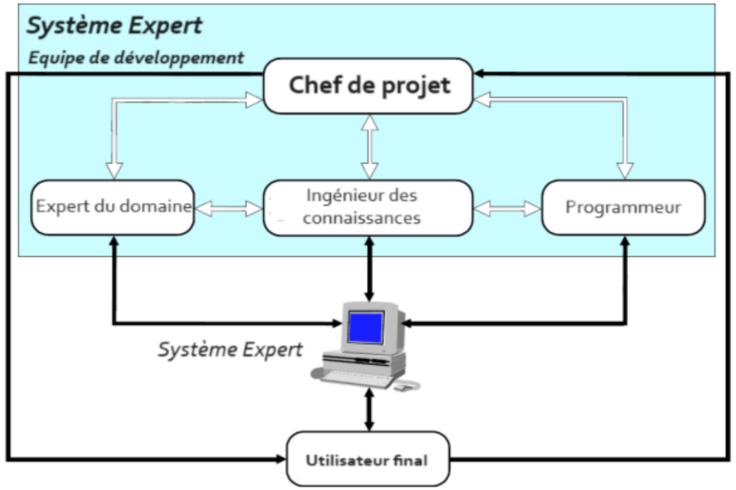
Probabilité et introduction à l’incertitude

Dr. NSENGE MPIA HERITIER, Ph.D

Précédemment

- Rôles dans le développement des systèmes experts
- Acquisition et ingénierie des connaissances
- Introduction à Experta
  - Exemple météorologique
  - Exemple de diagnostic de maladie




```python
from experta import *
# Définition des faits
class Symptome(Fact):
    pass
class Diagnostic(Fact):
    pass
# Définition des règles
class ExpertGrippe(KnowledgeEngine):
    @DefFacts()
    def initial_facts(self):
        yield Diagnostic(input("Saisir le nom de la maladie à diagnostiquer (ex. grippe): "))
    @Rule(Diagnostic('grippe'))
    def demande_symptomes(self):
        if input("Avez-vous la fièvre? ") == 'oui':
            self.declare(Symptome(name='fièvre', value=True))
        else:
            self.declare(Symptome(name='fièvre', value=False))
        if input("Avez-vous les douleurs corporelles? ") == 'oui':
            self.declare(Symptome(name='douleurs_corporelles', value=True))
        else:
            self.declare(Symptome(name='douleurs_corporelles', value=False))
    @Rule(Symptome(name='fièvre', value=True),
          Symptome(name='douleurs_corporelles', value=True))
    def regle_grippe(self):
        print("Vous avez la grippe.")
    @Rule(Symptome(name='fièvre', value=False),
          Symptome(name='douleurs_corporelles', value=False))
    def regle_pas_grippe(self):
        print("Vous n'avez pas de grippe.")
    @Rule(OR(Symptome(name='fièvre', value=False),
             Symptome(name='douleurs_corporelles', value=False)))
    def regle_inconnu(self):
        print("Maladie inconnue.")
# Création de l'instance de l'expert
expert_system = ExpertGrippe()
# Reinitialisation des faits
expert_system.reset()
# Lancement du système
expert_system.run()
```

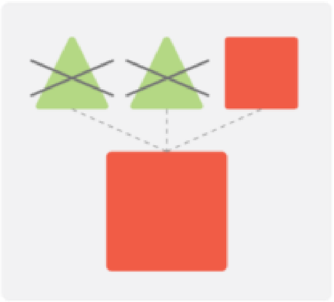
Plan de la leçon

- Théorie de base des probabilités
- Statistique inférentielle
- Hypothèses d'indépendance
- Rapports/Ratio de vraisemblance
- Facteurs de certitude
- Incertitude dans les systèmes experts

Qu’est-ce que et pourquoi l’ncertitude ?

- Pourquoi se soucier de l'incertitude ?
- Jusqu'à présent:
  - nous nous sommes concentrés exclusivement sur le raisonnement déductif.
  - En d'autres termes, les situations où les conclusions sont absolument vraies ou fausses.
- Alors que dans de nombreux problèmes du monde réel, nous avons besoin d'un moyen de conclure que quelque chose est probablement, mais pas nécessairement, vrai ou exclusivement vrai.
  - Par exemple, certaines observations, telles que des symptômes, sont généralement, mais pas toujours, associées à une cause ou à une maladie spécifique.
- C'est souvent le cas pour les applications de systèmes experts dans le domaine de la médecine.
- Dans ce cours, nous définissons simplement l'incertitude comme la ***qualité*** ou l'***état*** d'une information qui n'est pas clairement connue.
  - C'est cette idée d'incertitude qui distingue la connaissance déductive, par exemple ce que nous voyons en mathématiques, de la croyance inductive qui est le fondement de la plupart des sciences.

Rappel: Raisonnement déductif



- Général → Spécifique   c'est-à-dire "Logique descendante"

- **Raisonner à partir d'une ou plusieurs affirmations pour parvenir logiquement à certaine conclusion**

- Le raisonnement va dans le même sens que celui des conditionnels
  - L'inverse est vrai pour le raisonnement abductif
- Une conclusion est obtenue de manière réductrice :
  - Appliquer des règles générales qui s'appliquent à l'ensemble d'un domaine fermé, en réduisant l'éventail considéré jusqu'à ce qu'il ne reste que la (les) conclusion(s).
- Les syllogismes sont une forme courante de raisonnement déductif.
- Caractéristique des systèmes experts

Rappel: Raisonnement inductif

- Spécifique → général, c'est-à-dire "logique ascendante".

- **Raisonner à partir de preuves concernant certains membres d'une classe afin de formuler une conclusion concernant tous les membres d'une classe.**

- Une conclusion tirée par un raisonnement inductif est appelée "hypothèse".
- Toujours moins sûre que la preuve elle-même
  - c'est-à-dire que la conclusion est probable mais incertaine
- Le raisonnement inductif n'est pas ici le même que l'induction dans les preuves mathématiques
  - L'induction mathématique est en fait une forme de raisonnement déductif.
- L'apprentissage automatique (ML) repose sur le raisonnement inductif
  - Utilisation d'un ensemble de cas spécifiques pour trouver des généralisations utiles

Rappel: Raisonnement inductif (Cont.)

- **Syllogisme statistique :**
  - **Lien correct entre les prémisses et la conclusion : Fort ou faible**
  - **Prémisses vraies : Convaincant/fiable vs Non convaincant/non fiable**

90 % des villes congolaises n’ont pas d’eau potable.

Kinshasa est une ville congolaise.

C'est tout ce que nous savons à ce sujet.

∴ Il est probable à 90% que Kinshasa n’ait pas d’eau potable

Sources de connaissances incertaines

- Tout d'abord, parlons de certaines sources d'incertitude.
- Nous pouvons avoir l'incertitude sur les inputs:
  - Par exemple, l'incertitude peut être introduite par:
    - des données manquantes
    - des données bruitées
    - ou l'utilisation d'un langage imprécis
      - Notre langage naturel est ambigu.
      - Les faits sont décrits avec des termes tels que "souvent", "parfois", "fréquemment" et "presque jamais".
- Connaissances incertaines : Une connaissance incertaine peut résulter de
  - combinaison des points de vue de différents experts
  - des causes multiples qui entraînent des effets multiples
  - l'énumération incomplète des conditions ou des effets (par exemple, tâches complexes)
  - connaissance incomplète de la causalité dans le domaine (implications faibles)
  - effets probabilistes/stochastiques (c'est-à-dire aléatoires)
- Nous pouvons avoir l'incertitude sur les outputs :
  - l'utilisation de l'abduction et de l'induction (raisonnement) qui sont intrinsèquement incertaines
  - l'utilisation du raisonnement par défaut, même déductif, est incertain
  - avoir une inférence déductive incomplète peut être incertaine.

Evaluation de l'imprécision du langage (0-100)

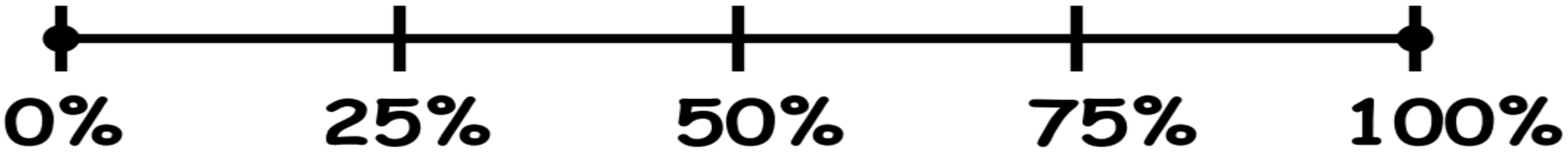
- En IA, lorsque nous réfléchissons sur l'incertitude, nous commençons généralement par essayer de quantifier la certitude des informations dont nous disposons.
- Dans le tableau ci-contre, deux études distinctes ont été menées pour demander aux participants de tenter de quantifier subjectivement le niveau de certitude impliqué par l'utilisation de ces mots.
  - Par exemple, toujours, très souvent, etc.
- Nous pouvons constater que ces deux études ont permis d'obtenir des résultats légèrement différents.
- Il est curieux de constater que, dans la première étude, le terme "toujours" n'a obtenu qu'une moyenne de 99 % sur 100 de certitude, d'après les données recueillies auprès des participants à l'étude.
- Dans cet exemple, nous pourrions imaginer que nous essayons de prendre un texte brut et de quantifier la certitude de l'information basée sur l'utilisation de certains de ces termes et, au lieu d'attribuer arbitrairement des valeurs de certitude nous-mêmes, nous enquêtons sur un groupe de sujets pour essayer de comprendre comment les gens perçoivent la certitude derrière l'utilisation de ces différents termes.

Ray Simpson (1944)		Milton Hakel (1968)	
Mot	Moyenne	Mot	Moyenne
Toujours	99	Toujours	100
Très souvent	88	Très souvent	87
En général	85	En général	79
Souvent	78	Souvent	74
Généralement	78	Plutôt souvent	74
Fréquemment	73	Fréquemment	72
Plutôt souvent	65	Généralement	72
Aussi souvent que possible	50	Aussi souvent que possible	50
De temps à autre	20	De temps à autre	34
Quelquefois	20	Quelquefois	29
Occasionnellement	20	Occasionnellement	28
De temps en temps	15	De temps à temps	22
Pas souvent	13	Pas souvent	16
Généralement pas	10	Généralement pas	16
Rarement	10	Rarement	9
Quasiment jamais	7	Quasiment jamais	8
Très rarement	6	Très rarement	7
Rarement	5	Rarement	5
Presque jamais	3	Presque jamais	2
Jamais	0	Jamais	0

Théorie de base des probabilités

Probabilité

- La probabilité est le langage de l'incertitude.
  - Il s'agit de savoir quelle est la probabilité qu'une chose se produise.
- La probabilité peut être exprimée mathématiquement sous la forme d'un indice numérique compris entre 0 (impossibilité absolue) et 1 (certitude absolue).
- On peut aussi, bien sûr, l'exprimer en termes de pourcentages compris entre 0 et 100.



0	$\frac{1}{4}$ ou .25	$\frac{1}{2}$ ou .5	$\frac{3}{4}$ ou .75	1
Impossible	**Peu probable**	**Également probable**	**Assez probable**	**Certain**

Probabilité (Cont.)

