CE 211
Mathematics for Engineers

Lecture 6

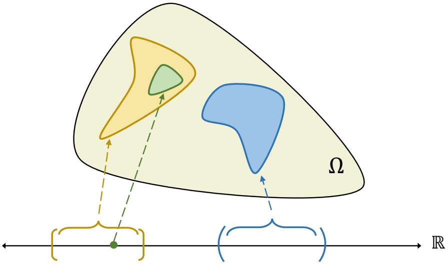
Multivariate Random Variables

Previously on Mathematics for Engineers

A random variable is an alternate way of constructing events. Defining random variables allows us to translate events of interest into probabilities more easily.

Definition (Random Variable)

A real-valued random variable is a function or mapping $X : \Omega \to \mathbb{R}$ such that for all $S \subset \mathbb{R}$, $X^{-1}(S) \in \mathcal{F}$.



*Technically, there are some restrictions on S just like valid events, but we'll ignore those details.

Previously on Mathematics for Engineers

Note the probability measure is a function $\mathbb{P} : \mathcal{F} \to [0,1]$ where you can think of $\mathcal{F}$ as 2^{Ω}, whereas the random variable X is another function $X : \Omega \to \mathbb{R}$.

As seen in the previous examples, for subsets $S \subset \mathbb{R}$, we can find an event $A \in \mathcal{F}$ such that $X^{-1}(S) = A = \{\omega \in \Omega | X(\omega) \in S\}$.

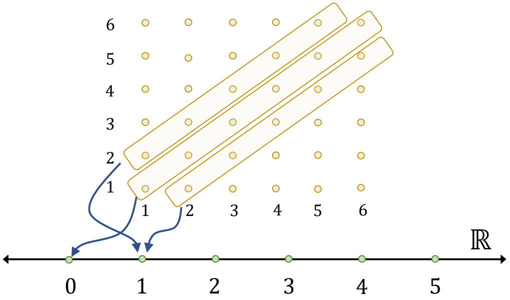
Hence, the following probabilities are the same

$$\mathbb{P}(X \in S) = \mathbb{P}(X^{-1}(S)) = \mathbb{P}(A) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in S\})$$

Be careful to not write $X(A)$ and $\mathbb{P}(S)$, where $A \in \mathcal{F}$ and $S \subset \mathbb{R}$ (unless of course $\Omega = \mathbb{R}$).

Previously on Mathematics for Engineers

Instead, if we define X as the absolute value of difference in the numbers on the dice. What is the event corresponding to $X = 7$? $X = 1$?



Previously on Mathematics for Engineers

Definition (Probability Mass Function)

The probability mass function (PMF) of a random variable X represents the probability of each outcome. It is denoted as $p_X(x)$ and is defined as

$$p_X(x) = \mathbb{P}(X = x)$$

Definition (Cumulative Distribution Function)

The cumulative distribution function (CDF) of a random variable X is denoted by $F_X(x)$ and is defined as

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x' \leq x} p_X(x')$$

Previously on Mathematics for Engineers

Definition (Expectation)

The expected value of a random variable X is denoted by $\mathbb{E}(X)$ or μ_X and is defined as

$$\mathbb{E}(X) = \sum_{x \in R_X} x p_X(x)$$

Definition (Variance)

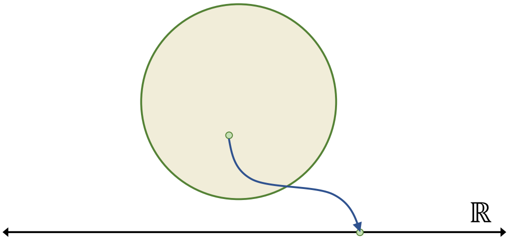
The variance of a random variable X is denoted by $V(X)$, $\text{Var}(X)$, or σ_X^2 and is defined as

$$V(X) = \mathbb{E}\left((X - \mu_X)^2\right) = \sum_{x \in R_X} (x - \mu_X)^2 p_X(x)$$

Previously on Mathematics for Engineers

Continuous random variables are ones which have uncountable images. (Their domain will hence be an uncountable sample space.)

