· Tabular CPDs

Table CPDs (Conditional probability tables) are an inherent part of the Bayesian Network

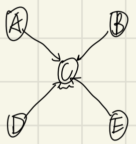
$\therefore \begin{cases} \text{two many parameters} \\ \text{ignores structure within the CPD} \end{cases}$

e.g. $P(x \mid \text{parents}(x))$ are the same where A takes some certain value

see a CPD as a function which takes in the value of $\text{parent}(x)$ and returns $P(X \mid \text{parent}(x))$, not a table that lists all combination

· Deterministic CPDs

$$P(X \mid \text{parents}(x)) = \begin{cases} 1 & X = f(\text{pa}(x)) \\ 0 & \text{otherwise} \end{cases}$$



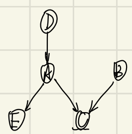
when C is a deterministic function of A and B we have $(D \perp_c E \mid A, B)$, which is not true only by d-separation

```
def deterministically_separated(G, D, X, Y, Z):
    Z+ = Z
    while ∃X such that { X_i ∈ D
                       { Parents(X_i) ⊆ Z+
        Z+ = Z ∪ {X_i}
    return de-sep_G{X; Y | Z+}
```

G: network
D: deterministic variable
$(X \perp Y \mid Z)$

let G be a network structure and D, X, Y, Z be sets of variables, if X is deterministically separated from Y given Z (i.e. $\text{det-sep}(G, D, X, Y, Z)$) then for all distribution that $P \models \mathcal{I}_\ell(G)$ and $P(X \mid \text{Pa}(x))$ is a deterministic CPD for $X \in D$, $P \models (X \perp Y \mid Z)$

if $\text{det-sep}(G, D, X, Y, Z)$ is false, there is a distribution P such that $P \models \mathcal{I}_\ell(G)$, $P(X \mid \text{Pa}(X))$ is deterministic $\forall X \in D$, but $P \not\models (X \perp Y \mid Z)$



particular deterministic CPDs may imply additional independencies

when $C = A \text{ Xor } B$, A is determined by BC

$\therefore (D \perp_c E \mid B, C)$, which is not true for some other deterministic CPDs $P(C \mid A, B)$

(independency holds for sepecific CPD)

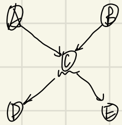
let X, Y, Z be pairwise disjoint sets of variables, C is a set of variable (might overlap with X∪Y∪Z)

X and Y are contextually independent given Z $(X \perp_c Y \mid Z, c)$ if

$P(X \mid Y, Z, C=c) = P(X \mid Z, C=c)$ whenever $P(Y, Z, c) > 0$

$(B \perp_c D \mid A = a^1)$

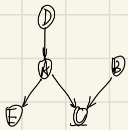
$(B \not\perp D \mid A = a^0)$



C = A or B

$(D \perp_c E \mid C = c^0)$

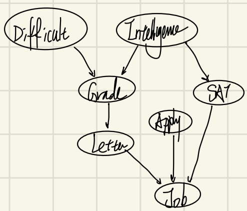
$(D \not\perp E \mid C = c^1)$



C = A or B

(independency holds for specific value of conditions)

· Context-Specific CPDs

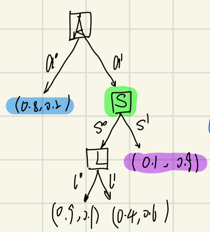


when a student doesn't apply but recruiter want employee the recruiter can't get SAT and letter.

∴ $P(J \mid L, S, A=a^0)$ is same for all L and s

· Context-Specific CPDs : Tree CPDs

A tree CPD for CPD $P(J \mid A, S, L)$



$(J \perp_c L, S \mid a^0)$

$(J \perp_c L \mid a^1, s^1)$ } $(J \perp_c L \mid s^1, a)$

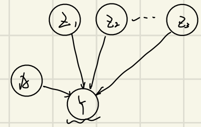
a leaf t-node, labeled with a distribution over J

an interior t-node, labeled with some r.v $Z \in$ parent(J)

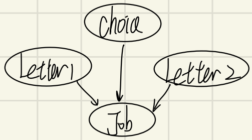
when a r.v depends on a bunch of r.v.

