# 

Contents

[Random Experiment 1](#_Toc18562581)

[Random Variables 1](#_Toc18562582)

[Conditional Probability 3](#_Toc18562583)

[Chain rue of conditional probability 3](#_Toc18562584)

[Independence 3](#_Toc18562585)

[Conditional Independence 3](#_Toc18562586)

[Expectation 4](#_Toc18562587)

[Probability Distribution 5](#_Toc18562588)

[Joint Probability Distribution 6](#_Toc18562589)

[Probability Distribution Function 9](#_Toc18562590)

[Cumulative Distribution Function 11](#_Toc18562591)

[Properties of CDF 12](#_Toc18562592)

[How to Manipulate among Joint, Conditional and Marginal Probabilities 12](#_Toc18562593)

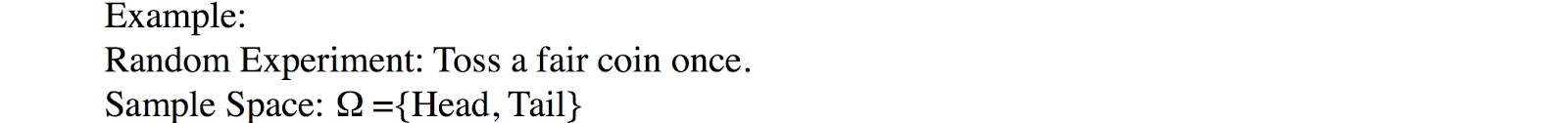
[Bayes’ Theorem 13](#_Toc18562594)

[Bayes’ Theorem Example 13](#_Toc18562595)

# Random Experiment

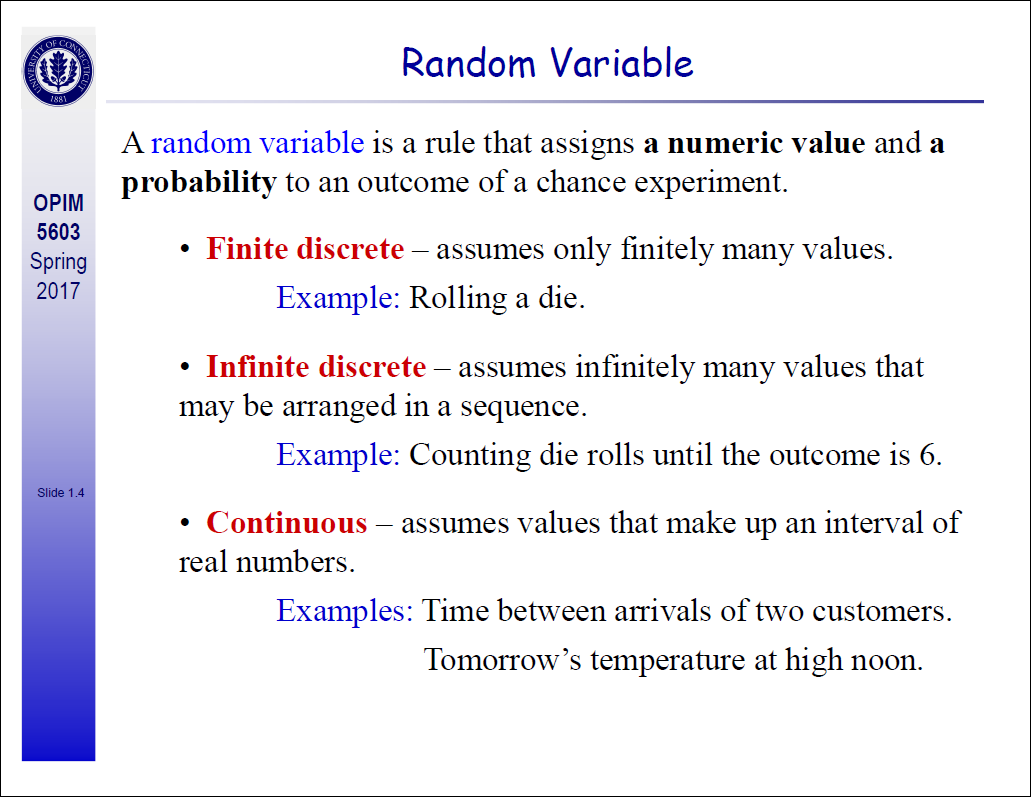
A random experiment is a physical situation whose outcome cannot be predicted until it is observed.

**Sample Space**A sample space, is a set of all possible outcomes of a random experiment.



# Random Variables

 A ***random variable***, is a variable whose possible values are numerical outcomes of a **random experiment**.



**Probability**Probability is the measure of the likelihood that an event will occur in a Random Experiment. Probability is quantified as a number between 0 and 1, where, loosely speaking, 0 indicates impossibility and 1 indicates certainty. The higher the probability of an event, the more likely it is that the event will occur.  
Example  
A simple example is the tossing of a fair (unbiased) coin. Since the coin is fair, the two outcomes (“heads” and “tails”) are both equally probable; the probability of “heads” equals the probability of “tails”; and since no other outcomes are possible, the probability of either “heads” or “tails” is 1/2 (which could also be written as 0.5 or 50%).

**Marginal probability:**

the probability of an event occurring (p(A)), it may be thought of as an unconditional probability.  It is not conditioned on another event.  Example:  the probability that a card drawn is red (p(red) = 0.5).  Another example:  the probability that a card drawn is a 4  (p(four)=1/13).

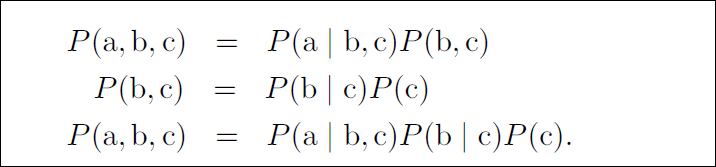
# Conditional Probability

Conditional Probability is a measure of the probability of an event given that (by assumption, presumption, assertion or evidence) another event has already occurred. If the event of interest is A and the event B is known or assumed to have occurred, “the conditional probability of A given B”, is usually written as P(A|B).

