



Éléments de statistiques

□ Intervalle de Confiance :

un **intervalle de confiance** encadre une valeur réelle que l'on cherche à estimer à l'aide de mesures prises par un procédé aléatoire. En particulier, cette notion permet de définir une marge d'erreur entre les résultats d'un sondage et un relevé exhaustif de la population totale. l'estimation d'une grandeur dépend à la fois du nombre d'expériences et de la variabilité des observations. On peut combiner élégamment la taille de l'échantillon et sa variabilité pour évaluer la différence qui peut exister entre l'estimation d'une grandeur et sa « vraie » valeur.

Définition

Un intervalle de confiance, au seuil de confiance $1 - \alpha$, pour une variable aléatoire Y , est un intervalle qui, avec une probabilité $1 - \alpha$, contient la valeur de l'espérance mathématique de Y

=> **Donc plus l'intervalle de confiance est petit, plus on peut avoir confiance en l'estimation de la grandeur à modéliser.**



Les Réseaux Bayésiens

□ Définition :

- ✓ Les réseaux bayésiens sont la combinaison **des approches probabilistes et la théorie de graphes**. Autrement dit, ce sont des modèles qui permettent de représenter des situations de raisonnement probabiliste à partir de connaissances incertaines. Connus sous le nom de "**belief networks**", "**causal networks**", qui est:

A. Modèle de représentation des connaissances ;

B. « Machine à calculer » des probabilités conditionnelles

C. Base pour des systèmes d'aide à la décision

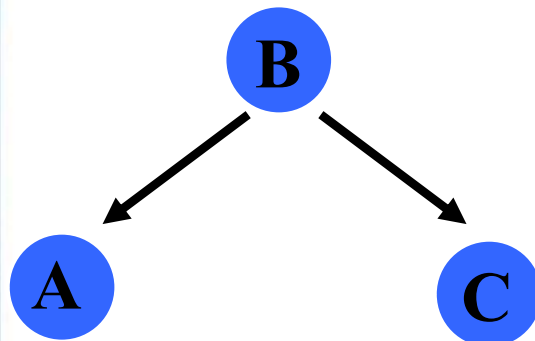
- ⇒ Un réseau bayésien est un graphe acyclique orienté i.e. c'est un graphe orienté sans circuit.
- ✓ Chaque nœud d'un réseau bayésien est une variable aléatoire qui porte une étiquette des attributs de problème.
- ✓ les arcs représentent des dépendances (**par exemple des causalités**) probabilistes entre les variables et des distributions de probabilités conditionnelles (locales) pour chaque variable étant donnés ses parents

Les Réseaux Bayésiens

□ Définition :

- On décrit **les relations causales** entre variables d'intérêt par un graphe. Dans ce graphe, **les relations de cause** à effet entre les variables ne sont pas déterministes, mais **probabilisées**. Les réseaux bayésiens sont surtout utilisés pour le diagnostic (médical et industriel), l'analyse de risques, la détection des spams et le data mining.

➤ Exemple :



A et C sont conditionnellement indépendants alors on peut dire
que la probabilité $P(A|B,C) = P(A|B)$

C'est-à-dire que la probabilité de A ne dépend que celle de B

La probabilité jointe de toutes les variables est

$$P(A,B,C) = P(A|B) \cdot P(B) \cdot P(C|B)$$



Les Réseaux Bayésiens

□ Formule de Bayes :

- ✓ Thomas Bayes, 1702-1761, un théologien et mathématicien britannique qui a écrit une loi de base de probabilité qui est maintenant appelée **le théorème de Bayes**.
- ⇒ **On sait que deux événements A et B sont dits indépendants si $P(A,B) = P(A) \cdot P(B)$**
- ✓ Deux événement **A** et **B** qui sont **conditionnellement dépendants** et un contexte **c**, le théorème de Bayes peut être représenté comme ci-dessous :

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

$P(B|A,c)$ est la probabilité a **posteriori** ou la probabilité de **B** après avoir pris en compte l'effet de **A** dans un contexte **c**,

$P(B|c)$ est la probabilité a **priori** de l'événement **B**,

$P(A|B,c)$ est la probabilité de **A** si l'on suppose que **B** est vrai dans un contexte **c**. Elle est appelée aussi la **"vraisemblance"**,

$P(A|c)$ est la **normalisation**.



Les Réseaux Bayésiens

□ Construction de réseaux bayésiens :

- ✓ Construire un réseau bayésien, c'est donc :
 1. Définir le graphe du modèle ;
 2. Définir les tables de probabilité de chaque variable, conditionnellement à ses causes.
- ✓ Le graphe est aussi appelé la « **structure** » du modèle, **et les tables de probabilités ses « paramètres »**. Structure et paramètres peuvent être fournis par des experts, ou calculés à partir de données, même si en général, la structure est définie par des experts et les paramètres calculés à partir de données expérimentales.

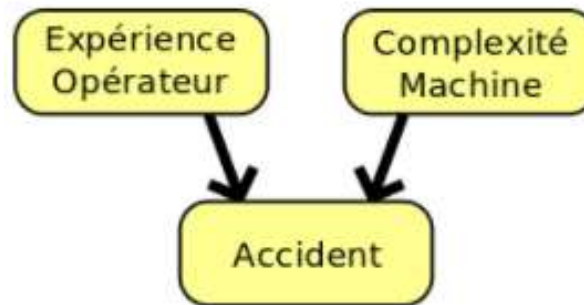
□ Utilisation des réseaux bayésiens :

L'utilisation d'un réseau bayésien s'appelle « **inférence** ». Il est une « **machine à calculer des probabilités conditionnelles** ». En fonction **des informations observées**, on calcule la probabilité des données non observées. Par exemple, en fonction des symptômes d'un malade, on calcule les probabilités des différentes pathologies compatibles avec ces symptômes. On peut aussi calculer la probabilité de symptômes non observés, et en déduire les examens complémentaires les plus intéressants.

Les Réseaux Bayésiens

➤ Exemple (la modélisation des risques) :

- ✓ Un opérateur travaillant sur une machine risque de se blesser s'il l'utilise mal. Ce risque dépend de l'expérience de l'opérateur et de la complexité de la machine. Alors « **Expérience** » et « **Complexité** » sont deux facteurs déterminants de ce risque.
- ✓ Bien sûr, ces facteurs ne permettent pas de créer un modèle déterministe. Quand bien même l'opérateur serait expérimenté et la machine simple, un accident reste possible. D'autres facteurs peuvent jouer : l'opérateur peut être fatigué, dérangé, etc. La survenance du risque est toujours aléatoire, mais la probabilité de survenance dépend des facteurs identifiés.



- ✓ On voit que la probabilité d'accident augmente si l'utilisateur est peu expérimenté ou la machine complexe.

Complexité	Faible			Moyenne			Elevée		
Expérience	Faible	Moyenne	Elevée	Faible	Moyenne	Elevée	Faible	Moyenne	Elevée
Accident	1.0	0.5	0.1	1.5	1.0	0.6	2.0	1.6	1.1
Pas d'accident	99.0	99.5	99.9	98.5	99.0	99.4	98.0	98.4	98.9



Les Réseaux Bayésiens

➤ Exemple (Possibilité d'alarme) :

Considérons la situation suivante :

- ✓ je suis au travail, et mes voisins Marie et Jean m'ont promis de m'appeler chaque fois que mon alarme sonne
- ✓ Mon voisin Jean m'appelle pour me dire que mon alarme sonne (parfois il confond l'alarme avec la sonnerie du téléphone)
- ✓ Par contre ma voisine Marie ne m'appelle pas toujours (parfois elle met la musique trop fort)
- ✓ parfois mon alarme se met à sonner lorsqu'il y a de légers séismes
- ✓ comment conclure qu'il y a un cambriolage chez moi?