but we have uncertainty about which r.v. it depends on

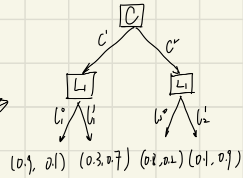
A CPD $P(Y \mid A, Z_1, \dots, Z_k)$ is a multiplexer CPD if $Val(A) = \{1, \dots, k\}$ and
$P(Y \mid a, Z_1, \dots, Z_k) = \mathbb{1}\{Y = Z_a\}$ where a is the value of A.
The variable A is called the selector variable of the CPD



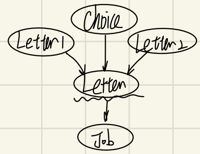
$(L_1 L_2 \mid J, c_1)$
$(L_1 L_2 \mid J, c_2)$
$(L_1 L_2 \mid J, C)$



Student can chose which RL to submit to the recruiter



The CPD $P(J \mid C, L_1, L_2)$

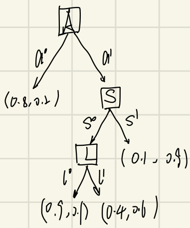


multiplexer CPD $P(L \mid C, L_1, L_2)$
where C is the selector variable

· Context-Specific CPDs : Rule CPDs

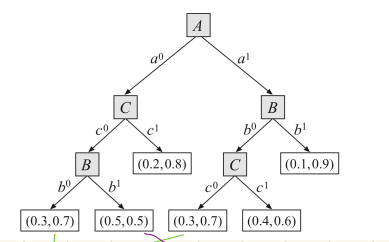
each rule corresponds to a single entry in the CPD

A rule ρ is a pair $\langle c; p \rangle$ where c is an assignment to some subset of variable C. and $p \in [0,1]$. C is the scope of ρ, denoted $Scope[\rho]$



$$\Longrightarrow \left\{ \begin{array}{l} \rho_1{:}\langle a^0, j^0; 0.8\rangle \\ \rho_2{:}\langle a^0, j^1; 0.2\rangle \\ \rho_3{:}\langle a^1, s^0, l^0, j^0; 0.9\rangle \\ \rho_4{:}\langle a^1, s^0, l^0, j^1; 0.1\rangle \\ \rho_5{:}\langle a^1, s^0, l^1, j^0; 0.4\rangle \\ \rho_6{:}\langle a^1, s^0, l^1, j^1; 0.6\rangle \\ \rho_7{:}\langle a^1, s^1, j^0; 0.1\rangle \\ \rho_8{:}\langle a^1, s^1, j^1; 0.9\rangle \end{array} \right\}$$

A rule-base CPD $P(X \mid \text{parents}(X))$ is a set of rules R such that
- $\forall\ \rho \in R$, $Scope[\rho] \subseteq \{X\} \cup Pa(X)$
- $\forall$ assignment (x, u) to $\{X\} \cup Pa(X)$, there's exactly one rule $\langle c; p \rangle \in R$ s.t. c is compatible with (x, u)
  we say $P(X = x \mid Pa(X) = u) = p$
- $\sum_x P(x \mid u) = 1$

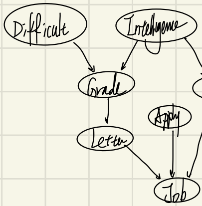


$$\Rightarrow \begin{array}{ll} \rho_1{:}\langle a^1, b^1, x^0; 0.1\rangle & \rho_2{:}\langle a^1, b^1, x^1; 0.9\rangle \\ \rho_3{:}\langle a^0, c^1, x^0; 0.2\rangle & \rho_4{:}\langle a^0, c^1, x^1; 0.8\rangle \\ \rho_5{:}\langle b^0, c^0, x^0; 0.3\rangle & \rho_6{:}\langle b^0, c^0, x^1; 0.7\rangle \\ \rho_7{:}\langle a^1, b^0, c^1, x^0; 0.4\rangle & \rho_8{:}\langle a^1, b^0, c^1, x^1; 0.6\rangle \\ \rho_9{:}\langle a^0, b^1, c^0; 0.5\rangle & \end{array}$$