For example, X could represent the location of a randomly thrown dart on the interval $[0,1]$ in which case it can be written as $X : [0,1] \to \mathbb{R}$ or on a two-dimensional circle of some radius, i.e., $X : C \to \mathbb{R}$, where $C = \{(x,y) | x^2 + y^2 \leq r\}$ etc.



Previously on Mathematics for Engineers

Definition (Probability Density Function)

The probability density function (PDF) of a continuous random variable is denoted as $f_X(x)$ and is defined as

$$f_X(x)dx = \mathbb{P}\big(X \in [x, x + dx]\big)$$

Thus, the probability that the random variable lies in a subset S is given by

$$\mathbb{P}(X \in S) = \int_{x \in S} f_X(x)dx$$

Definition (Cumulative Distribution Function)

The cumulative distribution function (CDF) of a random variable X is denoted by $F_X(x)$ and is defined as

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(x)dx$$

Previously on Mathematics for Engineers

Definition (Expectation)

The expected value of a random variable X is denoted by $\mathbb{E}(X)$ or μ_X and is defined as

$$\mathbb{E}(X) = \int_{x \in R_X} x f_X(x) dx$$

Definition (Variance)

The variance of a random variable X is denoted by $V(X)$, $\text{Var}(X)$, or σ_X^2 and is defined as

$$V(X) = \mathbb{E}\left((X - \mu_X)^2\right) = \int_{x \in R_X} (x - \mu_X)^2 f_X(x) dx$$

Lecture Outline

Lecture Outline

The Problem

The Problem

Buffon's Needle Experiment

Consider a floor with parallel strips of width D units. Needles of length L are randomly dropped on the floor. What is the probability that the needle will cross the line between two strips.

www.randomservices.org/random/apps/BuffonNeedleExperiment.html

In 1901, a mathematician Mario Lazzarini actually tossed 3408 needs to find an approximation of π to six decimal places.

Lecture Outline

Jointly Distributed Random Variables

Consider two random variables $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$. Suppose we are interested in the probability that

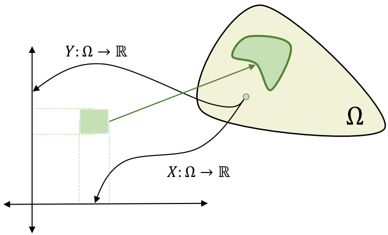
$$\mathbb{P}((X, Y) \in S)$$

We can compute the probability of such events by extending the concepts of PDF and CDF to situations involving two or more random variables *on the same sample space*.

Jointly Distributed Random Variables

Introduction

Note that finding the probability of $\mathbb{P}((X, Y) \in S)$ is different and more general than finding $\mathbb{P}(X \in S, Y \in T)$. Graphically, these imply



Note that $\mathbb{P}(X \in S, Y \in T)$ is not necessarily $\mathbb{P}(X \in S)\mathbb{P}(Y \in T)$. For example, imagine an experiment that involves picking a person from a population and measuring their height and weight.

The comma in the following notation is treated as *and*, i.e., the intersection of the events associated with $X = x$ and $Y = y$.

Definition

If X and Y are two random variables, the joint PMF is given by

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$$

- The probability that $(X, Y) \in S$ is simply $\sum\sum_{(x,y)\in S} p_{X,Y}(x, y)$.
- For the joint PMF to be valid $\sum_{x\in R_X} \sum_{y\in R_Y} p_{X,Y}(x, y) = 1$.

Let us look at two motivating examples, one for the discrete case and the other for continuous state spaces.

Jointly Distributed Random Variables

Discrete State Space

Suppose that 15% of the families in a certain community have no children, 20% have 1 child, 35% have 2 children, and 30% have 3 children. Suppose further that in each family each child is equally likely (independently) to be a boy or a girl.

