# IFT 615 – Intelligence artificielle

## Raisonnement probabiliste

Hugo Larochelle
Département d'informatique
Université de Sherbrooke
http ://www.dmi.usherb.ca/~larocheh/cours/ift615.html

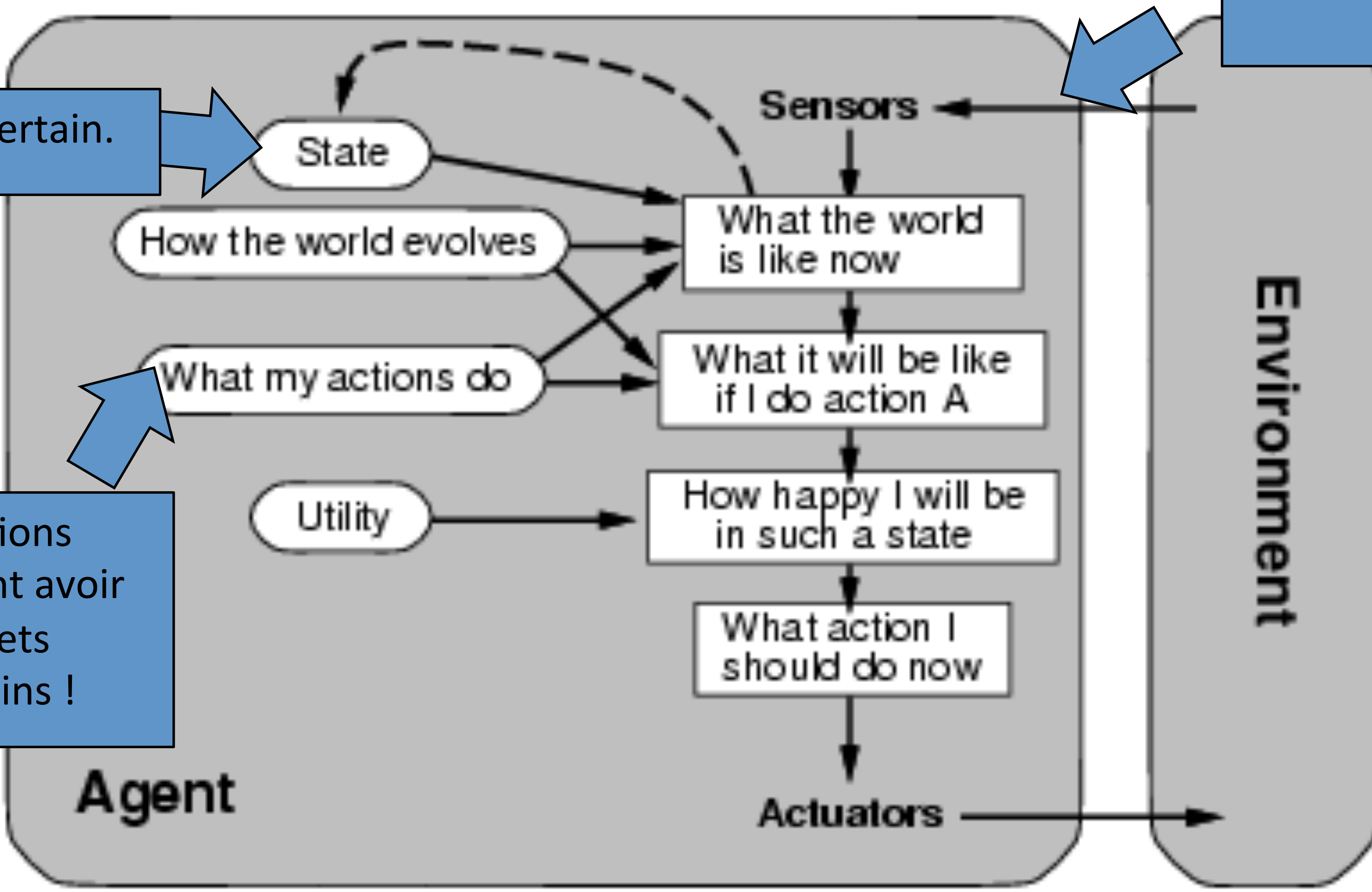
# Sujets couverts

- Introduction au raisonnement probabiliste
  - raisonnement avec incertitude
  - révision des concepts de base en théorie des probabilités

# Rappel : Utility-based agents

Les capteurs peuvent être bruités…

État incertain.