Example:  given that you drew a red card, what’s the probability that it’s a four

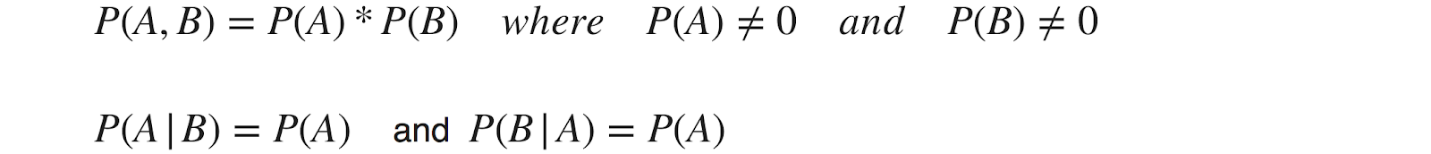
(p(four|red))=2/26=1/13.  So out of the 26 red cards (given a red card), there are two fours so 2/26=1/13.

## Chain rue of conditional probability



# Independence

Two events are said to be independent of each other, if the probability that one event occurs in no way affects the probability of the other event occurring, or in other words if we have observation about one event it doesn’t affect the probability of the other. For Independent events A and B below is true



Example  
Let’s say you rolled a die and flipped a coin. The probability of getting any number face on the die is no way influences the probability of getting a head or a tail on the coin.

# Conditional Independence

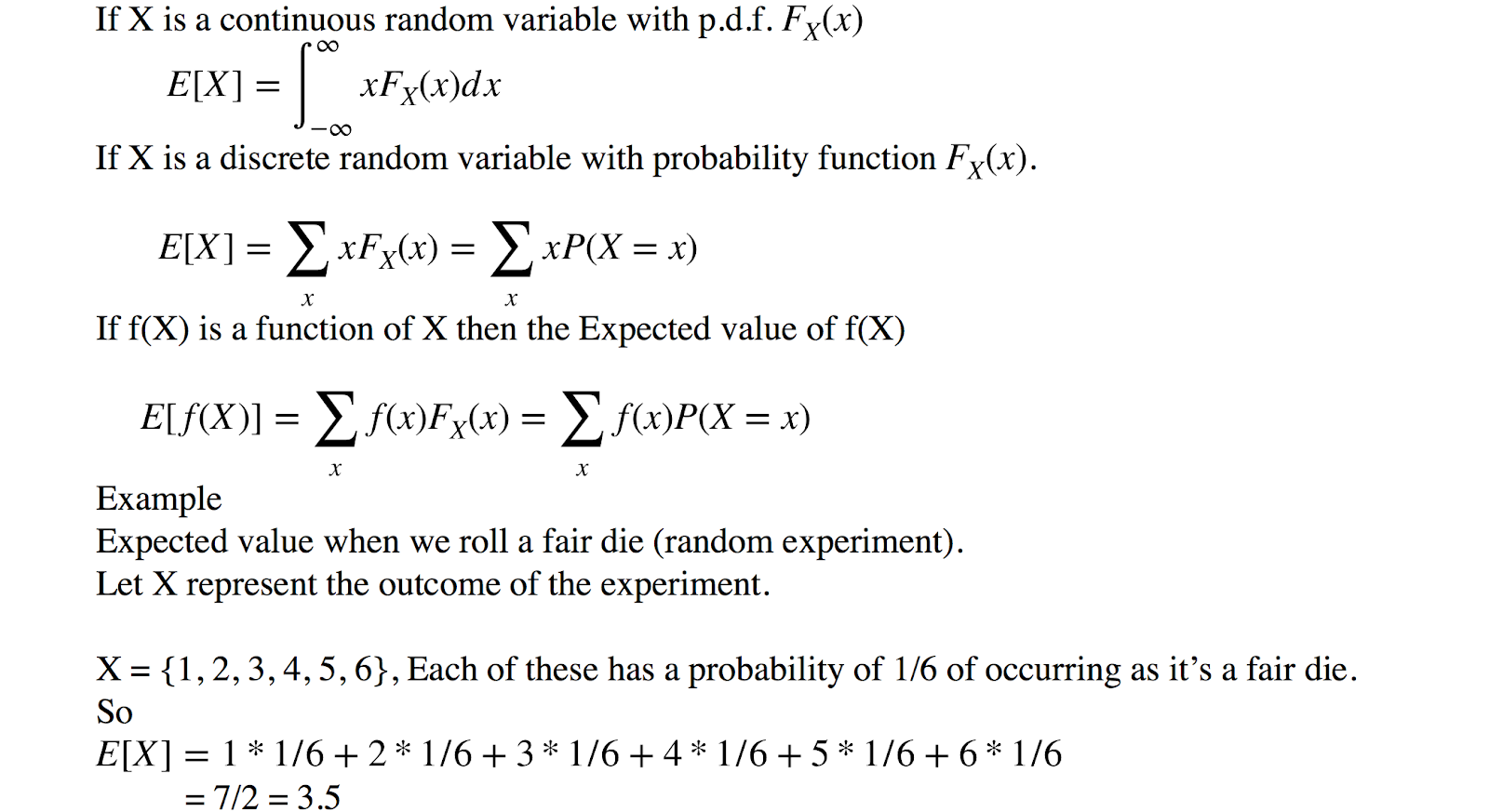
Two events A and B are conditionally independent given a third event C precisely if the occurrence of A and the occurrence of B are independent events in their conditional probability distribution given C. In other words, A and B are conditionally independent given C if and only if, given knowledge that C already occurred, knowledge of whether A occurs provides no additional information on the likelihood of B occurring, and knowledge of whether B occurs provides no additional information on the likelihood of A occurring.



Example  
A box contains two coins, a regular coin and one fake two-headed coin (P(H)=1P(H)=1). I choose a coin at random and toss it twice.   
Let  
A = First coin toss results in an HH.  
B = Second coin toss results in an HH.  
C = Coin 1 (regular) has been selected.  
If C is already observed i.e. we already know whether a regular coin is selected or not, the event A and B becomes independent as the outcome of 1 doesn’t affect the outcome of other event.