- La notation des probabilités ressemble souvent à quelque chose comme P(X), c'est-à-dire la probabilité de X.
- Par exemple, si nous avons une pièce de monnaie équitable, la probabilité de tirer pile ou face est toutes deux égales à 0,5.
- Il est également possible qu'une pièce de monnaie soit légèrement faussée et que la probabilité de tirer face soit de 0,51 et celle de tirer pile de 0,49.
  - P(face) = P(pile) = 0,5 pièce équitable
  - P(face) = 0,51, P(pile) = 0,49 pièce légèrement biaisée
- Juste pour revoir formellement comment cela serait illustré mathématiquement, la probabilité de succès (p) est le nombre de succès (s) divisé par le nombre de succès (s) et d'échecs (e)

$$P(succès) = \frac{\text{nombre de succès}}{le\ nombre\ de\ résultats\ possibles} = \boldsymbol{p} = \frac{s}{s+e}$$

et, bien sûr, la probabilité d'échec est exactement l'inverse:

$$P(échec) = \frac{\text{nombre d'échecs}}{le\ nombre\ de\ résultats\ possibles} = \boldsymbol{q} = \frac{e}{s+e}$$

- Notamment, lorsque nous avons deux résultats (p et q), la somme de nos deux probabilités doit être égale à 1.

$$\boxed{\boldsymbol{p} + \boldsymbol{q} = 1}$$

Table de probabilité

- Météo

Ensoleillé	Nuageux	Pluvieux
200/365	100/365	65/365

- Examinons maintenant les tables de probabilité.
- Nous avons ci-haut un exemple de météo à différents jours de l'année. 200 jours ont été ensoleillés, 100 nuageux et 65 pluvieux.
- Nous pouvons donc quantifier la probabilité que le temps soit ensoleillé. Dans ce cas, p(ensoleillé) est égal à 200/365.
- Ce qui peut s'écrire :

  P(Météo = ensoleillé) = P(ensoleillé) = 200/365
- Nous pouvons utiliser une notation similaire pour représenter la probabilité de tous les différents types de temps, qui pourraient alors être énumérés individuellement dans une liste comme celle-ci :

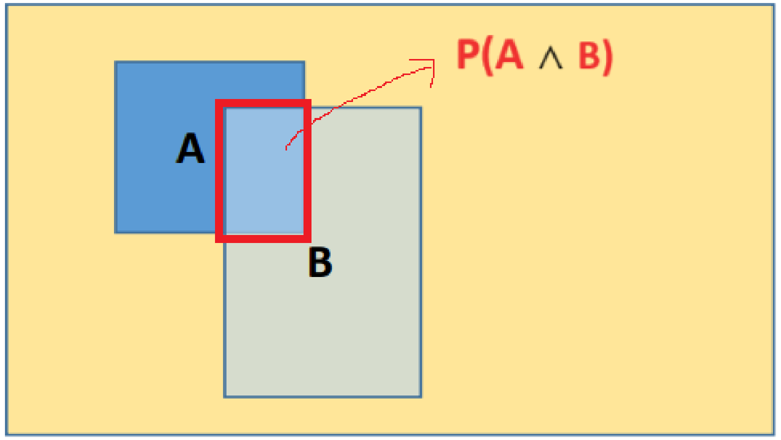
  P(Météo) = {200/365, 100/365, 65/365}
- À ce stade, nous nous contentons d'examiner la probabilité d'un seul résultat, en l'occurrence la météo.
  - Mais on pourra réfléchir à la longue à la possibilité d'avoir plusieurs résultats en même temps.

Probabilité disjointe

- C'est ici que nous entrons dans l'idée de probabilité disjointe.
- Il s'agit de la probabilité que l'un des deux événements possibles (A ou B) se produise.
- Une autre façon de l'exprimer est la probabilité que P(A=a ∨ B=b). Nous pouvons simplifier la notation ci-dessus en écrivant P(A ∨ B).
- Pour calculer la probabilité disjointe, on utilise cette formule:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

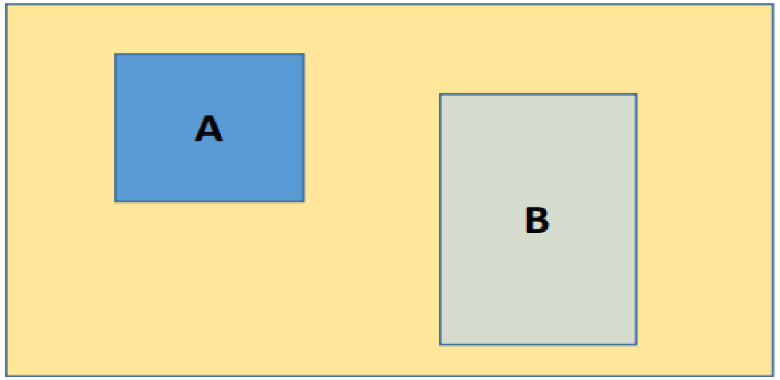
- Dans la formule ci-haut, nous devons soustraire la probabilité que les deux événements se produisent en même temps afin que les parties qui se chevauchent ne soient pas comptées deux fois.
- Here is an illustration of this disjoint probability situation



Espace des échantillons

Événements mutuellement exclusifs

- **Aussi connu comme**: disjoint
- Il existe des événements qui n'ont aucune issue de base en commun ou, de manière équivalente, leur intersection est l'ensemble vide.
  - C'est le cas lorsque l'on se trouve dans une situation telle que
    $$P(A \vee B) = P(A) + P(B).$$
  - En d'autres termes, il n'y a pas de chevauchement.
- Nous le voyons ici dans la même illustration de l'espace de probabilité, où A et B n'ont pas de région de chevauchement.



Espace des échantillons

- Nous appelons ces deux événements des événements mutuellement exclusifs, car ils ne se produisent pas en même temps.

Probabilité conjointe

- Il s'agit de la probabilité que deux événements se produisent en même temps.
- Nous pouvons utiliser cette annotation :
  - P(A,B) qui est une abréviation de P(A ∧ B).
  - Cela signifie la probabilité que A et B se produisent en même temps.
- On obtient cette probabilité en multipliant P(A) et P(B), mais seulement si "A et B sont indépendants".
- Voici l'exemple ci-contre de la météo; nous avons maintenant deux aspects de la météo que nous étudions (Temps et Température).
  - Il s'agit de ensoleillé, de nuageux ou de pluvieux, d'une part, et de la température (chaude ou froide), d'autre part.
  - Chacune des cellules du tableau représente une combinaison unique de l'aspect du Temps ou de la Température.
  - Par exemple, il y a eu 150 jours où il a fait à la fois chaud et ensoleillé.
  - Les valeurs dans chacune de ces cellules représentent la probabilité conjointe que cet événement se produise.
  - Dans ce cas, nous avons mesuré la probabilité conjointe directement à partir de l'environnement et n'avons pas eu à la calculer.

Temps

Température		Ensoillé	Nuageux	Pluvieux
	Chaude	150/365	40/365	5/365
	Froide	50/365	60/365	60/365

Probabilité conjointe (Cont.)

		Temps		
		Ensoillé	Nuageux	Pluvieux
Température	Chaude	150/365	40/365	5/365
	Froide	50/365	60/365	60/365

- Nous pouvons calculer ainsi la probabilité qu'il pleuve en additionnant les jours où il fait chaud et froid pour obtenir ceci:

P(Temps = pluvieux) = 5/365 + 60/365 = 65/365 = 0,178

- On peut aussi calculer la température chaude seule où l'on ajoute:

P(Temperature = chaude) = 150/365 + 40/365 + 5/365 = 195/365 = 0,534

- Enfin, on peut aussi chercher la probabilité de lorsqu'il fait chaud et il pleut:

P(pluvieux & chaude) = 5/365 = 0,014

- Nous pouvons maintenant nous demander si les variables Temps et Température sont indépendantes l'une de l'autre.

Probabilité conjointe (Cont.)

Température \ Temps	Ensoillé	Nuageux	Pluvieux
Chaude	150/365	40/365	5/365
Froide	50/365	60/365	60/365

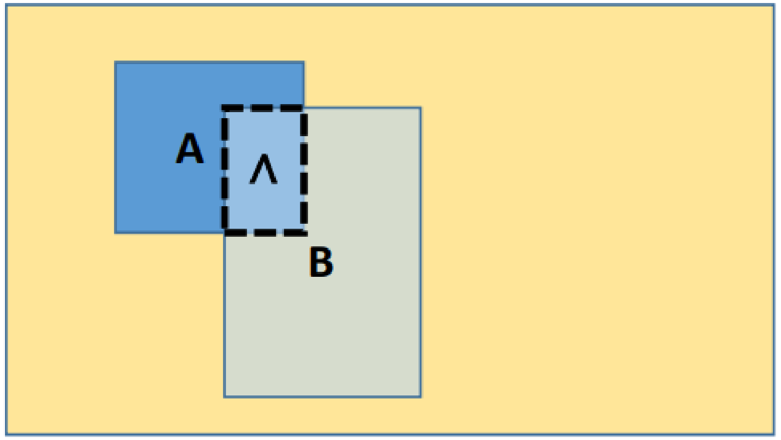
- Nous pouvons tester si les variables Temps et Température sont indépendantes l'une de l'autre en demandant si le produit des temps et température est égal à la probabilité conjointe que nous avons observée lorsque nous avons collecté les données.
- En d'autres termes, c'est demander si :

(P(Temps = pluvieux) = 0,178) * (P(Temperature = chaude) = 0,534) $\overset{?}{=}$ P(pluvieux & chaude) = 0,014.

- Ce qui donne : 0,178 * 0,534 = 0,095
- En vérifiant si P(temps = pluvieux) * P(température=chaude) = P(pluvieux & chaude), nous constatons que ce produit (0,095) est différent de la valeur de P(pluvieux & chaude) (0,014),
- nous concluons que ces deux variables (Temps et Température) ne sont pas indépendantes l'une de l'autre.

Probabilité conjointe (Cont.)

- Voici l'illustration de la probabilité conjointe:



Espace des échantillons

- Dans cette image, nous nous demandons quelle est cette région de chevauchement entre A et B.
- Note: Nous devons être prudents lorsque nous pensons aux probabilités conjointes et à la manière dont nous pouvons les calculer en fonction de notre connaissance de l'indépendance ou non de ces variables.

Probabilité conditionnelle

- Nous nous trouvons dans une situation où plusieurs événements se produisent, et nous pouvons exprimer la probabilité que l'un des événements se produise "étant donné" que l'autre est supposé s'être produit.
  - Avec les probabilités conditionnelles, nous supposons que les événements ne s'excluent pas mutuellement.
  - En d'autres termes, il y a un certain chevauchement.
- Nous pouvons exprimer la probabilité conditionnelle à l'aide de la syntaxe P(A|B).
- Nous pouvons calculer cette probabilité comme suit :

$$P(A|B) = \frac{\text{le nombre de fois où A et B peuvent se produire}}{\text{le nombre de fois où B peut se produire}} = \frac{P(A \cap B)}{P(B)}.$$

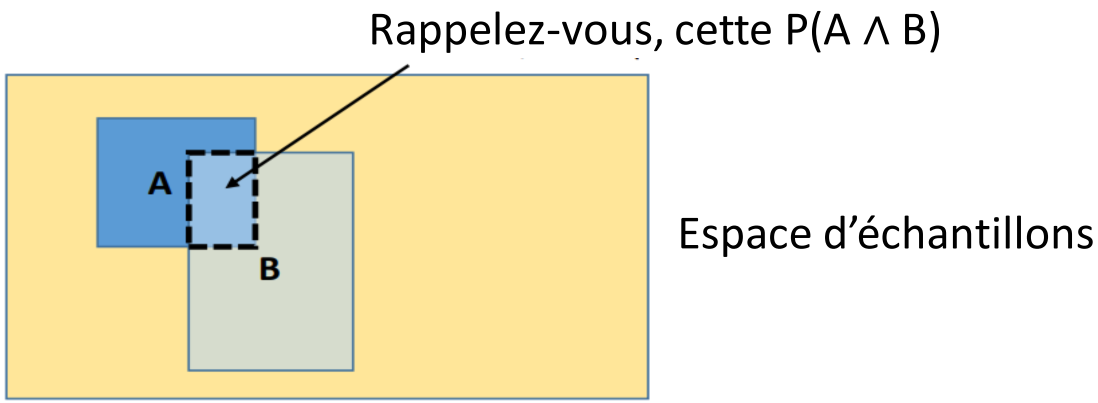
- De même, nous pouvons calculer la probabilité que B et A se produisent, c'est-à-dire P(B|A) comme:

$$\frac{P(B \cap A)}{P(A)}.$$

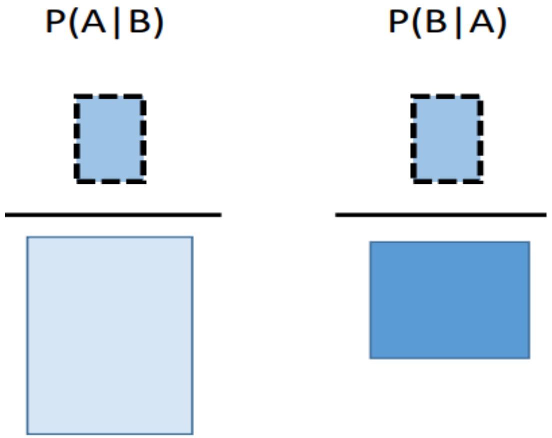
- Notez que si la probabilité de B et A était toujours nulle, en d'autres termes s'excluant mutuellement, la probabilité conditionnelle serait également nulle.