⇒ **Comment peut-on représenter ce problème par un Réseau Bayésien ??**

1. Extraction des variables aléatoires sont :

- | | |
|----------------|---------------|
| - JeanAppelle | - Cambriolage |
| - MarieAppelle | - Séisme |
| - Alarme | |

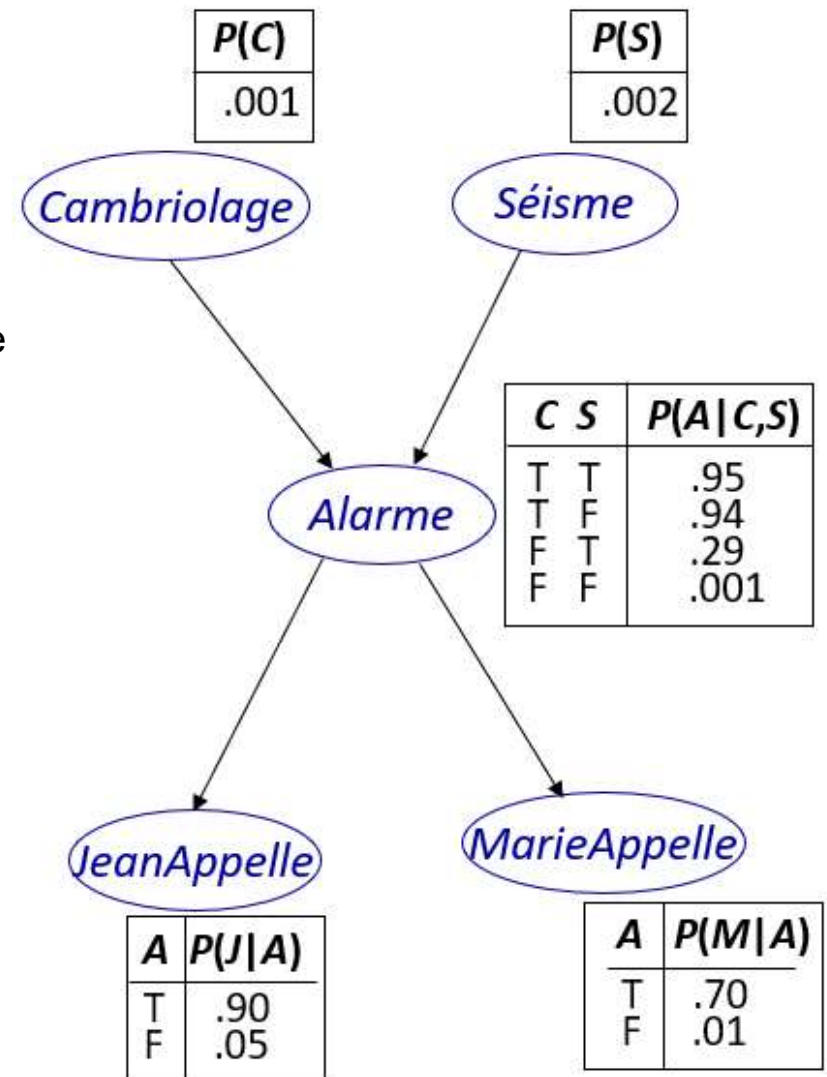
Les Réseaux Bayésiens

➤ Exemple (Possibilité d'alarme) :

- ❖ La topologie du Réseau Bayésien modélise les relations de causalité
- ❖ Un arc d'un nœud **X** vers un nœud **Y** signifie que la variable **X** influence la variable **Y**
- ❖ un cambriolage peut déclencher l'alarme
- ❖ un séisme le fait aussi
- ❖ l'alarme peut inciter Jean à appeler
- ❖ La même chose pour Marie

Une table de probabilités conditionnelles (TPC) donne la probabilité pour chaque valeur du nœud étant donnés les combinaisons des valeurs des parents du nœud (c'est l'équivalent d'une distribution)

2. Représentation des nœuds :

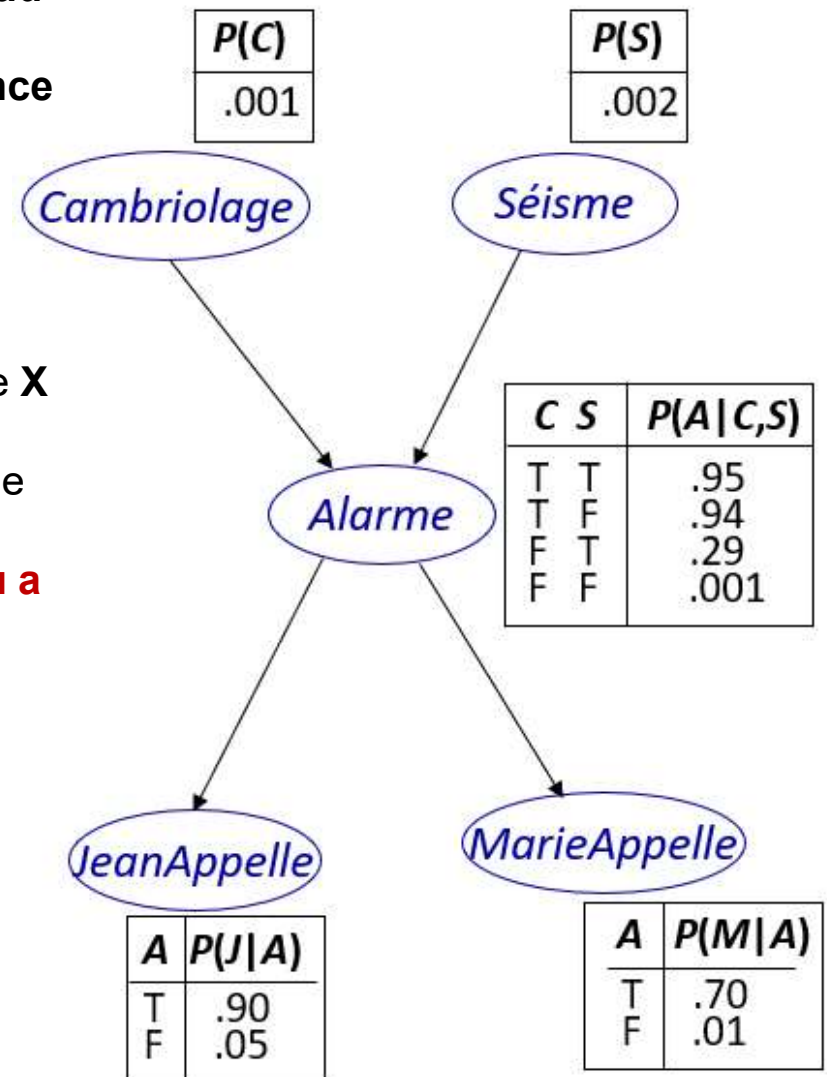


Les Réseaux Bayésiens

➤ Exemple (Possibilité d'alarme) :

- ❖ S'il y a un arc d'un nœud **Y** vers un nœud **X**, cela signifie que la **variable Y influence sur la variable X**
- ❖ Y est appelé **le parent de X**
- ❖ **Parents(X)** = l'ensemble des parents de **X**
- ❖ Si **X** n'a pas de parents, sa distribution de probabilités est dite **inconditionnelle ou a priori**
- ❖ Si **X** a des parents, **sa distribution de probabilités est dite conditionnelle**

2. Représentation des nœuds :





Les Réseaux Bayésiens

□ Formulation du réseaux bayésien :

Un Réseau Bayésien est une façon simple pour représenter des probabilités conjointes

- ❖ Par définition, la probabilité conjointe de **X1** et **X2** est donnée par la distribution **P(X1,X2)**, pour une valeur donnée de **X1** et **X2**
- ❖ La distribution conditionnelle de **X1** sachant **X2** est notée **P(X1 | X2)**

$$P(X1,X2) = P(X1 | X2) P(X2)$$

Soit **X = {X1, ..., Xn}**, l'ensemble des variables d'un Réseau Bayésien :

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Parents(X_i))$$

⇒ **En d'autres mots, la distribution conjointe des variables d'un Réseau Bayésien est définie comme étant le produit des distributions conditionnelles (locales)**

