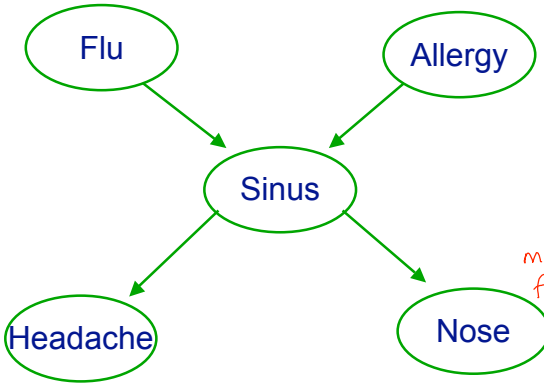
# Bayesian Networks – Representation

Machine Learning – 10701/15781
Carlos Guestrin
Carnegie Mellon University

October 31st, 2007

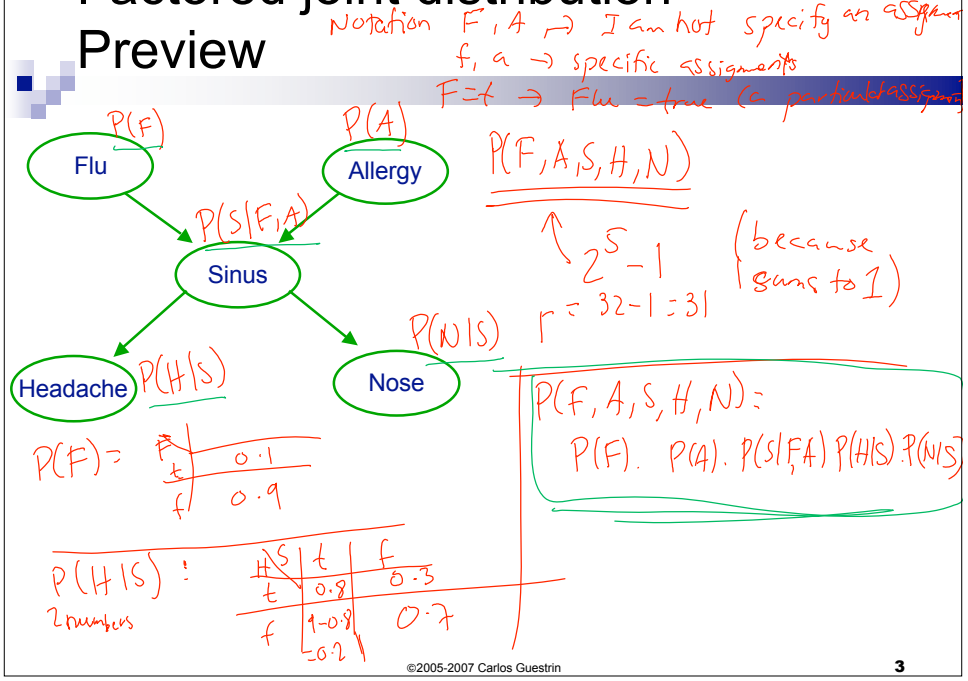



## Possible queries




- Inference
  $P(F=t \mid H=t, N=f)$
- Most probable explanation
  $\max_{f,a,s} P(f,a,s \mid H=t, N=f)$
- Active data collection
  what should I measure

