

## 9.3 Randomness, Probability and Statistics

a word about  
**what**  
probability and  
statistics **are** is  
in place here!!!

In this section we will present basic concepts for randomness, probability and statistics. We start with the simple univariate (one dimensional) case. Here we will introduce *random variables*, *probability distribution functions (PDF)*, the *mean* and the *variance*. Furthermore we will introduce the concepts of *covariance* and *correlation*. Covariance is the generalization of the variance for multiple variables and plays a very important role in the Computer Vision context. After this hopefully intuitive introductions we will move on to the multivariate (higher dimensional) case and adapt the entities presented for *multivariate random variables*. For multivariate random variables we will also present the so called *covariance matrix*, which is the generalization of the covariance for the n-dimensional case.

In this chapter we will always use the example of a measurement of a point in order to explain and motivate the basic mathematical entities. That is, we will talk about a series of experiment trials, in which repeat a certain measurement and thus obtain (randomly) different values. The experiment will be measuring the location of a point and the outcome will be the coordinates of the point.

This point of view automatically leads us to the *discrete case*, to which we will restrict ourselves in the following.

### 9.3.1 Univariate Probability Theory

The title of this and the next subsection are slightly overblown. We will not introduce the *probability theory* itself neither fully nor formally correct. We will merely present the concepts needed in the lecture in a hopefully intuitive way.

#### 9.3.1.1 Random Variables

Basically we can say that a one dimensional (univariate) random variable  $X$  is a variable that can take several random values. That is, we can think about a series of experiment trials. Each *trial* consists of a single measurement or observation, the so called *experiment* and the corresponding outcome value, the *sample point* value. In the discrete case, the experiments can be identified by their serial number, so they can be represented by natural numbers  $\mathbb{N}$ . So the random variable describes the outcome values that we get at the different trials. More formally speaking, we can define a random variable like following.

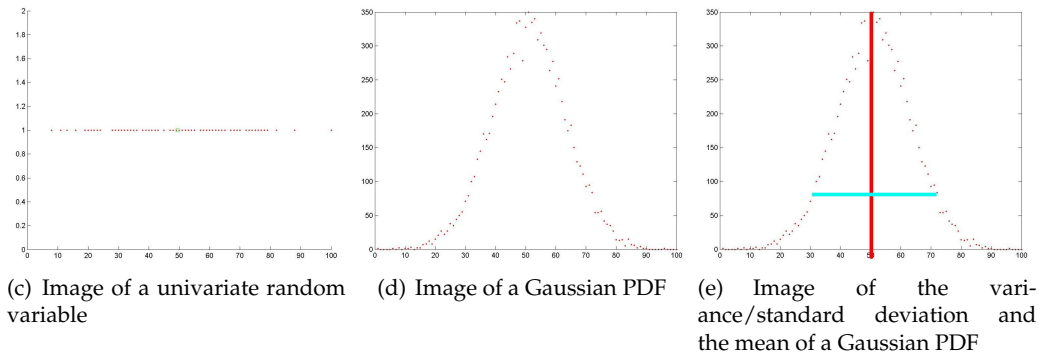
**Definition 9.8 (Univariate Discrete Random Variable)** *Random variable  $X$  is a function defined on the experiments, i.e.  $\mathbb{N}$ . Let us denote the sample space by  $\Omega$ . Let us further assume that the experiment can have only discrete outcomes so we can write the set of all outcomes  $\Omega = \{x_1, x_2, \dots\}$  with  $x_1, x_2, \dots$ . Then, we can say that the random variable  $X$  is a function  $X : \mathbb{N} \rightarrow \Omega$ .*

The most common Computer Vision usage of random variables will be that one random variable represents a set of measurements of one value. Measurement errors are reason for the “randomness” observed during the experiments.

#### 9.3.1.2 Probability Density Functions (PDF)

The set of all experiments where the random variable  $X$  assumes a certain value  $x_i$  are called an *event* and denoted  $X = x_i$ . It is important to keep in mind here that  $X = x_i$

introduce  
probability in  
the first place



represents a whole set of events. The probability that a certain event occurs is  $P(X = x_i)$ . Often, we will denote an event  $P(X = x_i)$  shortly as  $P(x_i)$ .