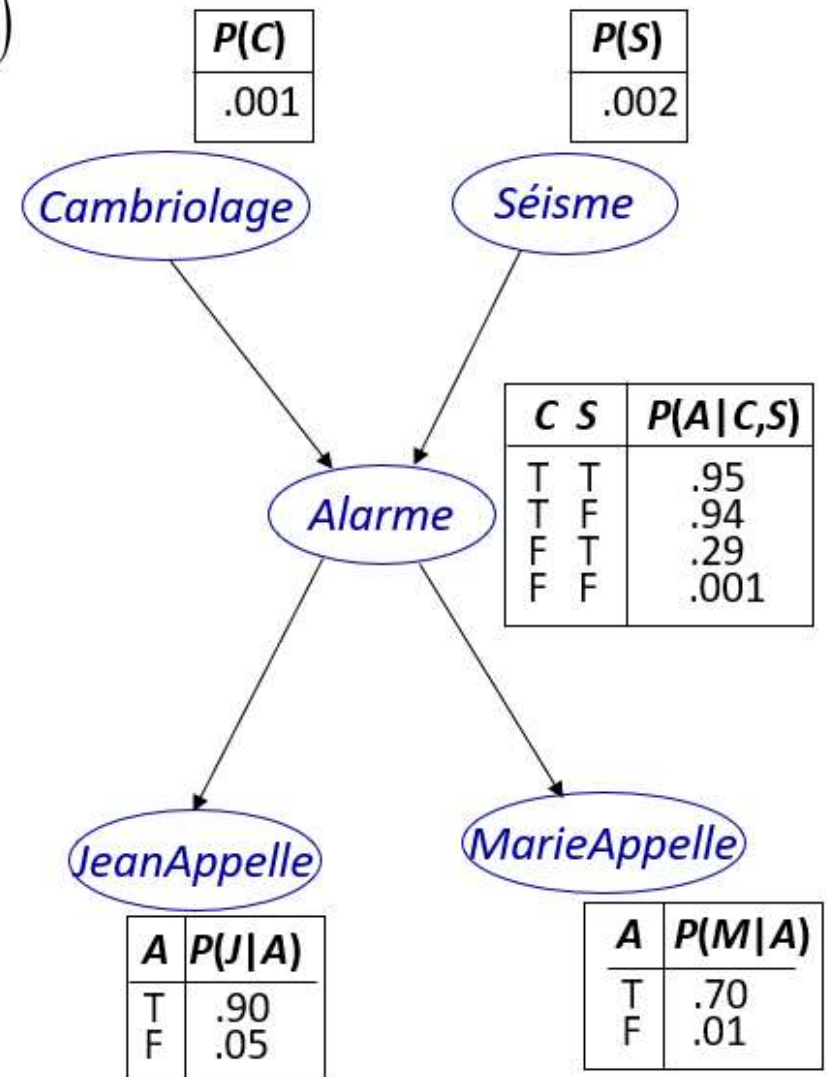
Les Réseaux Bayésiens

➤ Exemple (Possibilité d'alarme) :

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Parents}(X_i))$$

$$\begin{aligned} &P(j, m, a, \neg c, \neg s) \\ &= P(j | a) P(m | a) P(a | \neg c, \neg s) \\ &\quad P(\neg c) P(\neg s) \\ &= .90 * .70 * .001 * \\ &\quad .999 * .998 \\ &\approx .00062 \end{aligned}$$

2. Représentation des nœuds :



Les Réseaux Bayésiens

➤ Exemple (Possibilité d'alarme) :

Indépendance Conditionnelle

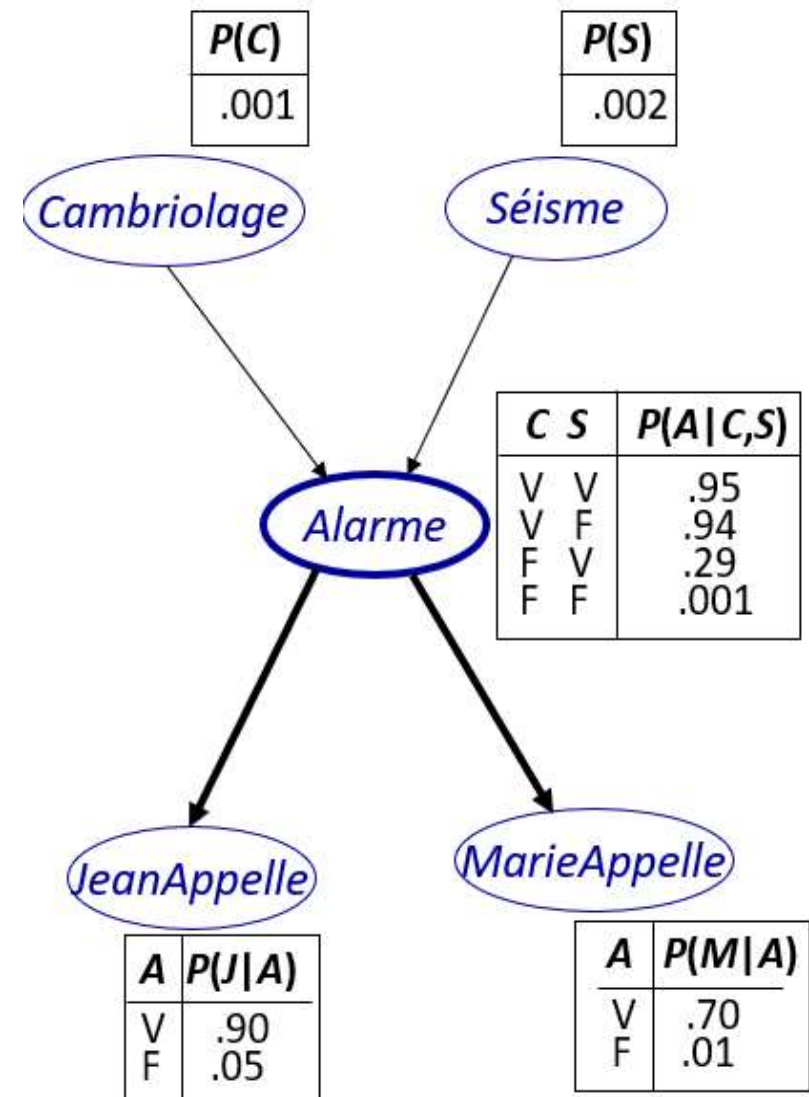
$$P(M|A,J) = P(M,A,J) / P(A,J)$$

$$= \frac{\sum_s \sum_c P(M,A,J,S=s,C=c)}{\sum_s \sum_c P(A,J,S=s,C=c)}$$

$$= \frac{\sum_s \sum_c P(J|A) P(M|A) P(A,S=s,C=c)}{\sum_s \sum_c P(J|A) P(A,S=s,C=c)}$$

$$= \frac{P(M|A) \sum_s \sum_c P(J|A) P(A,S=s,C=c)}{\sum_s \sum_c P(J|A) P(A,S=s,C=c)}$$

$$= P(M|A)$$



Les Réseaux Bayésiens

➤ Exemple (Diagnostique Médicale) :

- Est-ce que *Age* et *Gender* sont indépendants ?
 - ◆ chemin 1 est bloqué au niveau de *Smoking* → (N) ←
 - » *Smoking* et ses descendants *Cancer*, *Serum-Calcium* et *Lung-Tumor* ne sont pas observés
 - ◆ chemin 2 est aussi bloqué au niveau de *Smoking* → (N) ←
 - » même raisons
- Réponse : **oui**



Les Réseaux Bayésiens

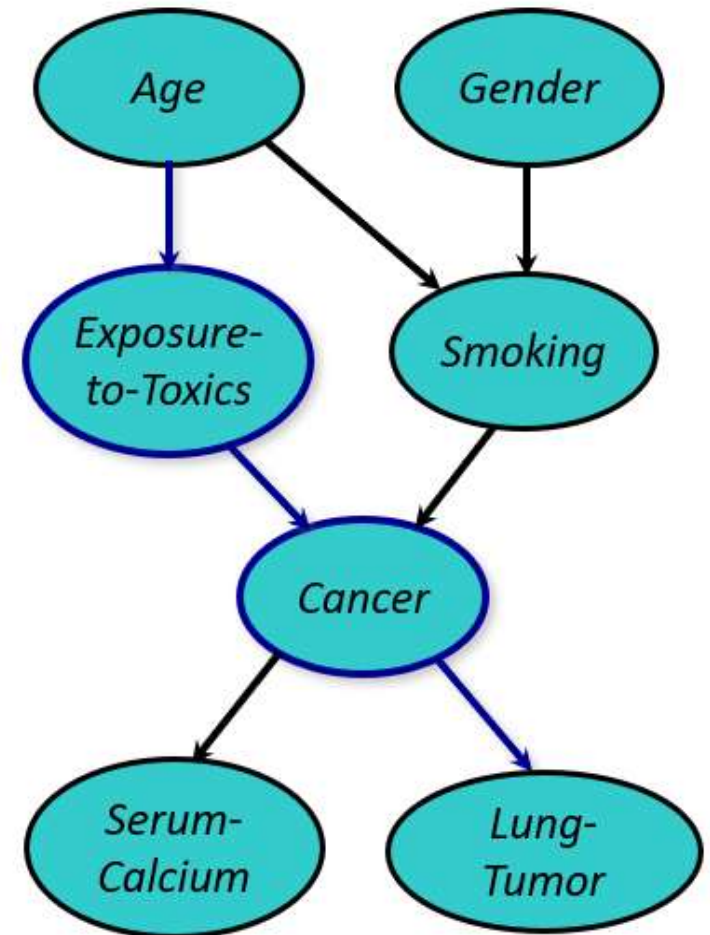
➤ Exemple (Diagnostique Médicale) :

- Est-ce que *Age* et *Lung-Tumor* sont indépendants sachant *Smoking* ?