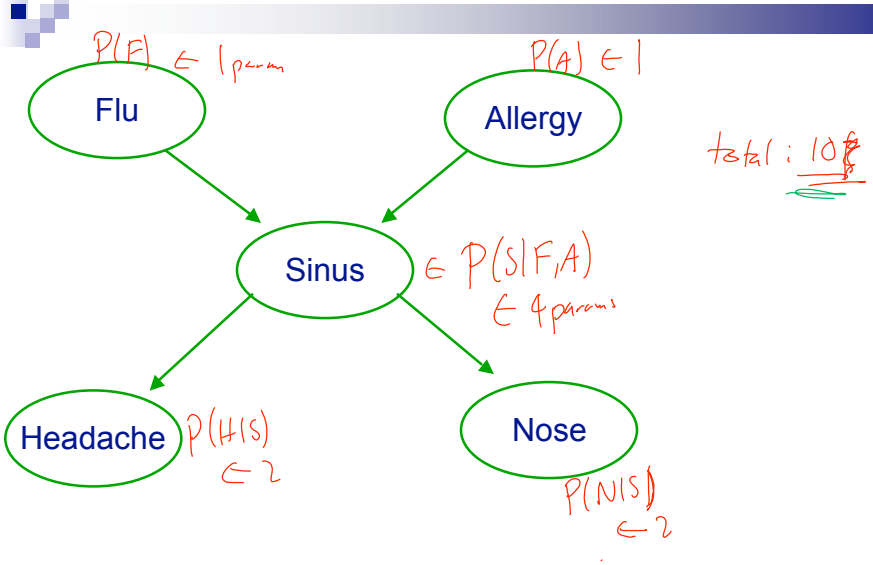
# Factored joint distribution - Preview



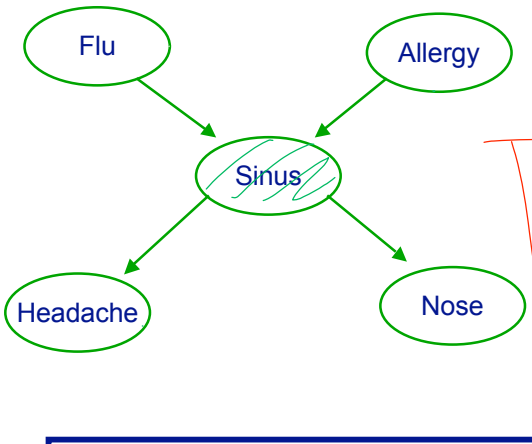




# Number of parameters



# Key: Independence assumptions




Knowing sinus separates the variables from each other



# (Marginal) Independence

- Flu and Allergy are (marginally) independent

$$P(F,A) = P(F) \cdot P(A)$$

- More Generally:

$$P(F|A) = P(F)$$

| | |
|---|---|
| Flu = t | 0.2 |
| Flu = f | 0.8 |

| | |
|---|---|
| Allergy = t | 0.3 |
| Allergy = f | 0.7 |

| | Flu = t | Flu = f |
|---|---|---|
| Allergy = t | 0.3 x 0.2 | 0.3 x 0.8 |
| Allergy = f | 0.2 x 0.7 | 0.7 x 0.8 |

# Marginally independent random variables

- **Sets** of variables **X**, **Y**
- X is independent of Y if
  - ~~$P^2$~~(**X**=**x**⊥**Y**=**y**), ~~8 **x2**Val(**X**), **y2**Val(**Y**)~~
- Shorthand:
  - **Marginal independence:** ~~$P^2$~~ (**X** ⊥ **Y**)
- **Proposition:** *P* statisfies (**X** ⊥ **Y**) if and only if
  - P(**X**,**Y**) = P(**X**) P(**Y**)




# Conditional independence

- Flu and Headache are not (marginally) independent
- Flu and Headache are independent given Sinus infection
- More Generally:

# Conditionally independent random variables

- **Sets** of variables **X**, **Y**, **Z**
- X is independent of Y given Z if
  - $P \models (\mathbf{X}=\mathbf{x} \perp \mathbf{Y}=\mathbf{y} | \mathbf{Z}=\mathbf{z}), \forall \mathbf{x} \in Val(\mathbf{X}), \mathbf{y} \in Val(\mathbf{Y}), \mathbf{z} \in Val(\mathbf{Z})$

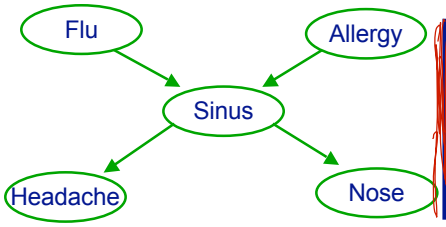
  $P(X=x | Y=y, Z=z) = P(X=x | Z=z)$
- Shorthand:
  - **Conditional independence:** $P \models (\mathbf{X} \perp \mathbf{Y} | \mathbf{Z})$
  - For $P \models (\mathbf{X} \perp \mathbf{Y} | \emptyset)$, write $P \models (\mathbf{X} \perp \mathbf{Y})$
- **Proposition:** *P* statisfies $(\mathbf{X} \perp \mathbf{Y} | \mathbf{Z})$ if and only if
  - $P(\mathbf{X},\mathbf{Y}|\mathbf{Z}) = P(\mathbf{X}|\mathbf{Z})\, P(\mathbf{Y}|\mathbf{Z})$