Les actions peuvent avoir des effets incertains !

# Incertitude

- Soit $A_t$ l'action d'aller à l'aéroport $t$ minutes avant le départ de l'avion
- $A_t$ me permettra-t-il d'arriver à temps?
- Problèmes :
  - observabilité partielle (conditions routières, etc.)
  - senseurs bruités (annonces du trafic, etc.)
  - incertitude dans l'effet des actions (crevaisons, pannes, etc.)
  - immense complexité pour modéliser les actions et le trafic
- Un raisonnement purement logique et déterministe :
  - risque de tirer des conclusions erronées
    - « $A_{25}$ me permettra d'arriver à temps » (impossible de faire cette garantie)
  - risque de tirer des conclusions peu exploitables du point de vue de la prise de décision
    - « $A_{25}$ me permettra d'arriver à temps, s'il ne pleut pas, s'il n'y a pas d'accident, si mes pneus ne crèvent pas, etc. »
    - « $A_{1440}$ me permettra presque certainement d'arriver à temps, mais je devrai passer une nuit à l'aéroport. »

# Modéliser l'incertitude à l'aide probabilités

- **Théorie des probabilités**
  - permet de modéliser la vraisemblance d'événements
    - l'information sur la vraisemblance est dérivée
      - des croyances/certitudes d'un agent, ou
      - d'observations empiriques de ces événements
  - donne un cadre théorique pour mettre à jour la vraisemblance d'événements après l'acquisition d'observations
    - après avoir observé qu'il n'y a pas de trafic, la probabilité que $A_{25}$ me permette d'arriver à temps doit changer comment ?
  - facilite la modélisation en permettant de considérer l'influence de phénomènes complexes comme du « bruit »

# Exemple : un patient a-t-il une carie ?

- On considère le patient d'un dentiste
  - on souhaite raisonner sur la possibilité qu'il ait une carie en tenant compte de l'incertitude associée à un tel diagnostique

| ***MalDeDent*** | ***Croche*** | ***Carie*** | **Probabilité** |
|---|---|---|---:|
| *vrai* | *vrai* | *vrai* | 0.108 |
| *vrai* | *vrai* | *faux* | 0.016 |
| *vrai* | *faux* | *vrai* | 0.012 |
| *vrai* | *faux* | *faux* | 0.064 |
| *faux* | *vrai* | *vrai* | 0.072 |
| *faux* | *vrai* | *faux* | 0.144 |
| *faux* | *faux* | *vrai* | 0.008 |
| *faux* | *faux* | *faux* | 0.576 |

# Variable aléatoire

- **Variables aléatoires** :
  - ***MalDeDent*** : est-ce que le patient a mal aux dents
  - ***Croche*** : est-ce que le dentiste a trouvé un « trou » dans la dent avec sa sonde
  - ***Carie*** : est-ce que le patient a une carie

| ***MalDeDent*** | ***Croche*** | ***Carie*** | **Probabilité** |
|---|---|---|---|
| *vrai* | *vrai* | *vrai* | 0.108 |
| *vrai* | *vrai* | *faux* | 0.016 |
| *vrai* | *faux* | *vrai* | 0.012 |
| *vrai* | *faux* | *faux* | 0.064 |
| *faux* | *vrai* | *vrai* | 0.072 |
| *faux* | *vrai* | *faux* | 0.144 |
| *faux* | *faux* | *vrai* | 0.008 |
| *faux* | *faux* | *faux* | 0.576 |

