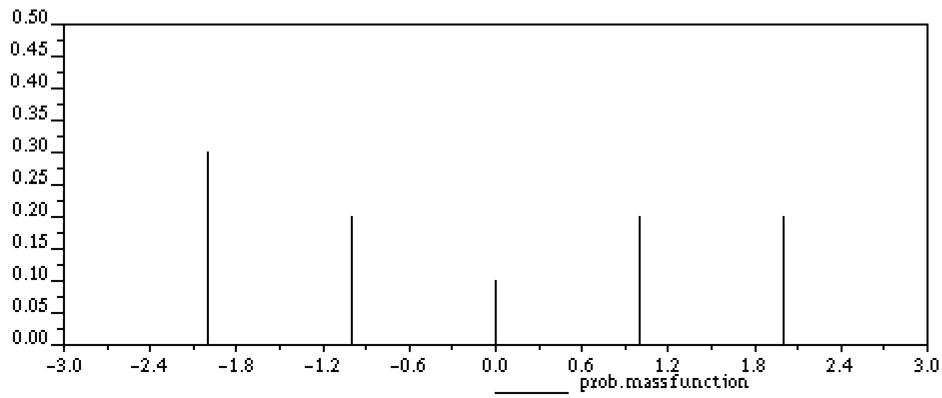
Probability Distribution for Discrete Random Variables

To describe the distribution of probabilities for a discrete random variable X we use a function called the *probability mass function* (*pmf*) defined as $f_X(x) = P(X = x)$. (The reference to this function as a "mass" function is in analogy to point mass distributions on the x-axis in mechanics). For example, the following table shows the probability mass function for a discrete random variable *X*:

x	-2	-1	0	1	2
$f_X(x)$	0.2	0.3	0.1	0.2	0.2

A plot of the *pmf*, produced by SCILAB, is shown below:

```
-->X = [-2:1:2];

-->fX = [0.3 0.2 0.1 0.2 0.2];

-->plot2d3('gnn',X',fX',1,'111','prob. mass function',[-3,0,3,0.5])
```



The probability mass function has the following properties:

$$0 \le f_X(x) \le 1$$

$$\sum_{all \quad x} f_X(x) = 1.0.$$

The *cumulative distribution* function (*cdf*) is defined as

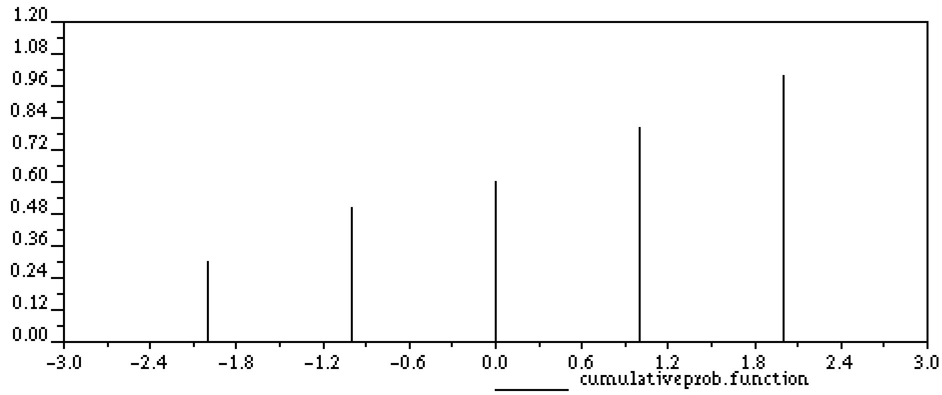
$$F_X(x) = P(X \le x) = \sum_{k \le x} P(X = k) = \sum_{k \le x} f_X(k).$$

For example, for the *pmf* defined by the table shown above, the *cdf* is given by the table:

x	-2	-1	0	1	2
$F_X(x)$	0.2	0.5	0.6	0.8	1.0

The graph for the *cdf* is shown below:

```
-->FX = [0.3 0.5 0.6 0.8 1.0]
 FX  =

!    .3     .5     .6     .8    1. !

-->xset('window',1)
 -->plot2d3('gnn',X',FX',1,'111','cumulative prob. function',[-3,0,3,1.2])
```

Most discrete random variables of interest are such that they take only integer values, say, *X = 0,1,2, …, n*. In such case, we can write the following formulas for calculating probabilities:

$$P(X = x) = f_X(x)$$

$$P(X \le x) = F_X(x) = \sum_{k=0}^{x} P(X = k) = \sum_{k=0}^{x} f_X(x)$$

$$P(X < x) = P(X \le x-1) = F_X(x-1) = \sum_{k=0}^{x-1} P(X = k) = \sum_{k=0}^{x-1} f_X(x)$$

$$P(X \ge x) = \sum_{k=x}^{n} P(X = k) = \sum_{k=x}^{n} f_X(k)$$

$$P(X > x) = P(X \ge x+1) = \sum_{k=x+1}^{n} P(X = k) = \sum_{k=x+1}^{n} f_X(k)$$

The next two formulas use the formula for the probability of the complement [*P(A') = 1-P(A)*]:

$$P(X \ge x) = 1 - P(X < x) = 1 - P(X \le x-1) = 1 - F_X(x-1) = 1 - \sum_{k=0}^{x-1} P(X = k) = 1 - \sum_{k=0}^{x-1} f_X(x)$$

$$P(X > x) = 1 - P(X \le x) = 1 - F_X(x) = 1 - \sum_{k=0}^{x} P(X = k) = 1 - \sum_{k=0}^{x} f_X(x)$$

The Poisson distribution

For example, suppose that discrete random variable *X* stands for the number of vehicles per hour visiting a service station on a road. This random variable follows the so-called Poisson distribution. The Poisson distribution has a probability mass function given by

$$f_X(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, \quad x = 0,1,2,\ldots,\infty$$

To indicate that a discrete random variable X follows the Poisson distribution, the notation *X ~ Poisson(λ)*, is sometimes used.

The parameter λ represents the mean value of vehicles per hour visiting the service station. Suppose that this value has been determined to be 8.5, then the *pmf* is

$$f_X(x) = \frac{e^{-8.5} \cdot 8.5^x}{x!}, \quad x = 0,1,2,\ldots,\infty$$

To define such a function in SCILAB we calculate the factorial through the use of the Gamma function, i.e., $\Gamma(n+1) = n!$. The Gamma function is available in SCILAB as *gamma(x)*. Thus, we can define the following function *fpoisson*, representing the probability mass function of the Poisson distribution, as

```
-->deff('[fX]=fpoisson(x,lambda)','fX = exp(-lambda)*lambda^x/gamma(x+1)')
```

Using this function and the formulas shown earlier for probability calculations of discrete variables, we proceed to calculate the following:

P(exactly 10 vehicles visit the service station in an hour) = P(X = 10) = $f_X(10) = e^{-8.5} \cdot 8.5^{10}/10! = 0.1104$

```
-->lambda = 8.5; fpoisson(10)
 ans  =

     .1103883
```

P(7 vehicles or less will visit the service station in an hour) =

$$P(X \le 7) = F_X(7) = \sum_{k=0}^{7} f_X(k) = \sum_{k=0}^{7} \frac{e^{-8.5} \cdot 8.5^k}{k!} = 0.3856$$

```
-->xx=[0:7], pp = fpoisson(xx), prob = sum(pp)
 xx  =

!   0.    1.    2.    3.    4.    5.    6.    7. !
 pp  =


         column 1 to 5

!    .0002035     .0017295     .0073503     .0208258     .0442549 !

         column 6 to 8

!    .0752333     .1065806     .1294192 !
 prob  =

     .3855971
```


P(less than 3 vehicles will visit the service station in an hour) =

$$P(X < 3) = P(X \le 2) = F_X(2) = \sum_{k=0}^{2} f_X(x) = \sum_{k=0}^{2} \frac{e^{-8.5} \cdot 8.5^k}{k!} = 0.00928$$

```
-->xx = [0:2]; pp = fpoisson(xx,lambda); sum(pp)
 ans  =

     .0092832
```

P(more than 5 will visit the service station in an hour) =

$$P(X > 5) = 1 - P(X \le 5) = 1 - \sum_{k=0}^{5} \frac{e^{-8.5} \cdot 8.5^k}{k!} = 1 - 0.1496 = 0.8504$$

```
-->xx=[0:5];pp=fpoisson(xx);1-sum(pp)
 ans  =

     .8504027
```

P(10 or more vehicles will visit the service station in an hour) =

$$P(X \ge 10) = 1 - P(X \le 9) = 1 - \sum_{k=0}^{9} \frac{e^{-8.5} \cdot 8.5^k}{k!} = 1 - 0.6530 = 0.3470$$

```
-->xx=[0:9];pp=fpoisson(xx);1-sum(pp)
 ans  =

     .3470263
```

SCILAB provides a variety of functions for calculating the cumulative distribution function and its inverse for a variety of discrete and continuous distributions. These will be presented in more detail in a later chapter. In this section we describe function *cdfpoi* corresponding to the Poisson distribution.