# Properties of independence

- **Symmetry:**
  - $(\mathbf{X} \perp \mathbf{Y} | \mathbf{Z}) \Rightarrow (\mathbf{Y} \perp \mathbf{X} | \mathbf{Z})$
- **Decomposition:**
  - $(\mathbf{X} \perp \mathbf{Y},\mathbf{W} | \mathbf{Z}) \Rightarrow (\mathbf{X} \perp \mathbf{Y} | \mathbf{Z})$
- **Weak union:**
  - $(\mathbf{X} \perp \mathbf{Y},\mathbf{W} | \mathbf{Z}) \Rightarrow (\mathbf{X} \perp \mathbf{Y} | \mathbf{Z},\mathbf{W})$
- **Contraction:**
  - $(\mathbf{X} \perp \mathbf{W} | \mathbf{Y},\mathbf{Z}) \,\&\, (\mathbf{X} \perp \mathbf{Y} | \mathbf{Z}) \Rightarrow (\mathbf{X} \perp \mathbf{Y},\mathbf{W} | \mathbf{Z})$
- **Intersection:**
  - $(\mathbf{X} \perp \mathbf{Y} | \mathbf{W},\mathbf{Z}) \,\&\, (\mathbf{X} \perp \mathbf{W} | \mathbf{Y},\mathbf{Z}) \Rightarrow (\mathbf{X} \perp \mathbf{Y},\mathbf{W} | \mathbf{Z})$
  - Only for positive distributions!
  - $P(\alpha)>0, \forall \alpha, \alpha \neq ;$

# The independence assumption




**Local Markov Assumption:**
A variable X is independent of its non-descendants given its parents

$F \perp A \mid \emptyset$

$S \perp ?$ nothing $\mid \{F, A\}$

$H \perp \{F, A, N\} \mid S$

$N \perp \{H, F, A\} \mid S$

$F \perp A$

$F \perp H \mid S$

$A \perp N \mid S$

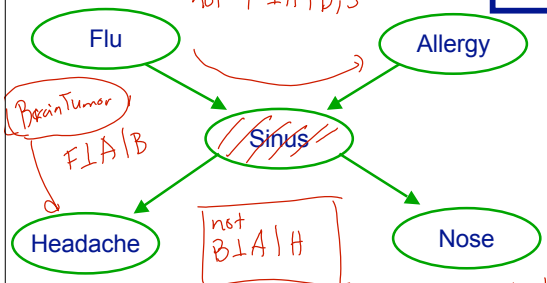
$H \perp N \mid S$ ..



# Explaining away

**Local Markov Assumption:**
A variable X is independent of its non-descendants given its parents and only its parents

not $F \perp A \mid B, S$




$F \perp A \Downarrow$ $F \perp A \mid S$ ← no!!

jumping ahead not: $F \perp A \mid N$

not $B \perp A \mid H$

$P(F=t) = 0.2$, $P(F=t \mid A=t) = 0.2 = P(F=t)$

$P(F=t \mid S=t) = 0.7 \neq P(F=t \mid S=t, A=t) = 0.4$ goes down!

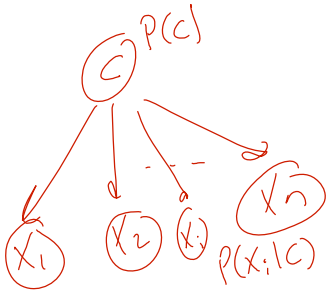
not: $F \perp A \mid S$ why? Allergies help explain $S=t$

