THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ JOSEPH FOURIER

Discipline : Mathématiques Appliquées

Présentée et soutenue publiquement

par

Franck CORSET

Le 22 janvier 2003

AIDE À L'OPTIMISATION DE MAINTENANCE À PARTIR DE RÉSEAUX BAYÉSIENS ET FIABILITÉ DANS UN CONTEXTE DOUBLEMENT CENSURÉ

Jury de thèse

Directeur de thèse : Gilles Celeux

M. Serge Dégerine	, Président
M. Antoine de Falguerolles	, Rapporteur
M. Bernard Dumon	, Rapporteur
M. André Lannoy	, Examinateur
M. Olivier Gaudoin	, Examinateur

Préambule

Durant les trois années de thèse, mon travail a porté sur la sûreté de fonctionnement, et plus particulièrement sur la fiabilité et l'aide à l'optimisation de maintenance. Ces trois domaines sont généralement très liés. Ce document est cependant divisé en deux parties distinctes, l'une comporte l'application des réseaux bayésiens à la maintenance et l'autre traite de la fiabilité dans un contexte doublement censuré. La raison de cette distinction vient de deux contrats avec la division Recherche et Développement d'Électricité De France. La première partie concerne le contrat ayant motivé cette thèse. La deuxième partie a été initiée lors de mon stage de D.E.A., qui concernait la fiabilité d'un composant dans un contexte doublement censuré, et s'est poursuivie pendant la thèse avec notamment l'encadrement d'un stage de D.E.A., concernant une modélisation d'un comportement de maintenance dans le même contexte.

Table des matières

Liste des tableaux

Table des figures

Première partie

Aide à l'optimisation de maintenance à partir de réseaux bayésiens

Introduction

Durant ces dernières années, un enjeu majeur du monde industriel est sans aucun doute la sûreté de fonctionnement. En effet, de nombreux accidents tragiques ont eu lieu causant pour la plupart des centaines de victimes civiles. Que ces accidents aient pour cause un mauvais fonctionnement, une erreur humaine, ou une cause externe (intempéries, attentats, etc), les industriels se doivent de prendre en compte tous ces risques afin de minimiser leurs conséquences. Cette démarche qui peut s'avérer extrêmement coûteuse pour l'industriel, reste indispensable vis-à-vis d'une opinion publique de plus en plus critique. Ainsi, de nombreuses études récentes tant académiques qu'industrielles ont eu pour objet la fiabilité, l'optimisation de la politique de maintenance et plus généralement la sûreté de fonctionnement. La première grande étape dans toutes ces études est la capitalisation des connaissances, que celles-ci proviennent d'experts ou de données de retour d'expérience (REX). Concernant le jugement d'expert dans le domaine de la sûreté de fonctionnement on peut se référer au livre récent de Lannoy et Procaccia [64]. Des modèles mathématiques permettent de décrire la politique optimale de maintenance à adopter. Généralement, cette politique est basée sur la fiabilité du système étudié. D'un point de vue plus général, ces approches consistent à établir une fonction d'utilité, prenant en compte les coûts de réparation, l'indisponibilité, la fiabilité, ou la sûreté de fonctionnement (faisant intervenir les coûts d'accidents), puis tente de maximiser cette fonction. On peut citer de façon non exhaustive par exemple les travaux de Pellegrin [80] et Koshimae et al. [63] sur le temps de délai (temps séparant une dégradation de la défaillance), Wetsberg and Klefjö [97], Aven [6], qui travaillent sur le taux de défaillance cumulé (taux de hasard cumulé), Bergman and Klefsjö [14] sur une transformée du Temps Total en Test (TTT), Barlow [10] pour les propriétés géométriques du Temps Total en Test, Berg [13] pour des travaux sur le coût marginal, où l'on décide de réparer lorsque les coûts de réparation dépassent un certain seuil, qui dépend lui-même de l'âge du composant. On peut également citer l'approche consistant à modéliser la dégradation d'un système par un processus de Markov (cf. Cocozza-Thivent [32], [33]). L'intérêt majeur de ces modélisations mathématiques, même si parfois les hypothèses sont lourdes et peu réalistes, est de permettre des calculs explicites dans de nombreux cas, comme par l'exemple des systèmes à l'architecture complexe (systèmes k sur n).
On peut également considérer l'optimisation de la maintenance en effectuant une analyse de données, visant à permettre une aide au diagnostic et une aide à la décision. Pour cela, on peut imaginer par exemple, une analyse de données en temps réel suite à l'observation de phénomènes, comme cela est couramment pratiqué en médecine. Enfin, le management

des risques doit également faire partie du processus de l'optimisation de la maintenance et de sûreté de fonctionnement. Dans ce travail de thèse, nous nous sommes attachés à la capitalisation des connaissances, avec notamment l'usage des réseaux bayésiens, qui nous ont également permis de proposer des outils d'aide au diagnostic et d'aide à la décision.

Objectifs de cette partie

Dans cette partie, nous nous intéressons à un système mécanique, jugé critique par l'industriel. L'objectif premier de ce travail est de modéliser le processus de dégradation de ce système, en introduisant de nouveaux modèles, encore peu connus et très peu utilisés en fiabilité et en optimisation de la politique de maintenance. Ces modèles sont les modèles graphiques introduits par Pearl [78], Lauritzen et Spiegelhalter [66]. Nous nous intéressons plus particulièrement à des modèles graphiques avec des arêtes orentiées sans cycle, noté DAGs (Directed Acyclic Graphs), plus communément appelés réseaux bayésiens. Cette appellation vient du fait qu'à partir de systèmes experts probabilistes, il est possible d'avoir une interprétation bayésienne sur ces modèles graphiques orientés et acycliques. Dans cette partie, notre intérêt porte principalement sur la construction et l'utilisation de réseaux bayésiens, comme outil d'aide à la décision en maintenance. Ainsi, les multiples algorithmes d'inférence sur les réseaux bayésiens sont ici utilisés mais ils ne sont pas l'objet de recherche. Notre objectif est donc d'utiliser cette théorie, bien sûr en maîtrisant les algorithmes, mais en mettant l'accent sur la méthodologie. On peut par exemple citer une phrase du livre de Cowell et al. [38] : *... no matter how sophisticated the inference procedures are for manipulating the knowledge in a knowledge base, if the content of the knowledge base is poor then the inferences will be correspondingly poor.* Certes, la recherche d'algorithmes de plus en plus performants est un enjeu majeur, à l'heure où les tailles des bases de données deviennent très grandes, mais la complexité de ces modèles nécessite de proposer une méthodologie simple visant à construire de tels modèles uniquement à partir d'avis d'experts.
L'objectif de cette partie est donc de présenter une méthodologie complète, afin de construire un réseau bayésien à partir de jugements d'experts, puis de proposer des outils d'analyse de ces modèles, et enfin de proposer des méthodes visant à enrichir le réseau bayésien (avec par exemple des données de retour d'expérience). Le but idéal de cette partie serait qu'elle permette de reproduire les mêmes schémas pour d'autres composants, et pas nécessairement dans le domaine nucléaire.

Plan de la première partie

Le **chapitre 1** vise à introduire les modèles graphiques et plus particulièrement les réseaux bayésiens. Ces modèles sont une représentation à la fois qualitative et quantitative des relations entre les variables du modèle. En effet, la structure du réseau traduit des indépendances conditionnelles entre les variables, tandis que les probabilités conditionnelles permettent de les quantifier. Nous présentons également, à travers un exemple, le célèbre algorithme d'arbre de jonction pour effectuer l'inférence statistique. Cet algorithme, bien que présentant quelques défauts, sera utilisé dans toute notre étude.

Le **chapitre 2** traite de la construction d'un réseau bayésien à partir uniquement d'avis d'experts. Après un rappel sur les enjeux de l'optimisation de la maintenance dans le domaine nucléaire, un protocole sur la construction de la structure du graphe est présenté. Puis, au vu de la complexité du réseau bayésien, et donc au vu du nombre de probabilités à renseigner par les experts, une première approximation par des modèles log-linéaires est faite. De plus, notre stratégie d'interrogation des experts prévoit de demander les probabilités les plus faciles à évaluer, et notamment de demander toutes les probabilités marginales. Ceci permet de vérifier la cohérence des avis d'experts et donc de prévoir un retour aux experts afin que ceux-ci puissent être plus précis et plus complets (par exemple en donnant plus de probabilités conditionnelles).

Le **chapitre 3** est consacré aux multiples possibilités des réseaux bayésiens. En premier lieu, nous effectuons une inférence statistique dans le réseau bayésien, ce qui consiste à calculer des probabilités marginales des variables du modèle, à rechercher les configurations les plus probables, à simuler des scénarios possibles et à mettre à jour des probabilités lorsque l'on observe des variables. Nous exposons également les méthodes d'analyse de sensibilité, ainsi que les outils d'aide à la décision. Puis, nous proposons plusieurs méthodes pour l'analyse des réseaux bayésiens. Enfin, nous proposons d'intégrer les actions de maintenance comme variables du modèle afin que toutes les méthodes que l'on vient de citer puissent être appliquées.

La motivation du **chapitre 4** est d'intégrer le retour d'expérience dans un réseau bayésien. En effet, notre réseau a été construit uniquement à partir d'avis d'experts. Or les données de retour d'expérience sont susceptibles de venir enrichir le réseau avec les années d'exploitation du système. Ainsi, nous proposons d'effectuer cette intégration du REX dans les réseaux bayésiens par une inférence bayésienne non paramétrique. Enfin, nous établissons une méthode simple pour paramétrer la confiance que l'on peut attribuer dans les avis d'experts.

Chapitre 1

Les réseaux bayésiens

Ce chapitre a pour objectif de présenter les réseaux bayésiens, leurs propriétés, et quelques algorithmes permettant de faire de l'inférence, ou de propager une observation, appelée ici évidence. Les réseaux bayésiens sont des cas particuliers de modèles graphiques. Les modèles graphiques ont fait l'objet de nombreuses études lors de ces dernières années et sont présents dans de nombreux articles. On peut citer par exemple Lauritzen et Spiegelhalter [66], Lauritzen [67], Cowell et al. [38], Jordan [60], Pearl [79], Whittaker [98], et dans le cadre de la maintenance Chatelain et Lannoy [30], Yildiz et al. [101]. Un ouvrage en français dédié aux réseaux bayésiens est le livre de Becker et Naïm [12]. Ces modèles se basent sur la théorie des graphes et la théorie des probabilités. En effet, la représentation des réseaux bayésiens est un graphe orienté acyclique où les variables sont les sommets du graphe et les relations entre ces variables sont les arcs du graphe. Dans la communauté de la théorie des réseaux bayésiens, le terme « sommet » est communément remplacé par le terme « nœud ». Dans tout le document, nous utiliserons le terme « nœuds » du graphe pour désigner les sommets du graphes. Pour notre étude, le réseau modélise l'évolution de l'état d'un composant en fonction de facteurs d'influence, où les arcs du graphe représentent la probable causalité entre deux variables. Cette probable causalité va être quantifiée par une probabilité conditionnelle. L'inférence bayésienne s'effectue facilement et avec un faible coût de calcul grâce à la théorie des graphes.

Le chapitre est organisé comme suit. Dans un premier temps, nous définissons les réseaux bayésiens et rappelons des notions de théorie des graphes. Puis, nous exposons la traduction d'indépendance conditionnelle en termes de séparation dans un graphe, et notamment les propriétés de Markov. Enfin, nous présentons l'algorithme d'inférence, l'arbre de jonction, à travers l'exemple fictif célèbre nommé ASIA (cf. [66]).

1.1 Définitions et notations

Nous rappelons dans un premier temps, les notions de théorie des graphes et notamment le vocabulaire employé généralement pour des graphes non orientés (voir le cours de D.E.A.

de G. Letac [70]).