For a family is chosen at random from this community, let X and Y be the number of boys and number of girls. What is the probability that $X = x$ and $Y = y$.

Both random variables X and Y can take values 0, 1, 2, or 3. The probability that the family has say x boys and y girls is

$$
\begin{aligned}
\mathbb{P}(X = x, Y = y) = &\mathbb{P}(X = x, Y = y | \text{no children})\mathbb{P}(\text{no children}) + \\
&\mathbb{P}(X = x, Y = y | \text{1 child})\mathbb{P}(\text{1 child}) + \\
&\mathbb{P}(X = x, Y = y | \text{2 children})\mathbb{P}(\text{2 children}) + \\
&\mathbb{P}(X = x, Y = y | \text{3 children})\mathbb{P}(\text{3 children})
\end{aligned}
$$

What are the answers for $X = 1, Y = 1$? $X = 2, Y = 1$?

Jointly Distributed Random Variables

Discrete State Space

x↓ y →	0	1	2	3	$\mathbb{P}(X = x)$
0	0.15	0.10	0.0875	0.0375	0.375
1	0.10	0.175	0.1125	0	0.3875
2	0.0875	0.1125	0	0	0.2
3	0.0375	0	0	0	0.0375
$\mathbb{P}(Y = y)$	0.375	0.3875	0.2	0.0375	

The row and column sums have a special significance but will they always be same?

Jointly Distributed Random Variables

Marginal Distributions

If we add the rows, we get the probability distribution of X, which is said to be the **marginal PMF** of X.

$$p_X(x) = \sum_y p_{XY}(x, y)$$

This follows directly from the law of total probability. Suppose we want the probability that $X = x$. Call this event B. We can condition it on events A_1, A_2 etc. associated with $Y = y_1$, $Y = y_2$, and so on respectively.

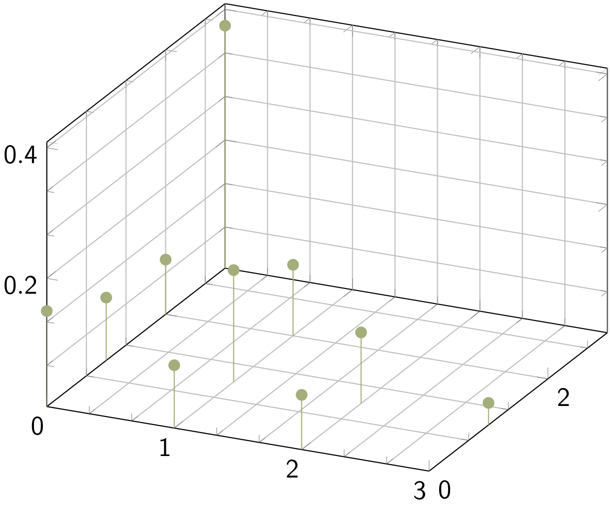
$$\mathbb{P}(B) = \mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B) + \dots$$

Likewise, if we add elements in each column, we get the probability distribution of Y, which is the marginal PMF of Y.

$$p_Y(y) = \sum_x p_{XY}(x, y)$$

Jointly Distributed Random Variables

Visualizing Joint PMF



Jointly Distributed Random Variables

Definition

The CDF of jointly distributed random variable is defined as

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$

x / y	0	1	2	3
0	0.15	0.10	0.0875	0.0375
1	0.10	0.175	0.1125	0
2	0.0875	0.1125	0	0
3	0.0375	0	0	0

x / y	0	1	2	3
0	0.15	0.25	0.3375	0.375
1	0.25	0.525	0.725	0.7625
2	0.3375	0.725	0.925	0.9625
3	0.375	0.7625	0.9625	1.0

Remember that expectation is a scalar quantity. When dealing with multiple random variables, we define the expectation of functions of the random variables.