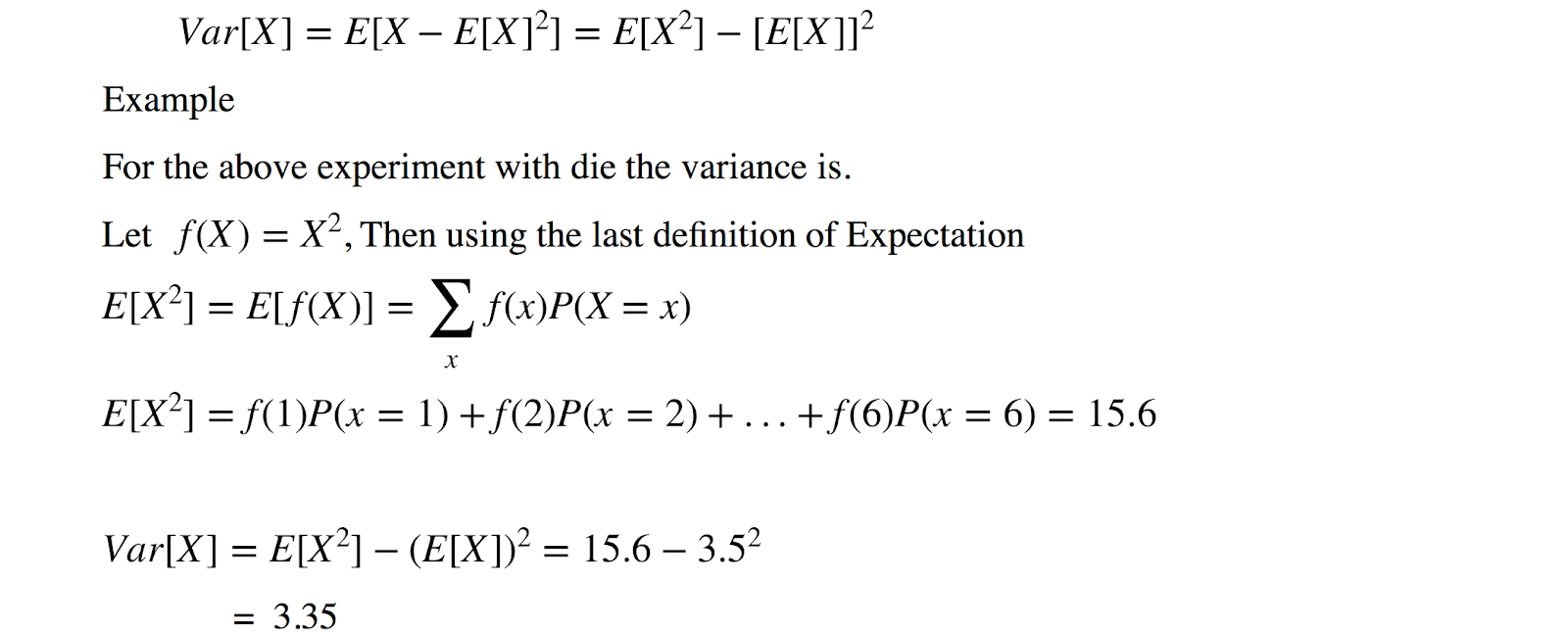
# Expectation

The expectation of a random variable X is written as E(X). If we observe N random values of X, then the mean of the N values will be approximately equal to E(X) for large N. In more concrete terms, the expectation is what you would expect the outcome of an experiment to be on an average if you repeat the experiment a large number of time.



So the expectation is 3.5 . If you think about it, 3.5 is halfway between the possible values the die can take and so this is what you should have expected.

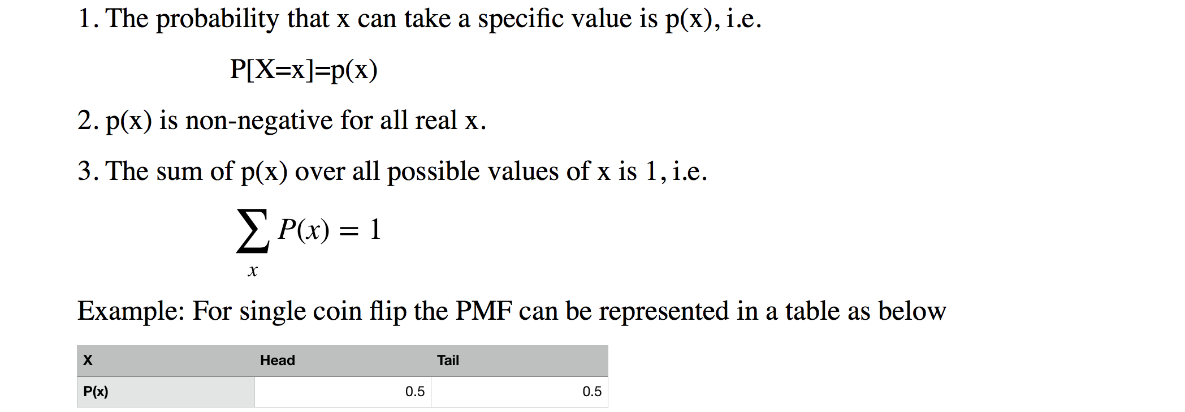
**Variance**The variance of a random variable X is a measure of how concentrated the distribution of a random variable X is around its mean. It’s defined as



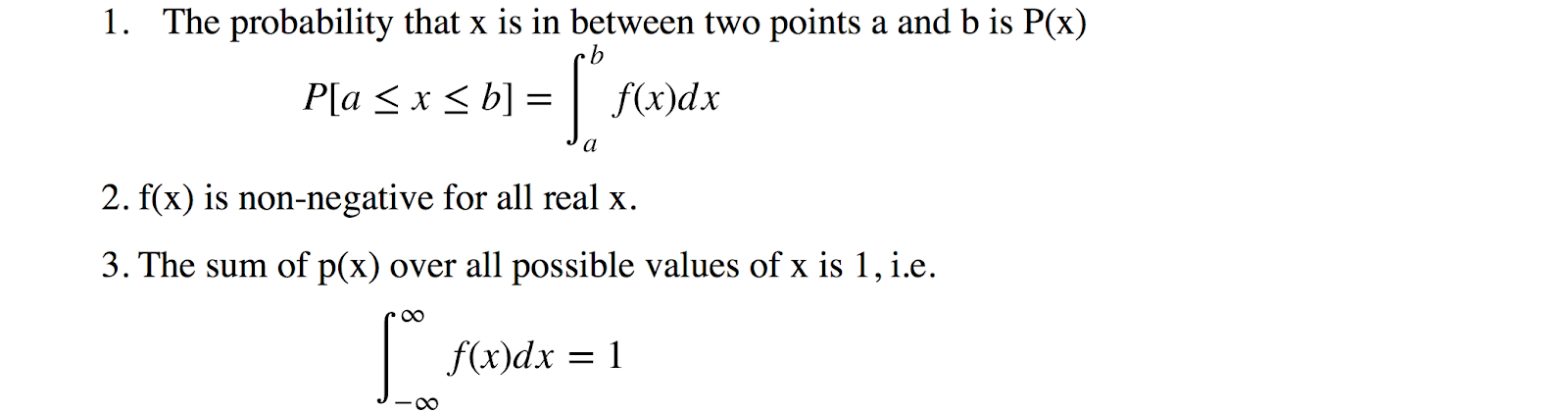
# Probability Distribution

Is a mathematical function that maps the all possible outcomes of an random experiment with it’s associated probability. It depends on the Random Variable X , whether it’s discrete or continues.