Function *cdfpoi*

Function *cdfpoi* stands for "*c*umulative *d*istribution *f*unction for the *Poi*sson distribution." A call to function *cdfpoi* may involve two or more parameters which we will identify as follows:

p = a probability representing the Poisson cdf, i.e., $p = P(X \le S)$
q = 1-p = the probability of the complement, i.e., $q = P(X > S) = P(X \ge S+1)$
S = the upper value of X in probability $p = P(X \le S)$
Xlam = the parameter λ of the Poisson distribution

Also, any of these parameters may be returned by the function call based on the values of the other parameters used as arguments in the call to function *cdfpoi*. The following are the three possible function calls specifying the arguments to use and the parameters returned:

1) *[p,q]=cdfpoi("PQ",S,Xlam)*
2) *[S]=cdfpoi("S",Xlam,P,Q)*
3) *[Xlam]=cdfpoi("Xlam",P,Q,S);*

- Function call number 1 returns the probabilities *p* and *q* given the value of *S*, and of the Poisson distribution parameter λ = *Xlam*
- Function call number 2 returns the value of *S* given the Poisson distribution parameter λ = *Xlam* and the probabilities *p* and *q*.
- Function call number 3 returns the value of the Poisson distribution parameter λ = *Xlam*, given the probabilities *p* and *q* corresponding to the value of *S*.

The string occupying the first position in the list of arguments specifies the type of parameter requested from the function call. Thus, the strings corresponding to the three different types of calls to *cdfpoi* are:

"*PQ*" - to request probabilities *p* and *q*
"*S*" - to request a value of the Poisson variable, from $P(X \le S) = p$ and $P(X>S) = q$.
"*Xlam*" - to request the parameter λ of the Poisson distribution

Care must be exercised when calling the function so that the arguments are in the appropriate order.

Applications of the function *cdfpoi*

Suppose that X follows the Poisson distribution with $\lambda = 5.5$. Calculate the following probabilities:

$P(X \le 6) = F_X(6)$:

```
-->cdfpoi("PQ",6,5.5)
 ans  =

     .6860360
```

$P(X<3) = P(X \le 2) = F_X(2)$:

```
-->cdfpoi("PQ",2,5.5)
 ans  =

     .0883764
```

$P(X > 4) = 1 - P(X \le 4) = 1 - F_X(4)$. The result is the *q* in the following call to *cdfpoi*:

```
-->[p,q] = cdfpoi("PQ",4,5.5)
 q  =

     .642482
 p  =

     .357518
```

$P(X \ge 2) = 1 - P(X<2) = 1 - P(X \le 1) = 1 - F_X(1)$. The result is the value of q in:

```
-->[p,q] = cdfpoi("PQ",1,5.5)
 q  =

     .9734360
 p  =

     .0265640
```


$P(3 \le X \le 5) = P(X \le 5) - P(X \le 2)$:

```
-->cdfpoi("PQ",5,5.5) - cdfpoi("PQ",2,5.5)
 ans  =

     .4405423
```

$P(X = 0) = F_X(0) = f_X(0)$, since $X = 0$ is the first possible value of X:

```
-->cdfpoi("PQ",0,5.5)
 ans  =     .0040868
```

$P(X=3) = P(X \le 3) - P(X \le 2) = F_X(3) - F_X(2) = f_X(3)$:

```
-->cdfpoi("PQ",3,5.5) - cdfpoi("PQ",2,5.5)
 ans  =

     .1133228
```

The function *cdfpoi* can be used also to determine the value of S in $P(X \le S) = p$, or $P(X>S) = q$, given λ. For example, if it is known that the average number of maintenance vehicles visiting a service station during a certain period of time is 3.2 vehicles per hour, what would be the number of vehicles, S, such that the probability that at most S vehicles will visit the service station in one hour is 0.85? The appropriate call to function *cdfpoi* is:

```
-->cdfpoi("S",3.2,0.85,1-0.85)
 ans  =

    4.5411371
```

To determine the Poisson distribution parameter λ given the probability $p = P(X \le S)$ and the value S, we use a call to *cdfpoi* as illustrated in the next example:

```
-->cdfpoi("Xlam",0.75,0.25,4)
 ans  =

    3.3686004
```

Probability Distribution for Continuous Random Variables

The calculation of probabilities for continuous random variables is based on a *probability density function* (*pdf*), $f_X(x)$, such that $f_X(x) \ge 0$. The probability that the random variable X is located between the values x_1 and x_2 is given in terms of the *pdf* by the equation:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f_X(x)dx.$$

The probability that the continuous random variable X takes a specific value is zero, i.e., $P(X = x)$. This can be checked by taking $x_1 = x - \Delta x$ and $x_2 = x + \Delta x$, where Δx is a small increment, and calculating the probability

$$P(X = x) = \lim_{\Delta x \to 0} P(x - \Delta x < X < x + \Delta x) = \lim_{\Delta x \to 0} \int_{x-\Delta x}^{x+\Delta x} f_X(x)dx \approx \lim_{\Delta x \to 0} 2 \cdot \Delta x \cdot f_X(x) = 0.$$

As a consequence, the following probabilities are equivalent for continuous random variables:

$$P(x_1 < X < x_2) = P(x_1 \le X < x_2) = P(x_1 < X \le x_2) = P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f_X(x)dx.$$

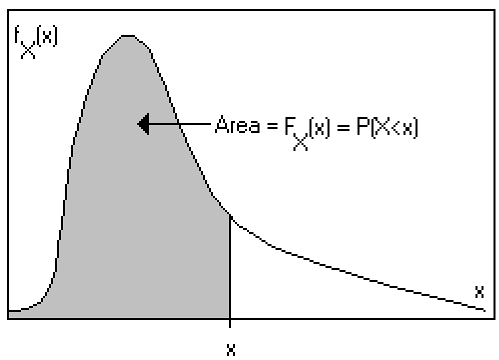
In general, the range of values of a continuous random variable encompasses all real numbers, i.e., $-\infty < x < \infty$. Therefore, the probability density function, $f_X(x)$, must satisfy the condition that

$$\int_{-\infty}^{\infty} f_X(x)dx = 1.0.$$

The *cumulative distribution function* (*cdf*) for a continuous random variable is defined by

$$F_X(x) = P(X \le x) = P(X < x) = \int_{-\infty}^{x} f_X(\xi)d\xi$$

If we plot the *pdf*, $f_X(x)$, we can illustrate the value of the *cdf* $F_X(x)$ as the area under the curve $f_X(x)$ within the limits $-\infty$ and x:



In terms of the *cdf* we can write the following probability formulas:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f_X(x)dx = \int_{-\infty}^{x_2} f_X(x)dx - \int_{-\infty}^{x_2} f_X(x)dx = F_X(x_2) - F_X(x_1).$$

$$P(X > x) = \int_{x}^{\infty} f_X(x)dx = 1 - P(X < x) = \int_{-\infty}^{x_2} f_X(x)dx = 1 - F_X(x).$$

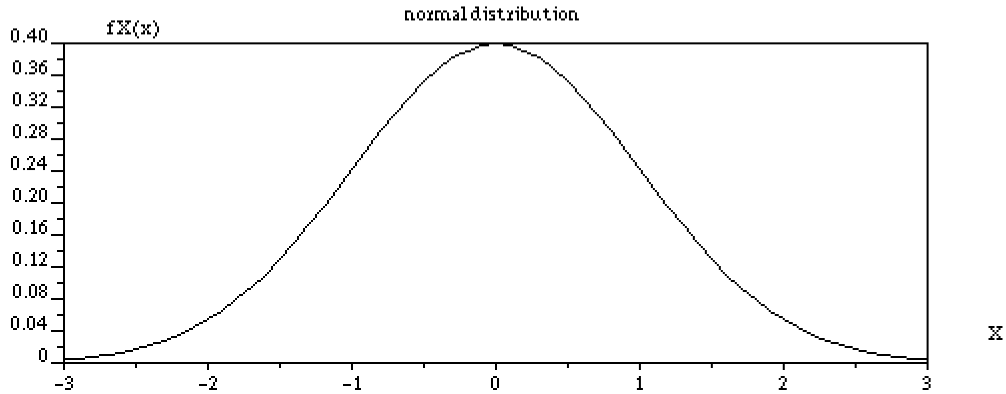
The standard normal distribution

Consider for example, the standard normal distribution, whose *pdf* is given by

$$f_X(x) = exp(-z^2)/(2\pi)^{1/2},\ -\infty < x < \infty.$$

A plot of the distribution follows:

```
-->x=[-3:0.1:3]; y=exp(-x.^2/2)./sqrt(2.*%pi);
-->plot(x,y,'X','fX(x)','normal distribution')
```

Note: In most textbooks, the standard normal variate is referred to as *Z*, with its *pdf* defined as

$$\phi(z) = exp(-z^2)/(2\pi)^{1/2},$$

and the corresponding CDF is referred to as $\Phi(z)$.

