# 02-680 Module 16
## Essentials of Mathematics and Statistics

### Aidan Jan

### November 18, 2025

## Graphical Representations of Random Variables

### Bayesian Networks

Sometimes, we need a more complicated conditional description. One way to do this is using a network (a graph). Bayesian networks decompose **joint distributions** into a product of simpler conditional probabilities, capturing variable dependencies and enabling more efficient computation.
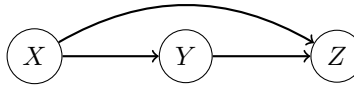


Bayesian networks have some convenient properties:

- They are a simple way to visualize the structure of a probabilistic model

- They can be used to design or motivate new kinds of statistical models.

- Inspection of the graph alone gives us insight into properties, e.g., conditional independence.

- Complex computations for inference and learning in statistical models can be expressed in terms of graphical manipulations.

Directed links (arrows) between two nodes (random variables) indicate conditional probabilities. For example, the arrow between $X$ and $Y$ gives the conditional probability $p(Y|X)$.

### Construction of Bayesian Networks

Starting with an example, consider the joint distribution:

$$p(a, b, c) = p(c|a, b)p(b|a)p(a)$$

of three random variables $a$, $b$, $c$. The factorization of the joint distribution tells us something about the relationship between the random variables:
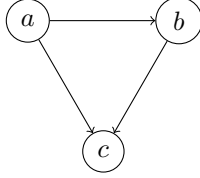
- $c$ depends directly on $a$ and $b$.

- $b$ depends directly on $a$.

- $a$ depends neither on $b$ nor on $c$.

Now, we can begin construction.

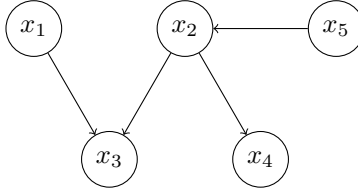1. Create a node for all random variables
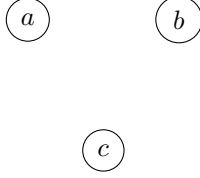
2. For each conditional distribution, we add a directed link (arrow) to the graph from the nodes corresponding to the variables on which the distribution is conditioned.



The graph layout depends on the choice of factorization of the joint distribution.

Now, for the reverse, consider the following network:



In this case, $x_1$ depends on nothing, $x_3$ depends on both $x_1$ and $x_2$, $x_2$ depends on $x_5$, $x_4$ depends on $x_2$, and $x_5$ depends on nothing. We can therefore extract

$$p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_5)p(x_2|x_5)p(x_3|x_1, x_2)p(x_4|x_2)$$

## Joint Distributions

In general, the joint distribution $p(x) = p(x_1, \ldots, x_k)$ is given as

$$p(x) = \prod_{k=1}^{K} p(x_k|\mathrm{Pa}_k)$$

where $\mathrm{Pa}_k$ means "the **parent** nodes of $a_k$". **Parent** nodes of $a_k$ are nodes that have arrows pointing to $a_k$.

In general for a network, we define the probabilities as:

$$p(\langle Q_1, Q_2, \ldots, Q_d \rangle) = \prod_{i=0}^{d} p\left(Q_i \,\middle|\, \bigwedge_{j \in N^{\mathrm{in}}(Q_i)} Q_j\right)$$

Note that $N^{\mathrm{in}}(v)$ is the set of in-neighbors of $v$ in a graph.

# Markov Chains

A **Markov chain** is a special type of **stochastic process**, defined in terms of the conditional distributions of future states given the **present** and **past** states.

Markov chains are usually used to model sequences of items. We saw when talking about conditional probabilities that:

$$p(X_n, X_{n-1}, \ldots, X_1) = p(X_n | X_{n-1}, \ldots, X_1) p(X_{n-1}, \ldots, X_1)$$
$$= p(X_n | X_{n-1}, \ldots, X_1) p(X_{n-1} | X_{n-2}, \ldots, X_1) p(X_{n-2}, \ldots, X_1)$$
$$\ldots$$
$$= p(X_n | X_{n-1}, \ldots, X_1) \ldots p(X_2 | X_1) p(X_1)$$

The **Markov assumption** is that the probability of $X_k$ depends on $X_{k-1}$ alone. This simplifies the computation, but at the cost of long term connections. Anecdotally the Markov assumption says *the future is independent of the past given the present.*

Therefore, we can rewrite the statement above as:

$$p(X_n, X_{n-1}, \ldots, X_1) = p(X_n | X_{n-1}) p(X_{n-1} | X_{n-2}) \ldots p(X_2 | X_1) p(X_1)$$

This is the basis for **Hidden Markov Models** which are used a lot in genetics.