1. **Discrete Probability Distribution:**The mathematical definition of a discrete probability function, p(x), is a function that satisfies the following properties. This is referred as **Probability Mass Function.**

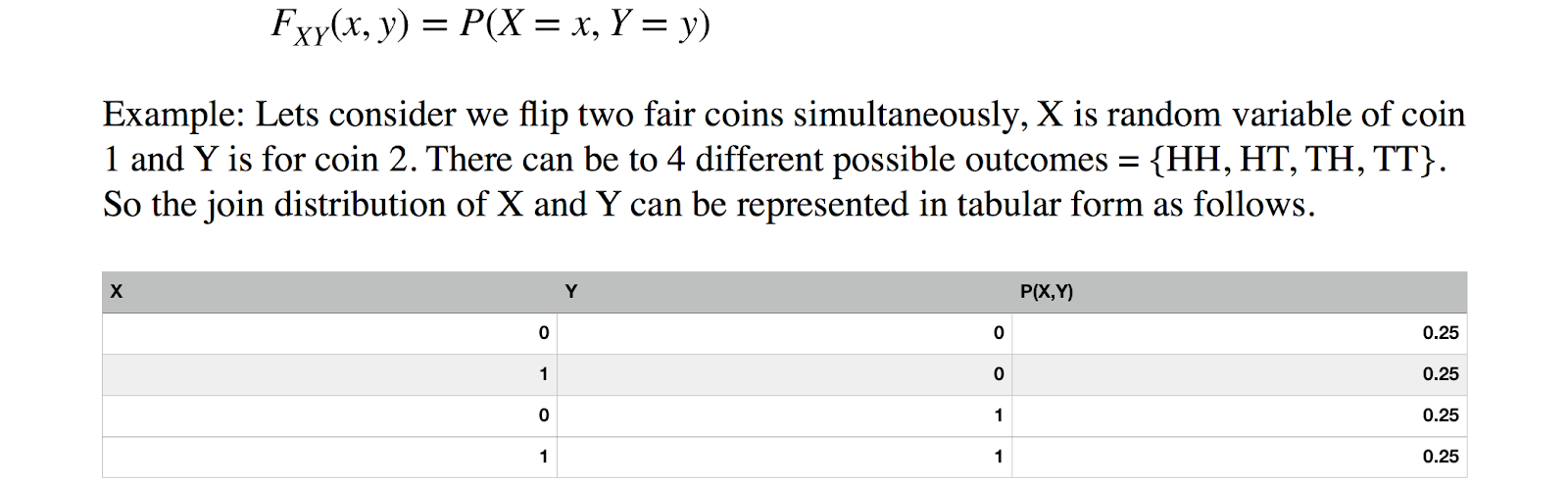


**2. Continuous Probability Distribution:**The mathematical definition of a continuous probability function, f(x), is a function that satisfies the following properties. This is referred as **Probability Density Function**.



# Joint Probability Distribution

If X and Y are two random variables, the probability distribution that defines their simultaneous behaviour during outcomes of a random experiment is called a joint probability distribution. Joint distribution function of X and Y ,defined as

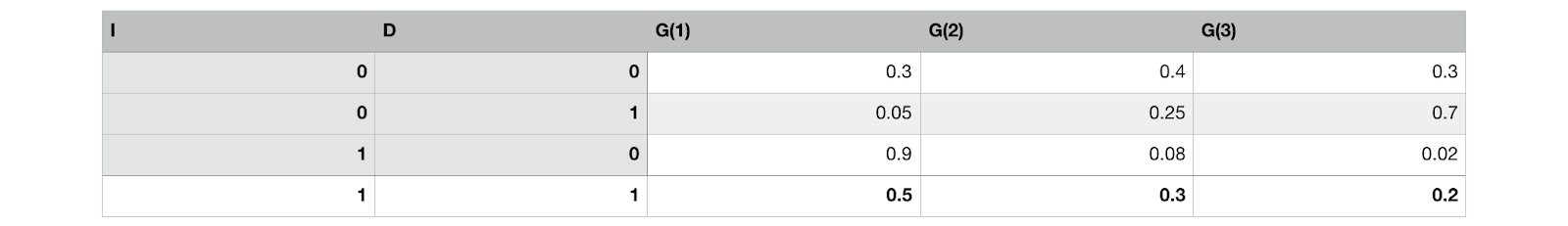


In general if there are n random variables and each can take values v1, v2 … vn different values then there will be total (v1)^n\*(v2)^n\*…(vn)^n rows in the table.

Joint probability**:** p(A and B).  The probability of event A and event B occurring.  It is the probability of the intersection of two or more events.  The probability of the intersection of A and B may be written p(A ∩ B).

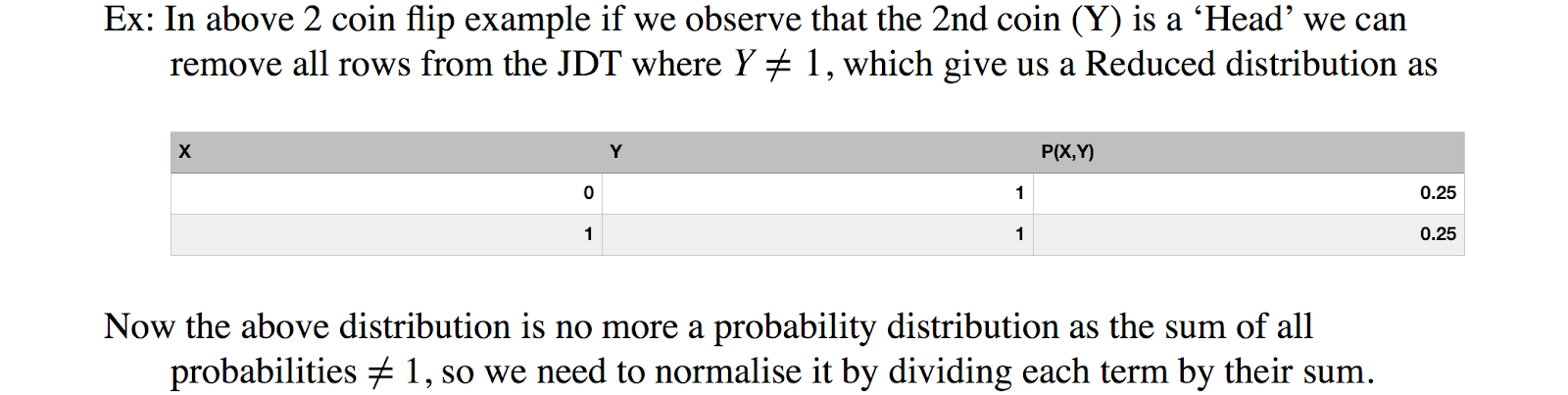
 Example:  the probability that a card is a four and red =p(four and red) = 2/52=1/26.  (There are two red fours in a deck of 52, the 4 of hearts and the 4 of diamonds).