« la probabilité que *MalDeDent=vrai* **et** *Croche=vrai* **et** *Carie=vrai* »

toutes ces probabilités somment à 1

# Univers et événement élémentaire

- **Événement élémentaire** $\omega$ : un état possible de l'environnement
  - c'est une rangée de la table ci-dessous, un événement au niveau le plus simple
- **Univers** $\Omega$ : l'ensemble des événements élémentaires possibles
  - c'est l'ensemble de toutes les rangées

| ***MalDeDent*** | ***Croche*** | ***Carie*** | **Probabilité** |
|---|---|---|---|
| *vrai* | *vrai* | *vrai* | 0.108 |
| *vrai* | *vrai* | *faux* | 0.016 |
| *vrai* | *faux* | *vrai* | 0.012 |
| *vrai* | *faux* | *faux* | 0.064 |
| *faux* | *vrai* | *vrai* | 0.072 |
| *faux* | *vrai* | *faux* | 0.144 |
| *faux* | *faux* | *vrai* | 0.008 |
| *faux* | *faux* | *faux* | 0.576 |

# Probabilité conjointe

- **Probabilités conjointes** : probabilité assignation de toutes la variables
  - *P*(*MalDeDent=vrai*, *Croche=vrai*, *Carie=vrai*) = 0.108 (10.8%)
  - *P*(*MalDeDent=faux*, *Croche=faux*, *Carie=vrai*) = 0.008 (0.8%)

| ***MalDeDent*** | ***Croche*** | ***Carie*** | **Probabilité** |
|---|---|---|---|
| *vrai* | *vrai* | *vrai* | 0.108 |
| *vrai* | *vrai* | *faux* | 0.016 |
| *vrai* | *faux* | *vrai* | 0.012 |
| *vrai* | *faux* | *faux* | 0.064 |
| *faux* | *vrai* | *vrai* | 0.072 |
| *faux* | *vrai* | *faux* | 0.144 |
| *faux* | *faux* | *vrai* | 0.008 |
| *faux* | *faux* | *faux* | 0.576 |

# Probabilité marginale

- **Probabilités marginales** : probabilité sur un sous-ensemble des variables
  - *P*(*MalDeDent=vrai, Carie=vrai*)
    = *P*(*MalDeDent=vrai, Croche=vrai, Carie=vrai*) + *P*(*MalDeDent=vrai, Croche=faux, Carie=vrai*)
    = $\Sigma_{x \in \{vrai,\, faux\}}$ *P*(*MalDeDent=vrai, Croche=x, Carie=vrai*) = 0.108 + 0.012 = **0.12**

| ***MalDeDent*** | ***Croche*** | ***Carie*** | **Probabilité** |
|---|---|---|---|
| *vrai* | *vrai* | *vrai* | 0.108 |
| *vrai* | *vrai* | *faux* | 0.016 |
| *vrai* | *faux* | *vrai* | 0.012 |
| *vrai* | *faux* | *faux* | 0.064 |
| *faux* | *vrai* | *vrai* | 0.072 |
| *faux* | *vrai* | *faux* | 0.144 |
| *faux* | *faux* | *vrai* | 0.008 |
| *faux* | *faux* | *faux* | 0.576 |

# Probabilité marginale

- **Probabilités marginales** : probabilité sur un sous-ensemble des variables
  - $P(Carie{=}vrai)$
    $= \Sigma_{x \in \{vrai,\, faux\}} \Sigma_{y \in \{vrai,\, faux\}}\ P(MalDeDent{=}vrai,\, Croche{=}x,\, Carie{=}y)$
    $= 0.108 + 0.012 + 0.072 + 0.008 =$ **0.2**

| *MalDeDent* | *Croche* | *Carie* | **Probabilité** |
|---|---|---|---:|
| *vrai* | *vrai* | *vrai* | 0.108 |
| *vrai* | *vrai* | *faux* | 0.016 |
| *vrai* | *faux* | *vrai* | 0.012 |
| *vrai* | *faux* | *faux* | 0.064 |
| *faux* | *vrai* | *vrai* | 0.072 |
| *faux* | *vrai* | *faux* | 0.144 |
| *faux* | *faux* | *vrai* | 0.008 |
| *faux* | *faux* | *faux* | 0.576 |