Now we have everything we need and can go on and define the so called *Probability Density Function (PDF)*.

**Definition 9.9 (Probability Density Function (PDF))** Given a univariate random variable  $X$ , then the function on the values  $\{x_1, x_2, \dots\}$

$$f(x_i) = P(X = x_i)$$

is called the Probability Density Function (PDF) of the random variable  $X$ .

Summed up, The PDF express the probability that the variable  $X$  takes the certain value  $x_i$ . So, for the univariate (one dimensional) case, the x-axis gives the sample points. The y-axis tells you how probable the sample point is - or in our measurements setting, how often this value was measured.

The most simple example for a PDF is the uniform PDF. Given  $n$  different outcome values  $x_1, \dots, x_n$  the uniform PDF tells us that every event is equally probable.

**Definition 9.10 (Uniform PDF)**

$$f(x) = \frac{1}{n}$$

Another very common example for a PDF is the so called *Gaussian* PDF:

**Definition 9.11 (Gaussian PDF)**

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right)$$

The Gaussian PDF is extremely important because it has numerous applications and will also be used quite often in the lecture.

*Normal distribution* is a synonym for the Gaussian distribution.

Actually the uniform and Gaussian are continuous. We use but their approximations. Explain!

Gauss properties, and visualization of Uniform and Gaussian

### 9.3.1.3 Expectation

The *expectation* expresses the expected value for a random variable  $X$ . In order to compute the expected value we obviously have to weight the value of the event  $x_i$  by the probability that this event occurs  $P(x_i)$ . Thus we get

**Definition 9.12 (Expectation)** *Given the univariate random variable  $X$  with the corresponding PDF  $f(x) = P(x)$ , we can calculate the expectation value  $E[X]$  of  $X$  by*

$$E[X] = \sum_{i=1}^n P(x_i)x_i$$

Synonyms for expectation are *mean* and *average*. Other common notations for  $E[X]$  are  $\mu_X$ , just  $\mu$ ,  $\bar{X}$  and  $\langle X \rangle$ .

**Example 9.13 (Uniform PDF Expectation)** *For the simple but nevertheless common case of uniform distribution with  $n$  different events, the mean is simply*

$$E[X] = \sum_{i=1}^n P(x_i)x_i = \sum_{i=1}^n \frac{1}{n}x_i = \frac{1}{n} \sum_{i=1}^n x_i$$

*For a fair dice we can assume that the probability of any of the possible outcomes  $\{1, 2, 3, 4, 5, 6\}$  is equal, i.e. we can assume a uniform distribution. With  $n = 6$  we get the expected value of*

$$E[X] = \frac{1}{6} \sum_{i=1}^6 i = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3.5$$

*Not really surprising, ha?*

### 9.3.1.4 Variance & Standard Deviation

*Variance* is a measure of how much the values of a random variable are spread around the expected value. For PDFs with a low variance the values similar to the expected value will be much more probable than others. To state it differently, if doing a measurement, most of the observed values will be similar to the mean. For measurements with a large variance the measured values will be widely spread around the mean.

**Definition 9.14** *Again, let  $X$  be a univariate random variable with a PDF  $f(x) = P(x)$  and the mean  $E(X)$ . Then it is intuitive to measure how much the values of  $X$  are spread around  $E(X)$  by computing  $x_i - E[X]$  for all events  $x_i$ . Then we just square the term  $x_i - E[X]$ , so that values for different  $x_i$  do not cancel each other out. Thus we get measure of how the random variable values are spread around the expected value. In order to compute the expected deviation from the mean we weight every squared deviation  $(x_i - E[X])^2$  by the probability that the sample  $x_i$  is actually measured and get  $P(x_i)(x_i - E[X])^2$ . Now we sum over all samples and get the so called variance*

$$\text{Var}(X) = \sum_{i=1}^n P(x_i)(x_i - E[X])^2$$

*We can compute the variance also using the expectation of  $X$  and  $X^2$  as follows*

$$\text{Var}(X) = \sum_{i=1}^n P(x_i)(x_i - E[X])^2 \quad (9.4)$$

$$= E[(X - E[X])^2] \quad (9.5)$$

$$= E[X^2] - E[X]^2 \quad (9.6)$$

The formula  $E[(X - E[X])^2]$  is also very intuitive since it lets us think about the variance as the expected value for  $(X - E[X])^2$  which can be interpreted as the (squared) deviation of the values of  $X$  from the expected value  $E[X]$ .