1.1.1 Les graphes

Définition 1 *On appelle graphe non orienté la paire $G = (V, E)$, où $V = X_1, \ldots, X_n$ représente les sommets du graphe, et E représente la famille des sous ensembles de taille 2 de V. Les éléments de E sont appelés arêtes du graphe.*

Avec cette définition, les arêtes ne sont pas orientés, entre deux sommets il ne peut y avoir au plus qu'une arête. Enfin, il n'y a pas d'arête qui va d'un sommet à lui même (pas de boucle). Si (X_i, X_j) appartient à l'ensemble des arêtes E, alors on représentera celle ci par une ligne entre les deux sommets, notée $X_i - X_j$.

Dans notre étude, nous nous intéressons à des graphes orientés sans cycle. Nous donnons la définition des graphes orientés ci-après.

Définition 2 *On appelle graphe orienté la paire $G = (V, E)$, où $V = X_1, \ldots, X_n$ représente les sommets du graphe, et $E = (e_1, \ldots, e_m)$ représente une partie du produit cartésien $V \times V$, dont les éléments sont appelés arcs du graphe.*

Si (X_i, X_j) appartient à l'ensemble E, alors on appellera cet élément, arc du graphe, noté $X_i \longrightarrow X_j$ et représenté par une flèche partant de X_i et pointant sur X_j, où X_i est appelé l'origine et X_j l'extrémité. Dans notre cadre de travail, comme nous le verrons plus tard, les nœuds du graphe orienté représentent les variables statistiques du modèle. Ces variables peuvent être continues ou discrètes. Pour notre étude, les variables du modèles sont toutes discrètes (cf. la notion de graphes purs dans Lauritzen [67]). Dans toute la suite du document, nous considérons des graphes simples, sans boucles ni arcs (ou arêtes pour les graphes non orientés) multiples entre deux mêmes nœuds.

On distingue trois types de graphes : les graphes orientés, les graphes non orientés, et les graphes mixtes où des arêtes et des arcs coexistent. Dans ce document, seuls les graphes orientés sont étudiés. Cependant, les graphes non orientés sont à la base de plusieurs algorithmes de calculs de probabilités, dont nous donnons la description à la fin de ce chapitre.

Pour les graphes orientés, il existe les notions de parents et d'enfants :

Définition 3 *S'il existe un arc partant de X_i vers X_j, alors X_i s'appelle le parent de X_j et X_j l'enfant de X_i. On note $pa(X_j)$ la suite de parents de X_j et $ch(X_i)$ la suite des enfants de X_i.*

Dans les graphes non orientés, la notion de parents et d'enfants n'existe plus, mais est remplacée par la notion de voisins. Deux nœuds sont dits voisins s'il existe une arête entre les deux nœuds. Dans les graphes orientés, il existe de plus des chemins orientés définis comme suit :

Définition 4 *Un chemin orienté est une suite de nœuds distincts $X_i, \ldots, X_j$ tel que (X_{k-1}, X_k) soit un arc pour tout $k = i+1, \ldots, j$. Ce chemin sera noté $X_i \longmapsto X_j$. On appelle cycle un chemin du graphe tel que $X_i = X_j$.*

Les graphes orientés ne possédant pas de cycles sont appelés DAGs (Directed Acyclic Graphs). Tout au long de ce manuscrit, les graphes étudiés, représentant les réseaux bayésiens, sont des graphes orientés sans cycles, notés DAGs.

Définition 5 *On appelle les ancêtres de X_j, noté $an(X_j)$, tous les nœuds X_i tels qu'il existe un chemin $X_i \longmapsto X_j$. De même, on appelle les descendants de X_i, notés $de(X_i)$, tous les nœuds X_j tel qu'il existe un chemin $X_i \longmapsto X_j$. Les non descendants sont définis comme $nd(X_i) = V \setminus (de(X_i) \cup X_i)$. Enfin deux nœuds sont dits voisins s'il existe une arête entre ces deux nœuds.*

La frontière d'un nœud X_i, noté $bd(X_i)$, est l'ensemble des parents et des voisins de X_i. Cette définition n'a de sens que dans le cas de graphes mixtes, car dans un graphe orienté, la frontière se limite aux parents, et dans le cas de graphes non orientés, elle se limite à l'ensemble des voisins. La fermeture de X_i, notée $cl(X_i) = X_i \cup bd(X_i)$ est constituée des voisins de X_i et de X_i lui-même. Enfin une suite de nœuds S sépare A et B si tous les chemins de A vers B (ou dans les deux sens pour le cas de graphes non orientés) croisent S.

1.1.2 Les réseaux bayésiens

Pour introduire et justifier le terme de réseau bayésien, nous rappelons le théorème de Bayes. Soient deux événements A et B, on a :

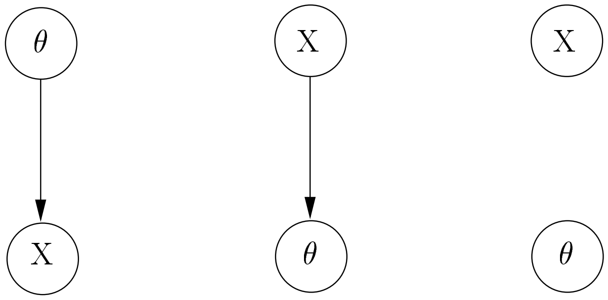
$$P(A|B) = \frac{P(B|A)p(A)}{p(B)}.$$

Dans la théorie bayésienne, la probabilité a priori est donnée par $p(A)$, les observations sont représentées par B, et la loi a posteriori est déduite du théorème de Bayes et donnée par $P(A|B)$. On peut bien entendu généraliser ce théorème non plus pour des événements mais pour des variables aléatoires. Si X est une variable aléatoire représentant le phénomène observé, et θ une variable aléatoire décrivant la connaissance a priori sur ce phénomène, la distribution a posteriori sachant les données $p(\theta|X)$ est proportionnelle au produit de la vraisemblance $P(X|\theta)$ et de la distribution a priori $P(\theta)$:

$$A\ Posteriori \propto Vraisemblance \times A\ Priori.$$

On peut représenter la probabilité jointe $P(X, \theta) = P(X|\theta)p(\theta) = p(\theta|X)p(X)$ par trois réseaux bayésiens simples donnés dans la figure 1.1.

Dans cette figure 1.1, le premier graphe (a) décompose la probabilité jointe en la probabilité a priori et la vraisemblance. Généralement, c'est le point de vue classique, où le paramètre est une cause du phénomène X, et l'arête représente ici une probable causalité.



(a) $p(X|\theta)p(\theta)$ (b) $p(\theta|X)p(X)$ (c) $p(\theta)p(X)$

FIG. 1.1 – Exemple de graphes pour une inférence bayésienne.

Dans le second schéma (b), moins commun, la probabilité jointe est décomposée en la probabilité a posteriori et en la probabilité prédictive. Enfin, le dernier schéma (c) est un exemple d'indépendance entre X et θ. Ces graphes ont ensuite été généralisés, notamment par Pearl [78], en gardant le terme de réseaux bayésiens, dont nous donnons la définition ci-après (cf Jensen [59]).

Définition 6 *Un réseau bayésien est défini par :*
- *Une suite de variables, notée V et une suite d'arcs entre les variables, notée E,*
- *Chaque variable possède un nombre fini d'états exclusifs,*
- *Les variables et les arcs forment un graphe orienté acyclique, noté $G = (V, E)$,*
- *À chaque variable Y avec ses parents $X_1, \ldots, X_n$, est associée une probabilité conditionnelle $P(Y|X_1, \ldots, X_n)$. Lorsque la variable ne possède pas de parent, la dernière quantité devient une probabilité marginale $P(Y)$.*

Remarque : L'utilisation du terme de probabilité a priori pour $P(Y)$ peut parfois être une source de confusion. En effet, dans un réseau bayésien, ces probabilités sont associées aux variables d'entrée sur lesquelles les experts ont des connaissances. Ces connaissances sont alors traduites par des lois a priori. Comme nous le verrons dans le chapitre suivant, notre étude porte sur un réseau bayésien construit uniquement avec des avis d'experts. Ainsi, toutes les probabilités associées au réseau bayésien, que celles-ci soient conditionnelles ou non, sont d'un point de vue bayésien des probabilités a priori. On remarquera que la définition des réseaux bayésiens ne réfère en aucune façon à des causalités entre variables. En revanche, nous allons voir dans le paragraphe suivant que la structure du graphe induit des dépendances conditionnelles entre les variables.

1.2 Propriétés des réseaux bayésiens

Dans ce paragraphe, nous présentons les propriétés des réseaux bayésiens, notamment la représentation de l'indépendance conditionnelle. Des articles pionniers sur l'indépendance conditionnelle sont, à notre avis, l'article de Dawid [41], et ceux de Geiger et al. [46, 47].

1.2.1 Indépendance conditionnelle

Définition 7 *Soit trois variables aléatoires discrètes, A, avec les modalités $(a_i)_{i \in I}$, B avec les modalités $(b_j)_{j \in J}$, et C avec les modalités $(c_k)_{k \in K}$. On dit que A et C sont conditionnellement indépendantes sachant B si pour tout i, j, k*

$$P(A = a_i, C = c_k | B = b_j) = P(A = a_i | B = b_j) P(C = c_k | B = b_j). \tag{1.1}$$

On notera cette indépendance conditionnelle par $A \perp C | B$.

L'équation (1.1) peut s'écrire dans le cas où $P(C = c_k | B = b_j) > 0$:

$$P(A = a_i | B = b_j, C = c_k) = P(A = a_i | B = b_j), \tag{1.2}$$

ce qui se traduit par le fait que sachant B, l'état de C n'influence en aucun cas l'état de A.

Ces propriétés d'indépendance conditionnelle peuvent être représentées par un graphe orienté d'indépendance, défini par Whittaker [98]. Dans un premier temps, il est nécessaire d'introduire un ordre dans les variables. Cet ordre se nomme *ordre topologique*, noté $\prec$, et stipule que pour chaque nœud X, tous ses parents $pa(X)$ le précédent dans l'ordre topologique. On montre facilement qu'un tel ordre existe si le graphe orienté est acyclique. On note $\mathcal{O} = 1, \ldots, n$ un ordre topologique (qui n'est pas unique) sur le graphe avec $1 \prec 2 \ldots \prec n$ et $o(i) = 1, \ldots, i$. Cet ordre peut par exemple être comparé avec le temps d'un processus de Markov, où $o(i)$ représente le passé et le présent de X_i. Dans un DAG, l'existence de l'ordre topologique signifie qu'il existe une direction dans laquelle le processus se dirige et qu'il est impossible de revenir en arrière car le graphe est acyclique.

On définit le graphe orienté d'indépendance par le graphe orienté $G^{\prec} = (K, E^{\prec})$ tel que (i, j) avec $i \prec j$ n'est pas un arc du graphe ($(i, j) \notin E^{\prec}$) si et seulement si $j \perp i | K(j) \setminus \{i, j\}$, avec $K(j) = (1, \ldots, j)$. Ce graphe va représenter les indépendances conditionnelles et grâce à sa définition possède de nombreuses propriétés, comme celle de Markov direct.

1.2.2 Propriétés de Markov

Le graphe orienté d'indépendance tel qu'il a été défini dans le paragraphe précédent possède la propriété de Markov direct, c'est-à-dire que tous les nœuds sont conditionnellement indépendants à leurs non descendants sachant leurs parents. Cette propriété permet d'écrire la probabilité jointe sous une forme récursive appelée *factorisation récursive* :

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | pa(X_i)), \tag{1.3}$$

où $pa(X_i) = \emptyset$ quand X_i est un nœud d'entrée et donc la probabilité correspondante n'est pas conditionnelle, mais marginale (ou a priori).