# Probabilité d'une disjonction

- **Probabilités de disjonction (« ou ») d'événements** :
  - *P*(*Carie=vrai* ou *MalDeDent=faux*)
    = *P*(*Carie=vrai*) + *P*(*MalDeDent=faux*) - *P*(*Carie=vrai, MalDeDent=faux*)
    = 1 - *P*(*Carie=**faux**, MalDeDent=**vrai***) = 1 – 0.016 – 0.064 = **0.92**

| ***MalDeDent*** | ***Croche*** | ***Carie*** | **Probabilité** |
|---|---|---|---:|
| *vrai* | *vrai* | *vrai* | 0.108 |
| *vrai* | *vrai* | *faux* | 0.016 |
| *vrai* | *faux* | *vrai* | 0.012 |
| *vrai* | *faux* | *faux* | 0.064 |
| *faux* | *vrai* | *vrai* | 0.072 |
| *faux* | *vrai* | *faux* | 0.144 |
| *faux* | *faux* | *vrai* | 0.008 |
| *faux* | *faux* | *faux* | 0.576 |

# Probabilité d'un événement en général

- On peut calculer la probabilité d'événements arbitrairement complexes
  - il suffit d'additionner les probabilités des éléments élémentaires associés
  - *P*( (*Carie=vrai, MalDeDent=faux*) ou (*Croche=faux, Carie=faux*) )
    = 0.064 + 0.072 + 0.008 + 0.576 = **0.72**

| ***MalDeDent*** | ***Croche*** | ***Carie*** | **Probabilité** |
|---|---|---|---|
| *vrai* | *vrai* | *vrai* | 0.108 |
| *vrai* | *vrai* | *faux* | 0.016 |
| *vrai* | *faux* | *vrai* | 0.012 |
| *vrai* | *faux* | *faux* | 0.064 |
| *faux* | *vrai* | *vrai* | 0.072 |
| *faux* | *vrai* | *faux* | 0.144 |
| *faux* | *faux* | *vrai* | 0.008 |
| *faux* | *faux* | *faux* | 0.576 |

# Probabilité conditionnelle

vrai seulement si
*P*(*MalDeDent=vrai*) ≠ 0

- **Probabilités conditionnelles** :
  - *P*(*Carie=faux* | *MalDeDent=vrai*)
    = *P*(*Carie=faux, MalDeDent=vrai*) / *P*(*MalDeDent=vrai*)
    = (0.016 + 0.064) / (0.016 + 0.064 + 0.108 + 0.012) = **0.4**

| ***MalDeDent*** | ***Croche*** | ***Carie*** | **Probabilité** |
|---|---|---|---|
| *vrai* | *vrai* | *vrai* | 0.108 |
| *vrai* | *vrai* | *faux* | 0.016 |
| *vrai* | *faux* | *vrai* | 0.012 |
| *vrai* | *faux* | *faux* | 0.064 |
| *faux* | *vrai* | *vrai* | 0.072 |
| *faux* | *faux* | *vrai* | 0.008 |
| *faux* | *faux* | *faux* | 0.576 |

En mots: « sachant que *MalDeDent=vrai*, quelle est la probabilité que *Carie=faux* »

# Distribution de probabilités

- **Distribution de probabilités :** l'énumération des probabilités pour toutes les valeurs possibles de variables aléatoires
- Exemples :
  - **P**(*Carie*) = [ *P*(*Carie=faux*), *P*(*Carie=vrai*) ] = [ 0.8, 0.2 ]
  - **P**(*Carie, MalDeDent*)
    = [ [ *P*(*Carie=faux* , *MalDeDent=faux*), *P*(*Carie=vrai*, *MalDeDent=faux*) ],
    [ *P*(*Carie=faux* , *MalDeDent=vrai*), *P*(*Carie=vrai*, *MalDeDent=vrai*) ] ]
    = [ [ 0.72, 0.08],
    [0.08, 0.12] ]
- La somme est toujours égale à 1
- J'utilise le symbole **P** pour les distributions et *P* pour les probabilités
  - *P*(*Carie*) désignera la probabilité *P*(*Carie*=*x*) pour une valeur *x* non-spécifiée
    - » c'est un élément quelconque de **P**(*Carie*)
- Le choix d'énumérer les probabilités dans un tableau 2D est arbitraire