**Conditional Probability Distribution (CPD)**If Z is random variable who is dependent on other variables X and Y, then the distribution of P(Z|X,Y) is called CPD of Z w.r.t X and Y. It means for every possible combination of random variables X, Y we represent a probability distribution over Z.  
**Example**There is a student who has a property called ‘**Intelligence**’ which can be either low(**I\_0**)/high(**I\_1**). He/She enrolls to a course, The course has property called ‘**Difficulty**’ which can take binary values easy(D\_0)/difficult(D\_1). And the student gets a ‘**Grade**’ in the course based on his performance, and grade can take 3 values **G\_1**(Best)/(**G\_2**)/(**G\_3**)(Worst). Then the CPD P(G|I,D) is as follow

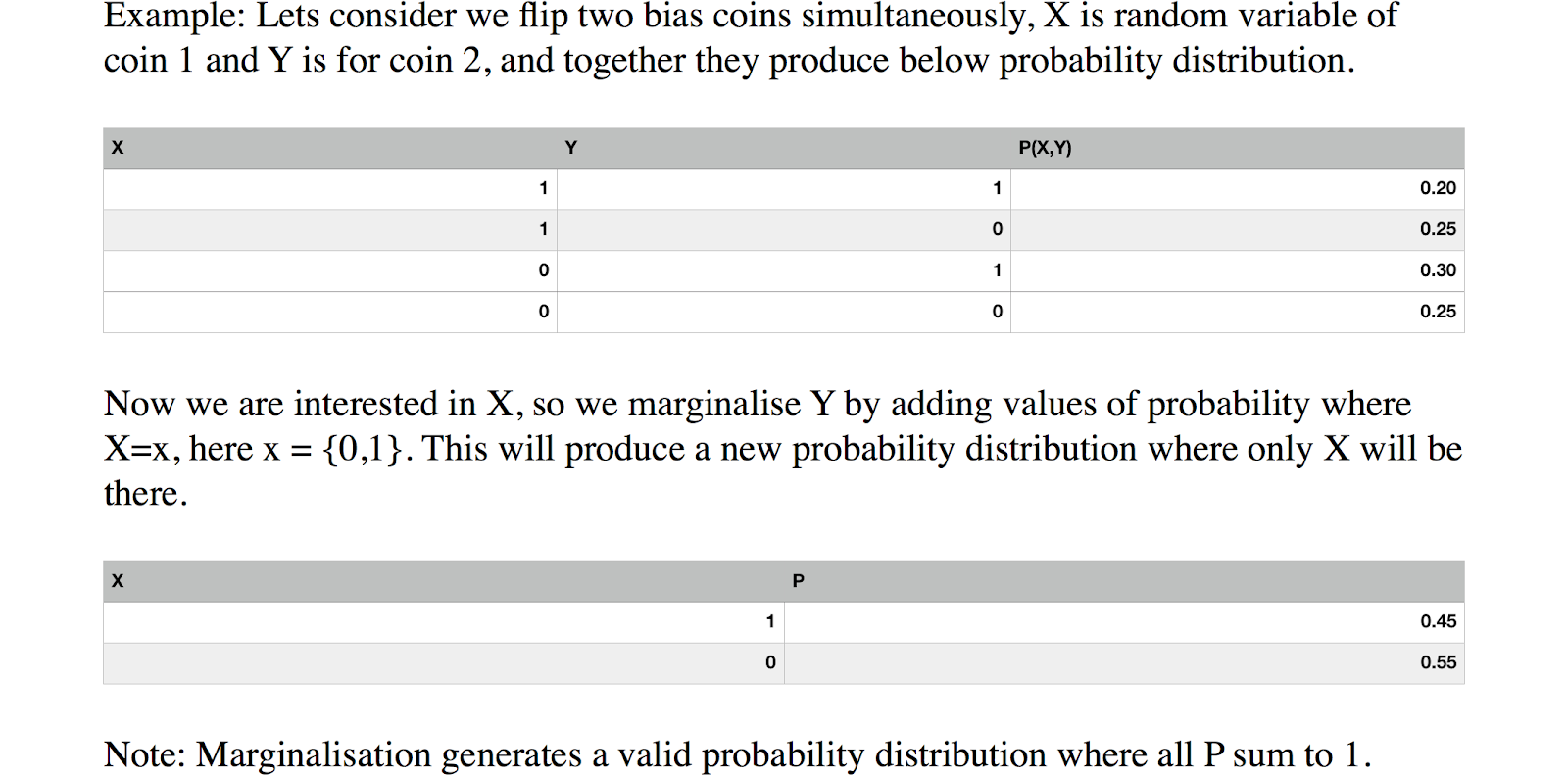


There are a number of operations that one can perform over any probability distribution to get interesting results. Some of the important operations are as below.

**Conditioning/Reduction**If we have a probability distribution of n random variables X1, X2 … Xn and we make an observation about k variables that they acquired certain values a1, a2, …, ak. It means we already know their assignment. Then the rows in the JD which are not consistent with the observation is simply can removed and that leave us with lesser number of rows. This operation is known as Reduction.