Avec cette propriété de Markov, on peut déduire une propriété très utile dans les réseaux bayésiens.

Proposition 1 *Soit X et Y deux variables d'un réseau bayésien G, si*

$$(X \cup an(X)) \cap (Y \cup an(Y)) = \emptyset$$

alors X et Y sont indépendants. Autrement dit, si X et Y n'ont pas d'ancêtres communs, alors ils sont indépendants.

Preuve :
La condition de la proposition induit que Y est non descendant de X. Ainsi Y et X sont conditionnellement indépendants sachant les parents de X. Soit $R(X)$, la suite des nœuds racines appartenant aux ancêtres de X, si $R(X) = \emptyset$ alors X est un nœud racine (car le graphe est orienté et acyclique) et la proposition est vérifiée (cf. les propriétés de Markov dans Lauritzen [67]). Si un parent de X, noté Z, appartient à $R(X)$, alors d'après la propriété de Markov, X et Y sont conditionnellement indépendants sachant Z et Z est indépendant de Y (reprendre le premier point de la preuve) et donc X et Y sont indépendants. Enfin si aucun des parents de X n'appartient à $R(X)$, alors les parents de X séparent (X, Y) et $R(X)$. De plus, comme le graphe moral (graphe représentant les indépendances, défini plus loin) engendré par $X \cup Y \cup an(X)$ est non connexe, i.e. se compose de deux graphes, le graphe engendré par X et celui engendré par Y, Y et $R(X)$ sont donc indépendants. On peut donc écrire :

$$\begin{aligned}
p(X, Y, R(X)|pa(X)) &= p(X, Y|pa(X))p(R(X)|pa(X)) \\
&= p(X|pa(X))p(Y|pa(X))p(R(X)|pa(X)) \\
&= p(X|pa(X))p(Y)p(R(X)|pa(X))
\end{aligned}$$

en multipliant par $p(pa(X))$ et en sommant sur $R(X)$ puis sur $pa(X)$

$$\begin{aligned}
p(X, Y) &= \sum_{pa(X)} \sum_{R(X)} p(X, Y, R(X), pa(X)) \\
&= \sum_{pa(X)} \sum_{R(X)} p(X|pa(X))p(Y)p(R(X), pa(X)) \\
&= \sum_{pa(X)} p(X|pa(X))p(Y)p(pa(X)) \\
&= p(X)p(Y)
\end{aligned}$$

et donc la proposition est vérifiée.

Nous avons choisi d'expliciter cette proposition et sa preuve, car bien que connues, nous ne l'avons jamais vu citer d'une manière simple.
L'exemple de Whittaker (cf. [98] page 73) avec 7 variables ordonnées suivant l'ordre topologique est donné dans la figure 1.2.

Dans ce graphe, on a les relations d'indépendances suivantes :

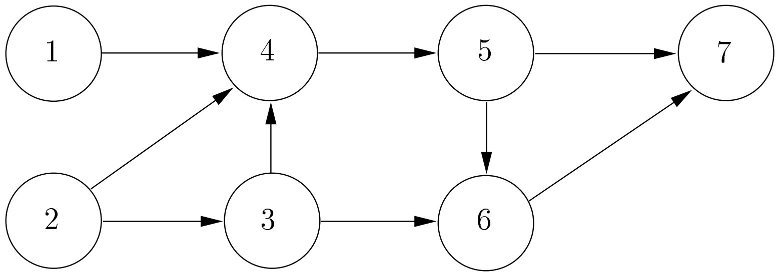


Fig. 1.2 – Exemple de graphe orienté d'indépendance.

- $1 \perp 2$,
- $3 \perp 1|2$,
- $5 \perp 1, 2, 3|4$,
- $6 \perp 1, 2, 4|3, 5$,
- $7 \perp 1, 2, 3, 4|5, 6$.

On peut remarquer que 1 et 3 n'ont pas d'ancêtres en commun et sont donc indépendants. En effet, les deux premières indépendances listées ci-dessus nous donnent :

$$p(3, 1|2) = p(3|2)p(1|2) = p(3|2)p(1),$$

soit en multipliant par $p(2)$ puis en sommant selon 2, on trouve $p(3, 1) = p(3)p(1)$.

Grâce à l'ordre topologique et les indépendances conditionnelles, on peut écrire la probabilité jointe comme

$$\begin{aligned}
P(1, 2, \ldots, 7) &= P(1)P(2, 3, \ldots, 7|1) \\
&= P(1)P(2|1)P(3, 4 \ldots, 7|1, 2) \\
&= P(1)P(2|1)P(3|1, 2)P(4, 5, 6, 7|1, 2, 3) \\
&\vdots \\
&= P(1)P(2|1)P(3|1, 2)P(4|1, 2, 3)P(5|1, 2, 3, 4)P(6|1, 2, 3, 4, 5)P(7|1, 2, 3, 4, 5, 6) \\
&= P(1)P(2)P(3|2)P(4|1, 2, 3)P(5|4)P(6|3, 5)P(7|5, 6).
\end{aligned}$$

L'écriture de la loi jointe sous sa factorisation récursive va permettre d'utiliser des propriétés de la théorie des graphes pour faire de l'inférence sur ce réseau bayésien, notamment pour calculer les probabilités marginales. Pour cela, il est nécessaire de construire le graphe moral, noté $G^m = (K, E^m)$, (nom donné par Lauritzen et Spiegelhalter [66] car cette procédure marie les parents ayant un enfant) qui consiste à unir les parents d'un nœud et à abandonner les directions des arcs. En effet, l'écriture de la loi jointe sous sa forme récursive fait intervenir des probabilités de variables et de leurs parents. Ainsi, le graphe moral permet de former des cliques comprenant exactement ces nœuds. Des potentiels sur les cliques sont ensuite définis à partir des probabilités initiales. Un autre objectif du graphe moral est de déterminer les indépendances conditionnelles plus facilement grâce au théorème de séparation dans les graphes. Soit A, B et S, trois ensembles non vide de nœuds, on dit que S sépare A de B si et seulement si pour tout nœud de A, tout nœud de B et tout chemin entre ces deux nœuds, ce chemin passe par un nœud de S. La propriété de Markov dans le graphe moral est alors la suivante : si une sous-suite de nœuds S sépare A et B alors A et B sont conditionnellement indépendants sachant S, ce que l'on note $A \perp B | S$. Par exemple, le graphe moral entier correspondant au graphe orienté d'indépendance de la figure 1.2 est présenté dans la figure 1.3. Cependant la réciproque n'est pas vraie et donc le graphe moral entier

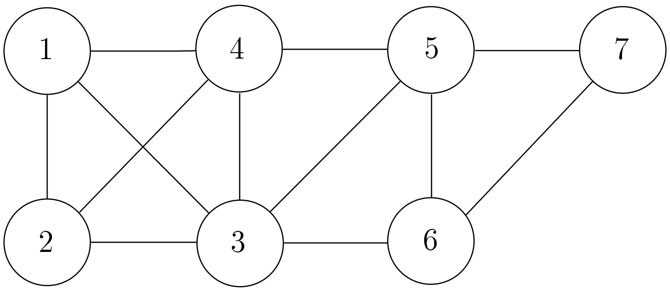


FIG. 1.3 – Le graphe moral entier correspondant au DAG de la figure 1.2.

peut cacher certaines indépendances conditionnelles. En effet dans ce graphe, connaissant 4, 5 et 3 n'apparaissent pas conditionnellement indépendants car non séparés dans le graphe moral. Pour une équivalence, il faut se restreindre au graphe moral généré par les ancêtres de l'union $A \cup B \cup S$. Cette propriété appelée *propriété de Markov global direct* est donnée dans le théorème suivant.

Théorème 1 *Si la probabilité jointe admet une factorisation récursive selon G alors*

$$A \perp B | S$$

quand A et B sont séparés par S dans $(\mathcal{G}_{an(A \cup B \cup S)})^m$*, le graphe moral de la plus petite suite des ancêtres contenant* $A \cup B \cup S$*.*

On peut illustrer ce théorème par le célèbre exemple fictif, nommé *ASIA* par Lauritzen et Spiegelhalter [66], et représenté dans la figure 1.4.

Les huit variables de ce modèle sont :
- A qui représente une récente visite en Asie,
- T la tuberculose,
- L le cancer des poumons,
- S le fait de fumer,

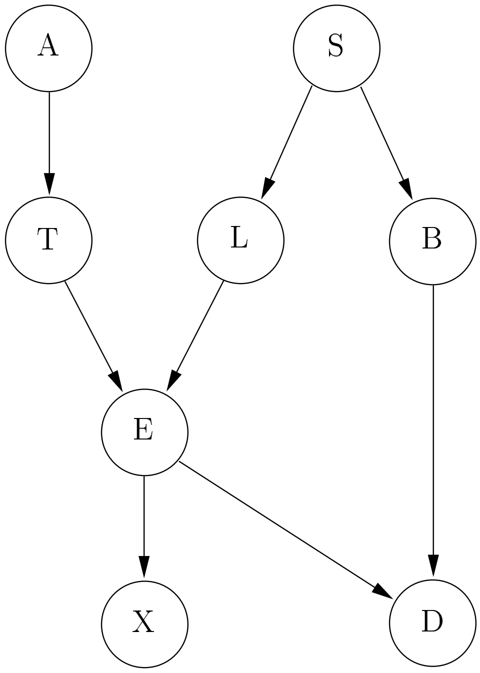


FIG. 1.4 – L'exemple fictif *ASIA* [66].

- B la bronchite,
- D la dyspnée,
- X une simple radio des poumons (détection aux rayons X),
- E une variable indicatrice qui prend la valeur 1 lorsque T et L sont "positifs" (valent 1).

Toutes ces variables sont binaires et aléatoires. A et S sont les variables d'entrée de notre problème. X et D sont les variables réponses, qui vont permettre l'interprétation des résultats. Toutes les autres variables permettent l'analyse des données, c'est-à-dire d'expliquer les variables réponses en fonction des variables d'entrée.

Une question intéressante peut par exemple être : "Sachant la variable L, A et S sont-ils conditionnellement indépendants ? " Pour répondre à cette question, on construit le graphe moral engendré par $\{A, L, S\}$ (cf. figure 1.5) qui permet de conclure à l'indépendance conditionnelle.

En revanche, pour répondre à la question : "Sachant la variable E, A et S sont-ils conditionnellement indépendants ? " il faut construire le graphe moral engendré par $\{A, T, E, L, S\}$ (cf. figure 1.6). Dans ce graphe, il est facile de voir qu'il existe un chemin dans le graphe moral entre A et S.

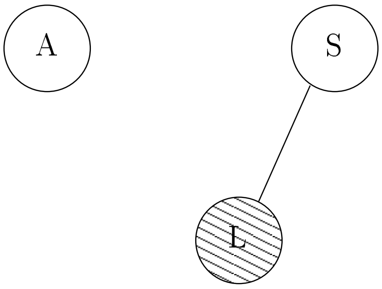


FIG. 1.5 – Connaissant L, A et S sont conditionnellement indépendants.

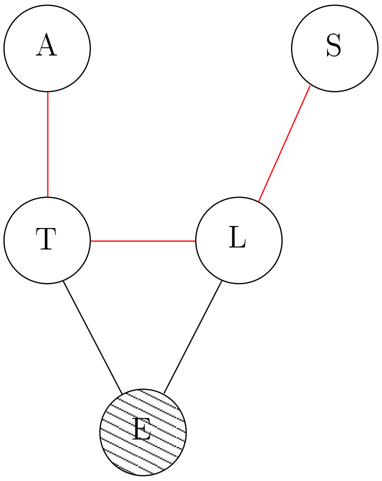


FIG. 1.6 – Connaissant E, A et S ne sont pas conditionnellement indépendants.