# Naïve Bayes revisited

$P(C, X_1, \ldots X_n) = P(C) \cdot \prod P(X_i|C)$

$X_i$'s are independent given C

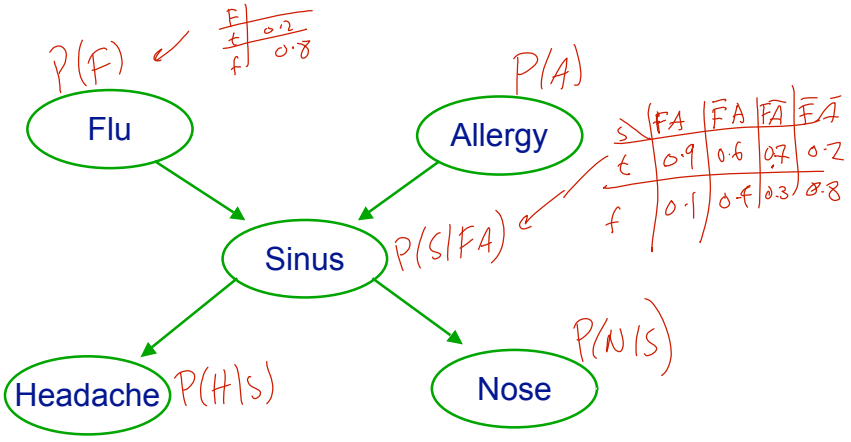
**Local Markov Assumption:** A variable X is independent of its non-descendants given its parents



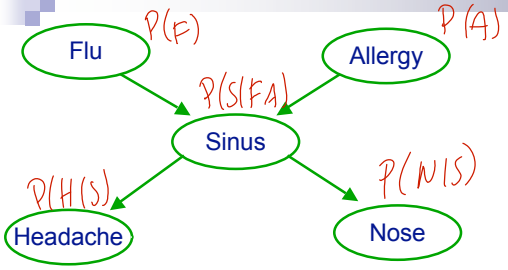

$X_i \perp \{X_1 \ldots X_{i-1}, X_{i+1} \ldots X_n\} | C$



# What about probabilities? Conditional probability tables (CPTs)



# Joint distribution




$P(F,A,S,H,N) = P(F) \cdot P(A) \cdot P(S|FA) \cdot P(H|S) \cdot P(N|S)$

**Why can we decompose? Markov Assumption!**

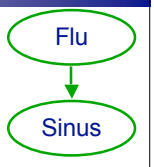



# The chain rule of probabilities

- $P(A,B) = P(A)P(B|A)$ exactly (no assumptions)

$P(F,S) = P(F) \cdot P(S|F)$

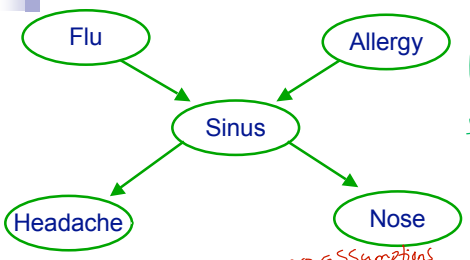
$P(S|F) = \frac{P(F,S)}{P(F)}$




- More generally:
  - $P(X_1,\ldots,X_n) = P(X_1) \cdot P(X_2|X_1) \cdot \ldots \cdot P(X_n|X_1,\ldots,X_{n-1})$

$P(X_1,\ldots,X_n) = P(X_1) \cdot P(X_2|X_1) \cdot P(X_3|X_2X_1) \cdots$

# Chain rule & Joint distribution




**Local Markov Assumption:**
A variable X is independent of its non-descendants given its parents

$P(F,A,S,H,N) \overset{\text{no assumptions}}{=} P(F) \cdot P(A|F) \cdot P(S|FA) \cdot P(H|FAS) \cdot P(N|FASH)$

With Markov Assumption

$= P(F)\,P(A)\,P(S|FA)\,P(H|S)\,P(N|S)$

$P(A|F) = P(A)$ | $P(H|F,AS) = P(H|S)$ | $P(N|FASH) = P(N|S)$

$A \perp F$ | $H \perp \{F,A\}|S$ | $N \perp \{A,H,F\}|S$



# Two (trivial) special cases

**Edgeless graph**

$X_1 \perp \{X_2 \ldots X_n\}$

$X_2 \perp \{X_1, X_3 \ldots X_n\}$

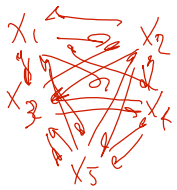
⋮

all possible independencies

lots of bias

red & green graphs will be "equivalent" can represent same distributions

**Fully-connected graph**




no assumptions!!

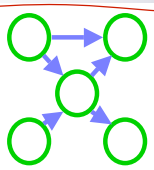
no bias

can represent any dist.