Function *cdfnor*

Function *cdfnor* stands for "*c*umulative *d*istribution *f*unction for the *nor*mal distribution." A call to function *cdfnor* may involve three or more parameters which we will identify as follows:

p = a probability representing the normal *cdf*, i.e., $P = P(X<x)$
$q = 1-p$ = the probability of the complement, i.e., $q = P(X>x)$
x = the value of X in the previous two expressions
mu = the mean value of the distribution (μ)
sigma = the standard deviation of the distribution (σ)

Also, any of these parameters may be returned by the function call based on the values of the other parameters used as arguments in the call to function *cdfnor*. The following are the four possible function calls specifying the arguments to use and the parameters returned:

1. *[p,q] = cdfnor("PQ",x,mu,sigma)*
2. *[x] = cdfnor("X",mu,sigma,p,q)*
3. *[mu] = cdfnor("Mean",sigma,p,q,x)*
4. *[sigma] = cdfnor("Std",p,q,x,mu)*

- Function call number 1 returns the probabilities *p* and *q* given the value of *x*, the mean, *mu*, and the standard deviation, *sigma*.
- Function call number 2 returns the value of *x* given the mean, *mu*, the standard deviation, *sigma*, and the probabilities *p* and *q*.
- Function call number 3 returns the value of the mean, *mu*, given the standard deviation, *sigma*, and the probabilities *p* and *q* corresponding to the value of *x*.
- Function call number 4 returns the standard deviation of the distribution, *sigma*, given the mean value, *mu*, and the probabilities *p* and *q* corresponding to the value of *x*.

The string occupying the first position in the list of arguments specifies the type of parameter requested from the function call. Thus, the strings corresponding to the four different types of calls to *cdfnor* are:

"*PQ*" - to request probabilities *p* and *q*
"*X*" - to request a value of the normal variable
"*Mean*" - to request the mean of the distribution
"*Std*" - to request the standard deviation of the distribution

Care must be exercised when calling the function so that the arguments are in the appropriate order.

Note: The probability density function, *pdf*, for a general normal distribution, *X*, with a mean value, μ, and a standard deviation, σ, is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \sigma > 0, -\infty < x < \infty.$$

The standard normal distribution has mean value $\mu = 0$ and standard deviation $\sigma = 1$.

Some probability calculations using the standard normal distribution ($\mu = 0,\ \sigma = 1$) follow. First, we define the values of the mean and standard deviation for the standard normal distribution:

```
-->mu = 0; sigma = 1.0;
```

The probability calculations are:

- $P(X<1.1) = F_X(1.1) = 0.8643$, calculated using

```
-->cdfnor('PQ',1.1,mu,sigma)
 ans  =

     .8643339
```

- $P(X>0.8) = 1 - P(X<0.8) = 1 - F_X(0.8) = 0.2118554$. This is the value of *q* calculated using:

```
-->[p,q] = cdfnor('PQ',0.8,mu,sigma)
 q  =

     .2118554
 p  =

     .7881446
```

- $P(-0.9 < X < 0.6) = P(X<0.6) - P(X<-0.9) = 0.5417$, calculated using:

```
-->cdfnor('PQ',0.6,mu,sigma) - cdfnor('PQ',-0.9,mu,sigma)
 ans  =

     .5416868
```


The following probability calculations correspond to a normal distribution with mean $\mu = 2.5$ and standard deviation $\sigma = 1.5$:

```
-->mu = 2.5; sigma = 1.5;
```

The probability $P(1.0 < X < 3.0) = P(X < 3.0) - P(X < 1.0) = 0.4719034$ is calculated using:

```
-->cdfnor('PQ',3.0,mu,sigma) - cdfnor('PQ',1.0,mu,sigma)
ans  =

     .4719034
```

To determine the mean given P(X<2) = 0.35 with standard deviation σ = 0.5 use

```
-->mu = cdfnor("Mean",0.5,0.35,0.65,2)
 mu  =

    2.1926602
```

To determine the standard deviation P(X<3) = 0.85 with mean μ = 2.5

```
-->sigma = cdfnor("Std",0.85,0.15,3,2.5)
 sigma  =

     .4824237
```

In the previous call to *cdfnor* we provided as arguments the values of the cumulative distribution function, $p = P(X<x) = F_X(x)$, as well as the value of the probability of the complement, q = $P(X>x) = 1 - F_X(x)$. By definition, these two probabilities should add to 1.0, i.e., p + q = 1.0. What would happen if the values of p and q do not add to 1.0:

```
-->mu = cdfnor("Mean",0.5,0.35,0.50,2)
                                           !--error   999
mu = cdfnor("Mean",0.5,0.35,0.50,2)
```

To avoid this problem you could simply write p as 0.35 and q as 1-0.35 in the call to function *cdfnor*:

```
-->mu = cdfnor("Mean",0.5,0.35,1-0.35,2)
 mu  =

    2.1926602
```

The expectation operator

Let X be a discrete random variable with mpf $f_X(x)$, and let $g(X)$ be a real-valued function of X, we define the mathematical expectation of the function $g(X)$ as

$$E[g(X)] = \sum_{all\ x} g(x_i) \cdot f_X(x_i)$$

For a continuous random variable X with *pdf* $f_X(x)$, the expectation of the function $g(X)$ is defined as

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot f_X(x) \cdot dx$$

For example, for the discrete random variable X whose *pmf* is given by the table

x	-2	-1	0	1	2
$f_X(x)$	0.2	0.3	0.1	0.2	0.2

We can calculate the expectation of the function $g(X) = X^2+X$ using SCILAB as follows:

```
-->X = [-2:1:2];

-->fX = [0.3 0.2 0.1 0.2 0.2];

-->g = X.^2+X;

-->EgX = g*fX';
 EgX  =

    2.2
```

The result is $E[g(X)] = E[X^2+X] = 2.2$.

Consider the continuous random variable X whose *pdf* is given by $f_X(x) = \lambda e^{-\lambda x}$, $\lambda > 0$, $x \geq 0$. This is called the *exponential distribution*. To obtain the expectation of $g(X) = X^2+X$, we need to calculate an improper definite integral (improper integral are those that have infinite integration limits), namely,

$$E[X^2 + X] = \int_0^{\infty} (x^2 + x) \cdot \lambda \cdot e^{-\lambda x} \cdot dx.$$

Such symbolic calculations are not possible in SCILAB, therefore, we need to perform them by hand or using a symbolic software such as Maple. You can verify that the integral under consideration results in

$$E[g(X)] = E[X^2+X] = (2+\lambda)/\lambda^2.$$

The mean value, variance, and standard deviation of the distribution

The *mean value* of a probability distribution is the expectation of X, i.e.,

$$\mu_X = E[X].$$

The *variance* of a probability distribution is the expectation of $(X-\mu_X)^2$, i.e.,

$$Var[X] = E[(X-\mu_X)^2] = E[X^2]-(E[X])^2 = E[X^2]-\mu_X^2.$$

The *standard deviation* of a probability distribution is simply the square root of the variance of the distribution, i.e.,

$$\sigma_X = (Var[X])^{1/2}.$$

We can define a *coefficient of variation* the same way we did for statistics of a sample, i.e.,

$$CV_X = \frac{\sigma_X}{\mu_X} \cdot 100\%$$

<u>*Calculation of mean, variance and standard deviation for a discrete random variable*</u>

For a discrete probability distribution with *pmf* $f_X(x)$ the following formulas apply:

$$\mu_X = \sum_{i=1}^{n} x_i \cdot f_X(x_i)$$

$$Var[X] = \sum_{i=1}^{n} (x_i - \mu_X)^2 \cdot f_X(x)$$

As an example, we use again the discrete probability distribution represented by the table:

x	-2	-1	0	1	2
$f_X(x)$	0.2	0.3	0.1	0.2	0.2

We calculate the mean, variance, standard deviation and coefficient of variation using SCILAB:

```
-->X = [-2:1:2];

-->fX = [0.3 0.2 0.1 0.2 0.2];

-->muX = X*fX'
 muX  =

  -   .2

-->VarX = (X-muX).^2*fX'
 VarX  =

    2.36

-->sigmaX = sqrt(VarX)
 sigmaX  =

    1.5362291

-->CVX = (sigmaX/muX)*100
 CVX  =

  - 768.11457
```

The results are $\mu_X = -0.20$, $Var[X] = 2.36$, $\sigma_X = 1.5362$, and $CV_X = -768.1146$.

Calculation of mean, variance and standard deviation for a continuous random variable

For a continuous random variable with *pdf* $f_X(x)$ the following formulas apply:

$$\mu_X = \int_{-\infty}^{\infty} x \cdot f_X(x) \cdot dx$$

$$Var[X] = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f_X(x) \cdot dx$$

For example, the exponential distribution with *pdf* $f_X(x) = \lambda e^{-\lambda x}$, $\lambda > 0$, $x \geq 0$, has the following values of mean, variance, standard deviation, and coefficient of variation:

$$\mu_X = 1/\lambda,\ Var(X) = 1/\lambda^2,\ \sigma_X = 1/\lambda,\ CV_X = 100\%.$$

If the continuous probability density function under consideration is defined in a finite range, it is possible to use the numerical integration methods developed or presented in Chapter … to calculate mean values, variances, and probabilities. For example, consider the probability density function given by

$$f_X(x) = k\,(1-x^2),\ 0 < x < 1$$

The value of k can be determined from the fact that $\int_0^1 f_X(x)\, dx = k \int_0^1 (1-x^2)\, dx = 1$. To calculate the integral $\int_0^1 (1-x^2)$ dx using SCILAB use the following commands:

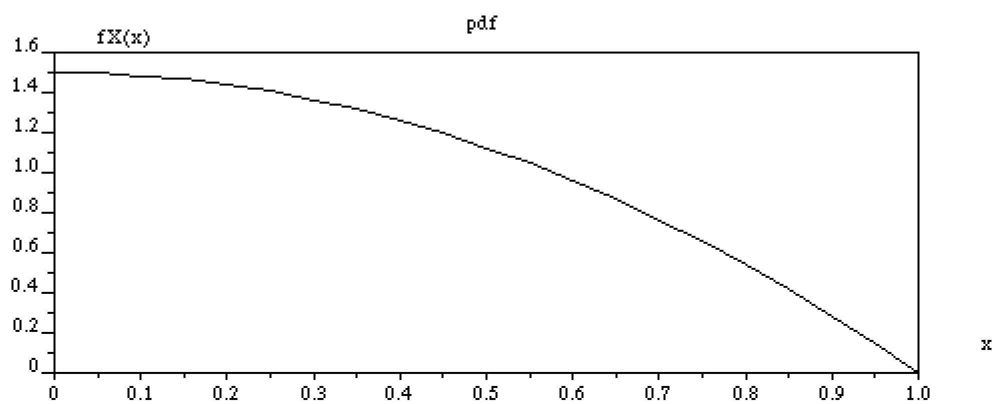
```
-->integrate('1-x^2','x',0,1)
 ans  =

     .6666667
```

The value of k turns out to be:

```
-->1/ans
 ans  = 1.5
```

A plot of the pdf is obtained by using:

```
-->deff('[y]=fX(x)','y=1.5*(1-x^2)')

-->xx=[0:0.05:1];yy=fX(xx);plot(xx,yy);
```



To calculate probabilities, we define the function *f(x)* and calculate the probabilities using SCILAB function *integrate* as indicated below.