1.3 Inférence dans un réseau bayésien

1.3.1 Introduction

Dans l'exemple d'*ASIA* la probabilité jointe $P(U)$ peut s'écrire sous sa factorisation récursive :

$$P(U) = \phi_A \phi_T \phi_S \phi_B \phi_L \phi_E \phi_D \phi_X, \tag{1.4}$$

où $\phi_A = p(A)$, $\phi_T = p(T|A)$, $\phi_S = p(S)$, $\phi_B = p(B|S)$, $\phi_L = p(L|S)$, $\phi_E = p(E|T, L)$, $\phi_D = p(D|E, B)$, $\phi_X = p(X|E)$ sont définis comme des potentiels. Nous verrons par la suite que ces potentiels, au cours des calculs, ne sont pas nécessairement des probabilités. Ces potentiels sont ensuite à la base du calcul des potentiels sur des ensembles de nœuds, appelés *cliques*.

La connaissance de la loi jointe par la formule (1.4) nécessite deux probabilités marginales, $p(A)$ et $p(S)$, et 16 probabilités conditionnelles, soit au total 18 probabilités. Ces probabilités constituent les données d'entrée du modèle, une fois la structure du réseau connue. Sans aucune hypothèse d'indépendances conditionnelles, il aurait fallu $2^8 - 1 = 255$ probabilités. Faire de l'inférence dans un réseau bayésien consiste principalement à calculer des probabilités marginales, à les mettre à jour après l'observation d'une variable, appelée *évidence*. Ce calcul peut s'avérer très coûteux. Ainsi l'objectif des procédures présentées ici vise à réduire ces coûts de calculs en utilisant notamment la théorie des graphes. Dans un premier temps nous nous intéressons au calcul d'une probabilité marginale seule, puis au calcul de marginales lorsque l'on observe une ou plusieurs variables. Puis, nous étudions en détail le célèbre algorithme d'arbre de jonction (junction tree en anglais, cf. Jensen [59]). L'objectif de ce paragraphe est de décrire les techniques d'inférence les plus souvent utilisées pour le calcul des probabilités dans les réseaux bayésiens. Ces techniques seront, tout au long de ce document, utilisées et notamment pour l'application industrielle via la boîte à outil Matlab, BNT [102, 103]).

1.3.1.1 Calcul d'une probabilité marginale

Supposons que l'on s'intéresse à la probabilité marginale de D :

$$P(D) = \sum_{A,T,S,B,L,E,X} \phi_A \phi_T \phi_S \phi_B \phi_L \phi_E \phi_D \phi_X.$$

Afin d'éviter le calcul de la probabilité jointe de toutes les variables, on peut réorganiser les sommes de la manière suivante :

$$\begin{aligned} P(D) &= \sum_A \phi_A(A) \sum_T \phi_T(T,A) \sum_S \phi_S(S) \sum_L \phi_L(L,S) \sum_E \phi_E(E,T,L) \\ &\quad \sum_B \phi_B(B,S) \phi_D(D,E,B) \sum_X \phi_X(X,E). \end{aligned} \tag{1.5}$$

On peut voir dans l'équation (1.5) que le nombre de probabilités nécessaires pour le calcul de $P(D)$ a considérablement diminué. En ordonnant les sommes (et donc les variables) d'une certaine manière, le calcul de la probabilité marginale est effectué en lisant l'équation de droite à gauche. Ainsi, premièrement, il faut calculer $\phi'_X(E) = \sum_X \phi_X(X,E)$, puis multiplier $\phi'_X(E)$ par $\phi_B(B,S)\phi_D(D,E,B)$ et calculer $\phi'_B(D,E,S) = \sum_B \phi_B(B,S)\phi_D(D,E,B)\phi'_X(E)$; $\phi'_B(D,E,S)$ est multiplié par $\phi_E(E,T,L)$, et ainsi de suite.

Nous verrons par la suite, que ces sommes bien ordonnées dépendent de variables qui forment des cliques dans un graphe non orienté triangulé, qu'il faudra construire. Ensuite, calculer les sommes sur ces variables revient à calculer des potentiels sur les cliques. Les multiplications entre les sommes seront représentées par des messages allant de cliques en cliques. Ces potentiels vont permettre de calculer non seulement la probabilité jointe mais aussi les probabilités marginales des variables, ceci étant possible, grâce à la réduction du nombre de variables (trois au maximum dans une clique d'un graphe rendu triangulé).

1.3.1.2 Après observations

Dans ce paragraphe, nous nous intéressons aux calculs des probabilités dans le cas où l'on observe une ou plusieurs variables. Cette situation est courante lors d'aide au diagnostic, où l'on veut calculer les probabilités de défaillance sachant que l'on a observé tel phénomène. Dans la suite, on appelle évidence, toute connaissance sur une ou plusieurs variables.

Afin de simplifier les calculs, on peut remarquer que le potentiel $\phi'_X(E) = \sum_X \phi_X(X,E)$ est égal à un. Ainsi, $\phi'_B(D,E,S) = \sum_B \phi_B(B,S)\phi_D(D,E,B)$. Cette variable X n'est pas prise en compte dans le calcul de cette probabilité marginale car ce nœud est appelé en anglais, « *barren node* », c'est-à-dire un nœud qui n'est pas observé ou tel que tous ses successeurs n'ont pas été observés. Ce type de nœud n'a donc aucun impact sur ses descendants. Supposons que l'on sache que la variable X prenne la modalité x. Les probabilités marginales peuvent être recalculées, sans marginaliser selon cette variable, mais en remplaçant son potentiel par un nouveau potentiel $\phi'_X(E) = P(X = x|E)$. De nouveau, nous verrons que l'utilisation d'un graphe non orienté triangulé va permettre de réduire considérablement les calculs.

1.3.2 Les graphes triangulés et l'arbre de jonction

À partir du graphe moral, on construit un graphe triangulé composé de cliques qui forment ensuite l'arbre de jonction (« *junction tree* » en anglais). Cet arbre possède de nombreuses propriétés et permet d'alléger considérablement le nombre d'opérations, via des potentiels sur les cliques. Nous définissons tout d'abord les cliques puis les graphes triangulés.

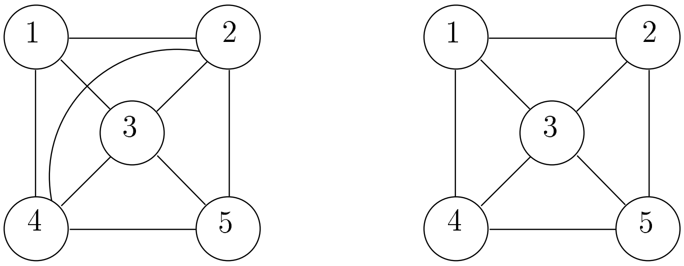
Définition 8 *Un graphe est dit complet, si toutes les paires de nœuds sont jointes. Une sous-suite de nœuds est dite complète si elle engendre un sous-graphe complet. Une clique est un sous-graphe complet maximal (au sens de l'inclusion).*

Définition 9 *Un graphe triangulé est un graphe non orienté, tel que tous les cycles de longueur $n \geq 4$ possèdent une corde, i.e. deux nœuds non consécutifs sont voisins.*

Dans le premier graphe (a) de la figure (1.7) on peut voir qu'il n'existe pas de cycle de longueur supérieure à quatre sans corde. Le deuxième graphe (b) quant à lui possède un cycle de longueur quatre, qui n'a pas de corde : $(1, 2, 5, 4)$.

Définition 10 *Soit G une suite de cliques d'un graphe non orienté, et les cliques de G organisées dans un arbre T. T est un arbre de jonction si pour chaque paire de nœuds V, W tous les nœuds sur le chemin entre V et W contiennent l'intersection $V \cap W$.*

Il est de plus prouvé que dans un graphe non orienté G triangulé, les cliques de G peuvent être organisées en arbre de jonction (cf. Jensen [59]). Ainsi, si le graphe moral n'est pas triangulé, on ajoute des lignes afin de le rendre triangulé. De cette manière, on construit un arbre de jonction avec des cliques. Les inférences dans le réseau bayésien vont se faire à partir de cet arbre de jonction grace à des passages de messages entre les cliques. Ceci



(a) Un graphe triangulé (b) Un graphe non triangulé

Fig. 1.7 – Exemple de graphes triangulés et non triangulés.

i	Cliques C_i	Résidus R_i	Séparateurs S_i	Potentiels ϕ_i
1	A, T	A, T	$\emptyset$	$p(A)p(T\|A)$
2	T, L, E	L, E	T	$p(E\|T, L)$
3	B, L, E	B	L, E	1
4	S, L, B	S	L, B	$p(S)p(B\|S)p(L\|S)$
5	D, B, E	D	B, E	$p(D\|B, E)$
6	X, E	X	E	$p(X\|E)$

Tab. 1.1 – Cliques ordonnées pour l'exemple *ASIA*.

permet de réduire considérablement le nombre de variables et donc le nombre d'opérations à effectuer lors de l'inférence. L'étape d'ajout des arêtes est toujours possible, le cas extrême correspondant à rendre le graphe complet. Toutefois il est judicieux de rendre triangulé le graphe moral avec peu d'arcs, et donc d'avoir un graphe de jonction avec peu de cliques. En effet, les potentiels sur les cliques sont à la base de l'inférence dans le graphe et donc limiter ce nombre allège le nombre d'opérations à effectuer.

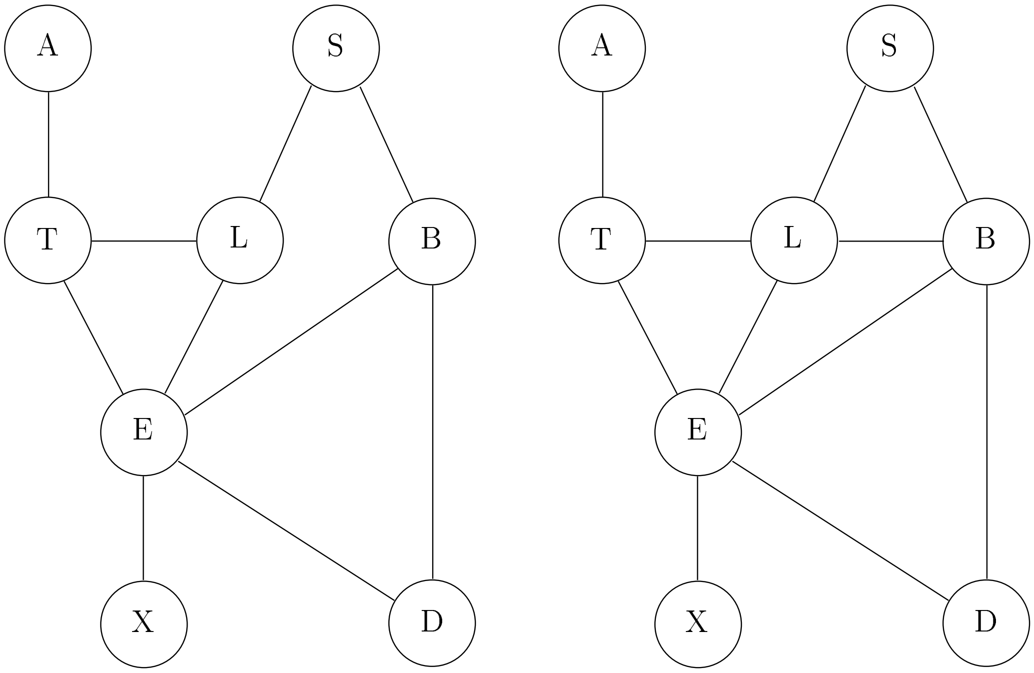
Ainsi, le graphe moral pour l'exemple *ASIA* et le même graphe rendu triangulé sont donnés figure 1.8. On peut remarquer que pour rendre le graphe triangulé, une autre possibilité consiste à ajouter une liaison entre S et E.