- ◆ chemin 1 est bloqué au niveau de *Smoking* → (N) →
» *Smoking* est observé

- ◆ chemin 2 n'est pas bloqué
» *Exposure-to-Toxics* → (N) →
» *Cancer* → (N) →
n'est pas observé

- Réponse : **non**





La chaine de Markov

□ Définition :

- N.B : un **processus stochastique** est un processus de probabilité qui représente **l'évolution discrète ou à temps continu d'une variable aléatoire**
- Une **chaîne de Markov** est un processus de Markov « c'est un processus stochastique possédant la propriété de Markov à temps discret, ou à temps continu et à espace d'états discret où l'information utile pour la prédiction du futur est entièrement contenue dans l'état présent du processus et non pas dans les états antérieurs c-à-d **(le système n'a pas de « mémoire »)**. »
- La chaine de Markov peut être représentée en plusieurs ordres :
 - ✓ **En ordre 0 (L'indépendance) :** $P(x_i | x_1, x_2, \dots, x_{i-1}) = P(x_i)$
 - ✓ **En ordre 1 :** $P(x_i | x_1, x_2, \dots, x_{i-1}) = P(x_i | x_{i-1})$
 - ✓ **En ordre 2 :** $P(x_i | x_1, x_2, \dots, x_{i-1}) = P(x_i | x_{i-1}, x_{i-2})$

La chaine de Markov

□ Formulation :

Definition – La matrice de Transition -

Le nombre $\mathbb{P}(X_1 = j \mid X_0 = i)$ est **appelé probabilité de transition de l'état i à l'état j** en un pas, On la note souvent : $p_{i,j} = \mathbb{P}(X_1 = j \mid X_0 = i)$.

La famille de nombres $P = (p_{i,j})_{(i,j \in E^2)}$ où (E est un espace des états) : est appelée **matrice de transition** ou **noyau de transition**, de la chaîne de Markov.

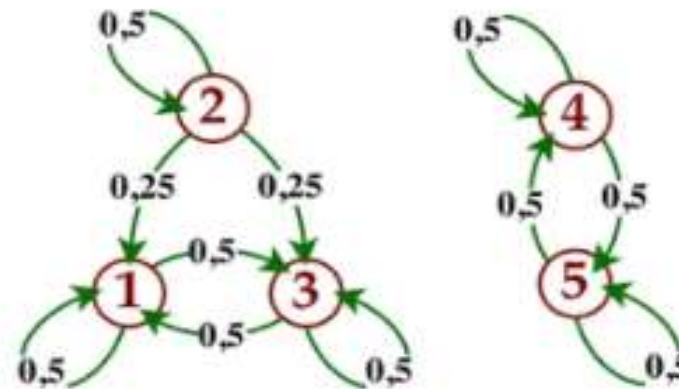
$$\begin{aligned}\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) \\&= \mathbb{P}(X_0 = x_0) \mathbb{P}(X_1 = x_1 \mid X_0 = x_0) \cdots \mathbb{P}(X_n = x_n \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\&= \mu_0(x_0) \prod_{k=1}^n P_{x_{k-1}, x_k}\end{aligned}$$

- La loi d'une chaîne de Markov $(X_n, n \in \mathbb{N})$ est entièrement caractérisée par sa matrice de transition, **P**, et la loi de X_0 , μ_0 . De plus, on a, pour tous $n \in \mathbb{N}^*$, $x_0, \dots, x_n \in E$,

La chaine de Markov

➤ Exemples 1 :

$$P = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$



Remarque :

La matrice de transition $P = (p_{i,j})_{(i,j) \in E^2}$ est **stochastique** : la somme des termes de n'importe quelle ligne de **P** donne toujours 1.

$$\forall i \in E, \sum_{j \in E} p_{i,j} = 1.$$

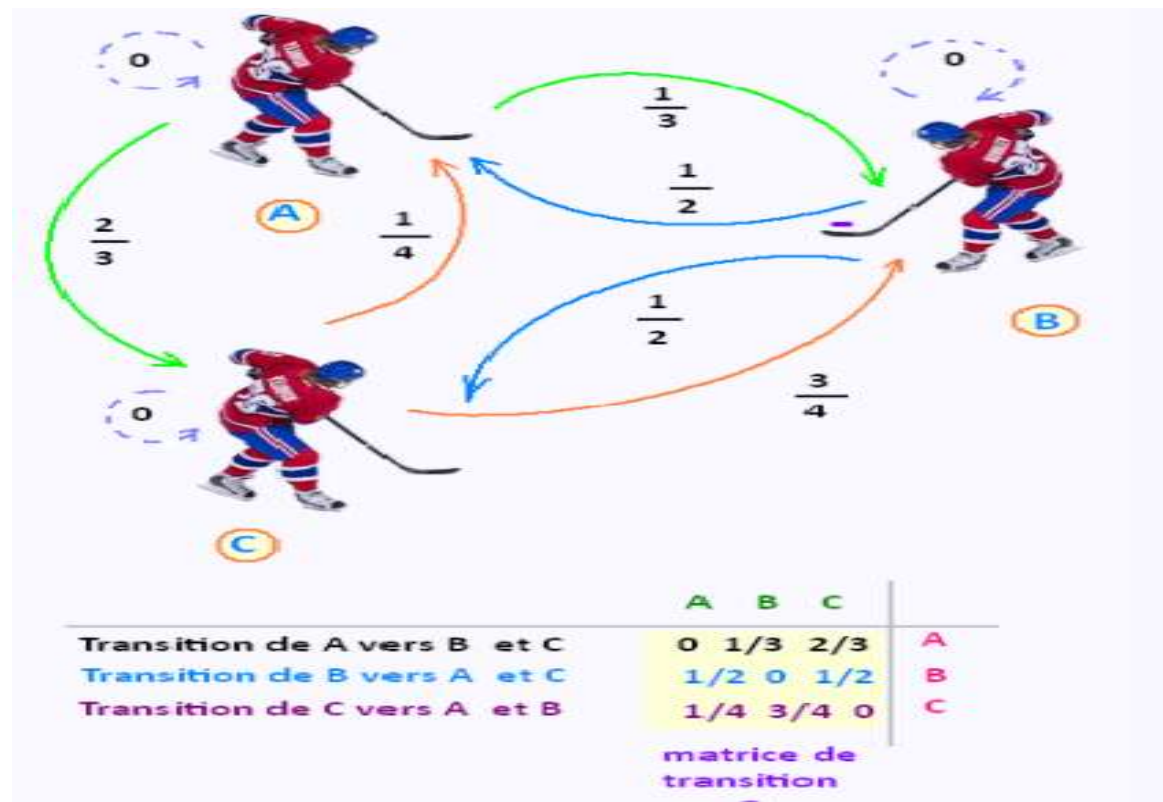
La chaine de Markov

➤ Exemples 2 :

Dans une équipe de Hockey, on étudie les passes de la rondelle que font les trois joueurs A , B et C entre eux.

Les probabilités qu'un joueur passe la rondelle à un autre sont représentées sur le graph suivant ci-contre.

À partir de ce graphe G , on en déduit la matrice de transition :





Le Modèle Markovien Caché (HMM)

□ Définition :

C'est un modèle statistique dans lequel le système modélisé est supposé être **un processus markovien de paramètres inconnus**, on ne peut pas observer directement la séquence d'états : **les états sont cachés**. Chaque état émet des "observations" qui, elles, sont observables. On ne travaille pas donc sur la séquence d'états, mais sur la séquence d'observations générées par les états. **HMM** sont massivement utilisés notamment en **reconnaissance de formes, en intelligence artificielle ou encore en traitement automatique du langage naturel**.

□ Formalisme :

Un **HMM** est un Quintuplets $\{N, M, S, A, B\}$ des ensembles décrits suivant :

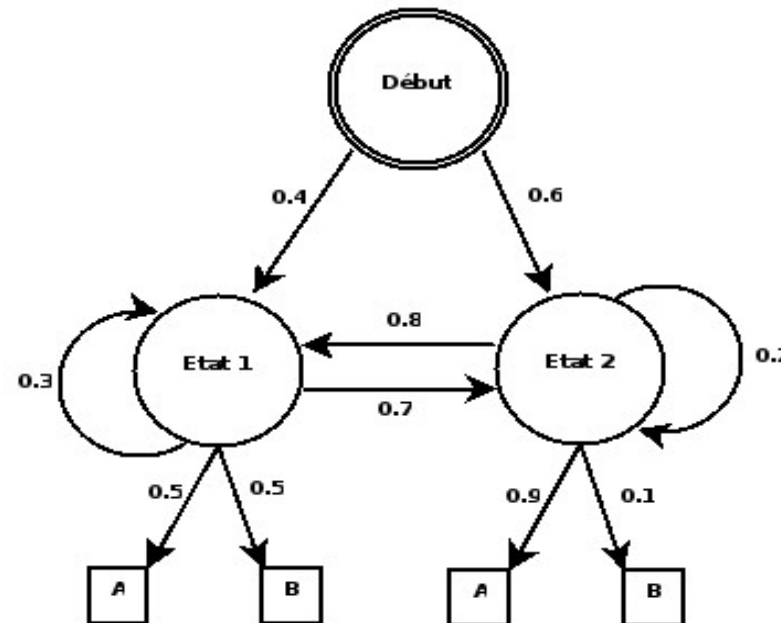
- **N** : le nombre d'états ;
- **M** : le nombre d'observations ;
- **S** : la matrice des probabilités de départ $[1 \times N]$, c'est à dire les probabilités de démarrer dans chacun des N états ;
- **A** : la matrice des probabilités de transition $[N \times N]$, c'est à dire les probabilités de passer d'un état à l'autre ;
- **B** : la matrice des probabilités d'émission $[N \times M]$, c'est à dire les probabilités pour chaque état d'émettre chacune des observations possibles ;

Le Modèle Markovien Caché (HMM)

➤ Exemple :

On retrouve bien le modèle de Markov : l'état "Début", les deux états "Etat 1" et "Etat 2" et les transitions avec leur probabilité associée. On y a ajouté deux **observations : "A" et "B"**.