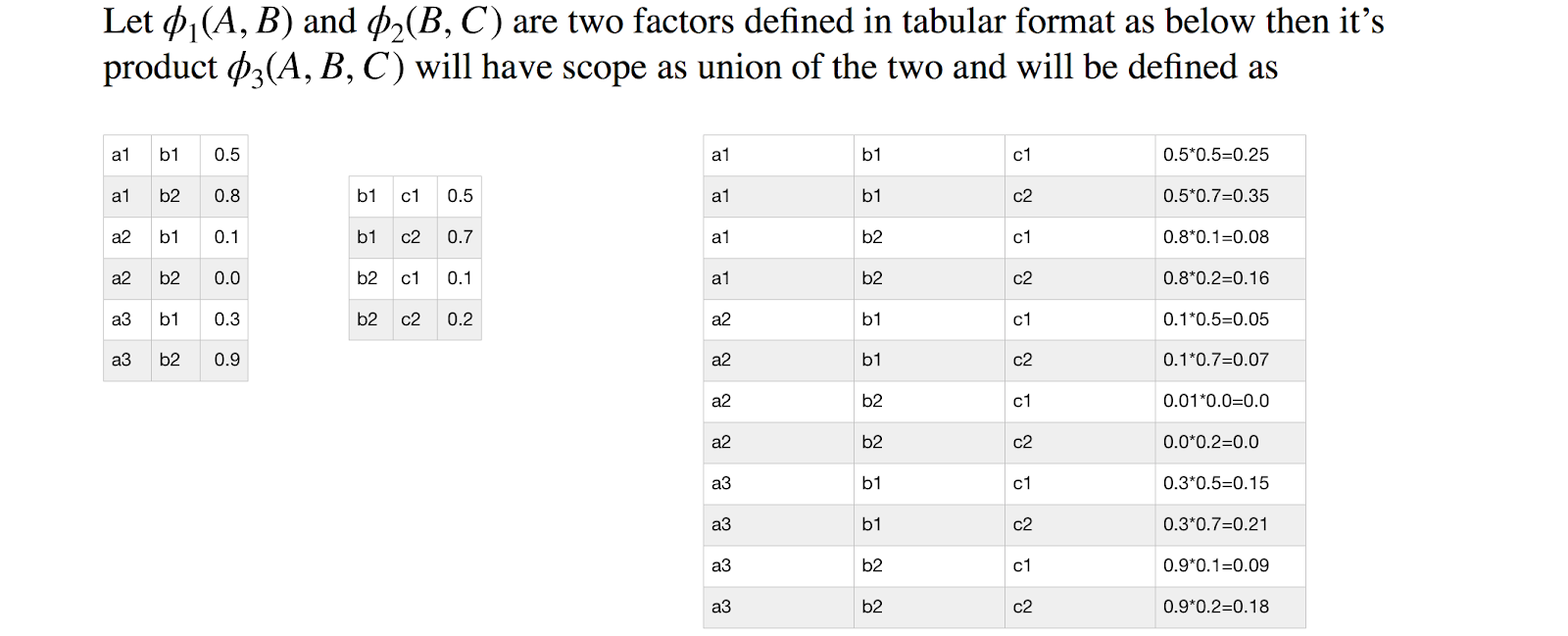
**2. Marginalisation**This operation takes a probability distribution over a large set random variables and produces a probability distribution over a smaller subset of the variables. This operation is known as marginalising a subset of random variables. This operation is very useful when we have large set of random variables as features and we are interested in a smaller set of variables, and how it affects output. For ex.



**Factor**A factor is a function or a table which takes a number of random variables {X\_1, X\_2,…,X\_n} as an argument and produces a real number as a output. The set of input random variables are called scope of the factor. For example Joint probability distribution is a factor which takes all possible combinations of random variables as input and produces a probability value for that set of variables which is a real number. Factors are the fundamental block to represent distributions in high dimensions and it support all basic operations that join distributions can be operated up on like product, reduction and marginalisation.



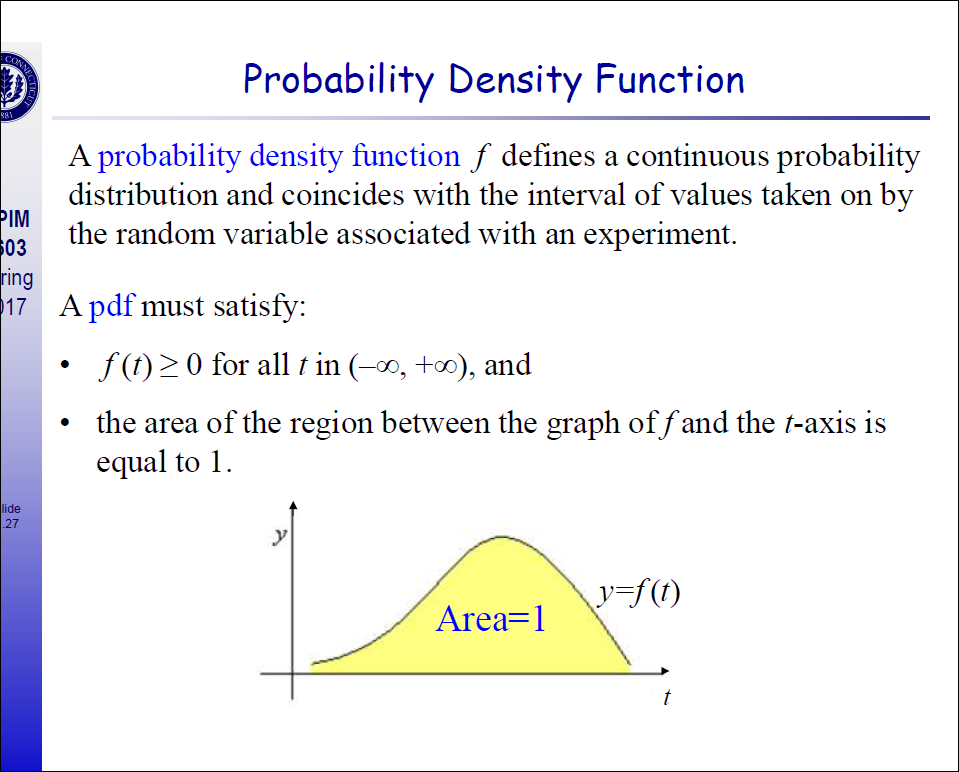
**Factor Product**We can do factor products and the result will also be a factor. For ex