Les cliques du graphe moral triangulé sont données dans le tableau (1.1), en les ordonnant suivant un algorithme décrit par Tarjan et Yannakakis [88]. Cet algorithme permet de « remplir » le graphe afin de le rendre triangulé et donc de s'affranchir des différents choix de « remplissage ». Cet ordre permet d'écrire la probabilité jointe comme :

$$P(U) = \frac{\prod_i p(C_i)}{\prod_i p(S_i)} = \prod_i p(R_i|S_i).$$

où $S_i = C_i \cap (C_1 \cup \ldots \cup C_{i-1})$ et $R_i = C_i \backslash S_i$. Les S_i sont appelés séparateurs de cliques et les R_i sont appelés les résidus.

Ces cliques sont représentées par des rectangles dans l'arbre de jonction (cf. figure (1.9)) et les séparateurs de cliques par des ellipses. Cet arbre n'est pas unique. En effet, la dernière



(a) Le graphe moral pour *ASIA* (b) Le graphe moral rendu triangulé

FIG. 1.8 – Le graphe moral de l'exemple *ASIA* et sa version triangulé.

clique (X, E) aurait pu être voisine de la deuxième clique (T, L, E) avec comme séparateur E. Le choix d'un arbre se fera selon les observations ou selon les probabilités marginales que l'on souhaitent calculer.

1.3.3 Propagation dans le graphe de jonction

À chaque clique du graphe de jonction est associé un potentiel. Ces potentiels sont initialisés par les probabilités conditionnelles du réseau bayésien. Ainsi les potentiels de la première clique C_1 sont ϕ_A, ϕ_T, où $\phi_A = p(A)$ et $\phi_T = p(T|A)$. Les autres potentiels du réseau bayésien sont $\phi_E = p(E|T, L)$ pour la clique C_2, $\phi_S = p(S)$, $\phi_B = p(B|S)$ et $\phi_L = p(L|S)$ pour la clique C_4, $\phi_D = p(D|E, B)$ pour la clique C_5 et enfin $\phi_X = p(X|E)$ pour la dernière clique C_6. La troisième clique $C_3 = (B, L, E)$ ne possède pas de probabilités conditionnelles associées au réseau bayésien. En effet, cette clique est née lorsque le graphe moral a été transformé en graphe triangulé. Ainsi, le potentiel de cette clique est initialisé à 1.

L'inférence dans le réseau bayésien se fait par passage de messages entre les cliques voisines via leurs séparateurs. Par exemple pour calculer la probabilité marginale $P(D)$, il faut trouver une clique contenant D, soit C_5. Il est donc nécessaire de choisir cette clique comme clique racine et d'envoyer des messages partant des feuilles de l'arbre en direction de $C_5 = (D, B, E)$, comme indiqué dans le graphe (1.10). On note $\phi^{\downarrow X}$ le potentiel résultant de la marginalisation selon X, c'est-à-dire après sommation sur les autres variables. La clique

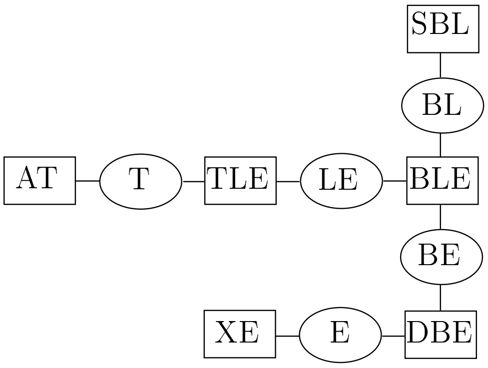


FIG. 1.9 – Un arbre de jonction pour *ASIA*.

$C_1 = (A, T)$ va envoyer le message

$$\psi_1 = \phi_1^{\downarrow T} = \sum_A \phi_A \phi_T = \sum_A p(A)p(T|A) = p(T).$$

Ce message est envoyé à $C_2 = (T, L, E)$, dont le nouveau potentiel est $\psi_1 \phi_E$, qui elle-même envoie un message à la clique $C_3 = (B, L, E)$:

$$\psi_2 = \phi_2^{\downarrow LE} = \sum_T \psi_1 \phi_E = \sum_T p(T)p(E|T, L) = \sum_T p(T|L)p(E|T, L) = \sum_T p(E, T|L) = p(E|L),$$

car T et L sont indépendants (voir proposition 1). La clique C_4 envoie le message à la clique $C_3 = (B, L, E)$:

$$\psi_3 = \phi_4^{\downarrow BL} = \sum_S \phi_S \phi_B \phi_L = \sum_S p(S)p(L|S)p(B|S) = p(B, L).$$

Cette même clique reçoit donc deux messages (ψ_2 et ψ_3) et a un nouveau potentiel égal à $\phi_3 = \psi_2\psi_3 = p(E|L)p(B, L) = p(E|L, B)p(B, L) = p(E, L, B)$ (car $E \amalg B|L$). En effet, le potentiel de cette clique a été initialisé à 1. Cette clique envoie comme message :

$$\psi_4 = \phi_3^{\downarrow BE} = \sum_L \psi_2\psi_3 = \sum_L p(E, B, L) = p(E, B).$$

La clique $C_6 = (X, E)$ va également envoyer un message

$$\psi_5 = \phi_6^{\downarrow E} = \sum_X \phi_X = \sum_X p(X|E) = 1.$$

Ainsi, la clique racine $C_5 = (D, B, E)$ reçoit deux messages ψ_4 et ψ_5, permettant de mettre à jour son potentiel

$$\phi_5 = \phi_D\psi_4\psi_5 = p(D|E, B)p(E, B) = p(D, E, B),$$

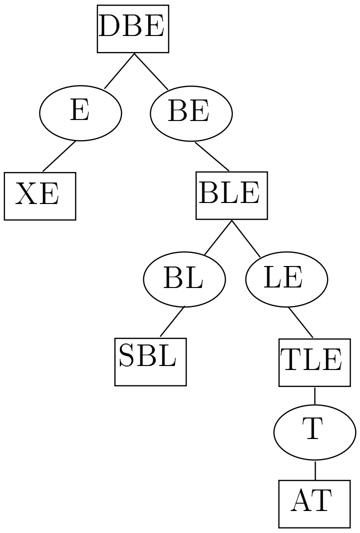


FIG. 1.10 – L'arbre de jonction avec la clique C_5 comme clique racine.

c'est-à-dire la probabilité marginale de la clique C_5 (cf. Jensen [59]). Il suffit de marginaliser selon B et E pour obtenir la probabilité marginale de D. L'algorithme décrit ci-dessus s'appelle la *collecte de l'évidence*. Une fois la collecte effectuée, il est possible de distribuer cette évidence (*distribution de l'évidence*) afin de calculer toutes les probabilités marginales des cliques. Pour cela, le départ de la propagation s'effectue de la clique racine vers les feuilles de l'arbre en envoyant des messages entre les cliques voisines. Les messages envoyés lors de la *distribution de l'évidence* sont notés ψ^i, pour les différencier des messages envoyés lors de la collecte. Ces messages peuvent être calculés selon une propagation dans le graphe, « *lazy propagation* » pour le terme anglais, donnée dans Jensen [59].

Définition 11 *Soit une clique V avec un potentiel ϕ_V et soit S un séparateur de V. On considère $S_1, \ldots, S_k$, les autres séparateurs de V. On suppose que chaque séparateur S_i envoie un message ψ_i vers V. Alors V peut passer le message $(\phi_V \cup \psi_1 \cup \ldots \cup \psi_k)^{\downarrow S}$ à S. On dit alors que la direction $S - V$ est déclenchée.*

Si toutes les directions sont déclenchées, alors l'arbre de jonction est dit plein. Nous énonçons le théorème permettant de calculer toutes les probabilités marginales.

Théorème 2 *Soit T un arbre de jonction d'un réseau bayésien avec les évidences e. Supposons que T soit plein. Soit $S_1, \ldots, S_k$ les k séparateurs de la clique C, avec un potentiel ϕ_C. Alors*

$$P(C, e) = \phi_C \phi_1 \ldots \phi_k.$$

Soit S un séparateur avec deux messages ψ_S et ψ^S. Alors,

$$P(S, e) = \psi_S \psi^S.$$

Reprenons l'exemple d'*ASIA*. La figure 1.11 montre le graphe de jonction et les messages entrants et sortants.

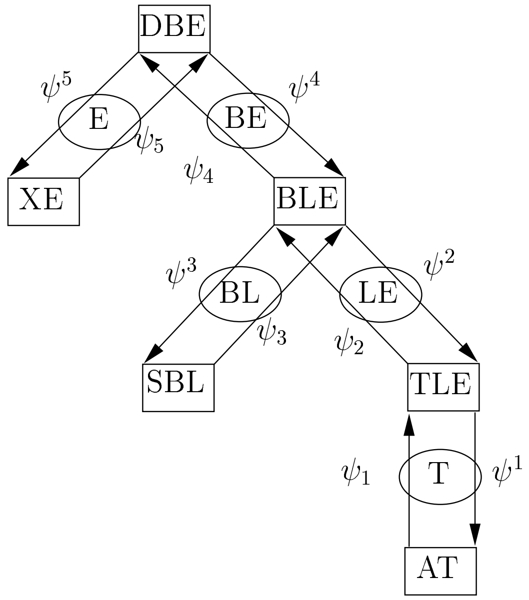


Fig. 1.11 – L'arbre de jonction avec les messages entrants et sortants.

Le message ψ^5 d'après ce qui est décrit ci-dessus est égal à $(\phi_5\psi_4)^{\downarrow E}$, soit

$$\psi^5 = \sum_{D,B} \phi_5\psi_4 = \sum_{D,B} p(D|B,E)p(E,B) = p(E).$$

La clique $C_5 = (D,B,E)$ envoie le message $\psi^4 = (\phi_5\psi_5)^{\downarrow BE}$ à la clique C_3

$$\psi^4 = \sum_{D} \phi_5\psi_5 = \sum_{D} p(D|B,E) = 1.$$

Le message $\psi^3 = (\phi_3\psi_2\psi^4)^{\downarrow BL}$ est

$$\psi^3 = \sum_{E} \phi_3\psi_2\psi^4 = \sum_{E} p(E|L) = 1.$$

Le message $\psi^2 = (\phi_3\psi_3\psi^4)^{\downarrow LE}$ est

$$\psi^2 = \sum_{B} \phi_3\psi_3\psi^4 = \sum_{B} p(B,L) = p(L).$$

Enfin le message $\psi^1 = (\phi_2\psi^2)^{\downarrow T}$ est égal à

$$\psi^1 = \sum_{LE} \phi_2\psi^2 = \sum_{LE} p(E|T,L)p(L) = \sum_{LE} p(E,L|T) = 1,$$

car T et L sont indépendants ($p(L) = p(L|T)$).

Une fois tous les messages envoyés dans les deux directions, l'arbre de jonction est plein et en appliquant le théorème 2, il est possible de calculer toutes les probabilités marginales des cliques et des séparateurs. Par exemple,

$$p(B, L, E) = \phi_3 \psi_2 \psi_3 \psi^4 = 1 * p(E|L) * p(B, L) * 1,$$

et la probabilité du séparateur (L, E) est

$$p(L, E) = \psi_2 \psi^2 = p(E|L)p(L).$$

Lorsqu'il y a une évidence, celle-ci est représentée par des potentiels $0-1$, qui sont inclus dans les cliques correspondantes. Il suffit alors de faire une propagation entière dans le graphe de jonction pour récupérer les probabilités marginales des cliques.