- P(X<0.3)

```
-->integrate('fX(x)','x',0,0.3)
 ans  =

     .4365
```

- P(X>0.7)

```
-->integrate('fX(x)','x',0.7,1)
 ans  =

     .1215
```

- P(0.2 < X < 0.6)

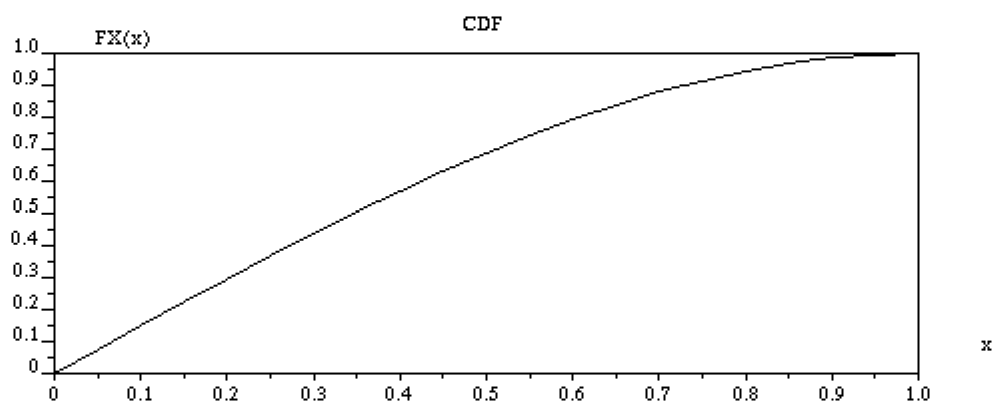
```
-->integrate('fX(x)','x',0.2,0.6)
 ans  =

     .496
```

The corresponding cumulative distribution function (cdf) is easily calculated by hand, as

$$F_X(x) = k\int_0^x (1-\xi^2)d\xi = k[\xi - \frac{\xi^3}{3}]\,|_0^x = k\left(x - \frac{x^3}{3}\right).$$

This function can be plotted using SCILAB:

```
-->deff('[y]=FX(x)','y=1.5*(x-x^3/3)')

-->YY = FX(xx);plot(xx,YY,'x','FX(x)','CDF')
```



Moments

The k-th moment of the distribution of a random variable *X* about the origin (i.e., about zero) is calculated as $\mu_k' = E[X^k]$. The mean value is the first moment about the origin, i.e., $\mu_X = \mu_1' = E[X]$.

The k-th moment of the distribution about the mean is calculated as $\mu_k = E[(X-\mu_X)^k]$. The variance is the second moment about the mean, i.e., $Var[X] = \mu_2 = E[(X-\mu_X)^2] = E[X^2]-(E[X])^2 = \mu_2' - \mu_1'^2$

Skewness and kurtosis

The skewness and kurtosis of a probability distribution are defined as

$$\sigma k = \mu_3/\mu_2^{1.5} = \mu_3/\sigma_X^3$$

and

$$\kappa = \mu_4/\mu_2^2 = \mu_4/\sigma_X^4$$

Calculating moments for a discrete distribution

The moments about the origin for a discrete distribution are calculated using

$$\mu'_k = \sum_{i=1}^{n} x_i^k \cdot f_X(x_i),$$

while the corresponding moments about the mean, $\mu_X = \mu'_1 = \Sigma x_i \cdot f_X(x_i)$, are calculated as

$$\mu_k = \sum_{i=1}^{n} (x_i - \mu_X)^k \cdot f_X(x_i).$$

To illustrate the calculation of these moments we use again the discrete probability mass function given by the table:

x	-2	-1	0	1	2
$f_X(x)$	0.2	0.3	0.1	0.2	0.2

Using SCILAB, we first enter the values of X and fX:

```
-->X = [-2:1:2]; fX = [0.3 0.2 0.1 0.2 0.2];
```

The following function, named *mukp(k)*, represents μ'_k:

```
-->deff('[mu_k_p] = mukp(k)','mu_k_p = sum(X^k.*fX)')
```

For example, the third moment about the origin, μ'_3, is calculated with *mukp(3),* i.e.,

```
->mukp(3)
 ans  =

  -  .8
```

The fourth moment about the origin for this distribution is:

```
-->mukp(4)
```



```
 ans  =

    8.4
```

The zero-th moment about the origin is by definition $\mu'_0 = \Sigma\, x_i^0 \cdot f_X(x_i) = \Sigma\, f_X(x_i) = 1$. Checking with function *mukp(0)*:

```
-->mukp(0)
 ans  =

    1.
```

The first moment about the origin is, of course, the mean value of the distribution, $\mu_X = \mu'_1$:

```
-->muX = mukp(1)
 ans  =

  -  .2
```

To calculate the moments about the mean we use function *muk(k)*, defined as:

```
-->deff('[mu_k] = muk(k)','mu_k = sum((X-mux)^k.*fX)')
```

The variance of the discrete distribution is the second moment about the mean, i.e., $Var(X) = \sigma_X^2 = \mu_2$:

```
-->muk(2)
 ans  =

    2.36
```

From the definition of the variance, $Var(X) = \Sigma\, (x_i - \mu_X)^2 \cdot f_X(x_i) = \Sigma\, x_i^2 \cdot f_X(x_i) - \mu_X^2 = \mu'_2 - (\mu'_1)^2$. Checking with SCILAB:

```
-->mukp(2)-mukp(1)^2
 ans  =

    2.36
```

The skewness and kurtosis of the discrete probability distribution under consideration are calculated from the definition of those parameters, i.e., skewness: $\sigma k = \mu_3/\mu_2^{1.5}$

```
-->sigmak = muk(3)/muk(2)^1.5
 sigmak  =

     .1721142
```

And, kurtosis: $\kappa = \mu_4/\mu_2^2$

```
-->kappa = muk(4)/muk(2)^2
 kappa  =

    1.4958345
```


Calculating moments for a continuous distribution

The moments about the origin for a continuous distribution are calculated using

$$\mu'_k = \int_{-\infty}^{\infty} x^k \cdot f_X(x) \cdot dx,$$

while the corresponding moments about the mean, $\mu_X = \mu'_1 = \Sigma x_i \cdot f_X(x_i)$, are calculated as

$$\mu_k = \int_{-\infty}^{\infty} (x-\mu_X)^k \cdot f_X(x) \cdot dx.$$

When the continuous pdf is defined in terms of symbolic parameters, e.g., the exponential distribution, $f_X(x) = \lambda e^{-\lambda x}$, $x>0$, the moments need to be calculated by hand or using a symbolic environment (e.g., Maple).

For continuous pdf given by specific functions in finite intervals, moments can be calculated using the numerical integration functions developed or presented in Chapter …. As an example, for the pdf given by $f_X(x)=1.5(1-x^2)$, $0<x<1$, we can calculate the first five moments about the origin using SCILAB as follows:

```
-->mukp0 = integrate('x^0*fX(x)','x',0,1)
 mukp0  =

    1.

-->mukp1 = integrate('x^1*fX(x)','x',0,1)
 mukp1  =

     .375

-->mukp2 = integrate('x^2*fX(x)','x',0,1)
 mukp2  =

     .2

-->mukp3 = integrate('x^3*fX(x)','x',0,1)
 mukp3  =

     .125

-->mukp4 = integrate('x^4*fX(x)','x',0,1)
 mukp4  =

     .0857143
```

The mean value is the first moment about the origin, thus $\mu_X = 0.375$:

```
-->muX = mukp1
 muX  =

     .375
```

The first five moments about the mean for the pdf under consideration are calculated using SCILAB as shown next:


```
-->muk0 = integrate('(x-muX)^0*fX(x)','x',0,1)
 muk0  =

    1.

-->muk1 = integrate('(x-muX)^1*fX(x)','x',0,1)
 !--error    24
convergence problem...
at line      26 of function integrate                called by :
muk1 = integrate('(x-muX)^1*fX(x)','x',0,1)


-->muk2 = integrate('(x-muX)^2*fX(x)','x',0,1)
 muk2  =

     .059375

-->muk3 = integrate('(x-muX)^3*fX(x)','x',0,1)
 muk3  =

     .0054688

-->muk4 = integrate('(x-muX)^4*fX(x)','x',0,1)
 muk4  =

     .0076381
```

Notice that there is a conflict in calculating the first moment. By definition, the first moment is equal to zero and function *integrate* has difficulties with its convergence check to obtain this result.