Chaque état peut émettre chacune des observations avec une certaine probabilité que nous appelons "**probabilité d'émission**". Cette probabilité peut éventuellement être nulle, ce qui signifie que l'état **ne peut pas émettre l'observation concernée**. Dans cet exemple, l'état 1 a 50% de chances (0.5) d'émettre un "A" et 50% de chances d'émettre un "B", tandis que l'état 2 a 90% de chances (0.9) d'émettre un "A" et 10% (0.1) d'émettre un "B". Comme pour les transitions partant d'un état, la somme des probabilités d'émission d'un état doit toujours être égale à 1.



La Méthode de Monte-Carlo

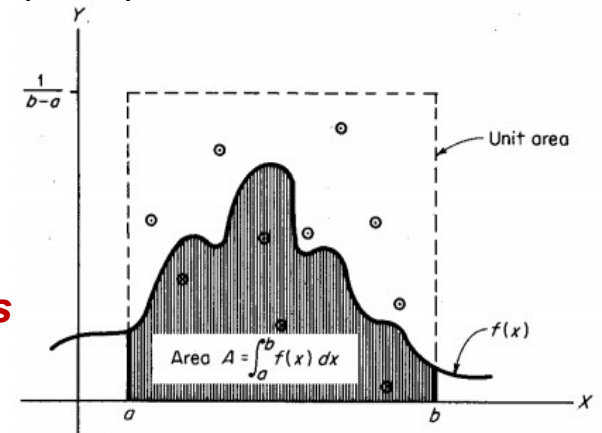
□ Définition :

- ✓ Souvent on trouve dans les modèles des équations qui ne peuvent pas être résolues par des techniques numériques standard. Il peut alors être plus efficace de construire un modèle stochastique analogue. La procédure de calcul est entièrement numérique et effectuée en introduisant des nombres aléatoires dans le système pour obtenir des réponses numériques, pour cela :
- ⇒ On évalue une intégrale multidimensionnelle. Si on veut évaluer l'aire d'une surface définie, on entoure la forme d'un carré normalisé pour qu'un coté soit une unité de longueur.

➤ Exemple :

- ❖ Si on prend un point au hasard dans la forme, la probabilité pour qu'il soit dans l'aire A est A.
- ❖ Si on prend un grand nombre de points au hasard, la proportion de points qui est dans la forme est une estimation de A.

On remarque que $A = \frac{4}{4+6}$ en se basant sur 10 points des nombres aléatoires





La Méthode de Monte-Carlo

□ Description :

- ✓ La méthode de Monte-Carlo désigne toute méthode visant à calculer une valeur numérique en utilisant des procédés aléatoires, c'est à dire des techniques probabilistes. Le nom de cette méthode fait allusion aux jeux de hasard pratiqués au **Casino Monte-Carlo**.
- ⇒ La description de cette méthode se caractérise par l'utilisation **du hasard** pour résoudre des problèmes centrés sur le calcul d'une valeur numérique. La réponse fournie sera une réponse statistique.
- ✓ La méthode de Monte Carlo a été préférée dans de très nombreux secteurs scientifiques et technologiques. La puissance développée des ordinateurs a permis à ces méthodes de devenir opérationnelles et de s'étendre dans des domaines différents tel que :
 - La finance, la biologie moléculaire et génétique.
 - Les mathématiques, la physique,
 - les télécommunications, les réseaux, la recherche opérationnelle.....etc



La Méthode de Monte-Carlo

□ Domaine d'application :

- ✓ D'une façon générale, l'utilisation de cette méthode recouvre tous les domaines où l'utilisation des méthodes scientifiques se heurte à des difficultés. Dans ce contexte, on distingue deux grands domaines où la méthode de Monte Carlo peut être utilisée avec succès :

❖ Problème déterministes:

Ce sont des problèmes de nature déterministe qui font appel aux calculs numériques. On cite comme exemple de ces problèmes :

- *Estimation des surfaces.*
- *Calculs d'intégrales multiples.*
- *Résolution d'équations différentielles.*
- *Résolution de systèmes d'équations algébriques.*
- *Résolution de problèmes d'optimisation combinatoire.*



La Méthode de Monte-Carlo

□ Domaine d'application :

❖ Phénomènes et processus aléatoires :

On cite comme exemple de ces problèmes :

- *Mouvement de particules.*
- *Systèmes stochastiques de gestion ou de production.*
- *Reconnaissance de formes (analyse d'images, de paroles, ...).*
- *Systèmes de commande décrits par des équations différentielles ordinaires ou des équations aux différences*

□ Principe de la méthode Monte Carlo :

- L'une des procédures pour calculer une quantité par la méthode de Monte-Carlo est de mettre tout d'abord sous la forme **d'une espérance**, à l'issue de cette étape, il reste à calculer cette quantité par une espérance $E(X)$ de la variable aléatoire X . Pour ce calcul, il convient de savoir simuler une variable aléatoire selon la loi de X .
- On dispose alors d'une suite $(X_i)_{1 \leq i \leq N}$ de N réalisations de la variable aléatoire X .

On approxime alors $E(X)$ par :
$$E(X) \approx \frac{1}{N} (X_1 + \dots + X_N)$$

La Méthode de Monte-Carlo

□ Calcul d'intégrale par la méthode de Monte-Carlo :

- ✓ Le calcul de l'intégrale $I = \int_{\Omega} f(x)dx$ par la méthode de Monte Carlo se ramène à résoudre l'intégrale suivante $I = \int_{[0,1]} f(x)dx$

Dans ce cas, la méthode de Monte Carlo consiste à écrire cette intégrale sous forme d'espérance de f qui est la généralisation de la moyenne de f sur $[0, 1]$, et avec u une variable aléatoire suivant la loi uniforme sur $[0, 1]$: $I = E(f(u))$

- ✓ On approxime l'intégrale comme suit : $I \approx I_N = \frac{1}{N} \sum_{i=1}^N f(x_i)$
- ✓ Les points x_i sont choisis dans l'intervalle Ω , et quand le nombre des points N augmente tend vers l'infini l'approximation sera plus précise et on aura :

$$I = \int_{\Omega} f(x)dx = \lim_{N \rightarrow \infty} I_N = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(x_i)$$

- ✓ Pour voir les procédés un peu plus claires, nous approchons le calcul de l'intégrale sur un intervalle $[a, b]$ de \mathbb{R} , c'est à dire calculer l'intégrale $I = \int_a^b f(x)dx$.
- ✓ Soient x_1, x_2, \dots, x_N des points aléatoires dans l'intervalle $[a, b]$, après le calcul des $f(x_1), f(x_2), \dots, f(x_N)$, on peut écrire la formulation générale suivante :

$$I = \int_a^b f(x)dx \approx \frac{b-a}{N} \sum_{i=1}^N f(x_i)$$

La Méthode de Monte-Carlo

□ Calcul d'intégrale par la méthode de Monte-Carlo :

⇒ Soient N réels, \mathbf{x}_i tirés aléatoirement sur l'intervalle $[\mathbf{a}, \mathbf{b}]$ avec une densité de probabilité uniforme, égale à : $p(x) = \frac{1}{b-a}$

⇒ A partir de cette expression qu'on a vu : $I = \int_a^b f(x) dx \approx \frac{b-a}{N} \sum_{i=1}^N f(x_i)$

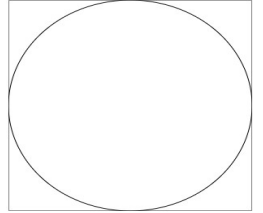
⇒ On introduit une variable aléatoire réelle \mathbf{X} sur l'intervalle $[\mathbf{a}, \mathbf{b}]$ avec une densité de probabilité $\mathbf{p}(\mathbf{x})$. Pour une fonction g , l'espérance de $g(X)$ est :

$$E(g(X)) = \int_a^b p(x) g(x) dx \quad \text{Donc si on pose } f(x)=p(x)g(x) \quad E\left(\frac{f}{p}\right) = \int_a^b f(x) dx$$

⇒ Une approximation de l'intégrale est donc obtenue en évaluant l'espérance de $\mathbf{f/p}$ avec \mathbf{N} échantillons \mathbf{x}_i tirés dans l'intervalle $[\mathbf{a}, \mathbf{b}]$ avec une densité de probabilité \mathbf{p} , et on aura selon **la méthode de Monte-Carlo :**

$$I \approx \frac{b-a}{N} \sum_{i=1}^N f(x_i) \approx \text{moy}\left(\frac{f}{p}, N\right) \approx \frac{1}{N} \sum_{i=1}^N \frac{f(x_i)}{p(x_i)}$$

La Méthode de Monte-Carlo



➤ Exemple 1 « Estimation de π » :

Prenons cet exemple, qui est la surface délimitée par un cercle de rayon 1, et vaut approximativement 3,14159, pour la déterminer par la méthode de Monte-Carlo ?