Probabilité conditionnelle (Cont.)

- Il convient toutefois de noter qu'en règle générale, P(A|B) n'est pas automatiquement égal à P(B|A).
  - Ainsi, pour considérer la probabilité d'un B donné, nous pouvons considérer qu'il s'agit également de la fraction des résultats B où A se produit également.
- Revenons à notre illustration de l'espace de probabilité comme illustré ci-bas et rappelons-nous que notre probabilité conjointe se situe dans la zone en pointillés ou la probabilité que A et B se produisent tous les deux:



- Par conséquent, notre probabilité conditionnelle P(A|B) est mieux visualisée dans l'image de juste à droite par la fraction ou la probabilité conjointe de A et B dans la zone en pointillés divisée par la probabilité de B.
- Et P(B|A) ressemblerait à l'image de l'extrême droite.
- Et, intuitivement, nous pouvons comprendre pourquoi ces valeurs ne sont pas toujours égales.



Vers l'obtention de la règle/théorème de Bayes

- La description de la probabilité conditionnelle est en fait un tremplin vers un élément que nous allons beaucoup utiliser ici, à savoir la règle ou le théorème de Bayes.
- Reprenons le calcul des probabilités conditionnelles.
  - $P(A|B) = \frac{\boldsymbol{P(A \cap B)}}{P(B)}$
  - $P(B|A) = \frac{P(B \cap A)}{P(A)}$
  - Nous savons également que la probabilité conjointe $P(B \cap A)$ sera toujours $P(A \cap B)$.
    - L'ordre des variables n'a pas d'importance ici.
  - Par conséquent, nous pouvons également obtenir les mêmes équations :
    - $P(B \cap A) = \mathrm{P(B|A)\ x\ P(A)}$
    - $P(A \cap B) = \mathbf{P(B|A)\ x\ P(A)}$
  - Cela nous permet de substituer l'expression $\boldsymbol{P(B|A)\ x\ P(A)}$ à $\boldsymbol{P(A \cap B)}$, ce qui nous donne la règle de Bayes :

$$P(A|B) = \frac{P(B|A)\ \mathrm{x}\ P(A)}{P(B)}$$

Vers l’obtention de la règle/théorème de Bayes (Cont.)

- La règle de Bayes mentionnée dans le slide précédent décrit une probabilité conditionnelle A (c’est-à-dire P(A|B)) en termes de probabilité conditionnelle B (c'est-à-dire P(B|A)) multipliée par la probabilité de A (c'est-à-dire P(A)) divisée par la probabilité de B (c'est-à-dire P(B)).
- Nous disposons maintenant d'un moyen de décrire les probabilités conditionnelles sans avoir besoin des probabilités conjointes (c'est-à-dire sans $P(A \cap B)$ et/ou $P(B \cap A)$).
- Nous verrons plus loin pourquoi il s'agit d'une simplification très utile.
- Il convient également de noter que cette formule de Bayes est largement utilisée pour soutenir le raisonnement probabiliste (c'est-à-dire raisonnement inductif).

La règle de la chaîne

- Ensuite, on peut se demander comment généraliser les probabilités conditionnelles à un nombre arbitraire d'événements?
- Jusqu'à présent, nous n'avons traité que deux événements à la fois.
- C'est ici que la règle de la chaîne entre en jeu.
  - La règle de la chaîne relie la probabilité conjointe des événements aux probabilités conditionnelles (càd elle permet de calculer la probabilité de deux événements conjoints (A et B) en fonction des probabilités conditionnelles des événements individuels).
  - En d'autres termes, si nous voulons connaître la probabilité conjointe de A, B, C tous ensemble au même moment, nous pouvons la calculer comme la probabilité de A étant donné B et C multipliée par la probabilité de B étant donné C multipliée par la probabilité de C seul.
    - $P(A,B,C) = P(A|B,C) * P(B|A) * P(C)$
      - N.B: $P(A|B,C) = \frac{P(A \cap B \cap C)}{P(B \cap A)} = \frac{P(A) \text{ x } P(B|A) \text{ x } P(C|A)}{P(B) \text{ x } P(C)}$
  - Cet exemple spécifique peut être généralisé comme suit:

P(A1 , A2 , ..., An) = P(A1) * P(A2 | A1) * P(A3 | A1 , A2) * ... * P(An | A1 ,A2 ... An - 1)

  - où au lieu de A, B et C, nous utilisons A1, A2, ..., An pour représenter le fait qu'il pourrait y avoir n'importe quel nombre d'événements pour lesquels nous sommes intéressés par leurs probabilités.

La règle de la chaîne (Cont.)

- Ainsi, dans le calcul P(A1 , A2 , …, An), nous pouvons voir le même type de schéma que $P(A, B, C)$, mais inversé.
  - Nous avons donc maintenant la probabilité de A1 seul au lieu de la probabilité de C seul.
- Nous pouvons appliquer cette règle de la chaîne pour étendre la règle de Bayes à une équation assez complexe
  - Potentiellement utile lorsque nous ne pouvons pas supposer l'indépendance (plus tard)
  - Peut être difficile à utiliser en fonction des probabilités que nous pouvons rassembler à l'avance.

Exemple d’application de la formule $P(A|B,C)$

- Données:
  - P(A) = 0.3 (Probabilité qu'il pleuve aujourd'hui)
  - P(B) = 0.6 (Probabilité que le ciel soit nuageux)
  - P(C) = 0.2 (Probabilité qu'il y ait des éclairs)
  - P(A|B) = 0.5 (Probabilité qu'il pleuve aujourd'hui sachant que le ciel est nuageux)
  - P(A|C) = 0.4 (Probabilité qu'il pleuve aujourd'hui sachant qu'il y a des éclairs)
- Nous voulons calculer P(A|B,C), la probabilité qu'il pleuve aujourd'hui sachant que le ciel est nuageux et qu'il y a des éclairs:

$$P(A|B,C) = \frac{P(A \cap B \cap C)}{P(B \cap C)}$$

$$P(A|B,C) = \frac{P(A) \times P(B|A) \times P(C|A)}{P(B) \times P(C)}$$

$$P(A|B,C) = \frac{0.3 \times 0.5 \times 0.4}{0.6 \times 0.2}$$

$$P(A|B,C) = \frac{0.06}{0.12}$$

$$P(A|B,C) = 0.5$$

- Donc, la probabilité qu'il pleuve aujourd'hui sachant que le ciel est nuageux et qu'il y a des éclairs est de 0.5.

Autres terminologies de la probabilité

- En guise d'interlude, examinons également d'autres terminologies relatives aux probabilités.
  - A priori : Ce terme est utilisé pour décrire la probabilité d'un événement seul ou la probabilité de cet événement avant que d'autres preuves ne soient disponibles pour le mettre en évidence.
    - Exemple P(A)
  - Vraisemblance: elle est utilisée pour décrire les probabilités conditionnelles.
    - Par exemple, P(B|A), en supposant le résultat A, quelle est la probabilité de la preuve B.
  - A postériori : Il est utilisé pour décrire la probabilité conditionnelle du résultat après avoir pris connaissance des preuves.
    - Dans ce cas, B ou la probabilité de A étant donné B.
  - Notez que tous les termes mentionnés ci-dessus sont liés à A, dans ce cas étant notre résultat cible et B une preuve à l'appui.
  - Inférence : Dans ce contexte, il s'agit de déduire des probabilités inconnues à partir de probabilités connues.

Inférence statistique

Inférence avec la règle de Bayes

- **Aussi connu comme**: Inférence statistique
- Nous voulons calculer la probabilité d'une requête ou d'un résultat.
- Mais la probabilité sera conditionnée par certaines preuves existantes obtenue via une enquête par exemple.
- Par exemple, en demandant quelle est la probabilité que nous ayons le résultat "grippe", nous pourrions vouloir prendre en considération la preuve que nous avons ou non "mal à la tête".
  - Avez-vous la grippe (G) étant donné que vous avez mal à la tête (M)?
  - Imaginons donc que nous disposions de la règle suivante :
    - IF mal de tête THEN grippe ?
      - Nous pouvons exprimer la probabilité de cette règle en utilisant la probabilité d'avoir la grippe si l'on a mal à la tête: P(G|M).
      - La règle de Bayes intervient dans cet exemple puisque nous avons une probabilité conditionnelle que nous voulons comprendre et calculer.

$$\boxed{p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}}$$

Règle de Bayes

Inférence avec la règle de Bayes (Cont.)

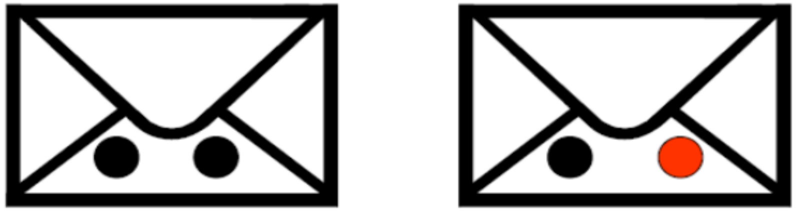
- Dans le cas de l'exemple du slide précédent, disons que nous disposons des "connaissances préalables/Prior knowledge" suivantes à partir des données collectées :
  - P(M) = 0,1 "une personne sur dix a mal à la tête".
  - P(G) = 0,01 "une personne (soit 1%) sur 100 a la grippe".
  - P(M|G) = 0,9 "90 % des personnes qui ont la grippe ont mal à la tête".

$$\boxed{p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}}$$

- Nous pouvons maintenant calculer P(G|M):
  - P(G|M) = (0.9 * 0,01) / 0,1 = 0,09
- Le résultat ci-dessus nous indique que si nous savons qu'une personne a mal à la tête, il n'y a que 9% de chances que ce mal de tête soit dû à la grippe.
- Cette probabilité est beaucoup plus faible que les 90% que nous avons observés à partir des données collectées et qui provenaient de la probabilité d'avoir un mal de tête si l'on a la grippe (cf. P(M|G) ci-haut).
  - Mais cette probabilité est surtout plus élevée que la probabilité d'avoir la grippe (c'est-à-dire P(G)) en général, qui n'est que de 1%.
  - Ainsi, dans ce cas, cet élément de preuve qu'une personne avait mal à la tête a fait passer la probabilité que cette personne ait la grippe d'une probabilité antérieure de 1% à une nouvelle probabilité conditionnelle de 9%

Inférence avec la règle de Bayes: Exemple

- Examinons un exemple classique d'inférence de probabilité à l'aide de la règle de Bayes.



- Dans cet exemple, nous avons un sac contenant deux enveloppes.
  - L'une des enveloppes contient une boule rouge d'une valeur de 100 $ ainsi qu'une **boule noire**,
  - tandis que l'autre enveloppe ne contient que deux **boules noires**.
- Supposons que vous participiez à ce jeu pour gagner 100$ et que vous preniez une enveloppe au hasard.
  - De cette enveloppe, vous tirez au hasard une boule et vous découvrez qu'elle est **noire**.
  - À ce stade, en tant que joueur essayant de gagner 100$, devriez-vous passer à l'autre enveloppe pour un autre tirage ou devriez-vous vous en tenir à l'enveloppe dont vous avez déjà tiré une boule ?