The skewness and kurtosis for this pdf are calculated as:

```
-->sigmaX = sqrt(muk2)
 sigmaX  =
     .2436699

-->sigmak = muk3/sigmaX^3
 sigmak  =
     .3779920

-->kappa = muk4/sigmaX^4
 kappa  =
    2.1666007
```

Multivariate probability distributions

Multivariate probability distributions are functions representing probabilities that depend on more than one random variable. The simplest case is, of course, the so-called *bi-variate probability distributions*, i.e., probability distributions that depend on two variables only.

Discrete random variables

If X and Y are two discrete random variables, we can define a joint probability mass function, $f_{XY}(x,y)$, representing the probability that $X = x$ and $Y = y$, i.e., $P[(X=x)\cap(Y=y)] = f_{XY}(x,y)$. The reason why the *pmf* is referred to as "joint" is because such function represents the intersection of two events, events ($X=x$) and ($Y=y$). An example of a joint probability distribution for two discrete variable is the table shown below. Variable X can take values of *0, 1, 2* and *3*, while variable Y takes values of *0, 1*, and *2*. The entries in the table represent values of $f_{XY}(x,y)$.

$f_{XY}(x,y)$	*Y = 0*	*Y = 1*	*Y = 2*
X = 0	*0.05*	*0.20*	*0.01*
X = 1	*0.05*	*0.10*	*0.05*
X = 2	*0.02*	*0.05*	*0.02*
X = 3	*0.15*	*0.10*	*0.20*

The joint probability mass function, $f_{XY}(x,y)$, satisfies

$$\Sigma_i \Sigma_j f_{XY}(x_i,y_j) = 1.0.$$

Marginal probability mass functions

Using the theorem of total probability we can calculate probabilities such as
$P(X=0) = P[(X=0)\cap(Y=0)] + P[(X=0)\cap(Y=1)] + P[(X=0)\cap(Y=2)]$,

which we can write as

$$f_X(0) = f_{XY}(0,0) + f_{XY}(0,1) + f_{XY}(0,2) = \Sigma_y f_{XY}(0,y) = 0.05 + 0.20 + 0.01 = 0.26.$$

A complete table of $f_X(x)$ for $x = 0, 1, 2$, and 3, is referred to as the marginal probability mass function of X, i.e.,

$$P(X = x) = f_X(x) = \sum_{all\ y_j} f_{XY}(x,y).$$

Similarly, a marginal probability mass function for Y is defined as:

$$P(Y = y) = f_Y(y) = \sum_{all\ x_i} f_{XY}(x,y).$$

The reason why these single-variable probability mass functions are referred to as marginal is because the corresponding probabilities can be obtained by adding rows and columns of the table representing the joint probability mass function. In other words, the values of the probability mass functions $f_X(x)$ and $f_Y(y)$ will appear in the margins of the table. This is illustrated in the following version of the table:

$f_{XY}(x,y)$	**Y = 0**	**Y = 1**	**Y = 2**	$f_X(x)$
X = 0	0.05	0.20	0.01	*0.26*
X = 1	0.05	0.10	0.05	*0.20*
X = 2	0.02	0.05	0.02	*0.09*
X = 3	0.15	0.10	0.20	*0.45*
$f_Y(y)$	*0.27*	*0.45*	*0.28*	***1.00***

Thus, the marginal probability mass function for X corresponds to the first and last columns of the expanded table, while the marginal probability mass function for Y corresponds to the first and last rows of the expanded table. The marginal probability mass functions for X and Y are univariate probability distributions with the properties that $\Sigma_i f_X(x_i) = 1.0$ and $\Sigma_i f_Y(y_j) = 1.0$. These results are highlighted in the shaded cell of the expanded table shown above.

The table below illustrates the construction of the marginal probability mass functions for both X and Y out of the original table:

$f_{XY}(x,y)$	Y = 0	Y = 1	Y = 2	$f_X(x)$
X = 0	0.05	0.20	0.01	*0.26*
X = 1	0.05	0.10	0.05	*0.20*
X = 2	0.02	0.05	0.02	*0.09*
X = 3	0.15	0.10	0.20	*0.45*
$f_Y(y)$	*0.27*	*0.45*	*0.28*	***1.00***

X	$f_X(x)$
0	0.26
1	0.20
2	0.09
3	0.45

Y	0	1	2
$f_Y(y)$	0.27	0.45	0.28

Joint cumulative distribution function

The joint cumulative distribution function, $F_{XY}(x,y)$, corresponding to a joint probability mass function, $f_{XY}(x,y)$, is defined as

$$F_{XY}(x,y) = P[(X \le x) \cap (Y \le y)] = \sum_{all\ x_i \le x} \sum_{all\ y_j \le y} f_{XY}(x,y).$$

For the joint probability mass function defined in the table shown above, we can construct a table corresponding to the joint cumulative distribution function as shown below:

$F_{XY}(x,y)$	Y = 0	Y = 1	Y = 2
X = 0	0.05	0.25	0.26
X = 1	0.10	0.40	0.46
X = 2	0.12	0.47	0.55
X = 3	0.27	0.72	1.00

X	$F_X(x)$
0	0.26
1	0.46
2	0.55
3	1.00

Y	0	1	2
$F_Y(y)$	0.27	0.72	1.00

The table also shows the marginal cumulative distribution functions for X and Y.

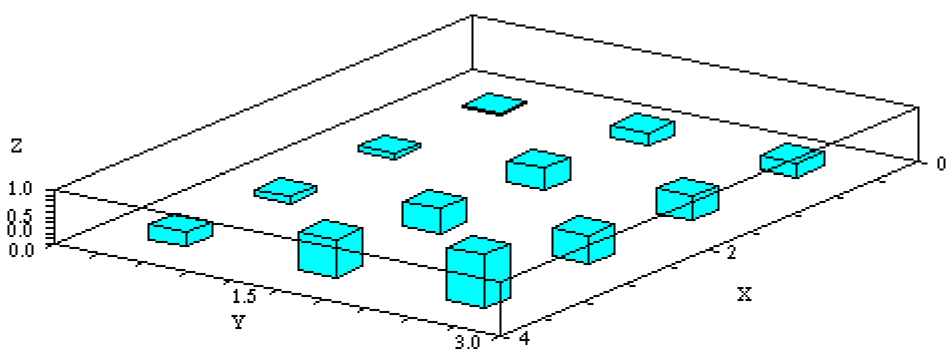
Working with tables of discrete bi-variate distributions in SCILAB

Due to the matricial nature of the SCILAB data structures, it is relatively simple to store a bi-variate probability mass function in SCILAB. We illustrate the handling of these tables using the table for the joint bi-variate probability mass function given earlier. To store the X, Y, and $f_{XY}(x,y)$ data use:

```
-->X = [0:3]; Y = [0:2];

-->fXY=[0.05,0.20,0.01;0.05,0.10,0.05;0.02,0.05,0.02;0.15,0.10,0.20]

 fXY  =
!    .05     .2      .01 !
!    .05     .1      .05 !
!    .02     .05     .02 !
!    .15     .1      .2  !
```

A three-dimensional plot of this joint probability mass function can be obtained by using the function *hist3d* (*hist*ogram in *3 d*imensions) with value *fXY*(*i,j*) representing a relative frequency in the three-dimensional histogram. Use:

```
-->hist3d(fXY)
```



The marginal distribution for *X*, $f_X(x)$, is obtained with:

```
-->fX = sum(fXY,'c')
 fX  =
!    .26 !
!    .2  !
!    .09 !
!    .45 !
```

The marginal distribution for Y, $f_Y(y)$, results from:

```
-->fY = sum(fXY,'r')
 fY  =
!    .27     .45     .28 !
```

To obtain the joint cumulative distribution function we use:

```
-->[n m] = size(fXY)
 m  =
    3.
 n  =
    4.
```

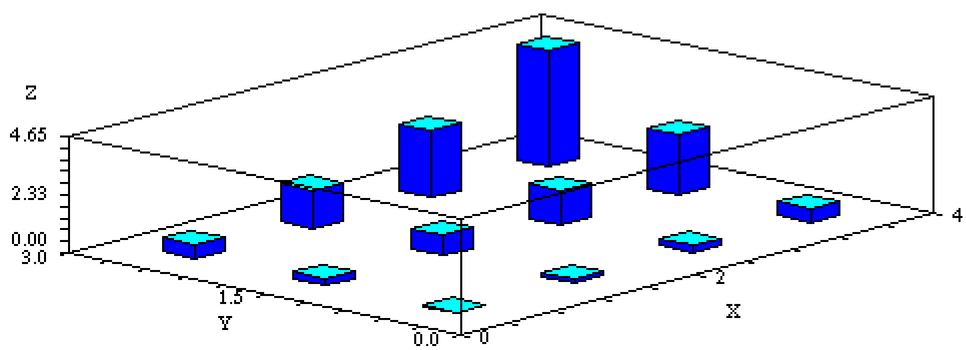


```
-->FXY = zeros(n,m);


-->for i = 1:n
-->    for j = 1:m
-->        FXY(i,j) = sum(fXY(1:i,1:j));
-->    end;
-->end;

-->FXY
 FXY  =
!    .05     .25     .26 !
!    .1      .4      .46 !
!    .12     .47     .55 !
!    .27     .72    1.   !
```

A three-dimensional histogram representing the joint cumulative distribution function of this example is produced in SCILAB through the use of function *hist3d*:

```
-->hist3d(FXY)
```

The three-dimensional histogram is shown below after rotating the axes:



Covariance and correlation coefficient

The covariance of a bi-variate probability distribution is a quantity that measures the joint variability of the two variables. The covariance, in general, is defined as the following expectation:

$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$.