# Distribution conditionnelle

- On peut faire la même chose pour le cas conditionnel
- Exemple :
  - **P**(*Carie* | *MalDeDent=vrai*)
    = [ *P*(*Carie=faux* | *MalDeDent=vrai*), *P*(*Carie=vrai* | *MalDeDent=vrai*) ]
    = [ 0.4, 0.6 ]
  - **P**(*Carie* | *MalDeDent*)
    = [ [ *P*(*Carie=faux* | *MalDeDent=faux*), *P*(*Carie=vrai* | *MalDeDent=faux*)],
    [ *P*(*Carie=faux* | *MalDeDent=vrai*), *P*(*Carie=vrai* | *MalDeDent=vrai*)] ]
    = [ [ 0.9, 0.1], → somme à 1
    [0.4, 0.6] ] → somme à 1
- **Chaque sous-ensemble de probabilités** associé aux mêmes valeurs des variables sur lesquelles on conditionne somme à 1

  **P**(*Carie* | *MalDeDent*) contient deux distributions de probabilités sur la variable *Carie* : une dans le cas où *MalDeDent=faux,* l'autre lorsque *MalDeDent=vrai*

# Distribution conditionnelle

- Une distribution conditionnelle peut être vue comme une distribution **renormalisée** afin de satisfaire les conditions de sommation à 1
- Exemple :
  - **P**(*Carie* | *MalDeDent=vrai*)
    = **P**(*Carie, MalDeDent=vrai*) / α
    = [0.08, 0.12] / α
    = [0.08, 0.12] / (0.08 + 0.12)
    = [ 0.4, 0.6 ]

  - **P**(*Carie* | *MalDeDent*)
    = [ **P**(*Carie, MalDeDent=faux*) / $\alpha_{faux}$,
    **P**(*Carie, MalDeDent=vrai*) / $\alpha_{vrai}$ ]
    = [ [ 0.72, 0.08] / (0.72 + 0.08),
    [0.08, 0.12] / (0.08 + 0.12) ]
    = [ [ 0.9, 0.1],
    [0.4, 0.6] ]

# Règle de chaînage

- **Règle du produit :**
  - $P(Carie{=}faux, MalDeDent{=}vrai)$
    $= P(Carie{=}faux \mid MalDeDent{=}vrai)\ P(MalDeDent{=}vrai)$
    $= P(MalDeDent{=}vrai \mid Carie{=}faux)\ P(Carie{=}faux)$
  - En général :
    $P(Carie, MalDeDent) = P(Carie \mid MalDeDent)\ P(MalDeDent)$
    $= P(MalDeDent \mid Carie)\ P(Carie)$

- **Règle de chaînage (*chain rule*)** pour $n$ variables $X_1 \ldots X_n$ :
  - $$\begin{aligned} P(X_1, \ldots, X_n) &= P(X_1, \ldots, X_{n-1})\ P(X_n \mid X_1, \ldots, X_{n-1}) \\ &= P(X_1, \ldots, X_{n-2})\ P(X_{n-1} \mid X_1, \ldots, X_{n-2})\ P(X_n \mid X_1, \ldots, X_{n-1}) \\ &= \ldots \\ &= \prod_{i=1..n} P(X_i \mid X_1, \ldots, X_{i-1}) \end{aligned}$$

# Règle de chaînage

- La règle du chaînage est vraie, quelle que soit la distribution de $X_1 \ldots X_n$
- Plutôt que de spécifier toutes les probabilités jointes $P(X_1, \ldots, X_n)$, on pourrait plutôt spécifier $P(X_1)$, $P(X_2|X_1)$, $P(X_3|X_1, X_2)$, ..., $P(X_n \mid X_1,\ldots,X_{n-1})$
- Exemple, on aurait pu spécifier :
  - *P*(*Carie=faux*) = 0.8, *P*(*Carie=vrai*) = 0.2
  - *P*(*MalDeDent=faux*| *Carie=faux*) = 0.9 , *P*(*MalDeDent=vrai*| *Carie=faux*) = 0.1
    *P*(*MalDeDent=faux*| *Carie=vrai*) = 0.4, *P*(*MalDeDent=vrai* | *Carie=vrai*) = 0.6
- On aurait tout les ingrédients pour calculer les *P*(*Carie*, *MalDeDent*) :
  - *P*(*Carie=faux*, *MalDeDent=vrai*) = *P*(*MalDeDent=vrai* |*Carie=faux*) *P*(*Carie=faux*)
    = 0.1 * 0.8 = 0.08
  - *P*(*Carie=vrai*, *MalDeDent=vrai*) = *P*(*MalDeDent=vrai*|*Carie=vrai*) *P*(*Carie=vrai*)
    = 0.6 * 0.2 = 0.12