1. Il faut donc représenter cette valeur comme la moyenne d'une variable aléatoire.
2. Une méthode parmi d'autres peut être de dessiner un cercle de rayon 1 et de l'englober dans le carré de côté 2, donc d'aire 4, qui touche le cercle en 4 points,
3. Dans le carré, on choisit un point aléatoirement et uniformément, ce qui signifie que chaque point a la même chance d'être choisi. Alors la probabilité qu'il se trouve à l'intérieur du cercle est la surface du cercle (π) divisée par la surface du carré (4), et vaut donc ($\pi/4$).
4. Cette probabilité est l'espérance de la variable aléatoire X :

$X=1$ si le point choisi aléatoirement dans le carré est à l'intérieur du cercle

$X=0$ si le point choisi aléatoirement dans le carré mais il est à l'extérieur du cercle

Pour estimer l'espérance de X , \Rightarrow on considère un grand nombre de tirages indépendants uniformes de points dans le carré, qui vont se répartir uniformément dans le carré. **Quand le nombre de points va tendre vers l'infini**, la proportion de points dans le cercle va converger vers $E(X)$, ici $E(X)=\pi/4$. **Le nombre π peut donc être estimé par 4 fois la proportion de points dans le cercle.**

La Méthode de Monte-Carlo

➤ Exemple 2 « La conception par la simulation de Monte Carlo »

Supposons qu'on veut concevoir un nouveau produit :

Ce produit doit être fabriqué selon :

- Les meilleurs niveaux de qualité
- Les meilleurs niveaux de fiabilité pour bénéficier d'une excellente réputation à long terme.

Le nouveau produit comprend de nombreuses améliorations et innovations afin qu'il soit très bien perçu par les clients potentiels.

- ⇒ On doit s'assurer à présent que le produit passera sans heurts de la phase de Recherche et Développement ➡ la phase de production de masse.
- ⇒ Pour pouvoir accroître rapidement les volumes de production,
- ⇒ l'équipe de conception devra fournir les spécifications optimales du produit final qui seront déclinées en spécifications pour les sous-systèmes et composants individuels.



La Méthode de Monte-Carlo

➤ Exemple 2 « La conception par la simulation de Monte Carlo »

Si cette phase de « tolérancement » n'est pas correctement effectuée au préalable, les ingénieurs en production devront recourir à leur propre ingéniosité pour résoudre les incohérences éventuelles et réajuster les

paramètres du produit. Ce n'est pas la meilleure façon de procéder car cela implique souvent de nombreuses altérations ultérieures du produit et des délais plus longs de mise sur le marché.

Malheureusement, tous les procédés sont affectés par des sources diverses de variations (fluctuations 'environnementales', variabilité naturelle des procédés...). Cette variabilité provoque souvent des problèmes majeurs de qualité.

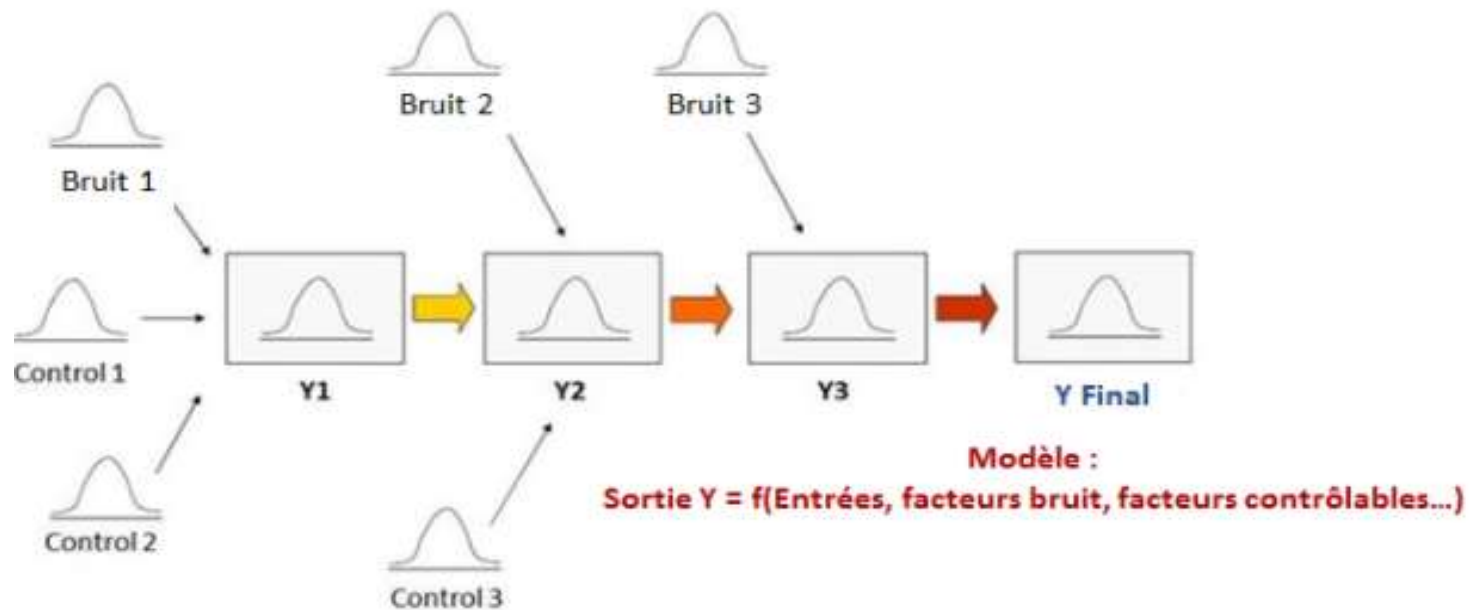


Si les spécifications du produit sont assez larges par rapport à la plage de fluctuation globale du procédé, vous obtiendrez des produits de haute qualité à faible coût (avec une valeur de capabilité élevée). Si ce n'est pas le cas, le pourcentage de produits hors-spécification augmentera rapidement

La Méthode de Monte-Carlo

➤ Exemple 2 « La conception par la simulation de Monte Carlo »

Avec plusieurs variables d'entrée et une seule réponse en sortie. Certaines entrées sont des paramètres contrôlables, d'autres sont des facteurs de bruit (incontrôlables).



A ce stade, seuls quelques prototypes sont probablement disponibles pour valider le concept produit. D'où la nécessité de simuler avec la méthode de Monte Carlo. C'est une technique probabiliste basée sur la génération d'un grand nombre de données aléatoires afin de simuler la variabilité d'un système. L'objectif est de simuler et tester le plus tôt possible afin d'anticiper les problèmes de qualité éventuels, éviter les changements de conception intempestifs et coûteux qui pourraient s'avérer nécessaires ultérieurement et de faciliter les opérations de fabrication.



La Méthode de Monte-Carlo

- Exemple 2 « La conception par la simulation de Monte Carlo »
- Chaque paramètre d'entrée dans le modèle est caractérisé par une moyenne et une variance. Il suffit d'identifier **une loi de distribution appropriée peut exiger une connaissance plus approfondie du comportement de ces entrées**. Pour simplifier ce choix, la distribution triangulaire peut être utilisée, il suffit simplement d'indiquer le minimum, le maximum et la valeur la plus probable du paramètre d'entrée.
- Lorsque vous effectuez **une simulation de Monte Carlo**, si la capacité prévue est insuffisante et si des améliorations supplémentaires sont nécessaires pour atteindre un niveau de qualité acceptable, la variabilité de certains paramètres d'entrée devra être réduite.



Modèles Statiques & Modèles Dynamiques

□ Définition du Modèle :

Un **modèle** est une **équation paramétrée** (ou un ensemble d'équations paramétrées) permettant de **calculer la valeur de la grandeur** (ou des grandeurs) à modéliser à partir des valeurs d'autres grandeurs appelées **variables** ou **facteurs**. On distinguera les **modèles statiques** & **des modèles dynamiques**, et **les modèles linéaires en leurs paramètres** & **des modèles non linéaires en leurs paramètres**.