# The Representation Theorem – Joint Distribution to BN

represent ≡ perfectly represent

**BN:**



**Encodes independence assumptions**

**If conditional independencies in BN are subset of conditional independencies in *P***

Obtain

**Joint probability distribution:**

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

the more independencies in the BN, the fewer Ps you can represent. But the easier it will be to learn the BN

you can represent P with this BN





# A general Bayes net

- Set of random variables $X_1, \ldots X_n$

  F, A, H, ...
- Directed acyclic graph
  - Encodes independence assumptions

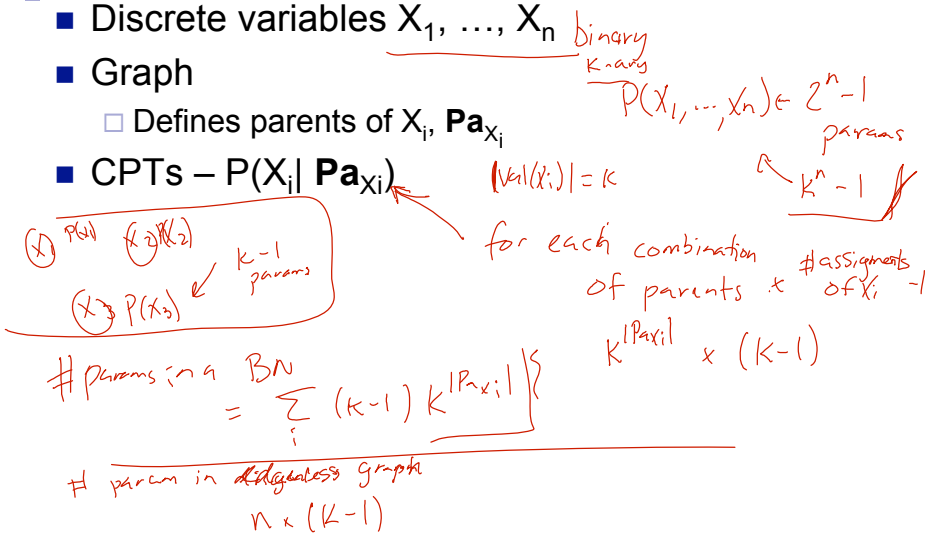
- CPTs $P(X_i \mid Pa_{X_i})$, e.g., P(S|FA), P(H|S), ...
- Joint distribution:

$$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

P(A), P(F), P(S|F,A), ...

# How many parameters in a BN?

- Discrete variables $X_1, \ldots, X_n$
- Graph
  - Defines parents of $X_i$, $\mathbf{Pa}_{X_i}$
- CPTs – $P(X_i | \mathbf{Pa}_{X_i})$




# Real Bayesian networks applications

- Diagnosis of lymph node disease
- Speech recognition
- Microsoft office and Windows
  - http://www.research.microsoft.com/research/dtg/
- Study Human genome
- Robot mapping
- Robots to identify meteorites to study
- Modeling fMRI data
- Anomaly detection
- Fault dianosis
- Modeling sensor network data

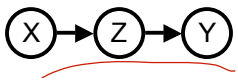
# Independencies encoded in BN

- We said: All you need is the local Markov assumption
  - $(X_i \perp \text{NonDescendants}_{Xi} \mid \mathbf{Pa}_{Xi})$, and only $Pa_{X_i}$
- But then we talked about other (in)dependencies
  - e.g., explaining away
- What are the independencies encoded by a BN?
  - Only assumption is local Markov
  - But many others can be derived using the algebra of conditional independencies!!!

# Understanding independencies in BNs – BNs with 3 nodes

**Local Markov Assumption:**
A variable X is independent of its non-descendants given its parents

Indirect causal effect:

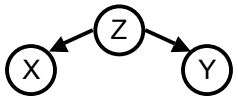



$Y \perp X \mid Z$

Indirect evidential effect:


$Y \perp X \mid Z$

Common cause:

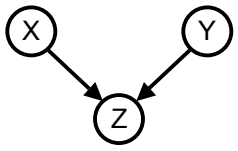



$Y \perp X \mid Z$

equivalent!!

different

Common effect:

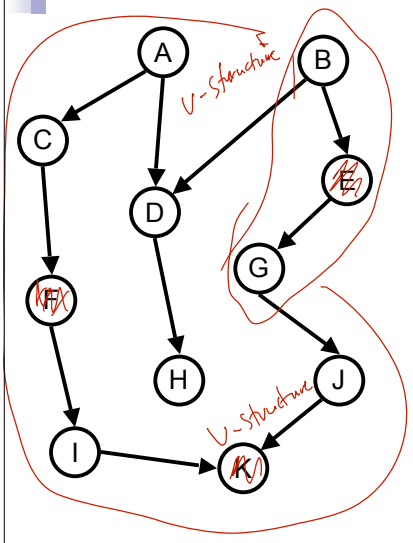



a V-structure

$X \perp Y$

not $X \perp Y \mid Z$

# Understanding independencies in BNs – Some examples




$A \perp B$

$B \perp G \mid E$

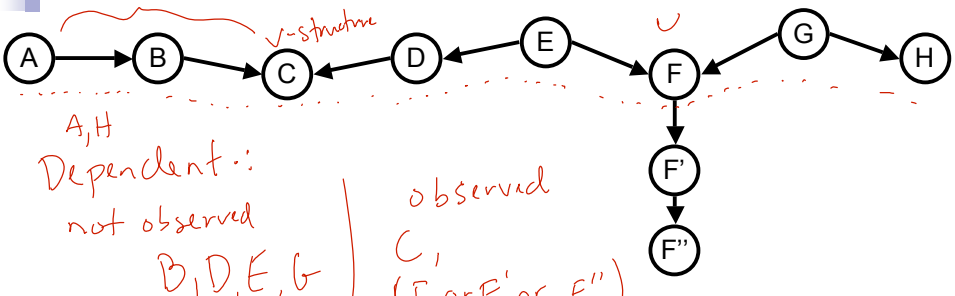
not $B \perp G \mid E, K, D$ or $H$?

$B \perp G \mid E, K, D, F$ ↺ true

V structure: stop flow of "dependency" unless K is observed or D or H observed

could influence flow through other path?

# An active trail – Example




A, H Dependent: not observed B, D, E, G | observed C, (F or F' or F'')

**When are A and H independent?**

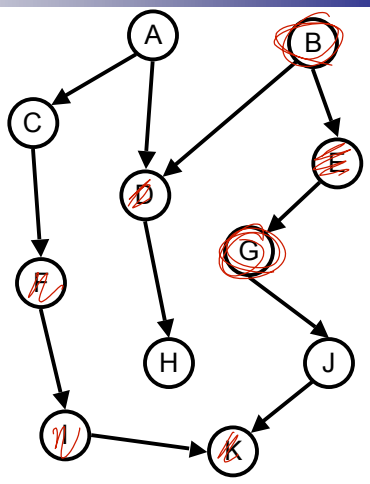
opposite: observe at least one of B D E G or not observe C, or (none of F, F', F'')

# Active trails formalized

- A path $X_1 - X_2 - \cdots - X_k$ is an **active trail** when variables $\boldsymbol{O} \subseteq \{X_1,\ldots,X_n\}$ are observed if for each consecutive triplet in the trail:
  - $X_{i-1} \rightarrow X_i \rightarrow X_{i+1}$, and $X_i$ is **not observed** ($X_i \notin \boldsymbol{O}$)
  - $X_{i-1} \leftarrow X_i \leftarrow X_{i+1}$, and $X_i$ is **not observed** ($X_i \notin \boldsymbol{O}$)
  - $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$, and $X_i$ is **not observed** ($X_i \notin \boldsymbol{O}$)
  - $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, and $X_i$ **is observed** ($X_i \in \boldsymbol{O}$), or **one of its descendents**

# Active trails and independence?

- **Theorem**: Variables $\boldsymbol{X_i}$ **and** $\boldsymbol{X_j}$ **are independent given** $\boldsymbol{Z} \subseteq \{X_1,\ldots,X_n\}$ if the is **no active trail** between $X_i$ and $X_j$ when variables $\boldsymbol{Z} \subseteq \{X_1,\ldots,X_n\}$ are observed



# The BN Representation Theorem

If conditional independencies in BN are subset of conditional independencies in ***P***

Obtain

Joint probability distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

**Important because:**
**Every *P* has at least one BN structure *G***

If joint probability distribution:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathbf{Pa}_{X_i}\right)$$

Obtain

Then conditional independencies in BN are subset of conditional independencies in ***P***

**Important because:**
**Read independencies of *P* from BN structure *G***