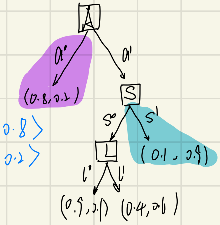
$P(X = x^0 \mid B = b^0, C = c^0, A = a^0) = 0.3$
$P(X = x^0 \mid B = b^0, C = c^0, A = a^1) = 0.7$

$P(X = x^0 \mid A = a^0, B = b^1, C = c^0) = 0.5$
$P(X = x^1 \mid A = a^0, B = b^1, C = c^0) = 0.5$

· Independencies in Context-specific CPDs



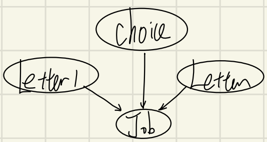
$(J \perp_c S, L \mid A = a^0)$



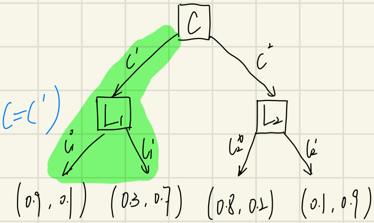
$\rho_1 = \langle a^0, x^0; 0.8 \rangle$
$\rho_2 = \langle a^0, x^1; 0.2 \rangle$
$Scope(\rho_1) = A$

$(J \perp_c L \mid S = s^1)$
$\{S = s^1\}$ is not the full assignment in this branch

$c = \{A = a^0\}$ is the context associated with a branch in tree-CPD for X then X is independent of remaining parents $(Parents(X) - Scope[C]) = \{S, L\}$

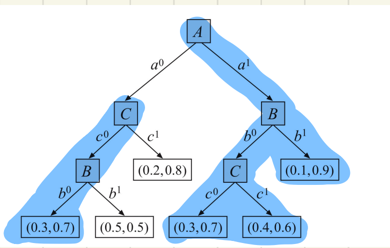


$(J \perp L_2 \mid C = c^1)$



$(J \perp L_1 \mid C = c^2)$
but $C = c^2$ is not the full assignment in the left branch

let $C = c$ be a context, $R[c] = \{\rho[c] : \rho \in R, \rho \sim c\}$

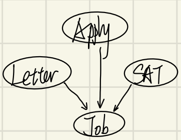


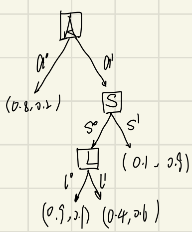
R

$\rho_1:\langle a^1, b^1, x^0; 0.1\rangle$	$\rho_2:\langle a^1, b^1, x^1; 0.9\rangle$
$\rho_3:\langle a^0, c^1, x^0; 0.2\rangle$	$\rho_4:\langle a^0, c^1, x^1; 0.8\rangle$
$\rho_5:\langle b^0, c^0, x^0; 0.3\rangle$	$\rho_6:\langle b^0, c^0, x^1; 0.7\rangle$
$\rho_7:\langle a^1, b^0, c^1, x^0; 0.4\rangle$	$\rho_8:\langle a^1, b^0, c^1, x^1; 0.6\rangle$
$\rho_9:\langle a^0, b^1, c^0; 0.5\rangle$	

$$R[a^1] = \begin{Bmatrix} \rho_1' : \langle b^1, x^0; 0.1 \rangle & \rho_2' : \langle b^1, x^1; 0.9 \rangle \\ \rho_5 : \langle b^0, c^0, x^0; 0.3 \rangle & \rho_6 : \langle b^0, c^0, x^1; 0.7 \rangle \\ \rho_7 : \langle b^0, c^1, x^0; 0.4 \rangle & \rho_8 : \langle b^0, c^1, x^1; 0.6 \rangle \end{Bmatrix}$$

Let R be the rules in the rule-based CPD for a variable X and R[c] be the rules in R that are compatible with c. $Y \subseteq Parents(X)$. If $Y \cap Scope[\rho] = \emptyset \ \forall \rho \in R[c]$, then $(X \perp_c Y \mid Pa(X) - Y, c)$