**Example 9.15 (Uniform PDF Variance)** For the above dice example we can now calculate the variance. For uniform distributions in general we get

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - E[X])^2$$

Applied to the dice example, with  $E[X] = 3.5$  as computed above we get

$$\text{Var}(X) = \frac{1}{6} \sum_{i=1}^6 (i - 3.5)^2 = \frac{(-2.5)^2 + (-1.5)^2 + (-0.5)^2 + 0.5^2 + 1.5^2 + 2.5^2}{6} = \frac{17.5}{6}$$

Another common measurement of the distribution of the values of a random variable around the mean is the *standard deviation*. It is basically just the square root of the variance. Since the square root is strictly increasing, the standard deviation can be interpreted simply as a different scaling of the variance.

**Definition 9.16 (Standard Deviation)** Let  $X$  be a univariate random variable with a variance  $\text{Var}(X)$ . Then the standard deviation is defined as

$$\sigma_X = \sqrt{\text{Var}(X)}$$

Equivalently we can now denote the variance as

$$\text{Var}(X) = \sigma_X^2$$

**Example 9.17 (One Dimensional Position Measurement)** Imagine the task of measuring the position of a certain point on a line. Unfortunately this simple task is something that we can not perform absolutely precisely, no matter how hard we try, because we will make measurement errors. Always! So we can assume that the error that we are making will be random. And thus also the result of every measurement will be. Hence we can model the measurements as a univariate random variable  $X$ . So the mean of the  $X$  would be the average of all measurements. It is reasonable to use this value as the final result of the measurement series. The variance on the other hand expresses how strongly the single measurements differed from the mean. It tells us, how large our measurement errors were. The larger the variance, the larger were the errors. We can see here that the variance somehow captures the concept of uncertainty for the one one dimensional case.

### 9.3.1.5 Covariance

Covariance is the generalization of the concept of variance for two variables. While the variance measures the average deviation from the mean of the random variable, the covariance measure average product of the deviations of two random variables  $X$  and  $Y$  from their respective means  $E[X]$  and  $E[Y]$ .

So more specifically, what we do is multiply the deviations of  $X$ , and  $Y$  from their means, i.e.  $(x_i - E[X])(y_i - E[Y])$ . In order to get the average value for  $(x_i - E[X])(y_i - E[Y])$  we do the following. We weight the term by the probability that the  $X$  and  $Y$  take on the values  $x_i$  and  $y_i$  at the same time, thus get  $P(x_i, y_i)(x_i - E[X])(y_i - E[Y])$  and finally sum over all  $i$ .

**Definition 9.18 (Covariance)** *Given two univariate random variables  $X$  and  $Y$  with their respective expectations  $E[X]$  and  $E[Y]$ . We define the covariance as*

$$\text{Cov}(X, Y) = \sum_{i=1}^n P(x_i, y_i) (x_i - E[X])(y_i - E[Y]) \quad (9.7)$$

$$= E[(X - E[X])(Y - E[Y])] \quad (9.8)$$

$$= E[XY - XE[Y] - E[X]Y + E[X]E[Y]] \quad (9.9)$$

$$= E[XY] - E[XE[Y]] - E[E[X]Y] + E[E[X]E[Y]] \quad (9.10)$$

$$= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \quad (9.11)$$

$$= E[XY] - E[X]E[Y] \quad (9.12)$$

Random variables can be dependent or independent on each other. If we say that two variables  $X$  and  $Y$  are dependent in some way on each other, then the knowledge of the one variable helps you to predict the values of the other.

One very important fact about the covariance is that it can give us information about the dependency of two variables.