1.4 Conclusions

Les réseaux bayésiens sont des outils souvent utilisés en médecine, en intelligence artificielle, en méthode d'apprentissage. Plus précisément, ils sont étudiés pour l'aide au diagnostic [1], l'aide à la décision [91], pour des modèles de Markov caché [43], pour la reconnaissance de formes. Dans tous ces domaines, l'algorithme de l'arbre de jonction présenté ci-dessus est très utilisé et fait souvent référence. Toutefois, quelques critiques entourent cet algorithme d'arbre de jonction. Premièrement, l'algorithme est basé sur le calcul de potentiels et non sur des probabilités, d'où un manque d'interprétation. De plus, si l'on effectue plusieurs inférences à la suite, l'algorithme va refaire de nombreux calculs inutiles. Cependant, nous appliquons cet algorithme pour un réseau bayésien de taille raisonnable (22 variables), et ils semblent que les nombreuses inférences effectuées ont été rapides. Nous nous sommes plus attachés à la construction de tels modèles et à leurs possibilités, plutôt qu'à la performance des algorithmes d'inférence. Cet algorithme et ses dérivés sont inclus dans une boite à outil Matlab (toolbox BNT). Elle peut être téléchargée gratuitement sur le web (cf. [102, 103]). Des logiciels commerciaux ou académiques dédiés uniquement aux réseaux bayésiens existent également. On peut citer par exemple les plus connus, Netica [104], Hugin [105], Java Bayes [106] développé en Java. Des études au sein d'Électricité De France [30] ont été menées avec le logiciel Netica. Généralement, la plupart de ces logiciels sont gratuits tant que le nombre de variables n'est pas trop élevé.

Chapitre 2

Construction de réseaux bayésiens dans le cadre de la maintenance

Pour notre étude, le réseau bayésien représente l'évolution de dégradation d'un système mécanique d'une centrale nucléaire. Le premier travail fut donc de choisir un système adapté à cette étude. Ici, le système mécanique étudié est le joint 1 de la pompe primaire 900 MW (cf. annexe). Le réseau bayésien donne une représentation causale du phénomène, d'une manière qualitative et quantitative. La partie qualitative est assurée par la structure du graphe, et donc par les relations d'indépendances conditionnelles associées au graphe. La partie quantitative relève de l'écriture de la loi jointe avec les probabilités associées au graphe. La construction d'un réseau bayésien se fait en deux étapes, la première consistant à construire la structure du graphe et la deuxième s'attachant à estimer les probabilités correspondantes. On peut distinguer deux cas de figures, le cas où il y a des données de retour d'expérience et le cas où il y en a pas. S'il existe une base de données, les principales méthodes utilisées pour la construction du graphe (voir par exemple [56]) sont basées sur la vraisemblance ou sur la vraisemblance pénalisée. Elle consiste par exemple à considérer le graphe complet, où tous les noeuds sont connectés deux à deux, puis à éliminer certaines arêtes en utilisant par exemple le critère AIC (cf. [2]). La deuxième étape vise alors à estimer les probabilités associées aux arêtes retenues lors de la première étape. Dans le cas où il n'existe pas de retour d'expérience, la structure du graphe est alors totalement déterminée par les avis d'experts (voir également [56]). Une fois la structure du graphe déterminée, la seconde étape vise à évaluer les probabilités conditionnelles associées à cette structure. Une nouvelle fois, les probabilités seront données par les avis d'experts.

Il est possible de rencontrer des cas intermédiaires, comme par exemple le cas où il y a peu de données, ou dans un cas avec des données mais également des avis d'experts. Dans ces cas, la construction du graphe devra faire appel aux avis d'experts. De plus, les réseaux bayésiens complexes avec beaucoup d'arêtes vont nécessiter énormément de données. Ainsi, dans la plupart des cas, les avis d'experts seront nécessaires pour la construction du graphe et pour l'évaluation des probabilités. En effet, le nombre de probabilités à évaluer est fonction exponentielle du nombre d'arêtes du graphe. L'étape de construction du graphe doit donc intégrer cette composante, afin que l'intégration des probabilités soit une étape faisable. Cette étape souvent difficile, nécessite la plupart du temps des avis d'experts, comme c'est le cas dans le cadre de l'étude présentée. Puis, le nombre de probabilités à leur demander étant trop important, nous avons approximé le réseau bayésien par un modèle log-linéaire, qui consiste à exprimer les logarithmes des probabilités en fonction de termes d'interactions entre les variables du modèle (voir Whittaker [98]). Cette approximation a donc permis de diminuer le nombre de probabilités nécessaires à l'inférence. De plus, nous avons décidé de demander toutes les probabilités marginales, ce qui nous permet de vérifier la cohérence des avis d'experts et de proposer un retour vers les experts lorsqu'une incohérence est détectée (voir [37]).

Le chapitre est organisé comme suit. Dans un premier temps, nous présentons les enjeux de l'optimisation des politiques de la maintenance, tant sur les coûts que sur la sûreté, puis nous verrons les principaux intérêts de l'utilisation des réseaux bayésiens pour ce type d'applications, et pour quel type de système ces modèles peuvent-ils être utilisés. Dans un second temps, nous nous intéresserons à la construction de la structure graphique, ainsi qu'à l'évaluation des probabilités associées, et notamment l'approche par les modèles log-linéaires afin de réduire le nombre d'opérations de calcul, et qui peut permettre aussi de vérifier la cohérence des avis d'experts. Cet angle d'approche nous permet également de maîtriser les hypothèses faites, et de les relâcher très facilement en questionnant à nouveau les experts. Elle s'inspire mais est assez différente de la célèbre méthode Delphi ([71]), dont le but principal est d'extraire les avis d'experts, puis de les comparer et de les questionner de nouveau.

2.1 L'optimisation de la maintenance

2.1.1 Enjeux

Depuis plus d'un siècle, le monde industriel n'a cessé de gagner en productivité, et en fiabilité des systèmes. Actuellement, les principaux enjeux sont la sûreté, la disponibilité, les coûts (en particulier de maintenance), et la durée de vie. Pour les entreprises, on peut résumer ces enjeux par la compétitivité et la sûreté. Il faut donc diminuer les coûts, en optimisant la maintenance tout en maintenant voire en améliorant les objectifs de sûreté. C'est pour cela qu'ont été développées les méthodes OMF (Optimisation de la Maintenance par la Fiabilité) ou RCM (Reliability Centered Maintenance) qui optimisent les programmes de maintenance sur la base de l'analyse fonctionnelle et du retour d'expérience : la maintenance la mieux adaptée au bon endroit. L'OMF est un processus vivant, que l'on réactualise tous les cinq ans pour EDF. Cependant, à part pour le retour d'expérience, lorsqu'on choisit une action de maintenance préventive, on ne connaît pas son efficacité et son impact. L'objectif est, à partir des connaissances et des observations sur les comportements et les dégradations :

- de modéliser la durée de vie d'un matériel et de quantifier la probabilité de dégradation ou de défaillance,

- de détecter les variables importantes agissant sur la dégradation et trouver des actions afin de différer ou éliminer le vieillissement,
- et de calculer l'impact d'une action de maintenance sur le comportement du matériel.

2.1.2 Définitions

Nous commençons par donner les définitions de composants et de systèmes données dans le livre d'Ascher et Feingold [5]. On appelle composant, un produit non démontable et jeté la première fois où il défaille. On appelle système , un assemblage de composants liés entre eux, et destinés à accomplir une ou plusieurs fonctions. On distingue deux types de systèmes : les systèmes réparables et les systèmes non réparables. On appelle système non réparable, un système jeté dès qu'il cesse de fonctionner correctement. Un système réparable est un système qui, après avoir perdu une ou plusieurs fonctionnalités, peut être restauré, afin de recouvrer toutes ses fonctionnalités, par toute méthode autre que le remplacement du système entier. Il existe trois états possibles pour le système, l'état sain, l'état dégradé et l'état défaillant. Dans les deux premiers cas, toutes les fonctions du système sont encore réalisées, le système est alors en état de bon fonctionnement. L'état dégradé correspond au cas où le système possède toutes ses fonctionnalités, mais pas de façon optimale. Dans l'état de défaillance, qui est un cas particulier d'un état de mauvais fonctionnement, le système n'assure plus les fonctions requises. Une intervention de maintenance est alors nécessaire. Enfin, on appelle défaillance, la transition d'un état de bon fonctionnement vers un état de mauvais fonctionnement.

En ce qui concerne la maintenance, il en existe de trois types : la maintenance corrective, la maintenance préventive, et la maintenance mixte. La maintenance corrective s'applique sur un système dans un état de défaillance. Cette action de maintenance vise à réparer entièrement ou partiellement le système, afin de le ramener à un état de bon fonctionnement. La maintenance préventive vise à maintenir le système avant que celui-ci ne défaille. Celle-ci sert notamment à améliorer l'état du système, à réduire le nombre de défaillances, et à améliorer la disponibilité du système. La maintenance préventive a par conséquent comme principal objectif de réduire la probabilité de défaillance et donc à terme de réduire les coûts de maintenance. En effet, contrairement à une maintenance corrective, la maintenance préventive permet d'optimiser la gestion du stock, d'améliorer l'ordonnancement des actions de maintenance, ... Enfin, puisque le risque zéro n'existe pas et que l'on ne pourra jamais prédire exactement la date d'une défaillance, il est souvent commode de modéliser la maintenance par une combinaison de ces deux types de maintenance, que l'on appelle la maintenance mixte. On peut résumer les différentes politiques de maintenance selon la figure 2.1. Une bonne synthèse de ces différentes stratégies de maintenance se trouve par exemple dans la thèse de Bruno Castanier [24]. Le choix d'une politique est parfois imposé, comme c'est le cas dans le nucléaire en France. Ainsi, l'optimisation de la maintenance par la fiabilité détermine la maintenance préventive "optimale". Dans le cas où il n'y a pas de maintenance préventive, cela revient à attendre la défaillance, c'est-à-dire une maintenance

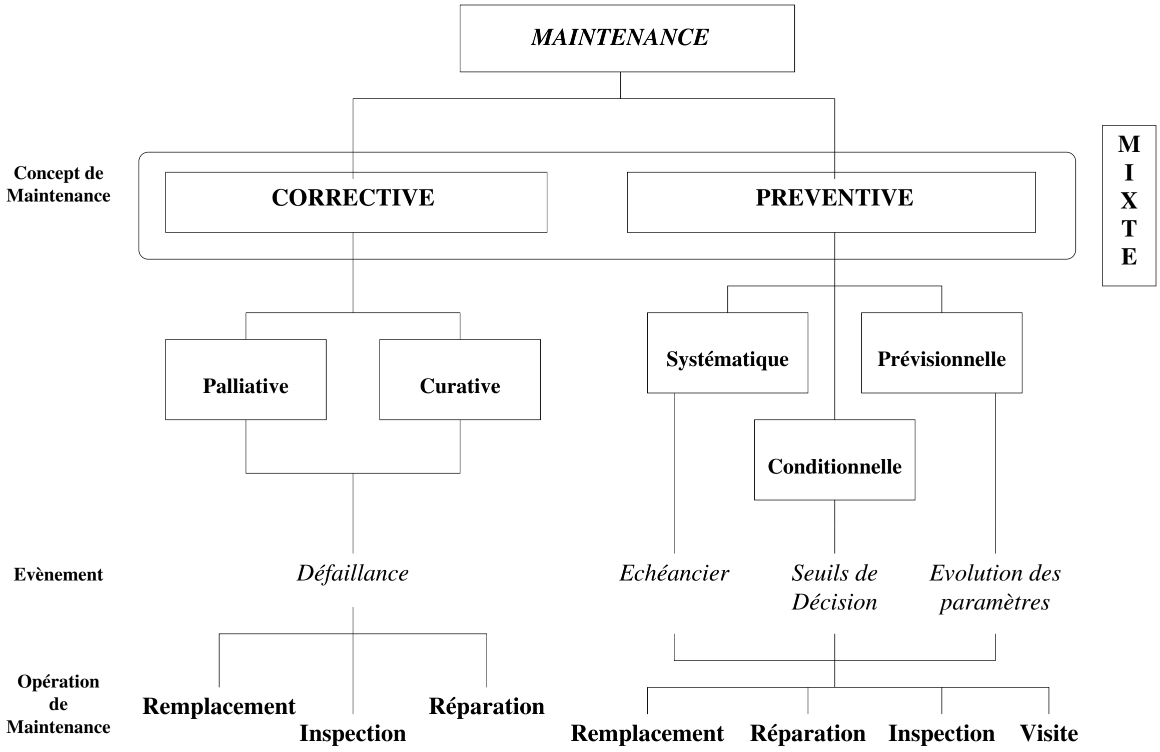


FIG. 2.1 – Exemple de politiques de maintenance.

corrective. De nombreuses études ont montré que les politiques de maintenance préventive (ou plutôt mixte) sont souvent les moins onéreuses sur le long terme.