Definition (Expectation)

Suppose X and Y are random variables with joint PMF $p_{X,Y}(x, y)$. Then, the expectation of $g(X, Y)$ is defined as

$$\mathbb{E}(g(X, Y)) = \sum_{x \in R_X} \sum_{y \in R_Y} g(x, y) p_{X,Y}(x, y)$$

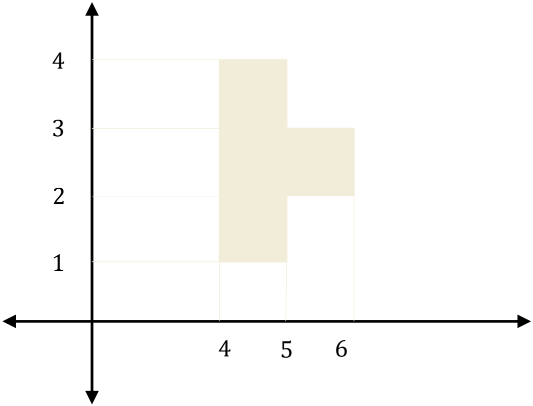
The interpretation is similar. If we perform this experiment repeatedly, observe the realizations of the random variables, estimate $g(x, y)$, and take its average over all repetitions, we get the expected value.

In the earlier example, what is the average number of children in the household? $\mathbb{E}(X + Y)$

Jointly Distributed Random Variables

Continuous State Spaces

Consider a uniformly distributed random variable over a 2D region shown below. Imagine you throw darts at this shape but in a random manner. If X denotes the x coordinate of the dart, what is the probability of $X \in [x, x + dx]$, $Y \in [y, y + dx]$? What is the joint PDF of (X, Y)



Definition

If X and Y are two random variables, the joint PDF is given by

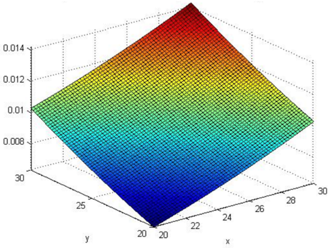
$$f_{X,Y}(x, y)dydx = \mathbb{P}(X \in [x, x + dx], Y \in [y, y + dy])$$

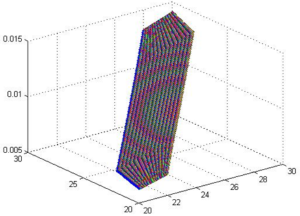
When clear from the context, we write $f_{X,Y}(x, y)$ as $f(x, y)$.

The probability that $(X, Y) \in S$ is thus given by

$$\int\int_{(x,y)\in S} f(x, y)dydx$$

Here is an example of a joint PDF $f(x,y) = k(x^2 + y^2)$, if $x \in [20, 30]$ and $y \in [20, 30]$ and is 0 otherwise.





The right panel shows the region above the event that the magnitude of the difference between the random variables is at most 2.

Jointly Distributed Random Variables

Joint CDF

The cumulative distribution function is defined just as done in the discrete case

Definition

The CDF of a jointly distributed continuous random variable is

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)dydx$$

The CDF must also satisfy

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Can you sketch the CDF for the 2D dart example? How about a 2D dart example on a unit square?

Jointly Distributed Random Variables

Marginal Densities and Expectation

We can also define **marginal densities** of X and Y by replacing sums with integrals. The marginal distribution of X is

$$\mathbb{P}(X \in [x, x + dx]) = \left(\int_y f(x, y) dy \right) dx$$

The marginal distribution of Y is

$$\mathbb{P}(Y \in [y, y + dy]) = \left(\int_x f(x, y) dx \right) dy$$

What are the marginal densities of X and Y in the 2D dart example? Be careful with the limits of integration!