For discrete random variables X and Y, the covariance is calculated as follows:

$$Cov(X,Y) = \sum_{all\ x_i}\sum_{all\ y_j}(x_i-\mu_X)\cdot(y_j-\mu_Y)\cdot f_{XY}(x,y)$$

Using the vectors *X*, *Y*, and the matrix *fXY* already loaded in SCILAB, we can calculate the co-variance of *X* and *Y* as follows:

```
-->CovXY = 0.0
 CovXY  =
```

```
    0.

-->for i = 1:n
-->    for j = 1:m
-->        CovXY = CovXY + (X(i)-mux)*(Y(j)-muy)*fXY(i,j);
-->    end;
-->end;

-->CovXY
 CovXY  =

     .1327
```

The correlation coefficient is another measure of the joint variability of X and Y, and it is defined, in general, as

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y},$$

where *Cov(X,Y)* is the covariance and σ_X and σ_Y are the standard deviations of *X* and *Y*, respectively.

Correlation coefficient for a discrete bi-variate distribution

Using the bi-variate distribution that we stored in SCILAB, the following sequence of SCILAB commands will produce the correlation coefficient for the distribution:

```
-->mux = X*fX                        //Mean of X
 mux  =

    1.73

-->muy = Y*fY'                       //Mean of y
 muy  =

    1.01

-->VarX = (X-mux)^2*fX         //Variance of X
 VarX  =

    1.6171

-->VarY = (Y-muy)^2*fY'         //Variance of Y
 VarY  =

     .5499

-->sigmax = sqrt(VarX)        //Standard deviation of X
 sigmax  =

    1.2716525

-->sigmay = sqrt(VarY)       //Standard deviation of Y
 sigmay  =

     .7415524
```


Using the covariance contained in variable *CovXY*, the correlation coefficient is:

```
-->rhoXY = CovXY/(sigmax*sigmay)
 rhoXY  =

     .1407216
```

The correlation coefficient of a bi-variate probability distribution, like its counterpart for a sample of values (x_i,y_i), is limited to the interval [-1,1].

Continuous random variables

Limiting the discussion to the bi-variate case, we can define the joint probability density function, $f_{XY}(x,y)$, of two continuous random variables, X and Y, as a continuous function having the properties:

$0< f_{XY}(x,y)$,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\cdot dy\cdot dx = 1.$$

and

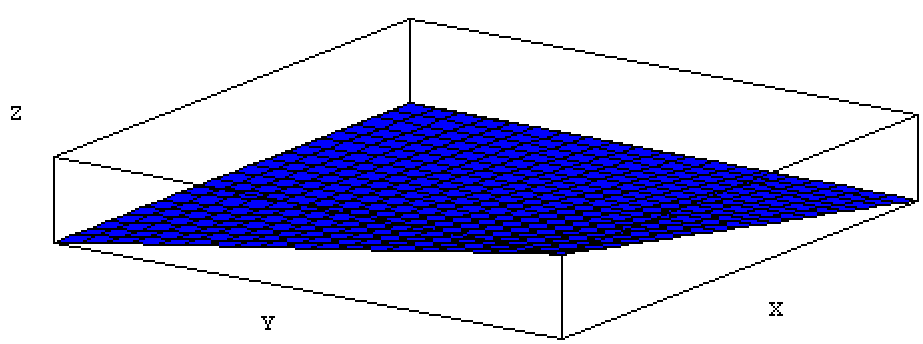
$$P[(x_1 < X < x_2)\cap(y_1 < Y < y_2)] = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f_{XY}(x,y)\cdot dy\cdot dx.$$

As an example, consider the joint probability density function given by the function

$$f_{XY}(x,y) = \begin{cases} \frac{1}{4}xy, 0\le x\le 2, 0\le y\le 2 \\ 0, \textit{elsewhere} \end{cases}$$

Notice that the function is non-zero only in the rectangle defined by $R = \{0 \le x \le 2,\ 0 \le y \le 2\}$. A plot of the distribution is produced by using function *fplot3d* in SCILAB evaluating the function $f_{XY}(x,y)$ in that region of the x-y plane:

```
-->deff('[z]=fXY(x,y)','z=1/4*x*y')
-->x = 0:0.1:2; y = x; fplot3d(x,y,fXY)
```



We can check that the joint probability density function just defined satisfies the properties listed above. For example,

$$\int_0^2\int_0^2 f_{XY}(x,y)\cdot dy\cdot dx = \int_0^2\int_0^2 \frac{1}{4}xy\cdot dy\cdot dx = \frac{1}{4}\int_0^2 \frac{1}{2}xy^2 \,|_0^2\, \cdot dx = \frac{1}{2}\int_0^2 x\cdot dx = \frac{1}{4}x^2\,|_0^2 = 1.0.$$

Joint cumulative distribution function

The joint cumulative distribution function $F_{XY}(x,y)$ corresponding to the cumulative density function $f_{XY}(x,y)$ for two continuous random variables X and Y is defined as

$$F_{XY}(x,y) = P[(X < x)\cap(Y < y)] = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(\xi,\eta)\cdot d\xi\cdot d\eta.$$

For the example of joint pdf defined above, the corresponding joint cumulative distribution function is given by

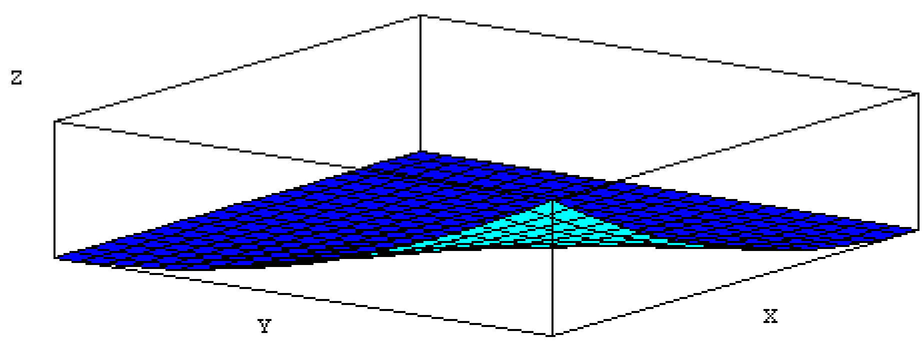
$$F_{XY}(x,y) = P[(X < x)\cap(Y < y)] = \int_0^x\int_0^y \frac{1}{4}\cdot\xi\cdot\eta\cdot d\xi\cdot d\eta = \frac{1}{16}x^2y^2.$$

The proper expression for the joint cumulative distribution is:

$$F_{XY}(x,y) = \begin{cases} \frac{1}{16}x^2y^2, 0 \le x \le 2, 0 \le y \le 2 \\ 0, \textit{elsewhere} \end{cases}$$

A plot of the joint cumulative distribution function on the region $R = \{0 \le x \le 2,\ 0 \le y \le 2\}$ is shown in the figure below:

```
-->deff('[z]=FXY(x,y)','z=1/16*x^2*y^2')
-->x = 0:0.1:2; y = x; fplot3d(x,y,FXY)
```



Covariance and correlation coefficient

The covariance for a bi-variate continuous distribution is defined as

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{XY}(x,y)dydx.$$

The correlation coefficient for a bi-variate continuous distribution is defined as that of a bi-variate discrete distribution, i.e.,

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y},$$

with the mean values and standard deviations of X and Y defined in terms of integrals of the marginal probability densities, i.e.,

$$\mu_X = \int_{-\infty}^{\infty} x \cdot f_X(x) \cdot dx,\ \mu_Y = \int_{-\infty}^{\infty} y \cdot f_Y(y) \cdot dy$$

$$\sigma_X = \sqrt{\int_{-\infty}^{\infty} (x-\mu_X)^2 \cdot f_X(x) \cdot dx},\ \sigma_Y = \sqrt{\int_{-\infty}^{\infty} (y-\mu_Y)^2 \cdot f_Y(y) \cdot dy}.$$

SCILAB calculations for bi-variate continuous distributions

We can use the functions for numerical integration developed or presented in Chapter … to calculate parameters or probabilities for specific bi-variate distributions. For example, using the bi-variate, joint probability density function, $f_{XY}(x,y)$, given earlier, namely,

$$f_{XY}(x,y) = \begin{cases} \frac{1}{4}xy, 0 \le x \le 2, 0 \le y \le 2 \\ 0, elsewhere \end{cases}$$

and the corresponding marginal probability density functions, we can calculate parameters such as the mean values and standard deviations, the co-variance of X and Y, and the correlation coefficient. Since the joint probability density function is defined within finite limits, double integrals related to this joint pdf are easily obtained by using the user-defined function *DoubleIntegral* first introduced in Chapter…

```
-->getf('DoubleIntegral')

-->deff('[z]=fXY(x,y)','z=x*y/4')
```