Inférence avec la règle de Bayes: Exemple (Cont.)

- Voyons comment calculer le meilleur choix dans cet exemple de problème.
- Commençons par définir nos variables:
- Nous utilisons E pour représenter une enveloppe, où E1 contient une boule rouge et une boule noire, et E2 deux boules noires.
  - E: enveloppe, E1=(R,B), E2=(B,B)
  - En réalité, nous ne connaissons pas les états sous-jacents des deux enveloppes, mais nous allons leur attribuer ces valeurs dans le seul but de démontrer comment ce problème peut être résolu.
- Nous avons également une variable B qui décrit l'événement du tirage d'une boule noire.
  - B: l'événement du tirage d'une boule noire
- Nous savons que notre "probabilité préalable" de tirer une boule noire est:
  - P(B)= ¾
    - puisque trois des quatre boules sont noires.
- Nous pouvons également calculer la probabilité conditionnelle de tirer une boule noire à partir de l'une ou l'autre enveloppe de cette façon:
  - $P(B|E1) = 0.5$
    - Car la probabilité de tirer une boule noire étant donné que nous avons E1 est de 0,5 puisque E1 contient à la fois une boule rouge et une boule noire
  - $P(B|E2) = 1$
    - la probabilité de tirer une boule noire pour E2 est de 100 % ou 1 car E2 ne contient que des boules noires.

Inférence avec la règle de Bayes: Exemple (Cont.)

- La probabilité que nous ayons choisi l'une ou l'autre des enveloppes est de 50 % ou 0,5 car le sac contient seulement deux enveloppes.
  - P(E1)=P(E2)=0.5
- Ensuite, nous voulons calculer la probabilité que vous ayez choisi l'enveloppe X (E_x) étant donné que vous avez tiré une boule noire.
  - Nous l'écrirons comme suit :
    - P(E_x|B) = P(B|E_x) * P(E_x) / P(B)
  - Ici, nous utilisons la règle de Bayes pour nous donner les variables avec lesquelles nous essayons de calculer cette probabilité.
  - Ensuite, nous voulons calculer cette P(E=1|B) par rapport à la P(E=2|B) pour les deux situations où vous avez E1 ou E2 et les comparer.

Inférence avec la règle de Bayes: Exemple (Cont.)

- Tout d'abord, quelle est la probabilité que nous ayons tiré E1 étant donné que nous avons tiré une boule noire ?
- D'après la règle de Bayes, cette probabilité est de 1/3
  - $P(E1|B)=1/3$
- Par ailleurs, quelle est la probabilité que nous ayons tiré E2 étant donné que nous avons tiré une boule noire ?
  - Dans ce cas, nous obtiendrions 2/3.
    - $P(E2|B)=2/3$
- Par conséquent, la probabilité que vous ayez tiré E2 avec deux boules noires est plus grande étant donné que vous avez tiré une boule noire lors de votre première tentative.
- Cela peut être contre-intuitif, car vous pourriez vous attendre à ce que, lors de votre premier tirage, vous ayez 1/4 de chances de choisir la boule rouge et que, lors de votre deuxième tirage, vous ayez maintenant 1/3 de chances de choisir une boule rouge.

Inférence avec la règle de Bayes: Exemple (Cont.)

- Cependant, la preuve que vous avez tiré une boule noire lors de votre premier essai modifie les probabilités.

- Par conséquent, la meilleure stratégie consiste à passer à l'autre enveloppe et à prendre le risque de tirer une boule rouge.

- Bien sûr, cette stratégie n'est pas certaine - il s'agit de probabilités - mais en moyenne, dans ce scénario, si vous avez tiré une boule noire lors de votre première tentative, vos chances de tirer une boule rouge lors de la seconde sont meilleures si vous changez d'enveloppe.

Pourquoi se préoccuper de la règle de Bayes ?

- Pourquoi s'embêter avec la règle de Bayes quand on construit un système expert ?
  - Ou pourquoi compliquer les choses à ce point ?
- Pourquoi ne pas simplement calculer P(G|M) (c'est-à-dire la probabilité d'avoir la grippe pour un mal de tête donné) à partir des données directement afin de l'utiliser dans un système expert ?
  - En fin de compte, il s'agit souvent de l'accessibilité de ces données.
    - Les patients que les cliniciens peuvent facilement observer seront identifiés comme souffrant d'une maladie donnée (en fin de compte), sinon ils ne seraient probablement pas venus à la clinique pour commencer.
      - À partir de ces patients, nous pouvons donc facilement obtenir la probabilité des symptômes, par exemple P(M|G) (c'est-à-dire la probabilité d'un mal de tête dans le cas d'une grippe).
      - Mais comment est-il possible de rassembler toutes les personnes souffrant de maux de tête et d'enquêter sur le fait que beaucoup d'entre elles ont la grippe ?
      - Dans les problèmes de diagnostic, nous disposons souvent de données sur les probabilités de divers symptômes de différentes maladies, par exemple la probabilité préalable d'avoir un mal de tête P(M).
      - Avec ces deux types de probabilités, P(M|G) et P(M), vous pouvez donc calculer la probabilité initiale P(G|M) qui nous intéresse à l'aide de la règle de Bayes.

Pourquoi se préoccuper de la règle de Bayes ? (Cont.)

- Vous pourriez également vous demander pourquoi ne pas calculer cette probabilité cible (c'est-à-dire $P(G|M)$) en utilisant la probabilité conjointe de la grippe et des maux de tête divisée par la probabilité des maux de tête (c'est-à-dire $P(G,M)/P(M)$) plutôt que la règle de Bayes ?
  - La probabilité conjointe de $P(G,M)$ peut également être difficile à évaluer.
  - Dans ce cas, il faudrait construire une table de probabilités conjointes, ce qui, dans certains cas, pourrait être possible,
    - mais qui est potentiellement très peu pratique à mesure que l'échelle du problème augmente.

- En fin de compte, lorsqu'il s'agit d'incertitude, de raisonnement et de probabilités, vous voudrez travailler avec les données les plus faciles à collecter ainsi qu'avec les données les plus fiables.

Manipulation supplémentaire de la règle de Bayes

- Examinons maintenant quelques autres possibilités légèrement plus avancées que nous offre la règle de Bayes.
- Par exemple, que se passe-t-il si nous ne disposons pas de données sur la probabilité d'un symptôme ou, en d'autres termes, sur P(B) ?
- Si la survenue d'un événement A ne dépend que de deux événements qui s'excluent mutuellement, B et NON B, nous obtenons une autre façon de calculer:

$$p(B) = \boxed{p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)}$$

- Nous pouvons alors substituer cette situation à la règle de Bayes, ce qui nous donne ce qui suit :

$$\boxed{p(A|B) = \frac{p(B|A) \times p(A)}{p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)}}$$

- Le dividende de l'équation P(B) dans la règle de Bayes est maintenant remplacé par l'expression entière que nous avons vue ci-dessus dans le dénominateur.

Manipulation supplémentaire de la règle de Bayes (Cont.)

- Notamment, la probabilité de ne pas avoir A (càd $P(\neg A)$) est la probabilité préalable que l'hypothèse A soit fausse.
  - Il s'agit donc peut-être de la probabilité de ne pas avoir mal à la tête.
- Il convient également de noter que la probabilité que B ne soit pas A (càd $P(B|\neg A)$ est la probabilité de trouver la preuve B même si l'hypothèse A est fausse.
- En fait, nous montrons la manipulation ci-dessus pour illustrer comment, lorsque certains éléments de preuve ne sont pas disponibles ou ne peuvent pas être facilement calculés, la probabilité de ne pas avoir mal à la tête peut être calculée.
- Par conséquent, il existe souvent d'autres moyens de les calculer sur la base des informations disponibles.

Hypothèses d'indépendance

Indépendance

- Nous réfléchissons maintenant à l'impact des hypothèses d'indépendance dans ces calculs de probabilité.
- Nous commençons à penser à l'indépendance lorsque nous sommes confrontés à une situation où nous voulons raisonner à partir de plusieurs éléments de preuve.
- Mais avant de pouvoir le faire, nous devons comprendre quels éléments de preuve sont indépendants les uns des autres.
- Comme nous l'avons mentionné précédemment, deux faits (par exemple E1 et E2) sont indépendants l'un de l'autre s'ils n'ont aucune influence l'un sur l'autre
- De façon claire, deux événements A et B sont indépendants si:
  - Leur probabilité conjointe est le produit des probabilités individuelles
    - P(A, B) = P(A) * P(B)
  - La probabilité de A ne change pas étant donné B
    - P(A | B) = P(A)
  - La probabilité de B ne change pas étant donné A
    - P(B | A) = P(B)

- Si les trois hypothèses sont satisfaites, vous savez que A et B sont indépendantes l'une de l'autre.

Considérations relatives à l'indépendance

- L'indépendance des variables est en soi une connaissance importante d'un problème.

- Si, dans cet exemple, A et B sont indépendants, la table de probabilité conjointe entre eux peut être calculée simplement.

- En fin de compte, il suffit d'ajouter k aux deux cellules, où k est le nombre de variables que vous avez.
  - C’est-à-dire cela possède k^2 cellules.

Indépendance insatisfaite/mal utilisée

- Avant d'examiner les situations où l'indépendance est mal utilisée, voyons comment l'hypothèse d'indépendance peut être utilisée à mauvais escient dans le calcul des probabilités.
- Voici donc l'anecdote d'un célèbre statisticien :
    - Un célèbre statisticien ne prenait jamais l'avion, parce qu'il avait étudié le transport aérien et estimé que la probabilité qu'il y ait une bombe sur un vol donné était de 1 sur 1 million, et qu'il n'était pas prêt à accepter cette probabilité.
    - Un jour, un collègue l'a rencontré lors d'une conférence loin de chez lui. Il lui demande : "Comment es-tu arrivé ici, en train ?".
    - Le statisticien répond : "Non, j'ai pris l'avion.
    - Son collègue lui pose à nouveau la question : "Et la possibilité d'une bombe ?".
    - Le statisticien répond : "Eh bien, j'ai commencé à penser que si la probabilité d'une bombe est de 1 sur un million, alors la probabilité de deux bombes est de (1/1 000 000) x (1/1 000 000). Il s'agit d'une probabilité très, très faible, que je peux accepter. Alors maintenant, j'apporte ma propre bombe !

Indépendance insatisfaite/mal utilisée (Cont.)

- Bien entendu, l'humour de cette anecdote provient de la mauvaise utilisation des hypothèses d'indépendance.

- Ici, le statisticien n'a pas reconnu qu'il y avait une indépendance entre la preuve de l'intention d'utiliser la bombe et la présence de la bombe.

- Ainsi, dans cette histoire, la probabilité initiale qu'une bombe ait été apportée dans l'avion reste de un sur un million malgré le fait que le statique ait apporté la sienne.

Indépendance conditionnelle

- Parlons du concept d'indépendance conditionnelle.
- Notamment, les ***variables aléatoires*** peuvent être dépendantes, mais conditionnellement indépendantes.
- Exemple :
  - imaginez que votre maison soit équipée d'une alarme et
  - que votre voisin Jean appelle lorsqu'il entend l'alarme.
  - Vous avez également une voisine Marie qui appellera lorsqu'elle entendra l'alarme.
  - Mais nous supposons que Jean et Marie ne se parlent pas.
- Dans ce cas, l'appel de Jean est-il indépendant de l'appel de Marie ?
- La réponse est non.
  - Si Jean a appelé, l'alarme s'est probablement déclenchée, ce qui augmente la probabilité que Marie appelle.
  - En d'autres termes, P(AppelDeMarie|AppelDeJohn) != (AppelDeMarie).
    - Cela signifie que la probabilité que Marie appelle si Jean a appelé n'est pas égale à la probabilité que Marie appelle seule.
  - Cependant, si nous connaissons l'état de l'alarme, la probabilité que Jean appelle ne devrait pas affecter du tout l'appel de Marie.

Indépendance conditionnelle (Cont.)