# Règle de Bayes

*Carie=faux* ⇔ ¬carie
*Carie=vrai* ⇔ carie

- On pourrait aussi calculer $P(Carie{=}faux \mid MalDeDent{=}vrai)$ :
  - $P(\neg carie \mid malDeDent)$
    $= P(\neg carie, malDeDent) / P(malDeDent)$
    $= P(\neg carie, malDeDent) / (\, P(malDeDent, \neg carie) + P(malDeDent, carie))$
    $= P(malDeDent \mid \neg carie)\, P(\neg carie) / \alpha$
    $= 0.08 / (0.08 + 0.12) = \mathbf{0.4}$
- On appelle $P(Carie)$ une probabilité a priori
  - c'est notre croyance p/r à la présence d'une carie **avant** toute observation
- On appelle $P(Carie \mid MalDeDent)$ une probabilité a posteriori
  - c'est notre croyance mise à jour après avoir observé que *MalDeDent*
- **La règle de Bayes** lie ces deux probabilités ensemble
  - $\underline{P(\neg carie \mid malDeDent)} = P(malDeDent \mid \neg carie)\, \underline{P(\neg carie)} / \alpha$
- Donne une probabilité **diagnostique** à partir d'une probabilité **causale** :
  - $P(Cause \mid Effect) = P(Effect \mid Cause)\, P(Cause) / P(Effect)$

# Indépendance

- Soit les variables *A* et *B*, elles sont **indépendantes** si et seulement si
  - $P(A|B) = P(A)$ ou
  - $P(B|A) = P(B)$ ou
  - $P(A, B) = P(A)\, P(B)$

- Exemple : $P(Pluie, Carie) = P(Pluie)\, P(Carie)$

***P(Pluie = vrai) = 0.3***

***P(Carie = vrai) = 0.1***

| ***Pluie*** | ***Carie*** | **Probabilité** | |
|---|---|---|---|
| *vrai* | *vrai* | 0.03 | $= P(pluie)\, P(carie) = 0.3 * 0.1$ |
| *vrai* | *faux* | 0.27 | $= P(pluie)\, P(\neg carie) = 0.3 * 0.9$ |
| *faux* | *vrai* | 0.07 | $= P(\neg pluie)\, P(carie) = 0.7 * 0.1$ |
| *faux* | *faux* | 0.63 | $= P(\neg pluie)\, P(\neg carie) = 0.7 * 0.9$ |

# Indépendance

- L'indépendance totale est puissante mais rare

- L'indépendance entre les variables permet de réduire la taille de la distribution de probabilités et rendre les inférences plus efficaces
  - dans l'exemple précédent, on n'a qu'à stocker en mémoire $P(Pluie = vrai) = 0.3$ et $P(Carie = vrai) = 0.1$, plutôt que la table au complet

- Mais il est rare d'être dans une situation où toutes les variables sont réellement indépendantes

# Indépendance conditionnelle

- Si j'ai une carie, la probabilité que la sonde accroche dans la dent ne dépend pas du fait que j'aie mal à la dent ou non :
  - *P*(*Croche* | *MalDeDents*, *Carie=vrai*) = *P*(*Croche* | *Carie=vrai*)
- Même chose si je n'ai pas la carie :
  - *P*(*Croche* | *MalDeDents*, *Carie=faux*) = *P*(*Croche* | *Carie=faux*)

- On dit que *Croche* est **conditionnellement indépendante** de *MalDeDents* étant donné *Carie*, puisque :
  - *P*(*Croche* | *MalDeDents*, *Carie*) = *P*(*Croche* | *Carie*)
- Formulations équivalentes :
  - *P*(*MalDeDents* | *Croche* , *Carie*) = *P*(*MalDeDents* |*Carie*)
  - *P*(*MalDeDents*, *Croche* | *Carie*) = *P*(*MalDeDents* |*Carie*) *P*(*Croche*|*Carie*)