First, we calculate a few probabilities using function *DoubleIntegral*. We use 20 sub-intervals in each of the coordinate directions x and y:

- P(X<1,Y<1.5)

```
-->DoubleIntegral(0,1,20,0,1.5,20,fXY)
 ans  =

    .140625
```

This probability is, by definition, the cumulative probability distribution, $F_{XY}(x,y) = x^2y^2/16$:

```
-->1/16*1^2*1.5^2
 ans  =

     .140625
```

- P(X>0.5,Y<1.2)

```
-->DoubleIntegral(0.5,2,20,0,1.2,20,fXY)
 ans  =

     .3375
```

- P(X>1, Y<0.5)

```
-->DoubleIntegral(1,2,20,0,0.5,20,fXY)
 ans  =

     .046875
```

- P(X<1.2,Y>0.7)

```
-->DoubleIntegral(0,1.2,20,0.7,2,20,fxy)
 ans  =
     .3159
```

Next, we need to obtain the marginal pdf's by symbolic integration, thus,

$$f_X(x) = \int_0^2 \frac{1}{4} xydy = \frac{1}{8} xy^2 \mid_{y=0}^{y=2} = \frac{1}{2} x,\ f_Y(y) = \int_0^2 \frac{1}{4} xydx = \frac{1}{8} x^2 y \mid_{x=0}^{x=2} = \frac{1}{2} y,$$

and calculate the mean values and standard deviations of each marginal distribution using SCILAB function *integrate* as follows:

```
-->deff('[p]=fX(x)','p=x/2'); deff('[p]=fY(y)','p=y/2');

-->muX = integrate('x*fX(x)','x',0,2)
 muX  =

    1.3333333

-->muY = integrate('y*fY(y)','y',0,2)
 muY  =

    1.3333333

-->VarX = integrate('(x-muX)^2*fX(x)','x',0,2)
 VarX  =

     .2222222

-->VarY = integrate('(y-muY)^2*fY(y)','y',0,2)

VarY  =

     .2222222

-->sigmaX = sqrt(VarX), sigmaY = sqrt(VarY)
 sigmaX  =

     .4714045
```



```
 sigmaY  =

     .4714045
```

Next, we use user-defined function *DoubleIntegral* to calculate the covariance of X and Y:

```
-->deff('[p]=g(x,y)','p = (x-muX)*(y-muY)*fXY(x,y)')

-->CovXY = DoubleIntegral(0,2,20,0,2,20,g)
 CovXY  =

     .0000028
```

Finally, the correlation coefficient is calculated as:

```
-->rhoXY=CovXY/(sigmaX*sigmaY)
 rhoXY  =

     .0000125
```

Exercises

[1]. The following table represents the probability mass function for a discrete random variable X.

x	0	1	2	3	4	5
$f_X(x)$	0.06	0.17	0.22	0.28	0.17	0.1

(a) Verify that the table represents indeed a probability mass function.
(b) Plot the probability mass function.
(c) Obtain and plot the cumulative distribution function for X.
(d) Determine the mean, variance, and standard deviation of X.
(e) Determine the skewness and kurtosis of the probability mass function.
(f) Determine the expectation $E[g(X)]$ for the function $g(X) = X^2+X$.

[2]. The following table represents the probability mass function for a discrete random variable X.

x	0.5	0.7	0.9	1.1	1.4	1.7	2.1	2.4	2.7
$f_X(x)$	0.05	0.16	0.21	0.25	0.15	0.1	0.05	0.02	0.01

(a) Verify that the table represents indeed a probability mass function.
(b) Plot the probability mass function.
(c) Obtain and plot the cumulative distribution function for X.
(d) Determine the mean, variance, and standard deviation of X.
(e) Determine the skewness and kurtosis of the probability mass function.
(f) Determine the expectation $E[g(X)]$ for the function $g(X) = ln(X)$.

[3]. Let X represent the number of hits that an engineering internet site receives per hour. X follows the Poisson distribution with parameter $\lambda = 1800$. Determine the following probabilities:

(a) $P(X > 1000)$ (b) $P(X \geq 1500)$ (c) $P(X < 2000)$ (d) $P(X \leq 1500)$

[4]. If X is a discrete random variable that follows the Poisson distribution with parameter λ and $P(X \le 10) = 0.80$, determine the value of λ.

[5]. If X is a discrete random variable that follows the Poisson distribution with parameter $\lambda = 12.5$, and $P(X \le x) = 0.65$, determine the value of x.

[6]. If X is a discrete random variable that follows the Poisson distribution with parameter λ and $P(X > 5) = 0.20$, determine the value of λ.

[7]. If X is a discrete random variable that follows the Poisson distribution with parameter $\lambda = 8.2$, and $P(X > x) = 0.85$, determine the value of x.

[8]. If X is a discrete random variable that follows the Poisson distribution with parameter λ and $P(X < 15) = 0.60$, determine the value of λ.

[9]. If X is a discrete random variable that follows the Poisson distribution with parameter $\lambda = 25.8$, and $P(X < x) = 0.35$, determine the value of x.

[10]. If X is a discrete random variable that follows the Poisson distribution with parameter λ and $P(X \le 25) = 0.80$, determine the value of λ.

[11]. If X is a discrete random variable that follows the Poisson distribution with parameter $\lambda = 82$, and $P(X \le x) = 0.25$, determine the value of x.

[12]. The function $f_X(x)$ shown below represents a probability density function

$$f_X(x) = \begin{cases} 3(1-\sqrt{x}), & 0 \le x \le 1 \\ 0, & elsewhere \end{cases}$$

(a) Verify that $f_X(x)$ represents indeed a probability density function.
(b) Plot the probability density function.
(c) Obtain and plot the cumulative distribution function for X.
(d) Determine the mean, variance, and standard deviation of X.
(e) Determine the skewness and kurtosis of the probability mass function.

[13]. The function $f_X(x)$ shown below represents a probability density function

$$f_X(x) = \begin{cases} (1/2)\exp(-x/2), & x \ge 0 \\ 0, & elsewhere \end{cases}$$

(a) Verify that $f_X(x)$ represents indeed a probability density function.
(b) Plot the probability density function.
(c) Obtain and plot the cumulative distribution function for X.
(d) Determine the mean, variance, and standard deviation of X.
(e) Determine the skewness and kurtosis of the probability mass function.

[14]. The distribution of the measurement of experimental data follows the normal distribution. Suppose that X represents measurements of the acceleration of gravity at a specific point on Earth. Thus, $X \sim N(\mu = 9.806\ m/s^2,\ \sigma = 0.005\ m/s^2)$. Determine the following probabilities:

(a) $P(X<9.85)$ (b) $P(X>9.78)$ (c) $P(9.80 < X < 9.81)$

[15]. Let X be a continuous random variable that follows the normal distribution with $\mu = 120.6$, $\sigma = 40$. Determine the value of x given the following probability conditions:

(a) $P(X<x) = 0.45$ (b) $P(X>x) = 0.25$
(c) $P(80 < X < x) = 0.20$ (d) $P(-x<X<x) = 0.60$

[16]. Let X be a continuous random variable that follows the normal distribution. Given the following probabilities, determine the mean or standard deviation as requested:

(a) $P(X<2.5) = 0.8$, $\sigma = 0.3$, determine μ
(b) $P(X>33.5) = 0.6$, $\sigma = 8.5$, determine μ
(c) $P(X<122) = 0.45$, $\mu = 12.22$, determine σ
(d) $P(X>33.5) = 0.3$, $\mu = 83.15$, determine σ

[17]. The table below shows a bivariate probability mass function for two discrete variables *X* and *Y*:

$f_{XY}(x,y)$	Y = 1	Y = 2	Y = 3	Y = 4
X = 0.5	0.05	0.03	0.01	0.01
X = 1.5	0.01	0.15	0.1	0.03
X = 2.5	0.08	0.12	0.18	0.05
X = 3.5	0.03	0.08	0.05	0.02

(a) Verify that the table indeed represents a bivariate probability mass function.
(b) Plot the probability mass function $f_{XY}(x,y)$ as a three-dimensional histogram using function *hist3d*.
(c) Produce tables of the marginal probability mass functions $f_X(x)$ and $f_Y(y)$.
(d) Determine the mean value, variance, and standard deviation for X.
(e) Determine the mean value, variance, and standard deviation for Y.
(f) Produce a table for the cumulative distribution function $F_{XY}(x,y)$, and plot this function using a three-dimensional histogram.
(g) Determine the covariance and correlation coefficient for variables *X* and *Y.*

[18]. The table below shows a bivariate probability mass function for two discrete variables *X* and *Y*:

$f_{XY}(x,y)$	Y = 0.1	Y = 0.2	Y = 0.3	Y = 0.4	Y = 0.5
X = 0.5	0.05	0.03	0.01	0.01	0.00
X = 1.5	0.01	0.11	0.09	0.03	0.03
X = 2.5	0.05	0.07	0.10	0.04	0.05
X = 3.5	0.03	0.06	0.05	0.02	0.02
X = 4.5	0.03	0.05	0.04	0.02	0.00

(a) Verify that the table indeed represents a bivariate probability mass function.
(b) Plot the probability mass function $f_{XY}(x,y)$ as a three-dimensional histogram using function *hist3d*.
(c) Produce tables of the marginal probability mass functions $f_X(x)$ and $f_Y(y)$.
(d) Determine the mean value, variance, and standard deviation for X.