- Ainsi, dans ce cas, la probabilité que Marie appelle étant donné que l'alarme a retenti et que Jean a appelé est la même que la probabilité que Marie appelle simplement étant donné que l'alarme a retenti.
  - P(AppelDeMarie|Alarme, AppelDeJohn) = P(AppelDeMarie|Alarme).
- Dans cet exemple, nous pouvons dire que l'appel de Jean et l'appel de Marie sont conditionnellement indépendants étant donné que l'alarme s'est déclenchée.
- Une autre façon de formuler cela est que ces deux variables sont indépendantes selon que l'alarme s'est déclenchée ou non.
- En général, A et B sont conditionnellement indépendants étant donné C si les situations suivantes sont vraies :
  - $P(A \mid B, C) = P(A \mid C)$
  - $P(B \mid A, C) = P(B \mid C)$
  - $P(A, B \mid C) = P(A \mid C) * P(B \mid C)$ (***N.B:*** $P(A, B)$ c'est la probabilité conjointe)

L'indépendance conditionnelle en pratique

- Voyons maintenant comment l'indépendance conditionnelle entre en jeu dans la pratique.
- Pour les tâches diagnostiques, nous nous intéressons à l'indépendance conditionnelle d'un ensemble d'éléments de preuve que nous appellerons (E1 ... En) compte tenu d'une hypothèse H.
- Cette hypothèse est souvent formulée même lorsqu'il est impossible de recueillir des données pour la confirmer, et ce pour un certain nombre de raisons pratiques.
  - En particulier, cela nous permet de simplifier considérablement le raisonnement probabiliste en réduisant le nombre de données dont nous avons besoin pour calculer quoi que ce soit.
- En pratique, nous pouvons faire cette hypothèse d'indépendance conditionnelle lorsqu'elle semble appropriée aux phénomènes modélisés.
  - En fin de compte, nous devons faire preuve de bon sens et de perspicacité pour avancer cette hypothèse d'indépendance conditionnelle.

L'indépendance conditionnelle en pratique (Cont.)

- À titre d'exemple, imaginons que nous examinions l'hypothèse de la pneumonie.

  - Nous pourrions nous attendre à une relation de cause à effet entre:
    - la présence d'une pneumonie et un résultat positif sur une radiographie pulmonaire,
    - ainsi qu'entre la présence d'une pneumonie et la présence d'une toux.

  - Par conséquent, une fois que l'on sait qu'il y a une pneumonie, la probabilité d'une radiographie positive sera probablement la même, qu'il y ait ou non une toux.

  - Il est donc raisonnable de supposer qu'une radiographie positive et une toux sont conditionnellement indépendantes de la présence d'une pneumonie.

L'indépendance conditionnelle en action

- Si nous nous permettons de supposer l'indépendance conditionnelle ou si nous vérifions qu'elle est effectivement vraie, nous obtenons la règle de Bayes suivante :

$$P(H | E_1 \land \ldots \land E_n) = \frac{P(E_1 \land \ldots \land E_n | H) * P(H)}{P(E_1) \land \ldots \land P(E_n)}$$

  - où nous avons la probabilité d'une certaine hypothèse (H) compte tenu d'un ensemble de preuves (E1, ..., En).
  - Nous avons écrit cette situation d'indépendance conditionnelle supposée dans l'équation ci-dessus, où nous avons la probabilité d'avoir conjointement tous les éléments de preuve étant donné l'hypothèse (c.-à-d. P(E1 ∧ … ∧ En |H)) multipliée par la probabilité préalable de l'hypothèse (c.-à-d. P(H)) divisée par la probabilité conjointe de toutes les preuves (c.-à-d. P(E1) ∧ … ∧ P(En)).
- Nous pouvons utiliser la forme de la règle de Bayes pour réécrire l'équation ci-dessus et la calculer comme suit :

$$= \frac{P(E_1 | H) * \ldots * P(E_2 | H) * P(H)}{P(E_1 \land \ldots \land E_n)}$$

L'indépendance conditionnelle en action Cont.)

$$= \frac{P(E_1|H)*...*P(E_2|H)*P(H)}{P(E_1 \wedge ... \wedge E_n)}$$

- La formule réécrite nous donne la probabilité conditionnelle de chaque élément de preuve *compte tenu* de l'hypothèse, multipliée par une autre et multipliée par la probabilité préalable de l'hypothèse, divisée à nouveau par la probabilité commune de tous les éléments de preuve.
- L'utilisation de cette équation est beaucoup plus simple que si nous devions nous fier à la règle de la chaîne que nous avons apprise plus tôt.
  - Rappelons que la règle de la chaîne peut être utilisée même lorsque nous ne pouvons pas supposer l'indépendance.
- Mais pour utiliser les équations, nous devons toujours connaître la probabilité commune de tous les symptômes, comme nous pouvons le voir dans le dénominateur ci-dessus

Manipulations similaires à la règle de Bayes

- Cependant, nous pouvons effectuer une manipulation similaire de la règle de Bayes si les hypothèses possibles sont exclusives et mutuellement exclusives.
  - En d'autres termes, si tout le monde est atteint d'une des maladies, mais pas plus d'une.
- Dans ce cas, nous pouvons manipuler la règle de Bayes de la manière suivante :

$$P(H|E_1 \wedge \ldots \wedge E_n) = \frac{P(E_1|H)*\ldots*P(E_2|H)*P(H)}{P(E_1|H)*\ldots*P(E_2|H)*P(H) + P(E_1|\neg H)*\ldots*P(E_2|\neg H)*P(\neg H)}$$

- où nous essayons de comprendre que la probabilité d'une hypothèse, compte tenu des éléments de preuve, est égale à l'expression droite de l'équation ci-dessus.

Raisonner à partir d'hypothèses multiples

- Voyons maintenant comment raisonner avec des hypothèses multiples.
- Dans ce cas, nous essayons de prendre en compte à la fois des hypothèses multiples (H1, ..., Hm) et des preuves multiples (E1, ..., En).
  - Ici, les hypothèses ainsi que toutes les preuves doivent être mutuellement exclusives et exhaustives.
  - Mais dans ce cas, nous revenons à l'absence d'hypothèse d'indépendance conditionnelle.

- Examinons tout d'abord l'équation permettant de traiter une seule preuve et plusieurs hypothèses.

$$p(H_i|E) = \frac{p(E|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E|H_k) \times p(H_k)}$$

- Différemment, nous avons l'équation suivante pour traiter des preuves multiples ainsi que des hypothèses multiples. Mais comme vous pouvez le constater, l'équation ci-dessus devient quelque peu complexe.

$$p(H_i|E_1 E_2 \ldots E_n) = \frac{p(E_1 E_2 \ldots E_n|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E_1 E_2 \ldots E_n|H_k) \times p(H_k)}$$

Hypothèses/preuves multiples simplifiées

- L'utilisation des équations de la dernière diapositive nécessite d'obtenir les probabilités conditionnelles de toutes les combinaisons possibles de preuves pour toutes les hypothèses, et impose donc une charge énorme à l'expert ou au chercheur qui tente de rassembler les données.
- Par conséquent, l'indépendance conditionnelle entre les différentes preuves est supposée d'une manière qui est confirmée par les données.
- Mais dans ce cas, nous pouvons utiliser l'équation suivante remplie d'un certain nombre de probabilités conditionnelles qui seraient généralement plus faciles à calculer :

$$p(H_i|E_1\,E_2\ldots E_n)=\frac{p(E_1|H_i)\times p(E_2|H_i)\times\ldots\times p(E_n|H_i)\times p(H_i)}{\sum_{k=1}^{m}p(E_1|H_k)\times p(E_2|H_k)\times\ldots\times p(E_n|H_k)\times p(H_k)}$$

  - En d'autres termes, la probabilité de la preuve 1 étant donné une hypothèse...
- Cette équation est celle qui est le plus souvent appliquée dans les systèmes experts qui s'appuient sur un cadre probabiliste plus formel pour calculer l'incertitude.

Classement des hypothèses potentiellement vraies

- Nous allons donc passer à la réflexion sur le classement (ranking) des hypothèses potentiellement vraies en fonction de différents éléments de preuve.

- Examinons l'exemple simple d'un expert qui dispose:
  - de trois preuves conditionnellement indépendantes E1, E2 et E3 et
  - d'hypothèses exhaustives H1, H2 et H3.
  - En outre, il fournit des probabilités préalables pour toutes ces hypothèses : P(H1), P(H2) et P(H3), respectivement.

- L'expert détermine également les probabilités conditionnelles d'observation de chaque preuve pour toutes les hypothèses possibles

Classement des hypothèses potentiellement vraies (Cont.)

- **Probabilités préalables et conditionnelles**
- Le problème de la dernière diapositive est résumé dans ce tableau.
- Ici, en haut, nous avons nos trois hypothèses i=1, i=2 et i=3.
- Dans la première colonne, nous avons différentes probabilités.
  - Il s'agit donc de la probabilité préalable d'une hypothèse donnée (c'est-à-dire P(Hi)).
  - La probabilité de la preuve 1 compte tenu de cette hypothèse (c'est-à-dire P(E1|Hi))
  - La probabilité de la preuve 2 compte tenu de cette hypothèse (c'est-à-dire P(E2|Hi))
  - La probabilité de la preuve 3 compte tenu de cette hypothèse (c'est-à-dire P(E3|Hi))...
- Ainsi, par exemple, la probabilité que vous voyiez la preuve 3 (E3) compte tenu du fait que vous avez la maladie 3 (i3) est de 90 % (0,9).

Probabilité	Hypothèse		
	$i = 1$	$i = 2$	$i = 3$
$p(H_i)$	0.40	0.35	0.25
$p(E_1 \mid H_i)$	0.3	0.8	0.5
$p(E_2 \mid H_i)$	0.9	0.0	0.7
$p(E_3 \mid H_i)$	0.6	0.7	0.9

Classement des hypothèses potentiellement vraies (Cont.)

- **Calcul des vraisemblances**
- Dans cet exemple, disons que nous avons d'abord observé la preuve 3 (E3) et que le système expert calcule maintenant les probabilités postérieures pour toutes les hypothèses.

Probabilité	Hypothèse		
	$i = 1$	$i = 2$	$i = 3$
$p(H_i)$	0.40	0.35	0.25
$p(E_1 \mid H_i)$	0.3	0.8	0.5
$p(E_2 \mid H_i)$	0.9	0.0	0.7
$p(E_3 \mid H_i)$	0.6	0.7	0.9

- Nous appliquons donc ici l'équation susmentionnée (cf. slide 52) dans laquelle nous attribuons une indépendance conditionnelle et nous avons trois hypothèses pour i.
- Par conséquent, nous avons la probabilité de chaque hypothèse étant donné que nous avons E3. Voilà l'équation et les hypothèses:

$$p(H_i \mid E_3) = \frac{p(E_3 \mid H_i) \times p(H_i)}{\sum_{k=1}^{3} p(E_3 \mid H_k) \times p(H_k)}, \qquad i = 1, 2, 3$$

Classement des hypothèses potentiellement vraies (Cont.)

- **Calcul des vraisemblances**
- Voici ces trois calculs pour chacune des hypothèses.

$$p(H_1|E_3) = \frac{0.6 \cdot 0.40}{0.6 \cdot 0.40 + 0.7 \cdot 0.35 + 0.9 \cdot 0.25} = 0.34$$

$$p(H_2|E_3) = \frac{0.7 \cdot 0.35}{0.6 \cdot 0.40 + 0.7 \cdot 0.35 + 0.9 \cdot 0.25} = 0.34$$

$$p(H_3|E_3) = \frac{0.9 \cdot 0.25}{0.6 \cdot 0.40 + 0.7 \cdot 0.35 + 0.9 \cdot 0.25} = 0.32$$

Probabilité	Hypothèse		
	$i = 1$	$i = 2$	$i = 3$
$p(H_i)$	0.40	0.35	0.25
$p(E_1\|H_i)$	0.3	0.8	0.5
$p(E_2\|H_i)$	0.9	0.0	0.7
$p(E_3\|H_i)$	0.6	0.7	0.9

- Rappelons qu'avant de connaître les éléments de preuve figurant dans le tableau, nos probabilités préalables de résultats pour l'une ou l'autre de ces hypothèses sont énumérées dans la première ligne du tableau, H1 (0,40) étant la plus probable.
- Après avoir calculé les probabilités conditionnelles de ces hypothèses compte tenu de E3, nous pouvons maintenant constater que H1 et H2 sont vraisemblables (0.34) et que H3 est la moins vraisemblable (0.32).