**Theorem 9.19** *If the two univariate random variables  $X$  and  $Y$  are independent, then it follows that their covariance is  $\text{Cov}(X, Y) = 0$ .*

**Proof 9.20** *Assume  $X$  and  $Y$  independent, then (and only then) it holds that  $E[XY] = E[X]E[Y]$ . We can use that to compute the covariance  $\text{Cov}(X, Y)$ .*

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] \quad (9.13)$$

$$= E[X]E[Y] - E[X]E[Y] \quad (9.14)$$

$$= 0 \quad (9.15)$$

It is important to keep in mind that the converse is not true. *That is, even if  $\text{Cov}(X, Y) = 0$ , then  $X$  and  $Y$  are not necessarily independent!* However we can deduce that if  $\text{Cov}(X, Y) \neq 0$ , then  $X$  and  $Y$  are dependent.

### 9.3.1.6 Correlation

At the end of this subsection we will present a slight modification of the covariance called *correlation*. For that we require the concept of so called *normalized variables*.

**Definition 9.21 (Normalized Variable)** *Given any random variable  $X$  with a mean  $\mu$  and variance  $\sigma^2$ . We can transform this variable to a new variable  $\tilde{X}$ , s.t.  $\tilde{X}$  has a mean of zero and its variance becomes 1. We define  $\tilde{X}$  as following*

$$\tilde{X} = \frac{X - \mu}{\sigma}$$

and call it the normalized variable corresponding to  $X$ .

Now we can define the correlation  $Cor(X, Y)$  for any two random variables  $X$  and  $Y$  as the covariance of the normalized variables corresponding to  $X$  and  $Y$ , i.e.  $\tilde{X}$  and  $\tilde{Y}$ .

**Definition 9.22 (Correlation)** *Given two random variables  $X$  and  $Y$  with standard deviations  $\sigma_X$  and  $\sigma_Y$  respectively. Furthermore, let  $\tilde{X}$  and  $\tilde{Y}$  be the normalized variables corresponding to  $X$  and  $Y$ . Then we define the correlation*

$$Cor(X, Y) = Cov(\tilde{X}, \tilde{Y}) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

$Cor(X, Y)$  is also often denoted as  $\rho(X, Y)$ .

Please note that the correlation is just the same as the covariance but for the scaling by  $\frac{1}{\sigma_X \sigma_Y}$ .

If you want to compute the correlation, you might also find the following formula interesting which just inserts the formula for the covariance in the definition for correlation.

$$Cor(X, Y) = \frac{1}{\sigma_X \sigma_Y} \sum_{i=1}^n P(x_i, y_i) (x_i - E[X])(y_i - E[Y])$$

Since the correlation is just a scaled covariance it inherits the properties of the covariance regarding the dependency of random variables. That is, also here holds that if  $X$  and  $Y$  are independent then  $Cor(X, Y) = 0$  or the other way round, if  $Cor(X, Y) \neq 0$  then  $X$  and  $Y$  are dependent.

Another property of the correlation is that its norm is always less or equal to one,  $\|Cor(X, Y)\| \leq 1$ . If  $Cor(X, Y) = \pm 1$  then it holds that  $X$  and  $Y$  are dependent s.t.  $Y = aX + b$  with  $a, b \in \mathbb{R}$ .

The correlation is very often used as a similarity measure in Computer Vision and other fields.

extend to  
subsubsection:  
Correlation as  
similarity  
measure!

### 9.3.2 Multivariate Probability Theory

In the first part of this section we have introduced the basic univariate probability concepts, such as random variable, PDF, event, mean, variance and standard deviation.

Now we will go on to the multivariate case. This case can also be interpreted as an experiment. This time the outcome of the experiment is not only on one value but many. Thus we speak about the *multivariate* instead of *univariate* case.

We will see that the multivariate concepts for the random variable, PDF, mean and variance and standard deviation are derived from the univariate case in a straight forward manner. Following we will introduce a very important new concept, the *covariance matrix*, which is the generalization of the covariance for the multivariate case. This will make up the most part of the following section.

#### 9.3.2.1 Multivariate Random Variables

Analogously to the univariate case, we can now reason about *multivariate random variables* as experiments. In contrast to the univariate case, where the outcome of the experiment is only one value from the sample space, now each experiment trial gives us several values - we perform several different measurements, so to say. An example for a univariate random variable would be a measurement of a point on the line. We always measure one value, let's say the X coordinate. In order to describe measurement of a point in a plane, we have to keep track of the X and Y coordinate we measure every time. Thus every sample point would be a vector containing the X and Y value for one trial measurement. We would use a multivariate random variable in order to describe all the measurements.