Definition (Expectation)

Suppose X and Y are random variables with joint PDF $f_{X,Y}(x, y)$. Then, the expectation of $g(X, Y)$ is defined as

$$\mathbb{E}(g(X, Y)) = \int_{x \in R_X} \int_{y \in R_Y} g(x, y) f_{X,Y}(x, y) dx dy$$

Jointly Distributed Random Variables

Quickercise

Consider two random variables X and Y representing the time to failure of hard disk and processor of a computer respectively. If their joint PDF is given by $f(x,y) = 2e^{-x}e^{-2y}$, what is the probability that the hard disk fails before the processor?

Conditional Random Variables and Independence

The idea of conditional probability of events can be extended similarly to random variables to define conditional random variables and their PMFs and PDFs.

For example, in the discrete setting, we can define the PMF of a conditional random variable as $p_{X|Y}(x|y)$ as

$$p_{X|Y}(x|y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

That is, we are conditioning X with events associated with $Y = y$.

Likewise, in the continuous setting, we can write

$$f_{X|Y}(x|y) = \frac{\mathbb{P}(X \in [x, x + dx], Y \in [y, y + dy])}{\mathbb{P}(Y \in [y, y + dy])} = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Conditional Random Variables and Independence

Independence

Definition

Two random variables X and Y are said to be independent if $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ or $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

In other words, knowing the value of realization of one of the random variables does not affect the density of the other random variable and vice versa.

We can also define independence using CDFs, i.e., X and Y are independent if

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

- Suppose we perform 5 coin tosses. If X is the number of heads in the first 3 coin tosses, and Y is the number of heads in the next 2 coin tosses. (Why? Can you construct the joint and marginal PMFs)
- If X and Y have a joint PDF $f(x, y) = 6e^{-2x}e^{-3y}$, where $x \geq 0$ and $y \geq 0$. Are they independent?
- If two random variables X and Y are independent, show that

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

More generally, $\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$.

Lecture Outline

Covariance and Correlation

Covariance and Correlation

Extending Variance

So far we have discussed the expectation of functions of random variables but extending the definition of variance is not straightforward.

Definition (Covariance)

The covariance of two random variables X and Y is denoted by $\text{Cov}(X, Y)$ and is defined as

$$\text{Cov}(X, Y) = \mathbb{E}\Big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\Big)$$

What happens when $X = Y$? The above equation is a bit unwieldy but there is an easier way to compute the covariance.

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Covariance and Correlation

Extending Variance

Unlike variance, covariance can be positive or negative.

Positive covariance implies that as when X is large, Y is likely to be large as well and vice versa. Negative covariance implies that when X is large, Y tends to be small.

What is $\text{Cov}(X, Y)$ when X and Y are independent? The converse is not true! It is possible to have dependent random variables with zero covariance.

The sign of covariance very informative, but its magnitude is not is not. To make sense of the magnitude of covariance between random variables, some normalization with the standard deviations of X and Y is needed.

Definition (Correlation)

The correlation coefficient of two random variables X and Y is denoted using $\rho_{X,Y}$ and is defined as

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

It turns out that $\rho_{X,Y} \in [-1, 1]$. A value of 1 indicates a linear relationship with positive slope between X and Y (i.e., $X = aY + b, a > 0$)

Likewise, a value of -1 indicates a linear relationship with negative slope. A value of 0 just implies no linear relationship between X and Y.

Lecture Outline

Multivariate Distribution

The concept of joint PMFs and PDFs can be trivially extended to situations involving more than 2 variables.

The joint PDF of a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ can be written as

$$\begin{aligned}\mathbb{P}(X_1 \in [x_1, x_1 + dx_1], \ldots, X_n \in [x_n, x_n + dx_n]) &= f(x_1, x_2, \ldots, x_n)dx_1 \ldots dx_n \\ &= f(\mathbf{x})dx_1 \ldots dx_n\end{aligned}$$

Thus, the probability that $(X_1, X_2, \ldots, X_n) \in S \subset \mathbb{R}^n$ can be obtained from an n-dimensional integral

$$\int\int \ldots \int_{(x_1, x_2, \ldots, x_n) \in S} f(x_1, x_2, \ldots, x_n)dx_1 dx_2 \ldots dx_n$$

Note: All vectors will be assumed to be column vectors.