Classement des hypothèses potentiellement vraies (Cont.)

- **Calcul des probabilités: Ajout de preuves**
- Supposons maintenant que nous obtenions la preuve E1 en plus de la preuve E3.
- Nous allons donc à nouveau calculer ces probabilités postérieures à l'aide de l'équation donnée:

$$p(H_i|E_1E_3) = \frac{p(E_1|H_i)\times p(E_3|H_i)\times p(H_i)}{\sum_{k=1}^{3} p(E_1|H_k)\times p(E_3|H_k)\times p(H_k)}, \qquad i = 1, 2, 3$$

Probabilité	Hypothèse		
	$i=1$	$i=2$	$i=3$
$p(H_i)$	0.40	0.35	0.25
$p(E_1\|H_i)$	0.3	0.8	0.5
$p(E_2\|H_i)$	0.9	0.0	0.7
$p(E_3\|H_i)$	0.6	0.7	0.9

- À partir de l'équation donnée, nous avons des éléments de preuve à prendre en compte (E1E3).

$$p(H_1|E_1E_3) = \frac{0.3\cdot 0.6\cdot 0.40}{0.3\cdot 0.6\cdot 0.40 + 0.8\cdot 0.7\cdot 0.35 + 0.5\cdot 0.9\cdot 0.25} = 0.19$$

$$p(H_2|E_1E_3) = \frac{0.8\cdot 0.7\cdot 0.35}{0.3\cdot 0.6\cdot 0.40 + 0.8\cdot 0.7\cdot 0.35 + 0.5\cdot 0.9\cdot 0.25} = 0.52$$

$$p(H_3|E_1E_3) = \frac{0.5\cdot 0.9\cdot 0.25}{0.3\cdot 0.6\cdot 0.40 + 0.8\cdot 0.7\cdot 0.35 + 0.5\cdot 0.9\cdot 0.25} = 0.29$$

- Nous examinons maintenant les probabilités conditionnelles mises à jour pour chacune des hypothèses et constatons que H2 est désormais beaucoup plus vraisemblable que les deux autres hypothèses.

Classement des hypothèses potentiellement vraies (Cont.)

- **Calcul des vraisemblances: Final**
- Supposons maintenant que nous obtenions finalement la preuve E2 et que nous souhaitions calculer nos probabilités postérieures finales pour toutes les hypothèses.
- Nous devons maintenant ajouter E3 pour avoir l'équation:

Probabilité	Hypothèse		
	$i=1$	$i=2$	$i=3$
$p(H_i)$	0.40	0.35	0.25
$p(E_1 \mid H_i)$	0.3	0.8	0.5
$p(E_2 \mid H_i)$	0.9	0.0	0.7
$p(E_3 \mid H_i)$	0.6	0.7	0.9

$$p(H_i \mid E_1E_2E_3) = \frac{p(E_1 \mid H_i) \times p(E_2 \mid H_i) \times p(E_3 \mid H_i) \times p(H_i)}{\sum_{k=1}^{3} p(E_1 \mid H_k) \times p(E_2 \mid H_k) \times p(E_3 \mid H_k) \times p(H_k)}, \qquad i = 1, 2, 3$$

$$p(H_1 \mid E_1E_2E_3) = \frac{0.3 \cdot 0.9 \cdot 0.6 \cdot 0.40}{0.3 \cdot 0.9 \cdot 0.6 \cdot 0.40 + 0.8 \cdot 0.0 \cdot 0.7 \cdot 0.35 + 0.5 \cdot 0.7 \cdot 0.9 \cdot 0.25} = 0.45$$

$$p(H_2 \mid E_1E_2E_3) = \frac{0.8 \cdot 0.0 \cdot 0.7 \cdot 0.35}{0.3 \cdot 0.9 \cdot 0.6 \cdot 0.40 + 0.8 \cdot 0.0 \cdot 0.7 \cdot 0.35 + 0.5 \cdot 0.7 \cdot 0.9 \cdot 0.25} = 0$$

$$p(H_3 \mid E_1E_2E_3) = \frac{0.5 \cdot 0.7 \cdot 0.9 \cdot 0.25}{0.3 \cdot 0.9 \cdot 0.6 \cdot 0.40 + 0.8 \cdot 0.0 \cdot 0.7 \cdot 0.35 + 0.5 \cdot 0.7 \cdot 0.9 \cdot 0.25} = 0.55$$

- Nous pouvons voir les probabilités calculées suivantes pour chaque hypothèse où, avec l'introduction de E3, la probabilité de H2 a complètement disparu (0) et nous constatons qu'en incluant les trois éléments de preuve, H3 est maintenant la plus probable.
- Ceci malgré le fait que H3 avait à l'origine (cf. tableau, dernière cellule de la ligne 1) la probabilité antérieure la plus faible (0,25).

Rapports de vraisemblance/ Likelihood ratios

Rapports de vraisemblance

- Le théorème de Bayes peut être gênant pour les problèmes de diagnostic
- C'est pourquoi il est souvent reformulé à l'aide de rapports de vraisemblance comme celui-ci :

$$O(H) = \frac{P(H)}{1 - P(H)}$$

  - ce qui signifie que les chances (O pour Odds en anglais) de H une hypothèse sont égales à la probabilité de l'hypothèse divisée par 1 moins la probabilité de cette hypothèse.
- C'est exactement comme les cotes des chevaux :
  - Si les chances de victoire de "speedy" sont de 3:2 (3 contre 2)
    - Il y a trois chances qu'il gagne pour deux chances qu'il ne gagne pas.
  - Une probabilité de 3:2 (ou 1,5) pour une maladie signifie que trois personnes l'auront pour deux qui ne l'auront pas.

Probabilités postérieures

- Ceci est lié aux probabilités conditionnelles.
- Il s'agit ici de la probabilité d'une hypothèse compte tenu des preuves,
  - qui peut être calculée comme la probabilité d'une hypothèse compte tenu des preuves, divisée par 1 moins cette probabilité:

$$O(H|E) = \frac{P(H|E)}{1 - P(H|E)}$$

- Ainsi, nous pourrions vouloir savoir que la probabilité qu'une personne soit atteinte d'une maladie est de 5 pour 2 (5:2) étant donné qu'elle a de la fièvre.

Exercice du problème du slide précédent

- On doit savoir que la probabilité qu'une personne soit atteinte de la maladie est de 5 pour 2 (5:2).
  - Cette notation est généralement interprétée comme un rapport ou une proportion, où le numérateur représente le nombre de cas favorables (dans ce cas, le nombre de personnes atteintes de la maladie) et le dénominateur représente le nombre total de cas (le total des personnes).
  - Ainsi, si nous avons 5 cas favorables (personnes atteintes de la maladie) pour 2 cas défavorables (personnes non atteintes de la maladie).
  - Cela signifie que la probabilité a priori de la maladie est P(Maladie)=$\frac{5}{5+2} = \frac{5}{7}$
    - Cela signifie que la probabilité qu'une personne soit atteinte de la maladie est de 5 sur 7, ce qui correspond à $\frac{5}{7}$.

Exercice du problème du slide précédent (Cont.)

- Nous avons donc les probabilités suivantes :
  - P(Maladie) = $\frac{5}{7}$ (probabilité a priori de la maladie)
  - P(Fièvre|Maladie) = 0,8 (probabilité de fièvre chez une personne atteinte de la maladie)
    - 0,8 est un exemple hypothétique que nous avons utilisé. Par exemple, si des études montrent que 80% des personnes atteintes de la maladie présentent de la fièvre, alors P(Fièvre|Maladie) serait de 0,8.
  - P(Fièvre|Non Maladie) = 0,2 (probabilité de fièvre chez une personne non atteinte de la maladie)
  - Mais on connais pas P(Non Maladie). On va alors le calculer avant de continuer.
    - La probabilité de non-maladie est simplement le complément de la probabilité de maladie, car les deux représentent l'ensemble complet des possibilités.
    - Donc, si P(Maladie)= $\frac{5}{7}$, alors
    - P(Non Maladie)= 1− P(Maladie)= 1 - $\frac{5}{7}$ = $\frac{2}{7}$.
  - P(Fièvre) n'est pas connu.
    - La formule pour P(Fièvre) peut sembler triviale, car elle peut être calculée en sommant les probabilités de fièvre dans les deux cas possibles : être atteint de la maladie et ne pas l'être.
    - Cela donne :
      - P(Fièvre)=(Fièvre|Maladie) × P(Maladie) + P(Fièvre|Non Maladie) × P(Non Maladie) = 0,8 x $\frac{5}{7}$ + 0,2 x $\frac{2}{7}$ = $\frac{4}{7}$.

Exercice du problème du slide précédent (Cont.)

- Connaissant maintenant:
  - P(Maladie) = $\frac{5}{7}$
  - P(Non Maladie)= $\frac{2}{7}$
  - P(Fièvre)= $\frac{4}{7}$
  - P(Fièvre|Maladie) = 0,8
  - P(Fièvre|Non Maladie) = 0,2

- Nous voulons calculer P(Maladie|Fièvre) et la probabilité postérieure *O*(Maladie|Fièvre).

Exercice du problème du slide précédent (Cont.)

- Calculons d'abord P(Maladie|Fièvre) à l'aide du théorème de Bayes :

  - $P(\text{Maladie|Fièvre}) = \dfrac{P(\text{Fièvre|Maladie})\ x\ P(\text{Maladie})}{P(Fi\grave{e}vre)}$

  - $\text{P(Maladie|Fièvre)} = \dfrac{0{,}8\ x\ \frac{5}{7}}{\frac{4}{7}}$

  - $\text{P(Maladie|Fièvre)} = \dfrac{\frac{4}{7}}{\frac{4}{7}}$

  - P(Maladie|Fièvre) = 1

- Donc, P(Maladie|Fièvre)=1, ce qui signifie que si une personne a de la fièvre, la probabilité qu'elle soit atteinte de la maladie est de 100%.

Exercice du problème du slide précédent (Cont.)

- Maintenant, calculons les probabilités postéreurs $O(\text{Maladie}|\text{Fièvre})$:

  - $O(\text{Maladie}|\text{Fièvre}) = \dfrac{P(\text{Maladie}|\text{Fièvre})}{1-P(\text{Maladie}|\text{Fièvre})}$

  - On connait déjà, à partir des résultats du slide précédent que $\mathrm{P}(\text{Maladie}|\text{Fièvre}) = 1$
  - Ce qui fait que:
    - $\mathrm{O}(\text{Maladie}|\text{Fièvre}) = \frac{1}{1-1} = \frac{1}{0}$

- Comme le dénominateur est de 0, cela signifie que les odds/probabilités postérieurs sont infinis, ce qui est un résultat logique étant donné que la probabilité postérieure est de 1 (ou 100%) (Cf. $P(\text{Maladie}|\text{Fièvre})$).
- Les odds postérieurs (posterior odds en anglais) sont le rapport entre la probabilité postérieure et la probabilité complémentaire postérieure d'un événement.
- En d'autres termes, les odds postérieurs représentent à quel point un événement est plus probable par rapport à son complément après avoir observé des données (dans ce cas, après avoir observé que la personne a de la fièvre).

Exercice du problème du slide précédent (Cont.)

- Dans le contexte de notre exemple, les odds postérieurs O(Maladie|Fièvre) représentent le rapport entre la probabilité que la personne soit atteinte de la maladie étant donné qu'elle a de la fièvre et la probabilité que la personne ne soit pas atteinte de la maladie étant donné qu'elle a de la fièvre.