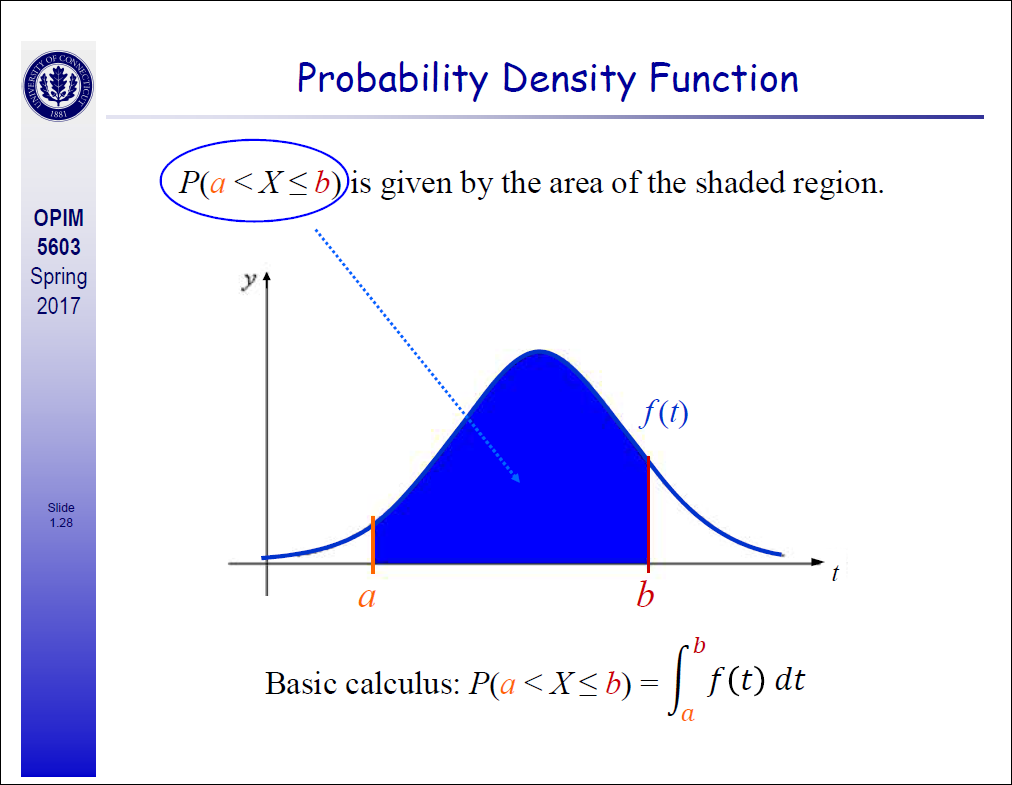


# Probability Distribution Function

The probability that a discrete random variable *X* takes on a particular value *x*, that is, *P*(*X* = *x*), is frequently denoted *f*(*x*). The function *f*(*x*) is typically called the **probability mass function**, although some authors also refer to it as the **probability function**, the **frequency function**, or **probability density function**.  We will use the common terminology — the probability mass function — and its common abbreviation —the **p.m.f.**

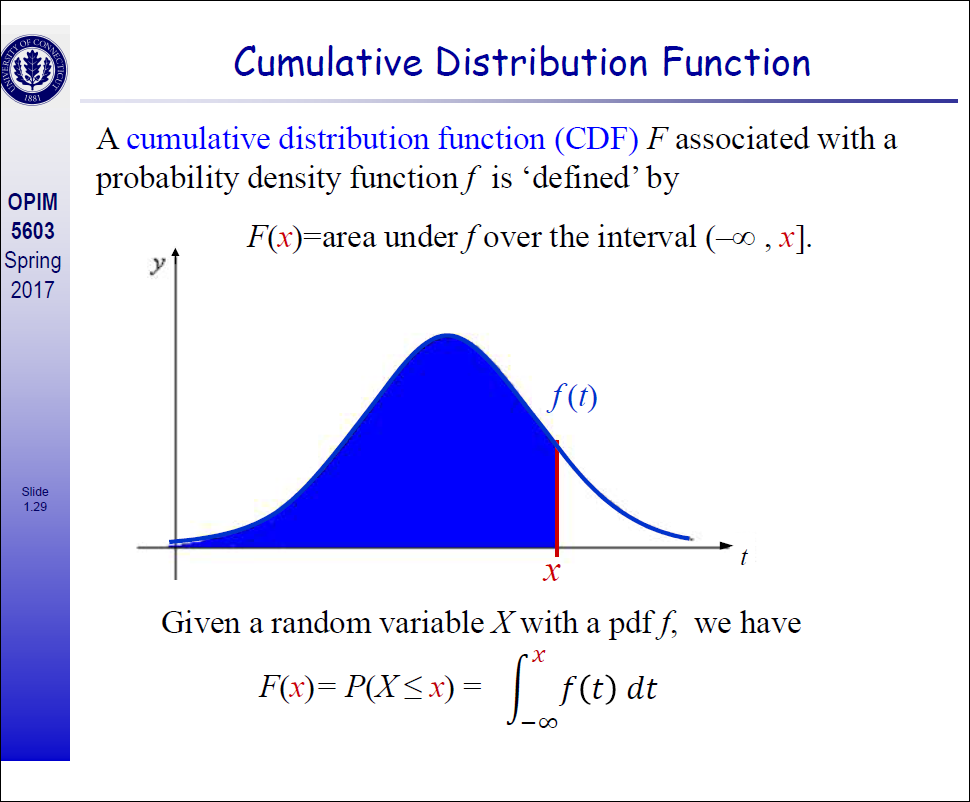
In [probability theory](https://en.wikipedia.org/wiki/Probability_theory), a **probability density function** (**PDF**), or **density** of a [continuous random variable](https://en.wikipedia.org/wiki/Continuous_random_variable), is a [function](https://en.wikipedia.org/wiki/Function_(mathematics)), whose value at any given sample (or point) in the [sample space](https://en.wikipedia.org/wiki/Sample_space) (the set of possible values taken by the random variable) can be interpreted as providing a *relative likelihood* that the value of the random variable would equal that sample