2.1.3 Le cas du nucléaire

Pour Électricité de France, optimiser les programmes de maintenance revient à gagner en disponibilité et diminuer les coûts de maintenance grâce à l'Optimisation de la Maintenance par la Fiabilité (OMF). D'autre part, dans ce travail de thèse, nous proposons une méthodologie permettant d'identifier les variables importantes agissant sur la dégradation ou la défaillance du matériel, mais également un outil d'aide au diagnostic et d'aide à la décision. Enfin, comme nous le verrons dans le chapitre suivant, il est possible de modéliser les actions de maintenance.

2.2 Pourquoi l'utilisation de réseaux bayésiens ?

Le début des réseaux bayésiens date probablement des travaux de Wright [99] en 1921. L'utilisation de tels modèles s'est faite grandissante en médecine. On peut citer les travaux de Warner et al. [95] et Cooper [34] par exemple. Avec l'arrivée de l'ordinateur et la capacité croissante des bases de données, ces modèles ont été de plus en plus fréquents dans les applications diverses, comme le diagnostic, la prédiction, le traitement d'images, l'analyse de sensibilité, la prospective, l'aide à la décision. Un état des lieux sur l'utilisation des réseaux bayésiens se trouve dans Heckerman et al. [54] et dans ses références. Les industries ont quelque peu tardé à appliquer ces modèles. Mais, depuis le début des années 90, une multitude d'applications industrielles a vu le jour. On peut en lister quelques unes, comme les travaux de Weber et al. [96], Chatelain et Lannoy [30], , Kang and Golay [61], et Bouissou et al. [20], Yildiz et al. [101] dans le cas du nucléaire, et Breese et al. [21] pour un outil d'aide à la décision sur des turbines à gaz.

Les nombreuses études proposant les réseaux bayésiens témoignent certainement de leurs qualités. Nous en listons quelques unes ci-après.

Tout d'abord, pour des non statisticiens, l'avantage d'un réseau bayésien, comparé par exemple à un diagramme d'influence, est la modélisation de l'aléatoire, et donc une plus grande finesse du modèle. Pour les experts du domaine sur lequel nous allons appliquer un réseau bayésien, la représentation "causale" du phénomène leur semble naturelle et plus facile à déterminer et le caractère synthétique d'un réseau bayésien est séduisant. Pour des statisticiens, l'attrait pour ces modèles réside dans la représentation naturelle du phénomène et faisant un usage puissant des informations conditionnelles, sa capacité à représenter une grande quantité de variables, ainsi que la prise en compte des indépendances conditionnelles.

2.2.1 Pour quels types de matériels et quels investissements ?

Dans ce paragraphe, nous essayons de donner les caractéristiques que devraient avoir un matériel, pour une étude de ce type. Nous nous limitons au cas d'un système mécanique pour une industrie nucléaire.

Le but de l'étude est de représenter le processus de dégradation d'un système réparable afin de définir les tâches de maintenance les mieux adaptées et mesurer leur impact. Ce système est constitué de composants élémentaires. Il a été choisi pour son ancienneté (plusieurs années d'utilisation) et pour sa fonction vitale de sûreté. Comme nous le verrons par la suite, il est possible (et peut être préférable) de choisir un matériel neuf qui va être mis en service et dont la connaissance provient des études sous tests. Dans notre cas, les experts du domaine connaissent bien le matériel, ainsi que ses défauts de fonctionnement. De plus, celui-ci étant essentiel dans le processus de bon fonctionnement, il est en général bien surveillé. On peut ainsi espérer récupérer une multitude d'informations. Il est possible que le nombre de données soit suffisant pour que l'on puisse par exemple se passer des avis d'experts. Ceci étant, le réseau bayésien est pour notre part essentiellement construit et quantifié par des avis d'experts. Nous verrons dans le chapitre 4 le moyen de pondérer les avis d'experts et d'intégrer les données de retour d'expérience.

Lors du choix du matériel, plusieurs questions doivent être posées :

- Quelle est la criticité du matériel ?
- Existe-t-il des données de retour d'expérience ?
- Existe-t-il des experts connaissant bien ce matériel ?
- Si oui, seront-il disponibles pour ce genre d'étude sachant que celle-ci risque d'être longue avec beaucoup de réunions ?
- Prévoir un médiateur suffisant neutre, mais ayant quelques connaissances du matériel et des experts.

Comme l'indiquent les travaux de Bouissou [20], ces études sont très difficiles, notamment lors du démarrage et vont demander beaucoup d'investissement pour tous les membres du projet. Enfin, ces travaux affirment également que ces modèles nécessitent des années de calibration. Cependant, la difficulté de la mise en place de tels modèles est récompensée par les nombreuses possibilités qu'offrent ces modèles. En effet, comme nous le verrons dans le chapitre suivant, il est possible d'intégrer par exemple les actions de maintenance comme variables du modèle, mais aussi d'intégrer les experts comme variables.

2.2.2 Pour quels types d'applications en maintenance ?

Pour ce type d'études, l'application la plus courante est l'aide au diagnostic, très utilisée en médecine. Ainsi, si les variables du modèle sont suffisamment nombreuses et décrivent bien le processus, il est alors possible de déterminer le défaut le plus probable du système.

Partant de là, cet outil peut également faire office d'outil d'aide à la décision, comme nous le verrons par la suite. En effet, les experts peuvent donner des variables modélisant les actions de maintenance. Ces actions vont être intégrées dans le réseau bayésien en tant que variables toujours observées. De nouveau, les experts devront être capables de quantifier le bénéfice de la maintenance proposée ainsi que son coût. Les réseaux bayésiens permettent également de faire de la prédiction, par exemple via des scénarios simulés (cf. Chatelain et Lannoy [30]). Enfin, ces études doivent obligatoirement mener à une confrontation entre les experts, pendant la construction de la structure du graphe, mais également pendant l'analyse des résultats. Les mêmes travaux [30] ont cherché à modéliser l'évolution des coûts de maintenance par un réseau bayésien, où les variables du modèle sont le vieillissement des composants, les normes de sûreté, la motivation du personnel, et les coûts de maintenance.

2.3 Mise en place d'un réseau bayésien

La mise en place d'un réseau bayésien nécessite deux grandes phases. La première phase consiste à construire la structure du graphe, c'est-à-dire identifier les variables du modèle, et indiquer les indépendances et les indépendances conditionnelles entre ces variables. La deuxième phase consiste à calculer ou évaluer les probabilités nécessaires au calcul de la probabilité jointe de toutes les variables, ou autrement dit à quantifier les dépendances conditionnelles données lors de la première phase. Ces deux phases peuvent être effectuées soit par une procédure de sélection de modèle avec des données de retour d'expérience, soit encore par des avis d'experts, ou soit par une combinaison des deux. Dans notre cas, très peu de données de retour d'expérience étaient disponibles. Ainsi, la construction du réseau bayésien, incluant les deux phases précédemment décrites, s'est faite uniquement avec des avis d'experts. Dans ce cas, avec un graphe complexe ayant beaucoup d'arêtes, le nombre de probabilités à donner peut être gigantesque au point qu'il devient déraisonnable d'envisager que des experts aussi motivés soient-ils, puissent les évaluer de manière sensée. Pour remédier à ce problème, nous allons proposer une approximation du réseau bayésien par un modèle log-linéaire, ce qui réduit considérablement le nombre d'inconnues (de probabilités à quantifier par expertise). Cela revient à supposer des indépendances conditionnelles entre certaines variables. Grâce à ces hypothèses, il sera possible de résoudre le problème et même d'identifier des incohérences dans les dires des experts.

2.3.1 Construire la structure du graphe

La méthode utilisée pour le choix des variables explicatives est basée sur un article d'Højsgaard (cf. [56]) et sur la méthode SERENE [20, 105]. Toutefois, elle diffère en quelques points que l'on citera dans ce paragraphe.

Cette phase s'est déroulée par des réunions entre quatre experts, deux statisticiens et un médiateur. On peut dès à présent noter une différence avec la méthode SERENE qui préconise de faire intervenir beaucoup plus d'experts. Nous nous sommes limités à ce nombre, car le projet était ambitieux et difficile. Il est important que lors des réunions, toutes les personnes concernées soient présentes. Pour anecdote, fixer une date de réunion avec cinq experts peut s'avérer une tâche elle aussi très compliquée.

La première étape consiste à déterminer les variables du modèle. Puis, les experts doivent donner :

- le nom de la variable,
- le label de la variable,
- son caractère (discret ou continu),
- dans le cas discret, donner les modalités avec le respect de l'exhaustivité (l'union de toutes les modalités représente tous les cas possibles) et l'incompatibilité (l'intersection des modalités est vide).

Dans un souci de simplicité, il est conseillé de choisir des variables discrètes (il est toujours possible de discrétiser une variable continue). De plus, le nombre de modalités ne doit pas être trop élevé, d'une part pour une meilleure compréhension et d'autre part pour que l'étape d'évaluation des probabilités conditionnelles ne soit pas infaisable.

Le choix des variables du problème s'est fait lors de plusieurs réunions entre des experts, des statisticiens, et a été piloté par un médiateur. Ces variables aléatoires sont ensuite représentées par des nœuds dans un graphe. Les arêtes du graphe donnent les probables dépendances ou les probables influences entre ces variables. On distingue le cas où l'on a des données et le cas où l'on ne fait intervenir que les avis d'experts.

2.3.1.1 Avec des données de retour d'expérience

Ce paragraphe décrit la méthodologie d'Højsgaard [56]. Connaissant les variables mises en jeu dans le modèle, l'idée est de partir d'un modèle complet (avec toutes les arêtes) et d'enlever des arêtes. Pour cela il faut se donner un critère (par exemple le critère AIC), et enlever les arêtes qui feront le plus baisser ce critère afin de choisir le meilleur modèle. On rappelle que le critère AIC (cf. [52] page 346) est égal à :

$$AIC = -2 \log \mathcal{L}_n(\hat{\theta}_n) + 2p,$$

où $\log \mathcal{L}_n(\hat{\theta}_n)$ est le maximum de la log-vraisemblance du modèle, et p est le nombre de variables indépendantes du modèle.

L'exemple donné dans l'article d'Højsgaard traite d'un problème d'agriculture. On veut connaître l'évolution de la production d'une plante, suivant les variables du modèle. Les variables sont désignées par les experts :

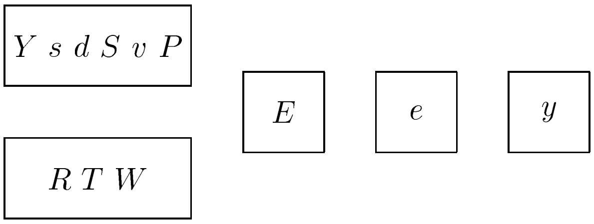
* y la production (variable continue),
* Y l'année de semence (variable discrète),
* S le type de sol (variable discrète),
* P les précédentes cultures faites sur ce sol (variable discrète),
* d la date de semence (variable discrète),
* s le traitement du sol (variable discrète),
* v la variété de la plante (variable discrète),
* W la dureté de l'hiver (variable discrète),
* R la pluie (variable discrète),
* T la température (variable discrète),
* E l'attaque de champignons au printemps (en %),
* e l'attaque de champignons en été (en %),
* a l'attaque d'autres champignons (en %).