- Dans notre calcul, nous avons trouvé que P(Maladie|Fièvre)=1,
- ce qui signifie que la probabilité que la personne soit atteinte de la maladie est de 100% étant donné qu'elle a de la fièvre.

- Par conséquent, les odds postérieurs sont infinis, ce qui indique que la maladie est extrêmement probable dans ce cas.

Rapport de vraisemblance positif

- Parlons du rapport de vraisemblance positif.
- **Aussi connu comme**: Niveau/Level de suffisance (LS).
- Avec le niveau de suffisance, nous voulons considérer à quel point une preuve E est adéquate pour conclure H.
- Le LS est alors calculé comme la probabilité que nous voyons la preuve E étant donné notre hypothèse H divisée par la probabilité de la preuve E étant donné que nous n'avons pas vu cette hypothèse ($\neg H$):

$$LS = \frac{P(E|H)}{P(E|\neg H)}$$

- Nous pouvons maintenant réécrire la probabilité postérieure ou la probabilité de l'hypothèse compte tenu des preuves en termes de niveau de suffisance multiplié par la probabilité de l'hypothèse elle-même:

$$\boxed{O(H|E) = LS * O(H)}$$

$$O(H) = \frac{P(H)}{1 - P(H)}$$

Le problème avec Bayes…

- Revenons rapidement sur Bayes où le cadre du raisonnement bayésien nécessite des valeurs de probabilité comme entrées/inputs principales:
  - Vous pouvez commencer par utiliser les données des patients pour déterminer les rapports de vraisemblance et les probabilités préalables de différentes maladies.
  - Et si les données n'étaient pas disponibles, un expert pourrait faire une supposition éclairée.
  - Cependant, il existe bien sûr de nombreuses sources d'erreur potentielles.
    - D'une part, les symptômes peuvent ne pas être indépendants de la maladie.
    - Les rapports de vraisemblance et les probabilités préalables peuvent être inexacts sur la base d'échantillons de patients non représentatifs ou trop petits ou de mauvaises suppositions d'experts.
  - Souvent, les systèmes simples qui traitent de l'incertitude peuvent donner l'illusion de la précision.
    - Par exemple, ils peuvent produire une probabilité élevée pour un résultat donné, telle que 96 %.
    - Mais ces valeurs peuvent contenir des erreurs critiques si le système n'a pas été configuré correctement ou si les hypothèses sur lesquelles reposent les calculs de probabilité n'ont pas été respectées.

Le problème avec Bayes... (Cont.)

- De plus, si nous ne faisons pas d'hypothèses sur l'indépendance conditionnelle, nous avons besoin d'énormes tables de probabilités conditionnelles.
  - Par exemple, quelle est la probabilité de chaque hypothèse $P(H|E1 \wedge E2 \wedge \ldots \wedge En)$ pour chaque considération unique de preuve possible.
  - En général, ce n'est pas pratique.
    - Par exemple, étant donné 16 symptômes possibles,
      - nous aurions besoin d'un tableau avec 2^16 = 65 536 entrées pour chaque maladie!
      - Il est peu probable que nous disposions de suffisamment de données sur les anciens patients pour obtenir pour obtenir des probabilités précises pour chaque maladie.
- Le message est donc que les performances des systèmes de raisonnement statistique bayésien se détériorent à mesure qu'un plus grand nombre de maladies et de symptômes sont pris en compte.
- Cela est principalement dû à la violation de l'indépendance conditionnelle dont nous avons discuté précédemment.

Facteurs de certitude

Facteurs de certitude (FC)

- FC est une grande simplification.
- Ces facteurs offrent une alternative populaire au raisonnement bayésien.
- Il s'agit en fait d'un simple chiffre qui mesure la conviction de l'expert,
- Où la valeur maximale du facteur de certitude étant, par exemple, est de +1,0 (tout à fait vrai) et la valeur minimale est de -1,0 (tout à fait faux).
- Par exemple, si l'expert déclare qu'une preuve est presque certainement vraie, une valeur FC de 0,8 sera attribuée à cette preuve.
- Il est important de noter que les facteurs de certitude ne sont pas les mêmes que les probabilités, mais ils sont bien sûr conceptuellement et intuitivement liés.

Facteurs de certitude dans le système expert MYCIN

Terme	Facteur de certitude
Certainement pas	-1
Presque certainement pas	-0,8
Probablement pas	-0,6
Peut-être pas	-0,4
Inconnu	-0,2 à + 0,2
Peut-être	+0,4
Probablement	+0,6
Presque certainement	+0,8
Certainement	+1

- Il est important de noter que les facteurs de certitude peuvent être interprétés différemment en fonction du système et du contexte spécifique dans lequel ils sont utilisés.
- Ces valeurs peuvent être adaptées pour correspondre aux besoins et aux exigences du système expert en question.

Facteurs de certitude dans les systèmes experts

- Dans les systèmes experts qui utilisent des facteurs de certitude, la base de connaissances consiste en un ensemble de règles qui ont la syntaxe suivante avec un facteur de certitude associé {FC} :

- IF <évidence>

  THEN <hypothèse> {FC}

- Là encore, ce facteur de certitude représente une croyance dans l'hypothèse H étant donné que la preuve E s'est produite.

Théorie des facteurs de certitude

- Elle repose sur deux fonctions.
  - La première est la mesure de confiance/belief MB(H,E), qui est la mesure de l'augmentation de la confiance en H due à E.
  - La seconde est la mesure de méfiance:disbelief MD(H,E), qui mesure l'augmentation de la méfiance en H due à E.
- Voici comment ces deux mesures peuvent être calculées:

$$MB(H,E) = \begin{cases} 1 & \text{if } p(H) = 1 \\ \dfrac{max\left[p(H|E), p(H)\right] - p(H)}{max\left[1,0\right] - p(H)} & \textbf{sinon} \end{cases}$$

$$MD(H,E) = \begin{cases} 1 & \text{if } p(H) = 0 \\ \dfrac{min\left[p(H|E), p(H)\right] - p(H)}{min\left[1,0\right] - p(H)} & \textbf{sinon} \end{cases}$$

- Dans la figure ci-dessus, la mesure de confiance (MB) en l'hypothèse H compte tenu de la preuve E peut être égale à 1 si la probabilité que l'hypothèse soit toujours vraie (c'est-à-dire P(H)=1) compte tenu de la preuve.
- Dans le cas contraire, nous pourrions la calculer en utilisant des maximums pour notre mesure de confiance et des minimums pour notre mesure de méfiance.
- P(H) est la probabilité préalable que l'hypothèse H soit vraie.
- P(H|E) est la probabilité conditionnelle, c'est-à-dire la probabilité que l'hypothèse H soit vraie compte tenu de la preuve E.

Calcul de facteur de certitude

- Les valeurs de MB(H, E) et MD(H, E) sont donc comprises entre 0 et 1.
- La degré de confiance ou de méfiance à l'égard de l'hypothèse H dépend du type de preuve E observée.
- Certains faits peuvent augmenter le degré de confiance, tandis que d'autres augmentent le degré de méfiance.
- Le degré total de confiance ou de méfiance à l'égard d'une hypothèse est calculé comme suit, où FC c'est facteur de certitude :

$$\text{FC} = \frac{MB(H,E) - MD(H,E)}{1 - min[MB(H,E), MD(H,E)]}$$

- Notamment, ces calculs de facteurs de certitude sont beaucoup plus simples d'un point de vue mathématique.

Utilisation de facteur de certitude

- Voyons donc un exemple d'utilisation des facteurs de certitude.
- Imaginons que nous ayons une règle simple :

  IF A est X

  THEN B est Y
- Un expert peut ne pas être absolument certain que cette règle est valable.
- Supposons également qu'il ait été observé que dans certains cas, même lorsque la partie IF de la règle est satisfaite et que l'objet A prend la valeur X, l'objet B peut acquérir une valeur différente Z.
- Nous avons donc ici si A est X alors B est Y avec un facteur de certitude de 0,7 ou B est Z avec un facteur de certitude de 0,2:

  IF A est X

  THEN B est Y {FC 0.7};

  B est Z {FC 0.2}

Facteur de certitude : Propagation des règles

- Nous avons maintenant examiné certains facteurs dans le cadre de règles individuelles.
- Mais comme nous l'avons vu jusqu'à présent, les règles fonctionnent souvent de concert les unes avec les autres.
- Nous allons donc commencer à examiner comment la certitude se propage lorsqu'une règle en déclenche une autre.
- Ces facteurs de certitude attribués aux règles se propagent dans la chaîne de raisonnement.
  - Il s'agit d'établir la certitude nette de la règle conséquente lorsque les preuves de l'antécédent de la règle sont incertaines.
  - En d'autres termes, il peut même y avoir certains facteurs associés aux composantes IF des règles ou notre degré de confiance de l'hypothèse compte tenu d'un élément de preuve (c'est-à-dire FC(H,e)) est égale à la confiance que nous avons dans cette preuve au départ (c'est-à-dire FC(E,e)) multipliée par la confiance que nous avons dans cette hypothèse compte tenu du fait que nous disposons d'un élément de preuve (c'est-à-dire FC(H,E)) :

$$FC(H,e) = FC(E,e) \times FC(H,E)$$

Facteur de certitude : Propagation des règles (Cont.)

- Encore un exemple simple :
- IF "le ciel est clair"

  THEN "les prévisions sont ensoleillées" {FC 0,8}
- et disons également que notre facteur de certitude actuel du fait que "le ciel est clair" n'est que de 50 % (c'est-à-dire 0.5),
- alors notre confiance dans la conclusion H compte tenu de la preuve E est la suivante :

$$FC(H,E) = 0,5 \times 0,8 = 0,4$$

- Cela signifie que le facteur de certitude total est de 0,4 pour cette conclusion.
- Ce résultat peut être interprété qualitativement comme "il est possible qu'il fasse beau".

Facteurs de certitude dans les règles conjonctives

- Examinons maintenant les facteurs de certitude dans les règles conjonctives ou les règles avec le nombre de AND entre les preuves, comme dans l'exemple de règles suivant :

IF <evidence E_1>
⋮
AND <evidence E_n>
THEN <hypothèse H> {FC}

- La certitude de l'hypothèse H est établie comme suit :

  FC(H,E1 ∧ E2 ∧ ... ∧ En) = min [FC(E1,e), FC(E2,e),..., FC(En,e)] x FC.

- Par exemple,
- IF "le ciel est clair"

  AND "les prévisions sont ensoleillées"

  THEN "porter des lunettes de soleil" {FC 0,8}.

- Maintenant si la certitude que "le ciel soit clair est de 0,9" et que la certitude que "la prévision soit ensoleillée " est de 0,7, alors nous pouvons conclure que la fiabilité de notre hypothèse H, compte tenu de ces deux preuves, est de 0,56, c'est-à-dire que nous prenons le minimum de la certitude de nos preuves et le multiplions par le facteur de confiance de nos règles :

  FC(H,E1 ∧ E2) = min [0,9, 0,7] × 0,8 = 0,7 × 0,8 = 0,56

Facteurs de certitude dans les règles disjonctives

- Examinons maintenant les facteurs de certitude dans les règles disjonctives ou les règles avec le nombre de OR entre les preuves, comme dans l'exemple de règles suivant :

IF <evidence E_1>
⋮
OR <evidence E_n>
THEN <hypothèse H> {FC}

- La certitude de l'hypothèse H est établie comme suit :

$$FC(H, E1 \vee E2 \vee ... \vee En) = \max[FC(E1,e), FC(E2,e), ..., FC(En,e)] \times FC.$$

- Par exemple,
- IF "le ciel est couvert"

OR "les prévisions annoncent de la pluie"

THEN "prendre un parapluie" {FC 0,9}.