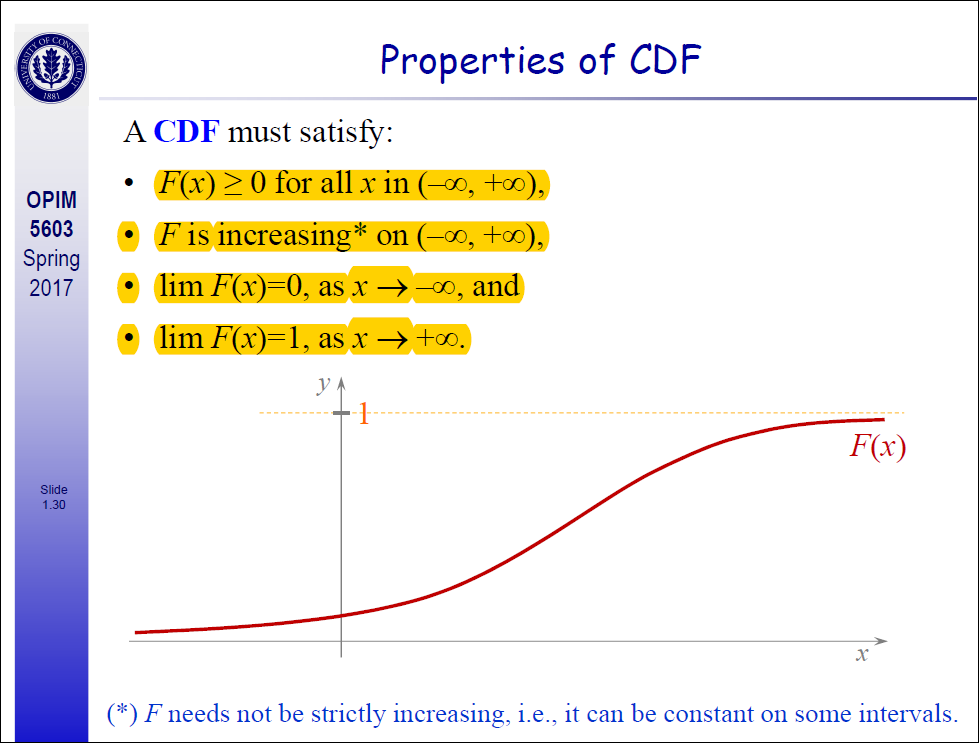


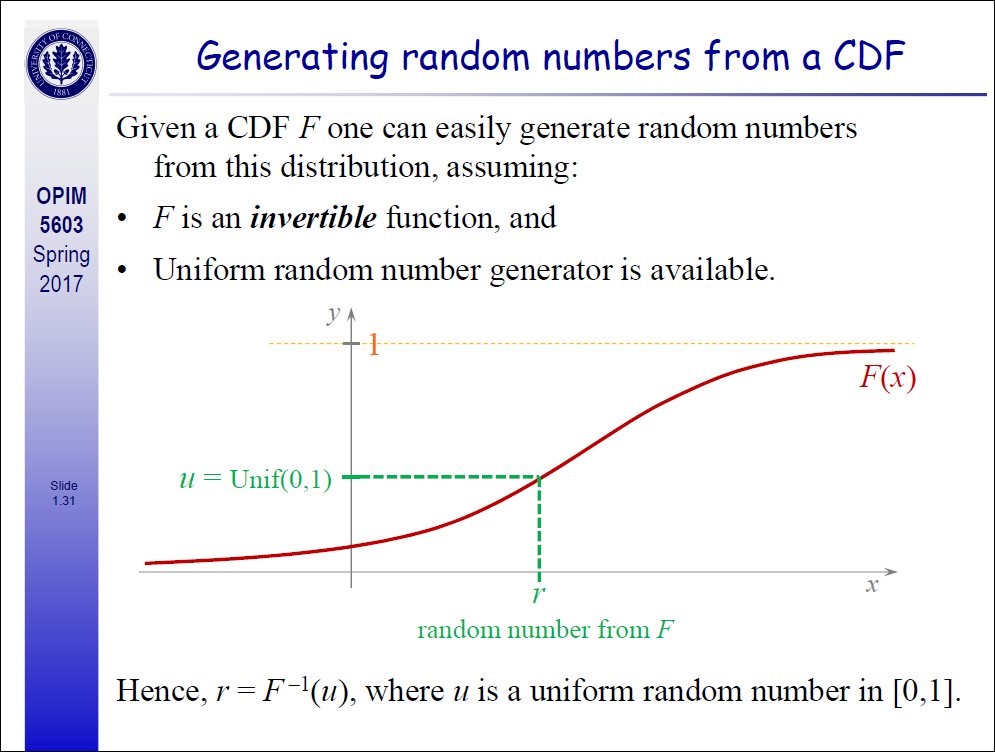
For a one-dimensional continuous random variable, the area under the curve between the points a and b is the probability that random variable is between a and b. The density bewteen a and b is represented by the height.

# Cumulative Distribution Function



## Properties of CDF

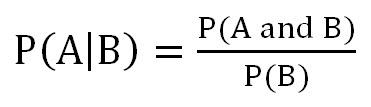




# How to Manipulate among Joint, Conditional and Marginal Probabilities

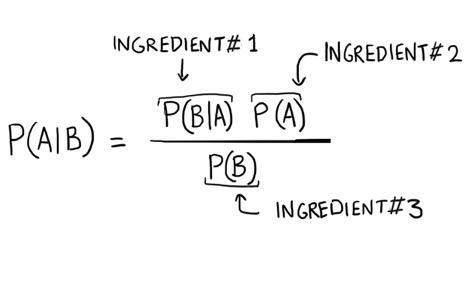
# 

The equation below is a means to manipulate among joint, conditional and marginal probabilities.  As you can see in the equation, the conditional probability of A given B is equal to the joint probability of A and B divided by the marginal of B.  Let’s use our card example to illustrate.  We know that the conditional probability of a four, given a red card equals 2/26 or 1/13.  This should be equivalent to the joint probability of a red and four (2/52 or 1/26) divided by the marginal P(red) = 1/2.  And low and behold, it works!  As 1/13 = 1/26 divided by 1/2.  For the diagnostic exam, you should be able to manipulate among joint, marginal and conditional probabilities.



# Bayes’ Theorem

Bayes’ theorem: an equation that allows us to manipulate conditional probabilities. For two events, A and B, Bayes’ theorem lets us to go from p(B|A) to p(A|B) if we know the  
marginal probabilities of the outcomes of A and the probability of B, given the outcomes of A. Here is the equation for Bayes’ theorem for two events with two possible outcome (A and not A). utcome (A and not A).



# Bayes’ Theorem Example

Q1

**Let’s assume we know that 1% of women over the age of 40 have breast cancer.**

**[p(cancer)=0.01]**

**Let’s assume that 90% of women who have breast cancer will test positive for breast cancer in a mammogram.**

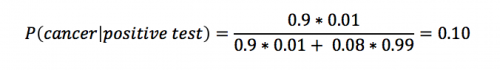
**[p(positive test|cancer)=0.9]**

**Eight percent ofwomen that do NOT have cancer will also test positive.**

**[p(positive test|no cancer)=0.08]**

**What is the probability that a woman has cancer if she tests positive [p(cancer|positive test)]?**

We will call p(cancer) = P(A), and the P(positive test) = P(B). We want to know P(A|B)–the probability of having cancer if you have a positive test.



Using Bayes’ theorem, we calculate that the likelihood that a woman has breast cancer, given a positive test equals approximately 0.10. This makes intuitive sense as (1) this result is greater than 1% (the percent of breast cancer in the general public).