$\rho_1 : \langle a^0, x^0; 0.8 \rangle$

$\rho_2 : \langle a^0, x^1; 0.2 \rangle$

$\rho_3 : \langle a^1, s^1, x^0; 0.1 \rangle$

$\rho_4 : \langle a^1, s^1, x^1; 0.9 \rangle$

⋮

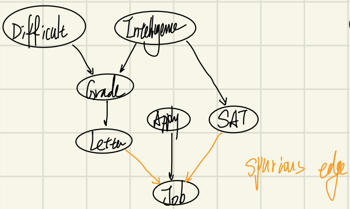
$$R[a^0] = \begin{Bmatrix} \langle a^0, x^0; 0.8 \rangle \\ \langle a^0, x^1; 0.2 \rangle \end{Bmatrix}$$

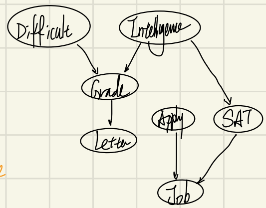
$(J \perp_c S, L \mid A = a^0)$

Let $P(X | Pa(X))$ be a CPD, $Y \in Pa(X)$, and c be a context.

the edge is spurious in the context c if $p(X | Pa(X))$ satisfies $(X \perp_c Y \mid Pa(X) - \{Y\}, c')$

where $c' = c[Pa(X)]$ is the restriction of c to variables in $Pa(X)$





$(J \perp_c L, S \mid A = a^0)$

$\therefore (J \perp_c I \mid A = a^0)$

$c' = \{A = a^0\}$

$(J \not\perp S \mid A = a^1)$

$(J \not\perp I \mid A = a^1)$

Algorithm for computing d-separation in the presence of context-specific CPDs

```
def context_specific_d_separation(G, c, X, Y, Z):
    G' = G
    for each edge Y->X in G'
        if Y->X is spurious given c in G
            Remove Y->X from G'
    return d-sep_G'(X; Y | Z, c)
```

let G be a network structure, let P be a distribution such that $P \models I_\ell(G)$.

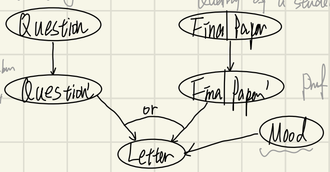
let c be a context, and X, Y, Z be sets of variables.

If X is CSI-separated from Y given Z in context c, then $P \models (X \perp_c Y \mid Z, c)$

Independence of Causal Influence : Noisy-Or Model

Student's participation by asking good question

Quality of a student's final paper



Prof may not remember the student

Prof might fail to recognize student's handwriting

leak probability

Q	F	l^0	l^1
q^0	f^0	1	0
q^0	f^1	0.1	0.9
q^1	f^0	0.2	.8
q^1	f^1	0.02	0.98

Assume $P(l^1 \mid q^1, f^0) = 0.8$ — prof is 80% likely to remember class participation

$P(l^1 \mid q^0, f^1) = 0.9$ — the student's handwriting is 90% readable

$P(l^1 \mid m^1) = 0.0001$ — prof gives a good letter he's happy

$\lambda_Q = P(q^{1\prime} \mid q^1) = 0.8 \quad \lambda_F = P(f^{1\prime} \mid f^1) = 0.9$ noise parameter

when $Q = q^1$, how likely Q is to turn on Q'

Q and F are two independent causal mechanisms for causing a strong letter.

the letter is weak iff neither of these succeeded.

when q^1 and f^1, the probability that both mechanisms fail (independently) is $0.2 \times 0.1 = 0.02$ $\quad P(l^0 \mid q^1, f^1) = 0.02$

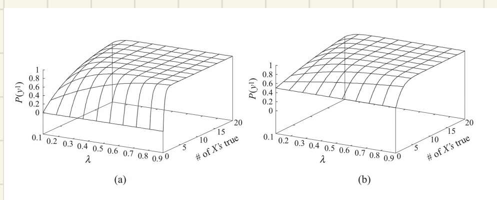
let Y be a binary-valued random variable with k binary-valued parents $X_1 \cdots X_k$

the CPD $P(Y \mid X_1 \cdots X_k)$ is a noisy or if there are k+1 noise parameters $\lambda_0 \cdots \lambda_k$

$$P(y^0 \mid x_1 \cdots x_k) = (1-\lambda_0) \prod_{x_i = x_i^1} (1-\lambda_i)$$ (all "pre-successful" mechanisms fail)