- Maintenant si la certitude que "le ciel soit couvert est de 0,6" et que la certitude que "la prévision soit la pluie" est de 0,8, alors nous pouvons conclure que la fiabilité de notre hypothèse H, compte tenu de ces deux preuves, est de 0,72, c'est-à-dire que nous prenons le minimum de la certitude de nos preuves et le multiplions par le facteur de confiance de nos règles :

$$FC(H, E1 \vee E2) = \max[0{,}6,\ 0{,}8] \times 0{,}9 = 0{,}8 \times 0{,}9 = 0{,}72$$

Différence entre le raisonnement bayésien et les facteurs de certitude

- La théorie des probabilités est la technique la plus ancienne et la mieux établie pour traiter les connaissances inexactes et les données aléatoires.
  - Elle fonctionne bien dans des domaines tels que la prévision et la planification, où les données statistiques sont généralement disponibles et où des déclarations de probabilité précises peuvent être faites.
  - La propagation des croyances bayésiennes est d'une complexité exponentielle et n'est donc pas pratique pour les grandes bases de connaissances.
- D'autre part, la théorie de la certitude n'est pas "mathématiquement pure", mais elle imite le processus de réflexion d'un expert humain.
  - L'utilisation de facteurs de certitude tend à surpasser le raisonnement bayésien subjectif dans des domaines tels que le diagnostic.
  - Les facteurs de certitude sont utilisés dans les cas où les probabilités ne sont pas connues ou sont trop difficiles ou coûteuses à obtenir.
  - Le mécanisme de raisonnement probatoire (évidence) peut gérer les preuves acquises progressivement, la conjonction et la disjonction d'hypothèses, ainsi que les preuves avec différents degrés de croyance.
  - L'approche des facteurs de certitude permet également de mieux expliquer le flux de contrôle par le biais d'un système expert basé sur des règles.
- En fin de compte, le raisonnement bayésien est plus précis et plus mathématiquement correct, tandis que les facteurs de certitude sont plus faciles à utiliser et à interpréter et imitent mieux la façon dont les humains pensent, que ce soit bien ou mal.

Incertitude dans les systèmes experts

Incertitude dans les systèmes experts

- Dans les systèmes experts, les probabilités requises pour résoudre un problème sont fournies par des experts.
- Un expert déterminera les probabilités préalables pour les hypothèses possibles, par exemple $P(H)$ et $P(\neg H)$, ainsi que les probabilités conditionnelles pour l'observation de la preuve E si l'hypothèse H est vraie, c'est-à-dire $P(E|H)$, et si l'hypothèse H est fausse, c'est-à-dire $P(E|\neg H)$.
- Les utilisateurs fourniront ainsi des informations sur les preuves observées et le système expert calculera $P(H|E)$ pour l'hypothèse H à la lumière des preuves E fournies par l'utilisateur.

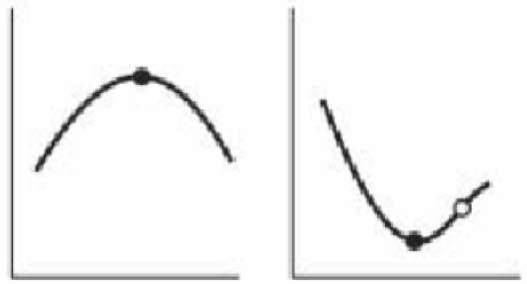
Systèmes experts monotones

- En bref, il existe deux grands types de systèmes experts.
  - Les systèmes experts les plus simples sont connus sous le nom de systèmes experts monotones.
  - Les systèmes experts plus difficiles à construire connus le nom de systèmes experts non monotones.
- Les systèmes experts basés sur des règles utilisent principalement une logique monotone, de sorte qu'ils ne peuvent pas revenir sur leurs conclusions.
  - Un exemple célèbre de système expert monotone ou la logique monotone ne fonctionne pas très bien
    - En 1982, Marvin Minsky a donné l'exemple suivant :
      - Tout d'abord, un humain indique que "tous les canards peuvent voler". Et Charlie est un canard".
      - Le système expert raisonne ainsi : "Alors Charlie peut voler".
      - Puis l'homme ajoute : "Mais Charlie est maintenant mort".
      - Le système expert doit alors être en mesure de modifier cette conclusion et de raisonner comme suit : "Charlie ne peut plus voler".
  - Malheureusement, les systèmes experts monotones ne peuvent pas apprendre des connaissances qui contredisent celles qui sont déjà connues.
  - En d'autres termes, un tel système ne peut jamais remplacer une affirmation par sa négation.
  - Par conséquent, la base de connaissances ne s'enrichit que de nouveaux faits de manière monotone et unidirectionnelle.
  - Ces systèmes experts monotones présentent deux avantages :
    - ils simplifient considérablement la gestion de la vérité
    - il existe une plus grande diversité de stratégies d'apprentissage que vous pouvez utiliser avec ces systèmes.

Systèmes experts non monotones

- En revanche, les systèmes experts non monotones peuvent apprendre des connaissances qui contredisent celles qu'ils connaissent déjà.
- Pour ce faire, ils remplacent les anciennes connaissances par de nouvelles.
- Avantages :
  - Meilleure applicabilité aux domaines réels
  - Plus grande liberté d'apprentissage.

- Ces systèmes sont beaucoup plus difficiles à construire, mais il existe aujourd'hui quelques systèmes basés sur des logiques non monotones.



Autres représentations de l'incertitude

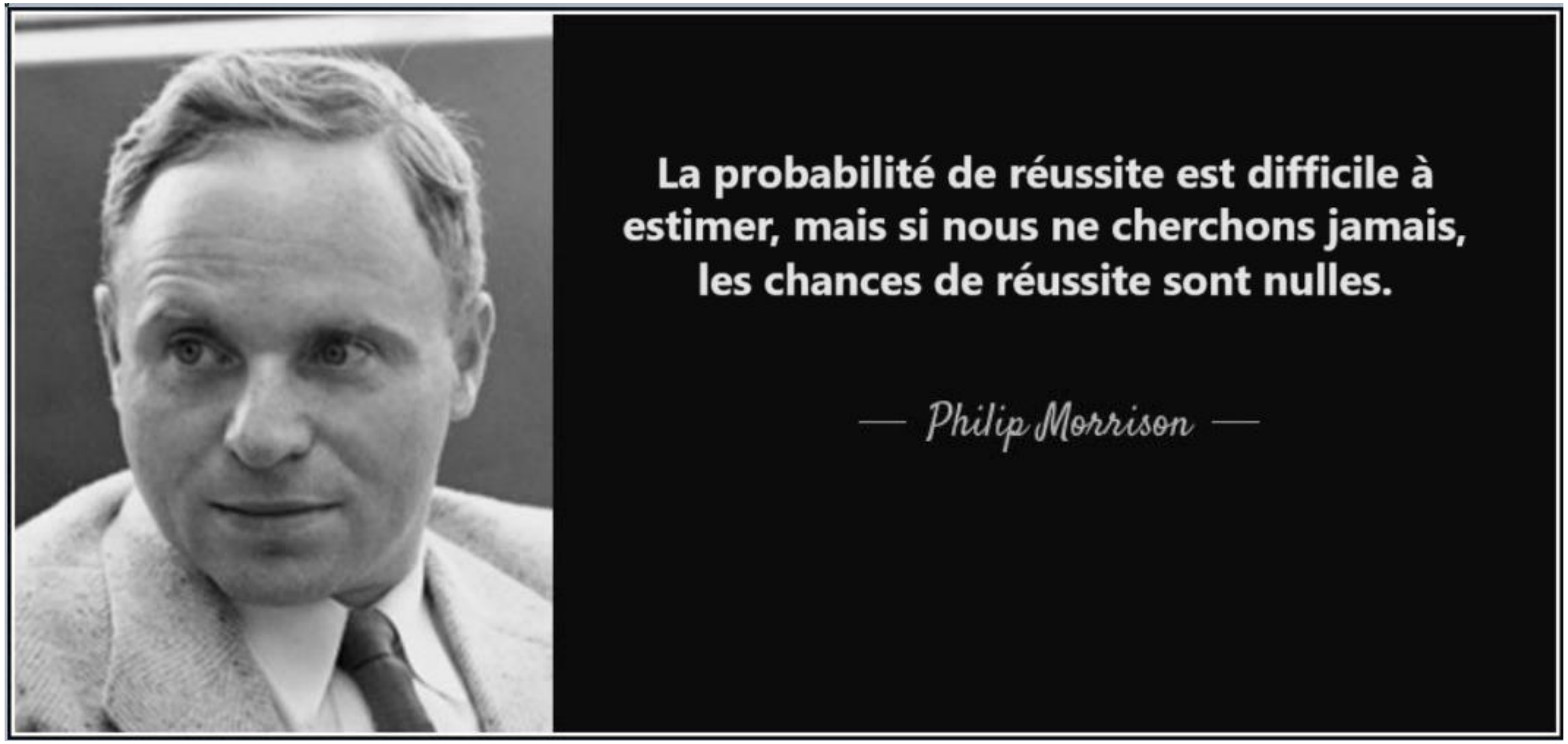
- Nous avons évoqué quelques façons d'envisager l'incertitude dans le contexte des probabilités et des systèmes experts.
- Mais il existe d'autres façons de représenter l'incertitude.
  - Raisonnement par défaut :
    - il s'agit ici d'une logique non monotone qui permet la rétractation des croyances par défaut si elles s'avèrent fausses.
  - Raisonnement par preuve :
    - Théorie de Dempster-Shafer : Bel(P) est une mesure de l'évidence pour P ;
    - Bel(¬P) est une mesure de l'évidence contre P ;
      - ensemble, ils définissent un intervalle de croyance (limites inférieures et supérieures de la confiance).
  - Raisonnement flou:
    - Ensembles flous : Dans quelle mesure un objet satisfait-il une propriété vague ?
    - Logique floue : "Dans quelle mesure un énoncé logique est-il vrai ?
  - Réseaux bayésiens (Lire ce texte)

Compromis en matière de représentation de l'incertitude

- Probabilité (Bayes) :
  - Grande complexité de l'inférence, mais mesures quantitatives de l'incertitude
- Facteurs de certitude :
  - Pas de fondement sémantique, les règles doivent être indépendantes, facteurs de certitude disponibles
  - Permet une inférence basée sur des règles, plus simple que les probabilités.
- Logique non monotone :
  - Représente le raisonnement de bon sens, mais peut être très coûteuse sur le plan informatique.
- Théorie de Dempster-Shafer :
  - Possède de belles propriétés formelles, mais elle est complexe et peut être très coûteuse sur le plan informatique, et les intervalles ont tendance à croître vers [0,1] (conclusion peu utile).
- Raisonnement flou :
  - La sémantique n'est pas claire (floue!), les chaînes d'inférence plus longues posent problème.
  - Facile à concevoir, règles intuitives, utile pour les applications commerciales.
- Réseaux bayésiens :
  - De bonnes propriétés théoriques combinées à un raisonnement efficace rendent les réseaux bayésiens très populaires ; l'expressivité limitée, les défis en matière d'ingénierie des connaissances peuvent limiter les utilisations.

Synthèse de la leçon

- Théorie des probabilités de base
  - Probabilités (individuelles, disjointes et conjointes)
  - Probabilité conditionnelle
  - Règle de Bayes
- Inférence statistique
  - Inférence avec la règle de Bayes
- Hypothèses d'indépendance
  - Indépendance conditionnelle
  - Raisonner avec des preuves et/ou des hypothèses multiples
  - Calcul des vraisemblances
- Rapports de vraisemblance
  - Probabilités/Odds postérieures
  - Degré de confiance
  - Problèmes avec Bayes
- Facteurs de certitude
  - Utilisation dans les systèmes experts
  - Calcul
  - Propagation
- Incertitude dans les systèmes experts


La probabilité de réussite est difficile à estimer, mais si nous ne cherchons jamais, les chances de réussite sont nulles.
— Philip Morrison —

Travail pratique 3: Rappel

- **Objectifs :**
  - Représentation des connaissances, systèmes experts et Experta
  - Pratiquer la représentation des cadres
  - Apprendre à installer et à utiliser Experta
  - Pratiquer la représentation des règles
  - Construire un système expert très simple pour le diagnostic
  - Apprendre à programmer des questions
  - Comprendre les chaînages avant et arrière dans les systèmes experts basés sur des règles