| |
|---|
| Statistical Techniques in Robotics (16-831, F10) Lectures 5 & 6 (9/7/10 and 9/9/10) |
| Graphical Models |
| *Lecturer: Drew Bagnell* *Scribe: Brian Coltin and Hugh Cover* [1] |

# 1 Graphical Models

Graphical models are a framework for reasoning about uncertain quantities and the structural relationships between them. They are a union of probability and graph theory. Nodes represent random variables and edges represent the links, or relationships between these random variables.

Graphical models can be viewed as a:

- **Communication tool** that helps to *compactly* express beliefs about a system.
- **Reasoning tool** that can be used to *extract* relationships that were not obvious when formulating the problem. In particular graphical models enable us to visualize conditional independence.
- **Computational skeleton** that helps organize how we perform computations on random variables.

We will examine three types of graphical models:

- **Bayes' Nets** (Directed Graphical Models)
- **Gibbs Fields** (Undirected Graphical Models)
- **Factor Graphs** (Undirected Graphical Models)

Graphical models are the equivalent of a circuit diagram — they are written down to visualize and better understand a problem.

# 2 Bayes' nets

One of the most common graphical models is called a Bayes' net. Bayes' nets are also known as Bayesian networks, belief networks, directed graphical models, and directed independence diagrams. In short, a Bayes' net is a directed acyclic graph with nodes representing uncertain quantities (random variables) and edges that encode relationships between them (often causal).

In Figure 1, we have uncertain quantities A, B, C, and we draw directed arrows between them to represent relationships (typically causal). A bayesian network encodes a joint probability distribution over all the nodes in the graph. In this case, our Bayes' net encodes the joint probability distribution, $P(A, B, C, D)$.

[1] Some content adapted from previous scribe: Byron Boots

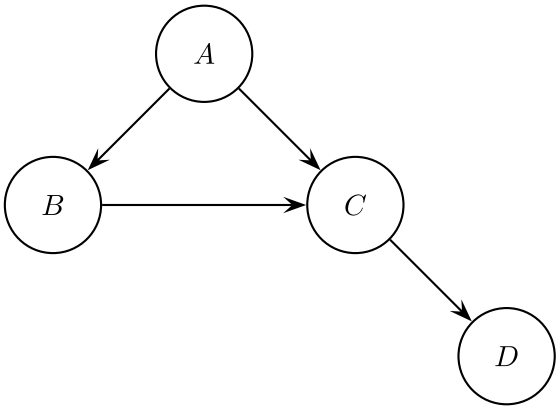


Figure 1: A Bayesian network.

The basic factorization of the probability distribution using the chain rule of probability is

$$P(A, B, C, D) = P(A)P(B|A)P(C|A, B)P(D|A, B, C).$$

This factorization always holds, and is not dependent on any particular graphical model.

In the network shown in Figure 1, we can use the edges in the graph to eliminate unnecessary conditional dependencies.

$$P(A, B, C, D) = P(A)P(B|A)P(C|A, B)P(D|C)$$

For an arbitrary Bayes' net with nodes $x_1, x_2, \ldots, x_n \in X$, we can derive the joint distribution $P(X)$ as the product of each node $x_i$ given its parents $\pi(x_i)$.

$$P(X) = \prod_{x_i} P(x_i|\pi(x_i))$$

Note that this factorization strategy only works if there are no cycles in the graph, and that Bayes' nets are acyclic by definition.

Bayes' net are often thought of as encoding causal relationships. However, these relationships are not necessarily causal. In our example, one should think of A as influencing B and C rather than A causing B and C. If all the arrows on a Bayes' net are flipped, then the resulting Bayes' net is equivalent to the original, since they both represent the same joint probability distribution.

In general, the absence of arrows is important in a Bayes net: *less* arrows mean *more* structure.