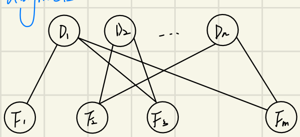
$$= (1-\lambda_0) \prod_i (1-\lambda_i)^{x_i}$$ if we interpret x_i^1 as 1 and x_i^0 as 0

$$P(y^1 \mid x_1 \cdots x_k) = 1 - \left[(1-\lambda_0) \prod_{x_i = x_i^1} (1-\lambda_i)\right]$$



(a) (b)

eg. medical diagnosis



independence assumption: assume that different disease use different causal mechanism

if any disease succeeds in activating its mechanism, the symptom is present

$$P(f_i^0 \mid Pa(F_i)) = (1-\lambda_{i,0}) \prod_{D_j \in Pa(F_i)} (1-\lambda_{i,j})^{d_j}$$

· Independence of Causal Influence : Generalized Linear Model

$Y \in \{0, 1\}$

examine a CPD $P(Y \mid X_1, \dots, X_k)$. Assume that the effect of the X_i's on Y can be summarized via a linear function $f(X_1, \dots X_k) = W_0 + \sum_{i=1}^{k} W_i X_i$

each invader adds to the burden of immune system

$$P(y^1 \mid X_1, \dots, X_k) = \mathrm{Sigmoid}\left(W_0 + \sum_{i=1}^{k} W_i X_i\right)$$

$$O(X) = \frac{P(y^1 \mid X_1 \cdots X_k)}{P(y^0 \mid X_1 \cdots X_k)} = \frac{\exp(w^T x) \,/\, (1 + \exp(w^T x))}{1 \,/\, (1 + \exp(w^T x))} = \exp(w^T x)$$

$Y \in \{y^1, y^2, \cdots y^m\}$

let Y be an m-valued random variable with k parents $X_1, \dots, X_k$ that take on numerical values

the CPD $P(Y \mid X_1, \dots, X_k)$ is a multinomial logistic if $\forall j = 1, \dots m$, there are $k+1$ weights $W_{j,0}, W_{j,1}, \cdots W_{j,k}$ such that

$$l_j(X_1, \dots X_k) = W_{j,0} + \sum_{i=1}^{k} W_{j,i} X_i$$

$$P(y^j \mid X_1, \dots X_k) = \frac{\exp(l_j(X_1 \cdots X_k))}{\sum_{j'=1}^{m} \exp(l_{j'}(X_1 \cdots X_k))}$$

· Independence of Causal Influence: General Formulation

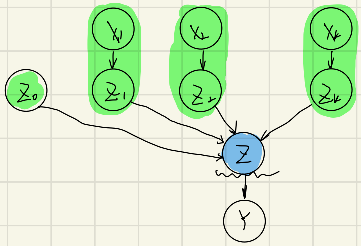
let Y be a random variable with parents $X_1, \dots, X_k$. The CPD $P(Y \mid X_1, \dots, X_k)$ exhibits independence of causal influence (ICI) if it's described via a network fragment of the structure below, where CPD of Z is a deterministic function

Generalized Linear Model

$Z_i = W_i X_i$

$Z = Z_0 + \sum_i W_i X_i$

$Y = \mathrm{sigmoid}(Z)$



each variable X_i can be transformed separately using its own individual noise model

we say $P(Y \mid X_1 \cdots X_k)$ exhibits symmetric ICI if the CPD of Z is a deterministic symmetric decomposable f

$P(Y \mid X_1 \cdots X_k)$ exhibits fully symmetric ICI if CPDs of different Z_i variables are identical

A deterministic binary function $x \diamond y$ is $\begin{cases} \text{commutative if } x \diamond y = y \diamond x \\ \text{associative if } (x \diamond y) \diamond z = x \diamond (y \diamond z) \end{cases}$

