Bayesian Networks

Malachy Campbell

11/5/2018

Introduction

- Goal: Model causation (or more appropriatly probabilistic dependance) between correlated variables/ traits
- Plant breeding example: Given a dataset of many (10+) agronomic traits (plant height, nitrogen use efficiency, grain protein, grain carbohydrates/starch, leaf area index, radiation use effiency, yeild), how does selection for one (e.g. grain protein) impact another (e.g. grain yeild)?

Graphical model terminology

- A graph ($G = (V, A)$) is composed of nodes (V) and vertices/arcs (A)
  - Nodes (u, v): variables or traits
  - Arcs ($a = (u, v)$: describe the relationships between nodes u and v
    - If u and v are ordered then the arc is directed, u and v unordered then the arc is undirected.
- Undirected graphs: no ordering between nodes; Directed graphs: all arcs are ordered; Partially directed graphs: some arcs are ordered
- Paths: sequences of arcs connecting two nodes; passes through each arc only once and all arcs follow the same direction
- Root and leaf nodes are the base and terminal nodes, respectively. No nodes are incoming for roots and no arcs leave leaf nodes.

PICTURE OF NETWORKS

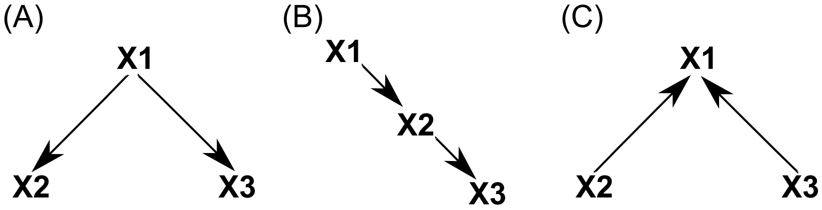
Bayesian background

- Chain rule: Joint probability distribution of i random variables (x_1, x_2, x_3, x_4) is
$\mathrm{P}(x_1, x_2, x_3, x_4) = \mathrm{P}(x_4|x_1, x_2, x_3)\mathrm{P}(x_3|x_1, x_2)\mathrm{P}(x_2|x_1)\mathrm{P}(x_1)$
- If we have no knowledge of any variables then variables are independent if $\mathrm{P}(x_1, x_2, x_3, x_4) = \mathrm{P}(x_4)\mathrm{P}(x_3)\mathrm{P}(x_2)\mathrm{P}(x_1)$
- If we have knowledge of some variable (x_4) then variables are independent if $\mathrm{P}(x_1, x_2, x_3|x_4) = \mathrm{P}(x_3|x_4)\mathrm{P}(x_2|x_4)\mathrm{P}(x_1|x_4)$

$$\mathrm{P}(x_1, x_2, x_3|x_4) = \frac{\mathrm{P}(x_3|x_4)\mathrm{P}(x_2|x_4)\mathrm{P}(x_1|x_4)\mathrm{P}(x_4)}{\mathrm{P}(x_4)}$$

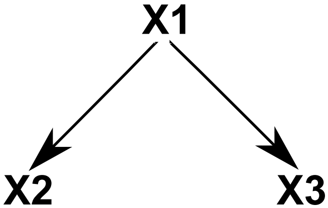
Bayesian Networks (BN)

- BN are a class of graphical models that represent the probabilistic dependencies between a set of random variables as a directed acyclic graph (no loops!). We can scale these concepts to a graphical model.
- For a divergent structure (A) the joint distribution is $\mathrm{P}(X1, X2, X3) = \mathrm{P}(X2|X1)\mathrm{P}(X3|X1)\mathrm{P}(X1)$; serial structure (B) the joint distribution is $\mathrm{P}(X1, X2, X3) = \mathrm{P}(X3|X2)\mathrm{P}(X2|X1)\mathrm{P}(X1)$; convergent structure (C) the joint distribution is $\mathrm{P}(X1, X2, X3) = \mathrm{P}(X2|X1, X3)\mathrm{P}(X1)\mathrm{P}(X3)$



Bayesian Networks (BN)

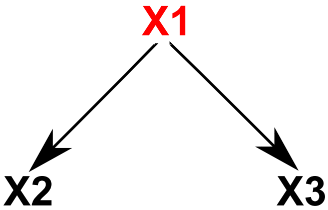
- **Example 1**: If we have no information for the divergent structure, are X2 and X3 independent?



$$\mathrm{P}(X2, X3) = \sum_{X3} \mathrm{P}(X2|X1)\mathrm{P}(X3|X1)\mathrm{P}(X1)$$

Bayesian Networks (BN)

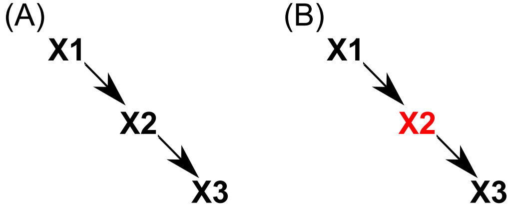
- **Example 2**: If we have information on X1 for the divergent structure, are X2 and X3 independent?



$$\mathrm{P}(X2, X3|X1) = \mathrm{P}(X2|X1)\mathrm{P}(X3|X1)$$

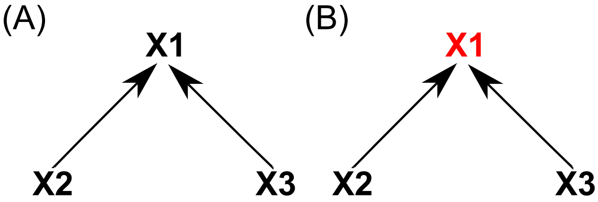
Bayesian Networks (BN)

- **Example 3**:
  1. If we don't have information on X2 for the serial structure (A), are X1 and X3 independent?
     $\mathrm{P}(X1, X3) = \sum_{X2} \mathrm{P}(X3|X2)\mathrm{P}(X2|X1)\mathrm{P}(X1)$
  2. If we do have information on X2 for the serial structure (B), are X1 and X3 independent? $\mathrm{P}(X1, X3|X2) = \mathrm{P}(X3|X2)\mathrm{P}(X1|X2)$



Bayesian Networks (BN)

- **Example 4**:
  1. If we don't have information on X1 for the convergent structure (A), are X2 and X3 independent? $\mathrm{P}(X2, X3) = \mathrm{P}(X2)\mathrm{P}(X3)$
  2. If we do have information on X1 for the convergent structure (B), are X2 and X3 independent?
     $\mathrm{P}(X1, X3|X2) = \frac{\mathrm{P}(X3)\mathrm{P}(X3)\mathrm{P}(X1|X2,X3)}{\mathrm{P}(X1)}$



Rules for *direct*-separation

- If we scale this up, each node can be considered a subset of nodes in a DAG
- Consider only the cases where we have information. Given some information for a group of nodes (**B**), the two subsets of nodes **A** and **C** are conditionally independant if there is a node *w* that has (1) converging arrows and *w* nor any of its descendants are part of **B** (e.g. does not carry information), or if *w* does not have converging arrows and is a part of **B**
- If these conditions are satisfied we can say that **B** *direct*-separates (*d*-separates) **A** from **C**