### 2.1 Determining Dependencies

Bayes' nets can be used to quickly determine whether pairs of variables are dependent on each other. This is done by following all available paths between the two variables and checking if the path is 'blocked'. A path is any sequence of edge connected nodes leading from the first variable to the second. The Bayes' net in Figure 2 has 2 paths from A to E.

$$A \to B \to D \to E$$

$$A \to C \to D \to E$$

Blockages are determined by visiting each node on a path and comparing the structure of surrounding nodes and edges to the 3 rule situations explained below.

Two variables are independent if *all* available paths between them are blocked. Any unblocked paths show a possibility of dependence. It should be noted that this analysis can only be as good as the Bayes' net it is based on; an incomplete net may be missing paths that show a dependence.

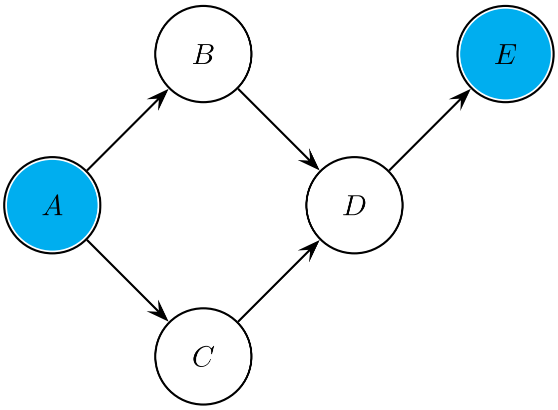


Figure 2: There are 2 paths from $A$ to $D$.

#### 2.1.1 Rule 1: Markov Chain

Figure 3 is a Bayes' net representation of a simple markov chain.

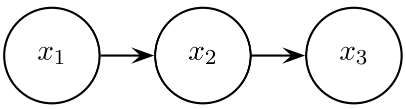


Figure 3: A Bayesian network representation of a Markov chain.

An example of such a chain is the process of robot localization, although the usual $z_i$ and $u_i$ terms have been omitted for simplicity. If the robot knows the current state, $x_2$, then it does not need any information about past states, $x_1$, in order to determine the next state, $x_3$. For example if the robot can only be on the $1_{st}$ or $2_{nd}$ floor of a building and it knows that it's previous state was that it was on the $1_{st}$ floor. As there were no inputs any previous information about the robot's state is irrelevant, it must still be on the $1_{st}$ floor. This is the same as saying that $x_1$ and $x_3$ are independent if $x_2$ is known.

$$P(x_3|x_2, x_3) = P(x_3|x_2)$$

In the case where $x_2$ is not known then knowledge of past states could provide information on the current state $x_3$. If our robot did not know it's state at $x_2$ but knew it was on the $1_{st}$ floor at $x_1$ then, given no inputs, $x_3$ has a high probability of being the $1_{st}$ floor. This is the same as saying that $x_1$ and $x_3$ could be dependent if $x_2$ is not known.

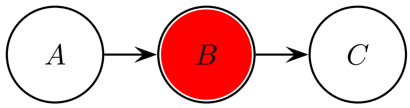


Figure 4: Markov chain is **BLOCKED** given $B$.

The rule is therefore that in a chain of nodes, as shown in Figure 4, C is independent from A if B is known. This means there is a blockage on any path passing through a Markov chain with a known middle node.

#### 2.1.2 Rule 2: Two Parents, One Child (The Bagpipes Case)

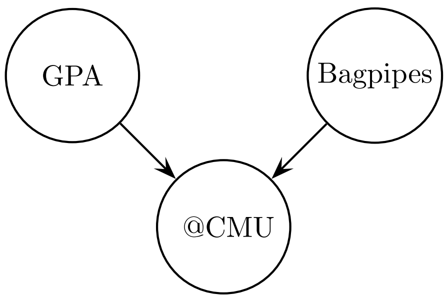


Figure 5: The Bagpipes Case.

Figure 5 is a simplified Bayes' net representation of the process of getting into CMU. Carnegie Mellon wants to admit students with a high GPA but it is also important to keep the school bagpipe band strong. A student's chances of getting into CMU can therefore be influenced by their GPA and also by their bagpipe playing skills.