A. Modèles Statiques :

Définition

Un modèle statique est une fonction paramétrée notée $g(\mathbf{x}, \mathbf{w})$, où \mathbf{x} est le vecteur dont les composantes sont les valeurs des variables, et où \mathbf{w} est le vecteur des paramètres du modèle.

Modèles Statiques & Modèles Dynamiques

A.1 Modèles Statiques linéaires en leurs paramètres :

Le modèle statique linéaire suppose que la relation entre **les chargements** et **la réponse induite** est **linéaire**. *Par exemple, si vous doublez l'intensité des chargements, la réponse (déplacements, déformations, contraintes, forces de réaction, etc.), sera également doublée.* Un modèle statique est linéaire en ses paramètres s'il est une combinaison linéaire de fonctions non paramétrées des variables ; il est de cette forme :

$$g(x, w) = \sum_{i=1}^p w_i f_i(x)$$

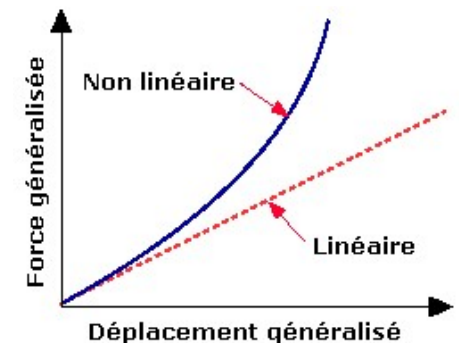
où f_i est une fonction connue. Ce modèle peut encore s'écrire sous la forme d'un produit scalaire :

$$g(x, w) = w \cdot f(x)$$

où $f(x)$ est le vecteur dont les composantes sont les fonctions $f_i(x)$

Un modèle affine est un modèle linéaire qui contient une constante additive w_0 :

$$g(x, w) = w_0 + \sum_{i=1}^p w_i x_i$$



Modèles Statiques & Modèles Dynamiques

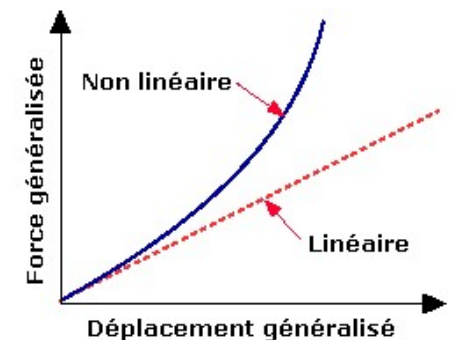
A.2 Modèles Statiques non linéaires en leurs paramètres :

Toutes les structures réelles se comportent **non linéairement** d'une façon ou d'une autre à un niveau donné de chargement. Dans certains cas, les analyses linéaires peuvent être adéquates. Dans beaucoup d'autres cas, la simulation linéaire peut produire des résultats erronés puisque les hypothèses sur lesquelles elle est basée ne sont pas respectées. La non linéarité peut provenir du comportement d'un matériau, de grands déplacements et d'interaction au niveau **de réseau de neurones**. Un modèle statique est non linéaire en ses paramètres s'il est une combinaison non linéaire de fonctions non paramétrées des variables ; il est de cette forme :

$$g(x, w) = \sum_{i=1}^p w_i f_i(x, w')$$

où f_i est une fonction non linéaire, paramétrées par les composantes du vecteur w' . Par exemple :

$$g(x) = x^3 + e^{-\frac{x}{2}} - 5\arcsin(6x)$$





Modèles Statiques & Modèles Dynamiques

B. Modèles Dynamiques :

Les modèles dynamiques, ont une forme de *mémoire* ils dépendent du temps : **la sortie du modèle à un instant donné dépend de ses sorties passées**. En conséquence, elle peut évoluer dans le temps, à partir d'un état initial, même si les variables x sont constantes.

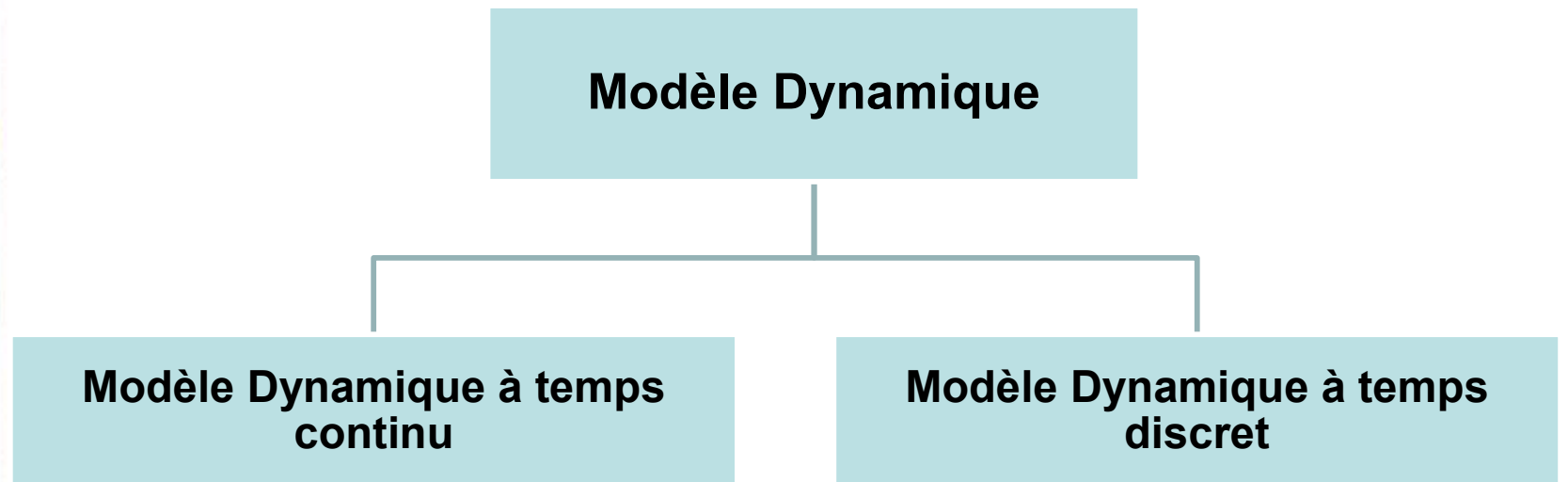
=> La très grande majorité des applications des modèles statistiques sont réalisées à l'aide d'ordinateurs, ou de circuits électroniques numériques. Dans les deux cas, les mesures des variables sont effectuées à intervalles réguliers, dont *la durée est appelée période d'échantillonnage*.

=> De même, les prédictions du modèle ne sont pas fournies de manière continue, mais à intervalles réguliers, généralement caractérisés par la même période d'échantillonnage que les mesures des variables.

=> Ces systèmes sont dits à **temps discret**, par opposition aux systèmes physiques naturels, qui sont des systèmes à **temps continu**.

Modèles Statiques & Modèles Dynamiques

B. Modèles Dynamiques :



Ce sont des équations différentielles (ou des systèmes d'équations différentielles du type :

$$\frac{dy}{dt} = g(y, x, w)$$

où **t** désigne le temps, **y** la prédiction effectuée par le modèle, **x** le vecteur des variables et **w** le **vecteur** des paramètres.

Le temps **t** n'est plus une variable continue :
t=kT

où **T** désigne la période d'échantillonnage et **k** est un nombre entier positif. La prédiction **y** de la valeur prise à modéliser à l'instant **kT**, connaissant les prédictions effectuées aux **n**, **n'** instants précédents, et les valeurs des variable, peut alors être mise sous la forme :

$$y(kT) = g[y((k-1)T), y((k-2)T), \dots, y((k-n)T), x((k-1)T), x((k-2)T), \dots, x((k-n')T), w]$$



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

Dans cette partie on présentera :

- I. Un formalisme général pour la modélisation par apprentissage.
- II. La classification
- III. Le classifieur de Bayes, le biais-variance et ses propriétés.
- IV. Regression Linéaire
- V. Modèle Linéaire de sélection et Régularisation
- VI. L'apprentissage supervisé,
- VII. L'apprentissage non supervisé

Eléments de Théorie D'Apprentissage Statistique / Machine Learning

I. Un formalisme général pour la modélisation par apprentissage :

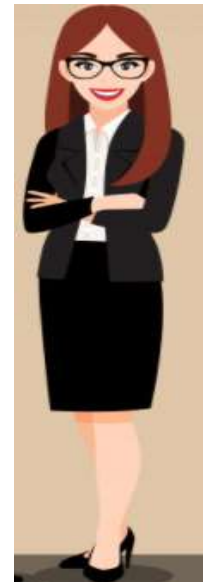
Pour vous motiver à étudier l'apprentissage statistique, supposons que vous êtes des consultants en statistique embauché par un producteur



Afin de fournir des conseils sur la façon d'améliorer les ventes d'un produit particulier.