**Definition 9.23 (Multivariate Discrete Random Variable)** *The multivariate discrete random variable  $X$  is a function defined on sample space  $S$ . An experiment outcome is now a  $n$ -dimensional vector, i.e. there are  $n$  different observations made during one experiment trial. Let us denote the  $n$ -dimensional sample space by  $\Omega$ . That means that  $\Omega$  represents all possible experiment outcomes. Let us further assume that the experiment can only have discrete outcomes, i.e. integers. Then the sample space can be written as  $\Omega = \mathbb{Z}^n$  being the different possible outcome values of an experiment.*

*Each experiment trial consists of an experiment itself and the corresponding outcome value, the sample point. Now we define the multivariate discrete random variable  $X$  as a vector valued function*

$$X : \mathbb{N} \rightarrow \Omega$$

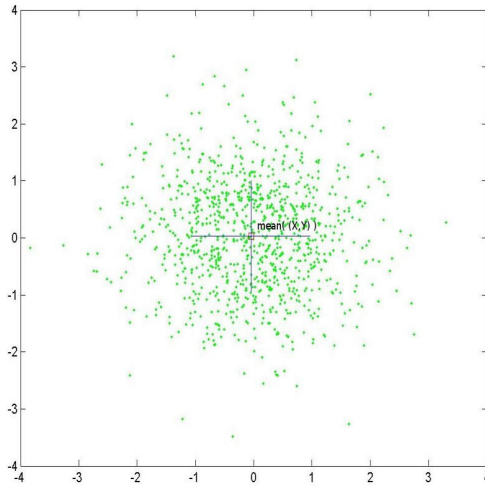
Simply speaking a discrete multivariate random variable  $X = (X_1, X_2, \dots, X_n)$  is vector of discrete univariate random variables  $X_1, X_2, \dots, X_n$ . For the example of the measurement of a 2D point the measurements in X and Y directions  $(X, Y)$  are univariate random variables. Together they give the position vector  $P = (X, Y)$  which can be interpreted as a two dimensional multivariate random variable - also known as bivariate random variable.

#### 9.3.2.2 Expectation, Mean, Variance and Standard Deviation

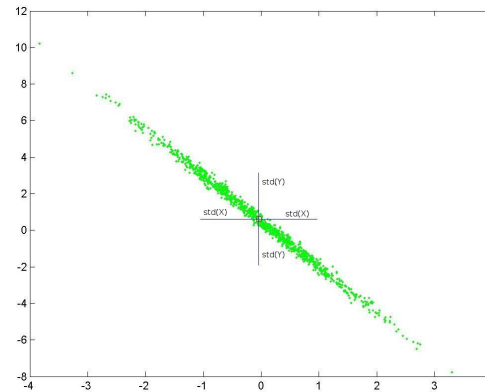
Extending PDFs and their properties to higher dimensions is straight forward. The respective univariate concept is simply applied component wise. For the *expectation*, the *variance*

and the *standard deviation* we get respectively

$$E[X] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}, \text{Var}(X) = \begin{bmatrix} \text{Var}(X_1) \\ \text{Var}(X_2) \\ \vdots \\ \text{Var}(X_n) \end{bmatrix}, \sigma_X = \begin{bmatrix} \sigma_{X_1} \\ \sigma_{X_2} \\ \vdots \\ \sigma_{X_n} \end{bmatrix}$$



(f) 2D random variable. The mean is visualized at the center of the hair cross. The X and Y components of the variable are independent



(g) 2D random variable. The standard deviation is visualized as hair cross. The X and Y components of the variable are dependent

To come back to the example of the measurement of a 2D point. In this context the mean is the *centroid* (mass center) of the cloud of the measured point. It would be reasonable to expect that it represents the true measured point the best. The variance expresses the variation of the measured points from the mean in the directions of the coordinate axes. It tells us, how confident our measurement is. The greater the variance, the larger are the measurement errors that we were making.

### 9.3.2.3 Covariance

In the previous section we introduced the covariance of two univariate random variables  $Cov(X, Y)$ . Now we will extend this to one multivariate random variable  $X = (X_1, X_2, \dots, X_n)$  instead by introducing the so called *covariance matrix*. The covariance will help us to judge whether the components of  $X$ , i.e. the univariate random variables  $X_i$ , are dependent on each other. In order to do that we will compute the covariance for every pair of  $X_i$  and  $X_j$ . Then, in order to handle these covariance in a convenient way, we will put them all into a matrix.