Given any student applying to CMU knowing that they have good grades doesn't tell us anything about their bagpipe skills, the two are independent. This is changed if we then discover that the student was admitted. Now if we know they are good at the bagpipes our expectation of their grades is reduced as their admittance has been 'explained away'. The reverse is true if we know they have particularly high grades. Thus knowledge about admittance creates a dependence between the student's bag piping skills and their GPA.

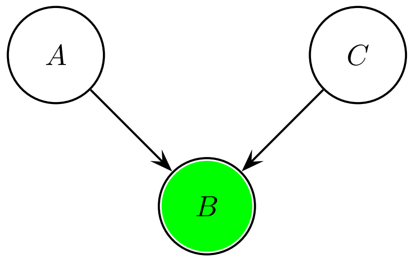


Figure 6: Two parents, one child case is **NOT BLOCKED** given B.

The rule is therefore that in a 'two parents, one child' case, as shown in Figure 6, A is dependent on C if B is known. The inverse is also true, A is independent from C if B is not known. This means that there is a blockage on a path passing through the 'two parent, one child' case if B is unknown.

#### 2.1.3 Rule 2 Extension: Addition of Further Children

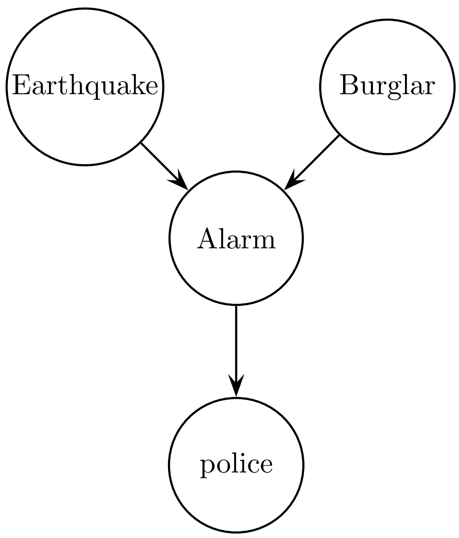


Figure 7: Home Alarm Example.

Rule 2 can be extended with the addition of children of the child. Figure 7 shows an example were an alarm can be set off by either an earthquake or a burglar and the police are called when the alarm goes off. As with the bagpipes example the presence of an earthquake and a burglar become dependent given the alarm going of. This is because if we know the alarm has been activated knowledge about a burglar reduces the liklyhood of there having been an earthquake. If however only the presence of the police is known the same dependency is formed as the police imply that the alarm has been activated.

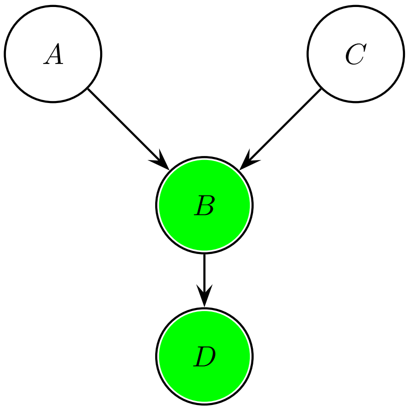


Figure 8: Path is **NOT BLOCKED** given either B or D.

The extension of rule 2 is therefore that if a child of the child is known the path is also unblocked. With reference to Figure 8 the path is only blocked if B and any children of B are unknown.

#### 2.1.4 Rule 3: One parent, Two Children

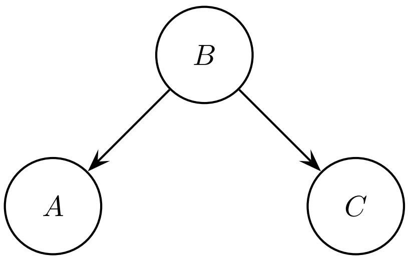


Figure 9: Rule 3, Anybot Example.

Figure 9 shows a Bayes' net representation of the uses of an Anybot in an office. An Anybot can be used to aid in teleconferenced meetings and can also be used as a racing vehicle. If we don't know that there is an Anybot and are told that there have been races the probability of teleconference meetings is increased. This is because knowledge about races increases the chance of there being an Anybot for use in meetings. If however we know there is an Anybot in the office the fact that people are racing it dosn't change the likelihood of there being any teleconference meetings.