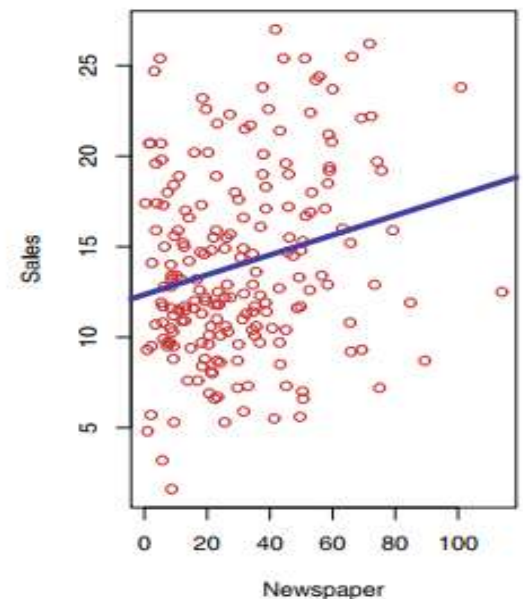
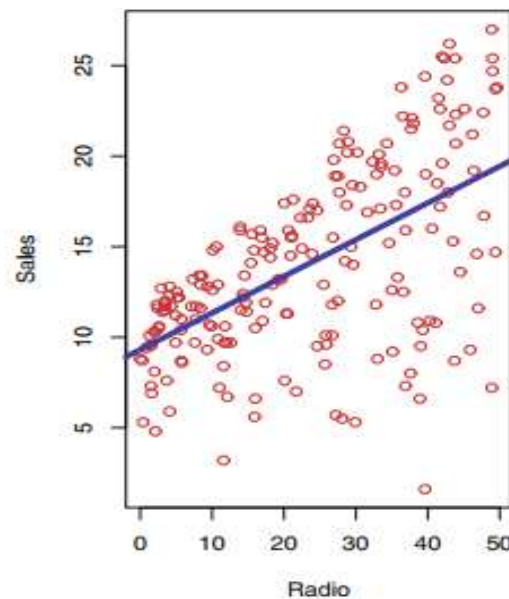
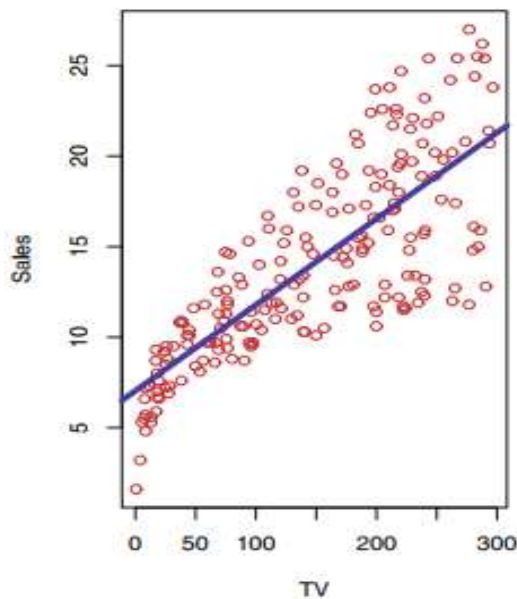
L'ensemble de données publicitaires comprend les ventes de ce produit dans 200 différents marchés, ainsi que les budgets publicitaires du produit dans chacun de ces marchés pour trois médias différents: **la télévision, la radio et les journaux.**



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

I. Un formalisme général pour la modélisation par apprentissage :

Au niveau de ces figures la ligne bleu signifie le simple model qui permet de prédire les ventes à travers **la TV, la radio et les journaux**.



=> Il n'est pas possible pour notre client d'augmenter directement ventes du produit. D'autre part, ils peuvent contrôler la publicité dépenses dans chacun des trois médias.



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

I. Un formalisme général pour la modélisation par apprentissage :

Par conséquent, si on détermine qu'il existe une association **entre la publicité et les ventes**, **alors nous pouvons instruire notre producteur pour ajuster les budgets publicitaires, augmentant ainsi indirectement les ventes**. En d'autres termes, notre objectif est de développer un modèle précis pouvant être utilisé pour prédire les ventes sur la base des trois budgets médias.

Dans ce cas :

➤ **Les budgets publicitaires sont des variables d'entrée** : (X_1 budget de la TV), (X_2 le budget de la radio), et (X_3 le budget du journal) [Les entrées utilisent des fois des noms différents, tels que prédictors, variables ou caractéristiques]***

tandis que

➤ **La vente est une variable de sortie notée par Y** : $Y = f(X) + \epsilon$, où ϵ désigne le terme d'erreur qui est aléatoire et indépendant de X et qui peut avoir une moyenne de 0



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

I. Un formalisme général pour la modélisation par apprentissage :

A. Fonction de perte :

Vu que l'apprentissage cherche à reproduire les données, il faut définir une « distance » entre **les prédictions du modèle** et **les données**.

=> on définit donc une fonction dite « fonction de perte » : $\pi[y^p, g(x, w)] \geq 0$

Où :

- ✓ y^p est la valeur souhaitée
- ✓ $g(x, w)$ est la valeur prédite par le modèle, dont les paramètres sont les composantes du vecteur w , étant donné le vecteur de variables x .

Exemple : Pour une tâche de classification à deux classes :

$y^p = +1$ pour un objet d'une classe

$y^p = -1$ pour un objet de l'autre classe.



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

I. Un formalisme général pour la modélisation par apprentissage :

A. Fonction de perte :

La fonction de perte la plus fréquemment utilisée pour la prédiction est :

$$\pi[y^p, g(x, w)] = [y^p - g(x, w)]^2$$

⇒ **Comment décrire mathématiquement la « qualité » du modèle ?**

On peut modéliser les résultats des mesures y^p comme des réalisations d'une variable aléatoire Y^p , et les vecteurs des variables x comme des réalisations d'un vecteur aléatoire X . Alors les valeurs de la fonction de perte π deviennent elles-mêmes des réalisations d'une variable aléatoire Π , fonction de Y^p et de X ,



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

I. Un formalisme général pour la modélisation par apprentissage :

B. Erreur de prédiction théorique :

⇒ Il est naturel de caractériser la performance du modèle par l'espérance mathématique de Π , $\mathbb{E}(\Pi)$ ou erreur de prédiction théorique, que nous noterons P^2 (cette quantité est toujours positive, d'après la définition de π) :

$$P^2 = \mathbb{E}(\Pi) = \iint \pi(y^p, g(x, w)) p_{Y^p, X} dy^p dx$$

Où $p_{Y^p, X}$ est la probabilité conjointe de la variable aléatoire \mathbf{Y}^p et du vecteur aléatoire \mathbf{X} , notons que, pour les modèles dont les paramètres \mathbf{w} sont déterminés par apprentissage, ces derniers dépendent aussi des réalisations de \mathbf{Y}^p présentées dans l'ensemble d'apprentissage.

L'erreur de prédiction théorique peut alors s'écrire : $P^2 = \mathbb{E}_{\mathbf{X}}[\mathbb{E}_{Y^p|X}(\Pi)]$
où $\mathbb{E}_{Y^p|X}(\Pi)$ désigne l'espérance mathématique de la variable aléatoire $\Pi(Y^p | X)$,
c'est-à-dire l'espérance mathématique de la fonction de perte pour les prédictions faites par le modèle pour chaque vecteur donné de variables \mathbf{X}



Eléments de Théorie D'Apprentissage Statistique / Machine Learning

I. Un formalisme général pour la modélisation par apprentissage :

B. Erreur de prédiction théorique :

Démonstration :

$$P^2 = \iint \pi(y^p, g(x, w)) p_{Y^p}(Y^p | X) p_X dy^p dx$$

$$\text{on sait que: } \mathbb{E}_{Y^p|X}(\Pi) = \left[\int \pi(y^p, g(x, w)) p_{Y^p}(Y^p | X) dy^p \right]$$

car par définition, on a: " $\mathbb{E}_{X|Y}(x) = \mathbb{E}(X|Y)(x) = \int x \cdot p_Y(X|Y) dy$ "

$$P^2 = \int \left[\int \pi(y^p, g(x, w)) p_{Y^p}(Y^p | X) dy^p \right] p_X dx$$

$$P^2 = \int \mathbb{E}_{Y^p|X}(\Pi) p_X dx$$

$$P^2 = \mathbb{E}_X[\mathbb{E}_{Y^p|X}(\Pi)]$$



**Le meilleur modèle est celui
qui a la plus petite erreur de
prédiction théorique.**