Multivariate Distribution

CDF and Expectation

The CDF and expected values can similarly be extended as follows.

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n$$

$$\mathbb{E}(g(\mathbf{X})) = \int_{x_1 \in R_{X_1}} \cdots \int_{x_n \in R_{X_n}} g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

How do we define the marginal densities of $X_1, \ldots, X_n$?

*Note that in defining the CDF we have been writing $F(x) = \int_{-\infty}^{x} f(x)dx$ instead of the more precise $F(x) = \int_{-\infty}^{x} f(t)dt$. Inconvenience caused to the observant and confused student is regretted.

We still define covariances between every pair of random variables in the vector.

$$\text{Cov}(X_i X_j) = \mathbb{E}\Big((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j)\Big)$$

How many such covariances terms can we define? The covariance terms can be represented in a matrix form and is usually denoted as $V(\mathbf{X}) = \mathbf{\Sigma}$. Does this matrix have any obvious property?

Suppose the vector of expected values $(\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_n))$ is written as $\boldsymbol{\mu}$. A shorthand notation for the covariance matrix is

$$\mathbf{\Sigma} = \mathbb{E}\big((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\big) = \mathbb{E}(\mathbf{X}\mathbf{X}^T) - \mathbb{E}(\mathbf{X})\mathbb{E}(\mathbf{X}^T)$$

The properties of variance and expectation that we have seen so far are applicable to vectors of random variables as well!

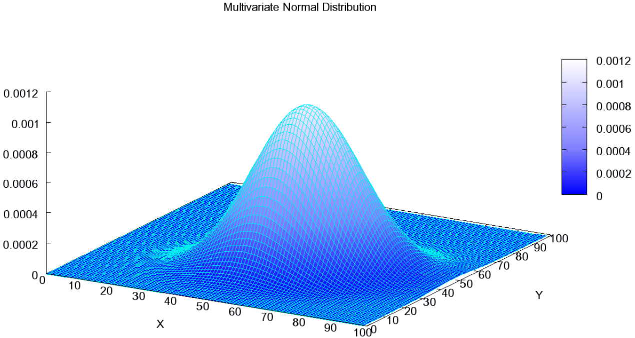
For example, assume **A** is an $n \times n$ matrix and **b** is an $n \times 1$ vector.

- $\mathbb{E}(\mathbf{AX} + \mathbf{b}) = \mathbf{A}\mathbb{E}(\mathbf{X}) + \mathbf{b}$
- $V(\mathbf{AX} + \mathbf{b}) = \mathbf{A}V(\mathbf{X})\mathbf{A}^T = \mathbf{A\Sigma A}^T$

Multivariate Distribution

Normal Distribution

One of the most popular multivariate distribution is the multivariate normal distribution.



If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is an n-dimensional random variable and suppose $\boldsymbol{\Sigma}$ is positive definite, then its joint PDF is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}) \right)$$

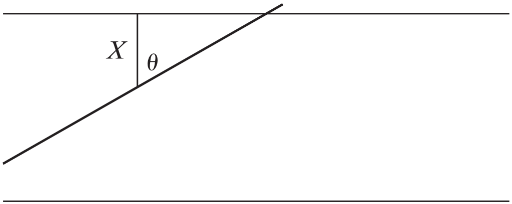
which in the degenerate case of a single random variable takes the form discussed in previous class

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

It can be shown that the marginal densities of the multivariate normal random variable is also normal.

A Solution

Define random variables X as the distance from the mid-point of the needle to the nearest parallel line and θ as the angle between the needle and the perpendicular from the mid-point as shown in the figure



The needle cuts the line if $X < L/2\cos\theta$. Find the probability of this using joint densities. The answer turns out to be $2L/\pi D$.

Your Moment of Zen