A function $f(x_1, \dots, x_k)$ is a symmetric decomposable function if there's a commutative associative function $x \diamond y$ such that $f(x_1 \cdots x_k) = x_1 \diamond x_2 \cdots x_k$

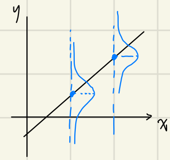
· Continuous Variables

let Y be a continuous variable with continuous parents $X_1, \dots, X_k$

Y has a linear Gaussian model if there are parameters $\beta_0, \dots, \beta_k$ and σ^2 such that

$$p(Y \mid x_1, \dots, x_k) = \mathcal{N}\left(\beta_0 + \sum_{i}^{k} \beta_i x_i \, ; \, \sigma^2\right)$$

$$Y = \beta_0 + \beta^T X + \epsilon \qquad \epsilon \sim \mathcal{N}(0, \sigma^2)$$



Continuous parent
Continuous child

· Hybrid Models

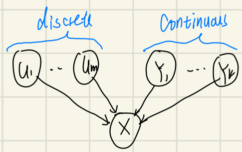
Let X be a continuous variable, $U = \{U_1, \cdots U_m\}$ be X's discrete parents and $Y = \{Y_1, \dots, Y_k\}$ be its continuous parents.

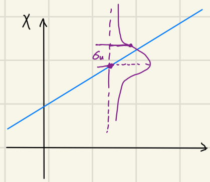
X has a conditional linear Gaussian (CLG) CPD if

$\forall u \in val(U)$, there're $k+1$ coefficients $a_{u,0} \cdots a_{u,k}$ and a variance σ_u^2 s.t.

$$p(X \mid u, y) = \mathcal{N}\left(a_{u,0} + \sum_{i=1}^{k} a_{u,i} \, y_i \, ; \, \sigma_u^2\right)$$

Continuous & discrete parents
Continuous child





$X = a_{u,0} + \sum_{i=1}^{k} a_{u,i} \cdot y_i$

A Bayesian network is called a CLG network if every discrete variable has only discrete parents and every continuous variable has a CLG CPD

· Conditional Bayesian Network

A conditional Bayesian network B over Y given X is a directed acyclic graph G whose nodes are $X \cup Y \cup Z$ where X, Y, Z are disjoint. X are inputs, Y are outputs, Z are encapsulated variables in X have parents in G. The variables in $Z \cup Y$ are associated with a CPD

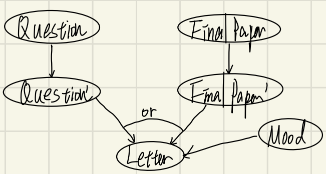
the network defines a CPD using a chain rule

$$P_B(Y, Z \mid X) = \prod_{U \in Y \cup Z} P(U \mid Pa(U))$$

Distribution $P_B(Y \mid X)$ is defined as marginal of $p(Y, Z \mid X)$

$$P_B(Y \mid X) = \sum_{Z} P_B(Y, Z \mid X)$$

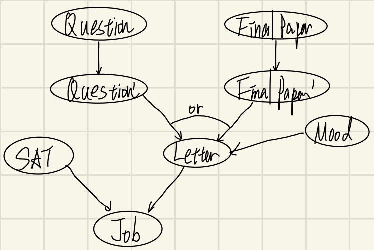
Conditional random field is the undirected analog of conditional Baysian Network



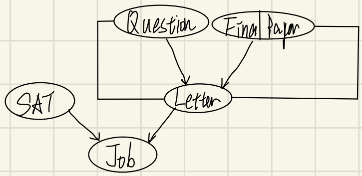
This network specifies not a joint distribution over r.v.s in the fragment (hence it's not a complete BN)

But a conditional distribution $P(L \mid Q, F)$

let Y be a random variable with k parents $X_1, \dots, X_k$. The CPD $P(Y|X_1, \dots, X_k)$ is an encapsulated CPD if it's represented using a Conditional Bayesian Network over Y given $X_1 \dots X_k$



$\Rightarrow$



Externally, to the rest of the network, we can still view Letter as a var with parent Question Final Paper