**Definition 9.24 (Covariance Matrix)** Let  $X = (X_1, X_2, \dots, X_n)$  be a multivariate random variable. The corresponding covariance matrix is then

$$COV(X) = \begin{bmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Cov(X_2, X_2) & \dots & Cov(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & Cov(X_n, X_n) \end{bmatrix}$$

We can build a *correlation matrix* in the same way as the covariance matrix.

**Definition 9.25 (Correlation Matrix)** Let  $X = (X_1, X_2, \dots, X_n)$  be a multivariate random variable. The corresponding correlation matrix is then

$$COR(X) = \begin{bmatrix} Cor(X_1, X_1) & Cor(X_1, X_2) & \dots & Cor(X_1, X_n) \\ Cor(X_2, X_1) & Cor(X_2, X_2) & \dots & Cor(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ Cor(X_n, X_1) & Cor(X_n, X_2) & \dots & Cor(X_n, X_n) \end{bmatrix}$$

**Example 9.26** Given the multivariate variable  $A = (X, Y, Z)$ . Let  $X, Y$  and  $Z$  be  $n$ -dimensional vectors, i.e. univariate random variables containing data of  $n$  experiment trials. Let  $\mu_X, \mu_Y$  and  $\mu_Z$  be the respective means. Further we assume that we have a uniform PDF, so that  $P(a, b) = \frac{1}{n}$  for all sample points  $(a, b)$ . Then the covariance matrix of  $A$  would look like this:

$$COV(A) = \frac{1}{n} \sum_{k=1}^n \begin{bmatrix} (X_k - \mu_X)(X_k - \mu_X) & (X_k - \mu_X)(Y_k - \mu_Y) & (X_k - \mu_X)(Z_k - \mu_Z) \\ (Y_k - \mu_Y)(X_k - \mu_X) & (Y_k - \mu_Y)(Y_k - \mu_Y) & (Y_k - \mu_Y)(Z_k - \mu_Z) \\ (Z_k - \mu_Z)(X_k - \mu_X) & (Z_k - \mu_Z)(Y_k - \mu_Y) & (Z_k - \mu_Z)(Z_k - \mu_Z) \end{bmatrix}$$

In the following we briefly sum up some useful properties of the covariance matrix.

- The covariance matrix is symmetric.
- The diagonal entries  $c_{ii}$  are the variances of the variable  $x^{(i)}$ .
- The covariance matrix entry  $COV(X)_{ij}$  represents the dependency between the variables  $X^{(i)}$  and  $X^{(j)}$ . If there is no dependency, then for normalized random variables  $X^{(i)}$  the diagonal entries should be 1 since this is exactly the variance of  $X^{(i)}$  and all other entries should be 0.
- We can also define an ellipsoid that contains the random variables with a certain probability. Then the eigenvectors of the covariance matrix represent the axes of the ellipsoid and the corresponding eigenvalues are the respective scalings.

### 9.3.3 References

Here are some references for the topic.

- Feller [2]
- Press et al. [5]

- <http://mathworld.wolfram.com>
- Hartley and Zisserman [3]
- Andrew W. Moore “Probability Densities in Data Mining”, [www.cs.cmu.edu/~awm/tutorials](http://www.cs.cmu.edu/~awm/tutorials)
- Schickinger and Steger [7]

### 9.3.4 MatLab Commands

Here comes a short list with useful MatLab commands in this setting.

MatLab Command	Usage
<code>X = randn([m n])</code>	Creates an $m$ by $n$ matrix $X$ with normally distributed random entries. Can be used to generate a multivariate random variable (e.g. $m$ -dimensional variable with $n$ measurements).
<code>X = rand([m n])</code>	Creates uniformly distributed entries within range 0-1.
<code>mean(X)</code>	Mean of the random variable $X$ .
<code>var(X)</code>	Variance of the random variable $X$ .
<code>std(X)</code>	Standard deviation of the random variable $X$ .
<code>cov(X)</code>	Covariance of the random variable $X$ .