# Learning Bayesian Networks

## Introduction to Bayesian Networks

## Joint Probability Distributions

***Definition 1.1*** Suppose we have sample space containing distinct elements. That is:

.

A function which assigns a real number to each event is called probability function on the set of subsets of if it satisfies the following conditions:

1. for
2. For each event E = {e

***Definition 1.8*** Let a set of random variables be specified such that each has a countably infinite space. A function, that assigns a real number to every combination of values of the ’s such that the value of is chosen from the space of , is called joint probability distribution of the random variables in if it satisfies the following conditions:

1. For every combination of values of the ’s:

1. We have:

Theorem 1.3 Let a set of random variables V be given and let a joint probability distribution of the variables in V be specified according to Definition 1.8. Let be the Cartesian product of the sets of all possible values of the random variables. Assign probabilities to elementary events in as follows:

These assignments result in a probability function on according to ***Definition 1.1***. Furthermore, if we let denote a function (random variable in the classical sense) on this sample space which maps each tuple in to the value of in that tuple, then the joint probability distribution of the ’s is the same as the originally specified joint probability distribution.

### Markov Condition

***Definition 1.9*** Suppose we have a joint probability distribution of the random variables in some set and a DAG = . We say that satisfies *the Markov condition* if for each variable conditionally independent of the set of all of its non-descendants given the set of all its parents. If we denote the sets of parents and non-descendants of by and , respectively, then

When satisfies *the Markov condition*, we say and satisfy Markov condition with each other.

If is a root, then its parent set is empty. So in this case it means that the Markov condition means is independent of - . But implies . We have . So we can rewrite the Markov condition as:

***Theorem 1.4*** If satisfies *the Markov condition*, then P is equal to the product of its conditional distributions of all nodes given values of their parents, whenever these conditional distributions exist.

*Proof*: We prove the case where is discrete. Order the nodes in their ancestral ordering. Let , , … , be the resultant ordering. For a given set of values , , … , let be the subset of these values containing the values of ’s parents. We need to show that whenever for ,

We show this using induction on the variables of the network. Assume for some combination of values of the for

*Induction base* :

Since is empty .

*Induction hypothesis* :

Suppose for this combination of values of the that:

*Induction Step*: Prove assuming that the hypothesis for is true

(1.7)

There are two cases:

*Case 1*: for this combination of values:

(1.8)

Clearly (1.8) implies

Furthermore, due to (1.8) and the induction hypothesis, there is some , where such that . So (1.7) holds.

Case 2: For this combination of values:

In this case

Theorem 1.5 Let a DAG be given in which each node is a random variable, and let a discrete conditional probability distribution