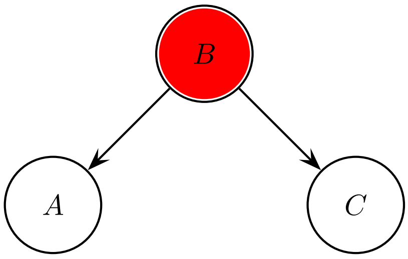


Figure 10: One parent, Two Children case is **BLOCKED** given B.

Rule 3 is therefore that in a 'one parent, two children' case, as shown in Figure 10, A is independent from C given B. This means that there is a blockage on a path passing through this case if the parent, B, is known.

#### 2.1.5 Example: Localization

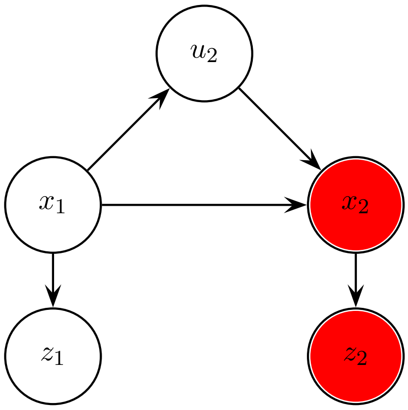


Figure 11: A Remote Controled Car.

Figure 11 shows a simple remote controlled car scenario with a human driver sending inputs based on the car's actual state. The derivation of Bayes Filter in "Probabilistic Robotics" assumes that $x_{t-1}$ is independant of $u_t$. To test that this is the case for the remote control example the two paths between $x_1$ and $u_2$ need to be tested. The path via $x_2$ is a case of rule 2 where both $x_3$ and $z_3$ are unknown and so is blocked, however the direct path cannot be blocked. Therefore there does exist a dependancy and the assumption is incorrect.

If the scenario is modified such that the input is now based on the previous observation and not a human who knows the actual state the Bayes' net looks like Figure 12. In this case the path via $x_2$ is still blocked and the path via $z_1$ is a case of rule 1 where $z_1$ is known and therfore blocked. All paths are blocked and the assumption that $x_{t-1}$ is independant from $u_t$ is valid.

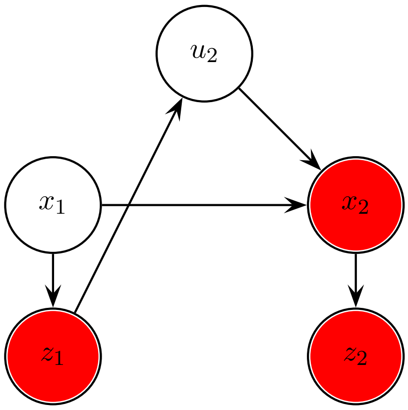


Figure 12: A Remote Controled Car.

#### 2.1.6 Example: Landmark Based Navigation

Figure 13 shows a bayesian network representation of a localisation scenario with internal states $x_i$, observations $z_i$ and landmarks $l_i$. If we pose the case that we see all the observations, $z_0, z_1, z_2, ...$, are the landmarks conditionally independent of each other? To take one case, is $l_1 \perp l_2 | Z$?

The converging arrows at $z_0$ are an example of rule 2 and the path remains unblocked as $z_0$ is known. Looking now at $x_0$, rule 3 can be used to show there is no blockage as $x_0$ is not known. The path extends through $x_1$ to $z_1$, a rule 1 case where $x_1$ is not known leaving the path unblocked. Finally $z_1$ to $l_2$ is another unblocked rule 2 case. Thus, $l_1$ and $l_2$ are not conditionally independent given $z_0$ and $z_1$.

The existence of conditional dependencies between landmarks introduces significant computational complexity due to high dimensionality. In this case if the values of X can be observed the dependency between $l_1$ and $l_2$, and all the landmarks, is removed. Using a particle filter, samples of X can be taken making each landmark independent and allowing for sperate filters to be run for each landmark, greatly reducing the dimensionality of the problem.