(e) Determine the mean value, variance, and standard deviation for Y.
(f) Produce a table for the cumulative distribution function $F_{XY}(x,y)$, and plot this function using a three-dimensional histogram.
(g) Determine the covariance and correlation coefficient for variables *X* and *Y.*

[19]. The following function represents the joint probability density function of two continuous random variables X and Y:

$$f_{XY}(x,y) = \begin{cases} 4(1-x)(1-y), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & elsewhere \end{cases}$$

(a) Verify that the table indeed represents a bivariate probability mass function.
(b) Plot the probability density function $f_{XY}(x,y)$ as a three-dimensional surface using function *plot3d*.
(c) Obtain the marginal probability density functions $f_X(x)$ and $f_Y(y)$.
(d) Determine the mean value, variance, and standard deviation for X.
(e) Determine the mean value, variance, and standard deviation for Y.
(f) Obtain the cumulative distribution function $F_{XY}(x,y)$, and plot this function using a three-dimensional surface.
(g) Determine the covariance and correlation coefficient for variables *X* and *Y.*

[20]. The following function represents the joint probability density function of two continuous random variables X and Y:

$$f_{XY}(x,y) = \begin{cases} \frac{1}{4}\exp(-\frac{x+y}{2}), & x \ge 0,\ y \ge 0 \\ 0, & elsewhere \end{cases}$$

(a) Verify that the table indeed represents a bivariate probability mass function.
(b) Plot the probability density function $f_{XY}(x,y)$ as a three-dimensional surface using function *plot3d*.
(c) Obtain the marginal probability density functions $f_X(x)$ and $f_Y(y)$.
(d) Determine the mean value, variance, and standard deviation for X.
(e) Determine the mean value, variance, and standard deviation for Y.
(f) Obtain the cumulative distribution function $F_{XY}(x,y)$, and plot this function using a three-dimensional surface.
(g) Determine the covariance and correlation coefficient for variables *X* and *Y.*

REFERENCES (for all SCILAB documents at InfoClearinghouse.com)

Abramowitz, M. and I.A. Stegun (editors), 1965,"*Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*," Dover Publications, Inc., New York.

Arora, J.S., 1985, "*Introduction to Optimum Design*," Class notes, The University of Iowa, Iowa City, Iowa.

Asian Institute of Technology, 1969, "*Hydraulic Laboratory Manual*," AIT - Bangkok, Thailand.

Berge, P., Y. Pomeau, and C. Vidal, 1984,"*Order within chaos - Towards a deterministic approach to turbulence*," John Wiley & Sons, New York.

Bras, R.L. and I. Rodriguez-Iturbe, 1985,"*Random Functions and Hydrology*," Addison-Wesley Publishing Company, Reading, Massachussetts.

Brogan, W.L., 1974,"*Modern Control Theory*," QPI series, Quantum Publisher Incorporated, New York.

Browne, M., 1999, "*Schaum's Outline of Theory and Problems of Physics for Engineering and Science*," Schaum's outlines, McGraw-Hill, New York.

Farlow, Stanley J., 1982, "*Partial Differential Equations for Scientists and Engineers*," Dover Publications Inc., New York.

Friedman, B., 1956 (reissued 1990), "*Principles and Techniques of Applied Mathematics*," Dover Publications Inc., New York.

Gomez, C. (editor), 1999, "*Engineering and Scientific Computing with Scilab*," Birkhäuser, Boston.

Gullberg, J., 1997, "*Mathematics - From the Birth of Numbers*," W. W. Norton & Company, New York.

Harman, T.L., J. Dabney, and N. Richert, 2000, "*Advanced Engineering Mathematics with MATLAB® - Second edition*," Brooks/Cole - Thompson Learning, Australia.

Harris, J.W., and H. Stocker, 1998, "*Handbook of Mathematics and Computational Science*," Springer, New York.

Hsu, H.P., 1984, "*Applied Fourier Analysis*," Harcourt Brace Jovanovich College Outline Series, Harcourt Brace Jovanovich, Publishers, San Diego.

Journel, A.G., 1989, "*Fundamentals of Geostatistics in Five Lessons*," Short Course Presented at the 28th International Geological Congress, Washington, D.C., American Geophysical Union, Washington, D.C.

Julien, P.Y., 1998,"*Erosion and Sedimentation*," Cambridge University Press, Cambridge CB2 2RU, U.K.

Keener, J.P., 1988, "*Principles of Applied Mathematics - Transformation and Approximation*," Addison-Wesley Publishing Company, Redwood City, California.

Kitanidis, P.K., 1997,"*Introduction to Geostatistics - Applications in Hydogeology*," Cambridge University Press, Cambridge CB2 2RU, U.K.

Koch, G.S., Jr., and R. F. Link, 1971, "*Statistical Analysis of Geological Data - Volumes I and II*," Dover Publications, Inc., New York.

Korn, G.A. and T.M. Korn, 1968, "*Mathematical Handbook for Scientists and Engineers*," Dover Publications, Inc., New York.

Kottegoda, N. T., and R. Rosso, 1997, "*Probability, Statistics, and Reliability for Civil and Environmental Engineers*," The Mc-Graw Hill Companies, Inc., New York.

Kreysig, E., 1983, "*Advanced Engineering Mathematics - Fifth Edition*," John Wiley & Sons, New York.

Lindfield, G. and J. Penny, 2000, "*Numerical Methods Using Matlab®,*" Prentice Hall, Upper Saddle River, New Jersey.

Magrab, E.B., S. Azarm, B. Balachandran, J. Duncan, K. Herold, and G. Walsh, 2000, "*An Engineer's Guide to MATLAB®*", Prentice Hall, Upper Saddle River, N.J., U.S.A.

McCuen, R.H., 1989,"*Hydrologic Analysis and Design - second edition*," Prentice Hall, Upper Saddle River, New Jersey.

Middleton, G.V., 2000, "*Data Analysis in the Earth Sciences Using Matlab®,"* Prentice Hall, Upper Saddle River, New Jersey.

Montgomery, D.C., G.C. Runger, and N.F. Hubele, 1998, "*Engineering Statistics*," John Wiley & Sons, Inc.

Newland, D.E., 1993, "*An Introduction to Random Vibrations, Spectral & Wavelet Analysis - Third Edition*," Longman Scientific and Technical, New York.

Nicols, G., 1995, "*Introduction to Nonlinear Science*," Cambridge University Press, Cambridge CB2 2RU, U.K.

Parker, T.S. and L.O. Chua, , "*Practical Numerical Algorithms for Chaotic Systems*," 1989, Springer-Verlag, New York.

Peitgen, H-O. and D. Saupe (editors), 1988, "*The Science of Fractal Images*," Springer-Verlag, New York.

Peitgen, H-O., H. Jürgens, and D. Saupe, 1992, "*Chaos and Fractals - New Frontiers of Science*," Springer-Verlag, New York.

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, 1989, "*Numerical Recipes - The Art of Scientific Computing (FORTRAN version)*," Cambridge University Press, Cambridge CB2 2RU, U.K.

Raghunath, H.M., 1985, "*Hydrology - Principles, Analysis and Design*," Wiley Eastern Limited, New Delhi, India.

Recktenwald, G., 2000, "*Numerical Methods with Matlab - Implementation and Application*," Prentice Hall, Upper Saddle River, N.J., U.S.A.

Rothenberg, R.I., 1991, "*Probability and Statistics*," Harcourt Brace Jovanovich College Outline Series, Harcourt Brace Jovanovich, Publishers, San Diego, CA.

Sagan, H., 1961,"*Boundary and Eigenvalue Problems in Mathematical Physics*," Dover Publications, Inc., New York.

Spanos, A., 1999,"*Probability Theory and Statistical Inference - Econometric Modeling with Observational Data*," Cambridge University Press, Cambridge CB2 2RU, U.K.

Spiegel, M. R., 1971 (second printing, 1999), "*Schaum's Outline of Theory and Problems of Advanced Mathematics for Engineers and Scientists*," Schaum's Outline Series, McGraw-Hill, New York.

Tanis, E.A., 1987, "*Statistics II - Estimation and Tests of Hypotheses*," Harcourt Brace Jovanovich College Outline Series, Harcourt Brace Jovanovich, Publishers, Fort Worth, TX.

Tinker, M. and R. Lambourne, 2000, "*Further Mathematics for the Physical Sciences*," John Wiley & Sons, LTD., Chichester, U.K.

Tolstov, G.P., 1962, "*Fourier Series*," (Translated from the Russian by R. A. Silverman), Dover Publications, New York.

Tveito, A. and R. Winther, 1998, "*Introduction to Partial Differential Equations - A Computational Approach*," Texts in Applied Mathematics 29, Springer, New York.

Urroz, G., 2000, "*Science and Engineering Mathematics with the HP 49 G - Volumes I & II*", www.greatunpublished.com, Charleston, S.C.

Urroz, G., 2001, "*Applied Engineering Mathematics with Maple*", www.greatunpublished.com, Charleston, S.C.

Winnick, J., , "*Chemical Engineering Thermodynamics - An Introduction to Thermodynamics for Undergraduate Engineering Students*," John Wiley & Sons, Inc., New York.