# Indépendance conditionnelle

- Réécrivons la distribution conjointe en utilisant la **règle de chaînage** (***chain rule***) :

$$
\begin{aligned}
&P(MalDeDents, Croche, Carie) \\
&\quad = P(MalDeDents \mid Croche, Carie)\, P(Croche, Carie) \\
&\quad = P(MalDeDents \mid Croche, Carie)\, P(Croche \mid Carie)\, P(Carie) \\
&\quad = P(MalDeDents \mid Carie)\, P(Croche \mid Carie)\, P(Carie)
\end{aligned}
$$

- C-à-d., 2 + 2 + 1 = 5 **paramètres individuels/distincts**

- Dans des cas idéals, l'exploitation de l'indépendance conditionnelle réduit la complexité de représentation de la distribution conjointe de exponentielle ($O(2^n)$) en linéaire ($O(n)$)
- En raisonnement probabiliste, l'indépendance conditionnelle est le concept de représentation des connaissances le plus basique et utile

# Autres types de variables aléatoires

- On s'est concentré sur des variables aléatoires Booléennes ou binaires
  - le **domaine**, c.-à-d. l'ensemble des valeurs possibles de la variable, était toujours {*vrai*,*faux*}

- On pourrait avoir d'autres types de variables, avec des domaines différents :
  - **Discrètes** : le domaine est énumérable
    - *Météo* ∈ {*soleil*, *pluie*, *nuageux*, *neige*}
    - lorsqu'on marginalise, on doit sommer sur toutes les valeurs :
      $P(Température) = \Sigma_{x \in \{soleil,\ pluie,\ nuageux,\ neige\}} P(Température, Météo{=}x)$
  - **Continues** : le domaine est continu (par exemple, l'ensemble des réels)
    - exemple : *PositionX* = 4.2
    - le calcul des probabilités marginales nécessite des intégrales

# En bref

- **Probabilité jointe :** $P(X_1, \ldots, X_n)$
- **Probabilité marginale :** $P(X_i)$, $P(X_i, X_j)$, etc.
- **Probabilité conditionnelle :** $P(X_1, \ldots, X_k \mid X_{k+1}, \ldots, X_n) = \dfrac{P(X_1, \ldots, X_k, X_{k+1}, \ldots, X_n)}{P(X_{k+1}, \ldots, X_n)}$
- **Régle de chaînage :** $P(X_1, \ldots, X_n) = \prod_{i=1..n} P(X_i \mid X_1, \ldots, X_{i-1})$
- **Indépendance :** $X_i$ et $X_j$ sont indépendantes si

  $P(X_i, X_j) = P(X_i)\, P(X_j)$, ou $P(X_i \mid X_j) = P(X_i)$ ou $P(X_j \mid X_i)$
- **Indépendance conditionnelle :** $X_i$ et $X_j$ sont indépendante sachant $X_k$ si

  $P(X_i, X_j \mid X_k) = P(X_i \mid X_k)\, P(X_i \mid X_k)$ ou $P(X_i \mid X_j, X_k) = P(X_i \mid X_k)$ ou $P(X_j \mid X_i, X_k) = P(X_j \mid X_k)$
- **Règle de Bayes :** $P(X_1, \ldots, X_k \mid X_{k+1}, \ldots, X_n) = \dfrac{P(X_{k+1}, \ldots, X_n \mid X_1, \ldots, X_k)\, P(X_1, \ldots, X_k)}{P(X_{k+1}, \ldots, X_n)}$

# Résumé

- La théorie des probabilités est un formalisme cohérent pour raisonner avec l'incertitude
- Une distribution conjointe spécifie les probabilités pour toute valuer des variables aléatoires
- Pour les domaines d'application réalistes, on doit trouver une façon de réduire la taille de la distribution conjointe
- L'indépendance et l'indépendance conditionnelles nous fournissent les outils de base pour simplifier les distributions conjointes

# Vous devriez être capable de...

- À partir d'une distribution conjointe ou des distributions conditionnelles et a priori nécessaires :
  - calculer une probabilité conjointe
  - calculer une probabilité marginale
  - déterminer si deux variables sont indépendantes
  - déterminer si deux variables sont conditionnellement indépendantes sachant une troisième