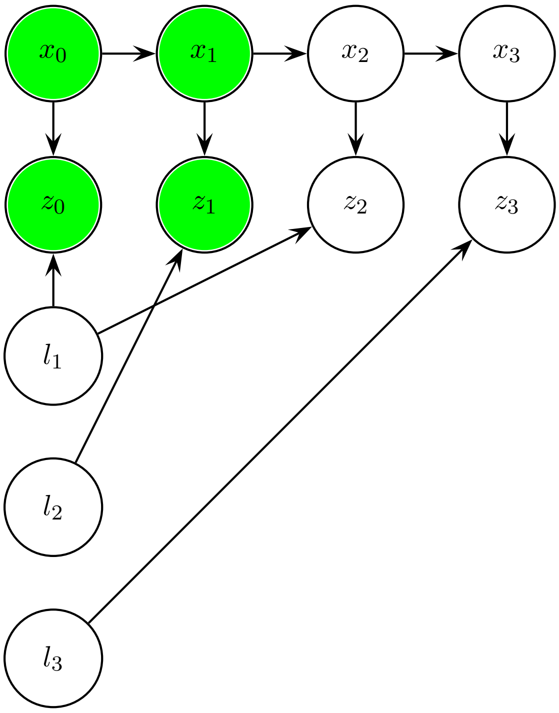


Figure 13: A bayesian network representation of a localisation scenario.

# 3 Gibbs Field

A Gibbs Field is a collection of nodes that have undirected edges between them, this can be seen in Figure 14. There is no causal link along edges rather it shows that the connected nodes are related in some way, they 'move together'. As with a Bayes' net less connections means more structure.

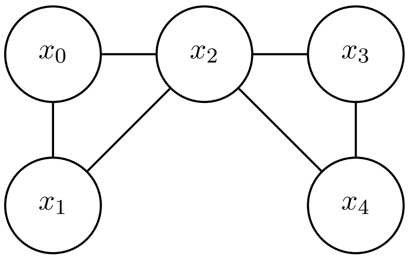


Figure 14: A Gibbs Field with nodes $x_0, x_1, x_2, x_3, x_4$.

A clique is a fully connected subset of the the graph, i.e. all nodes in a clique must be connected to all other nodes in the clique. In Figure 14 $x_0$, $x_1$, and $x_2$ form a clique and $x_2$, $x_3$, and $x_4$ also form a clique. These two cliques are known as maximal cliques because they are not part of any larger clique, however they are not the only cliques. The full set includes pairs of connected nodes, e.g. $x_0$ and $x_1$ and individual nodes. It is only nessasry to pay attention to maximal cliques as characteristics of smaller cliques can be absorbed into the larger maximal cliques.

The joint probability distribution can be represented by the product of a set of clique potential functions $\phi$ (see Equation 1).

$$P(\vec{x}) = \frac{1}{Z} \prod_{i \in cliques} \phi_i(X_i) \tag{1}$$

The first clique has functions representing the probability distribution of the set of three variables $x_0, x_1, x_2$: $\phi_i(x_0, x_1, x_2)$, similarly there exists functions for the other maximal clique, $\phi_i(x_2, x_3, x_4)$. Each potential function $\phi_i$ must be positive and unlike probability distribution functions they do not need to sum to 1. Because the potential functions are not normalized a normalization constant, Z (see Equation 2) is required in Equation 1 to create a valid probability distribution.

$$Z = \sum_{x} \prod_{i \in cliques} \phi_i(x_i) \tag{2}$$

The potential functions can also be thought of as energy functions, $f_i(X_i)$, with a probability function as shown in Equation 3. The energy assigned by the functions $f_i(X_i)$ is an indicator of the likelihood of the corresponding relationships within the clique. Higher energies lead to a lower probability and vis-versa.

$$P(\vec{x}) = \frac{1}{Z} e^{-\sum_{i \in cliques} f_i(X_i)} \tag{3}$$