Avec 789 observations, beaucoup de variables manquantes et un grand nombre d'états possibles (10692), il est difficile de repérer le vrai modèle (s'il existe !). Les réseaux bayésiens représentent des phénomènes complexes, il est donc courant que le nombre d'observations soit insuffisant. L'important est de choisir :

* un modèle raisonnable avec les avis d'experts,
* un modèle fournissant une description raisonnable des données,
* un modèle facilement interprétable,

* un modèle ouvert à discussion.

L'idée est tout d'abord de regrouper les variables et d'ordonner les groupes de variables :



Ce schéma représente un ordre naturel entre des groupes de variables. Nous considérons ici qu'une variable peut être expliquée par d'autres variables. Ceci ne veut pas dire que toutes les variables sont directement reliées les unes aux autres, mais que la plupart des relations ont une interprétation en termes de causalité.

Ici, les attaques des champignons au printemps sont vues comme une réponse du type de sol utilisé, de la date de semence, du traitement du sol, et de la variété. Les attaques de champignons en été sont vues comme une réponse des attaques de champignons au printemps et donc comme une réponse des variables précédemment citées. Finalement, la production est vue comme une réponse de toutes les autres variables.

Ainsi, dans cette représentation, les variables d'un bloc sont les réponses des blocs précédents (c'est-à-dire à gauche de ce bloc). Dans cet exemple, la variable E est expliquée selon les experts par Y, s, d, S, v, P, R, T et W. Une analyse de la variance, ou un modèle de Cox [39] permettent de ne garder que les variables explicatives significatives pour la mise en place d'un réseau bayésien. Cette représentation permet d'exprimer l'indépendance entre les variables d'un même bloc. Cette méthode permet également de réduire le nombre de connexions possibles entre les variables. Les variables sont regroupées dans un même bloc si l'on considère qu'elles ne dépendent pas les unes des autres. Enfin, les deux blocs de gauche sont séparés, car ils sont considérés être de nature différente. Ainsi, la température, la pluie et la dureté de l'hiver ont été regroupées dans un même bloc, que l'on peut qualifier de bloc météorologique ; l'autre bloc étant un bloc de caractéristiques agricoles.

Les arêtes peuvent être testées par un critère AIC, pour les variables discrètes, l'analyse de variance pour les variables continues. Cependant il est également possible d'utiliser un modèle de Cox pour connaître les variables explicatives, et donc les liaisons significatives. Ces estimations nécessitent cependant un nombre suffisant de données.

2.3.1.2 À partir d'avis d'experts

Dans ce paragraphe, nous nous inspirons de la méthode SERENE (cf. [29] pour le logiciel de construction de réseaux bayésiens). La construction du réseau se fait en premier lieu par un ensemble de sous-graphes reliés entre eux. Dans la méthode SERENE, la relation entre deux sous-graphes se fait par l'intermédiaire d'un nœud commun. Ici, nous proposons comme dans l'article d'Højsgaard de regrouper les variables par famille, et de construire le graphe de manière hiérarchique. En effet, l'ordre naturel est généralement trouvé en termes de causalité, ou de probable causalité. Une variable est une réponse d'un groupe de variables explicatives. Cette variable réponse fait également partie d'un groupe explicatif d'une autre variable. Ainsi, on peut distinguer trois groupes de variables. Les variables d'entrée sont toutes indépendantes entres elles et sont en amont. Les variables intermédiaires sont les réponses des variables d'entrée ou des variables intermédiaires. Enfin, les variables de sortie sont les variables d'intérêt, et sont les réponses de toutes les autres variables. Cette construction est appelée construction hiérarchique. Contrairement à la méthode SERENE, les sous-graphes représentés par les groupes de variables ne sont pas reliés via un nœud commun. En effet, le nœud commun a le désavantage de posséder un grand nombre de parents, et donc le nombre de probabilités conditionnelles requises devient trop important. De plus, un tel nœud représente un grand ensemble d'état de variables. Il a donc l'avantage de simplifier les dépendances entre variables, mais par conséquent il est moins précis.

Le regroupement des variables permet de simplifier et de clarifier les relations entre les variables. En effet, en reprenant l'exemple du paragraphe précédent, il est impossible d'avoir un arc allant de E vers le groupe R, T, W. En revanche, il se peut qu'il y ait des arcs entre des variables d'un même groupe. Ainsi, le nombre d'arcs possibles est considérablement réduit.

De plus, même s'il existe une base de données de REX, il est préférable de faire ce travail préliminaire afin que la sélection de modèle soit simplifiée. En effet, la représentation causale des groupes de variables a permis de réduire considérablement le nombre d'arêtes possibles. Une fois la structure du graphe déterminée, l'étape suivante consiste à évaluer les probabilités correspondantes.

2.3.2 Évaluation des probabilités

On suppose dans ce paragraphe que les probabilités sont déduites des avis d'experts.

2.3.2.1 Quelles probabilités et comment les calculer ?

Comme nous l'avons vu dans le premier chapitre, la probabilité jointe peut s'écrire dans le cas de graphe orienté acyclique comme le produit des probabilités des variables sachant les parents de ces variables (factorisation récursive). Ainsi, pour chaque nœud, il est nécessaire d'évaluer la probabilité de ce nœud sachant ses parents. Si la variable est une variable d'entrée, donc sans parent, alors la probabilité requise est une probabilité a priori. Pour les autres variables, les probabilités conditionnelles sont requises. On déduit toutes ces probabilités, en questionnant les experts. Le travail des statisticiens est donc de préparer un questionnaire par expert. Ainsi, à chaque probabilité correspond une question. Ces questions doivent être écrites avec des phrases en français. Ces interviews se sont déroulés en tête à tête entre un statisticien et un expert. Pour éviter toute corrélation entre les avis d'experts,

ces interviews se sont déroulés à des jours différents, et sans aucune communication entre les experts.
La principale difficulté pour un expert non statisticien, ou peu à l'aise avec le calcul des probabilités, réside en la conception des probabilités conditionnelles. Prenons par exemple un réseau bayésien avec trois variables discrètes binaires, S, A et C, respectivement le fait de fumer, d'être alcoolique, et d'avoir un cancer de la gorge. L'expert est capable de donner la probabilité d'avoir le cancer sachant que l'on est fumeur, et la probabilité marginale d'avoir le cancer. En revanche, la probabilité d'avoir le cancer sachant que l'on n'est pas un fumeur est plus difficile à donner. En effet, la plupart des experts donne pour cette probabilité, la probabilité marginale d'avoir le cancer, considérant que le fait d'être non fumeur n'est pas une information. Dans ce même exemple, l'expert doit être capable de donner 4 probabilités conditionnelles, comme par exemple la probabilité d'avoir le cancer sachant que la personne est non fumeur et alcoolique. Cette combinaison de variables est encore assez simple avec deux nœuds parents, mais il est difficile d'imaginer demander une combinaison avec un grand nombre de variables.

Lors de l'établissement du questionnaire, il est important de ne pas négliger les événements rares. Considérons par exemple, le cas d'un composant très fiable, dont la probabilité d'être défaillant est de l'ordre de 10^{-6}. Supposons un réseau bayésien avec deux nœuds, une variable binaire "maladie" (oui/non) influant la variable "état" (sain/défaillant). Afin de connaître la probabilité jointe, il est nécessaire de connaître la probabilité marginale de la maladie et deux probabilités conditionnelles, les deux autres étant déduites par complémentarité. Ainsi, afin de prendre en compte l'évènement rare, il est préférable de demander la probabilité d'être défaillant sachant la maladie, et la probabilité d'être défaillant sachant la non maladie. En effet, si l'on demande à un expert la probabilité d'être sain sachant la maladie ou la non maladie, ces deux probabilités risquent de n'avoir que très peu de différences. Ainsi, s'il existe un facteur multiplicatif, celui-ci risque d'être amoindri voire gommé.

L'établissement d'un réseau bayésien se heurte à la multiplicité des probabilités a priori et surtout des probabilités conditionnelles à renseigner par les experts. Il suffit de quelques variables comportant de nombreux liens pour rendre cette tâche, indispensable, irréalisable. Il est donc important de définir des règles simplificatrices pour laisser des proportions humaines à cette étape préliminaire (cf. [30]). Nous exposons ci-après les règles que nous avons choisi.

Règle 1 : L'analyse demande aux experts les probabilités marginales de toutes les variables. Considérant que les probabilités marginales étaient les plus faciles à évaluer, nous demandons ces probabilités même pour des variables intermédiaires. Or, dans la procédure classique seules les probabilités marginales des variables d'entrée sont demandées. Cet ajout de probabilités évaluées ne permet toutefois pas de calculer la probabilité jointe. En effet, l'écriture de la loi jointe nécessite la connaissance de probabilités conditionnelles (sauf dans le cas où toutes les variables sont indépendantes, ce qui réduit l'intérêt de l'étude). Les probabilités conditionnelles pour une variable binaire sont au nombre de 2^n, où n est le nombre de parents (dans le cas où tous les parents sont aussi des variables binaires). Il est impossible de donner ces probabilités quand le nombre de parents devient grand. En pratique, un nombre de parents supérieur à trois est déjà problématique.

Règle 2 : Nous proposons donc de demander uniquement les probabilités conditionnelles du premier ordre. Par exemple considérons un réseau bayésien, avec trois noeuds parents A, B, C et un noeud fils D. Toutes les variables sont supposées discrètes et binaires. A, B et C sont indépendants, mais non conditionnellement indépendants sachant D. Pour faire de l'inférence sur ce modèle, nous avons besoin de $p(A)$, $p(B)$, $p(C)$ et $p(D|ABC)$, soit 3 probabilités marginales et 8 probabilités conditionnelles. Notre méthodologie consiste à demander aux experts les 4 probabilités marginales, à savoir $p(A)$, $p(B)$, $p(C)$ et $p(D)$, et les probabilités conditionnelles suivantes : $p(D|A), p(D|\bar{A}), p(D|B), p(D|\bar{B}), p(D|C), p(D|\bar{C})$. Ces probabilités sont les plus faciles à donner pour les experts. En demandant ces probabilités, on peut remarquer qu'il y a plusieurs redondances. Cette méthode va ainsi nous permettre de pointer les grandes incohérences qui peuvent exister dans les dires des experts. En effet, par sommation on obtient

$$p(D) = \sum_A p(D|A)p(A) = \sum_B p(D|B)p(B) = \sum_C p(D|C)p(C). \tag{2.1}$$

En réalité, ces équations ne sont jamais strictement vérifiées. Le statisticien doit donc décider d'en éliminer. Pour l'exemple, il peut choisir d'éliminer trois probabilités conditionnelles. En gardant en mémoire que les probabilités conditionnelles sachant un "non-évènement" sont difficile à conceptualiser pour un expert, le statisticien éliminerait donc $p(D|\bar{A})$, $p(D|\bar{B})$ et $p(D|\bar{C})$. Cependant, cette démarche ne doit pas constituer une règle absolue. Nous proposons quelques règles qui peuvent permettre au statisticien de choisir les probabilités à éliminer.

Règle 3 : Cette règle consiste à retenir toutes les probabilités provenant des bases de données. En effet, il est considéré que le retour d'expérience est beaucoup plus fiable que les avis d'experts, ceci est encore plus vrai pour les probabilités conditionnelles. De plus, lorsqu'un retour d'expérience existe, les avis d'experts sont en général, basés sur ce retour d'expérience. Dans ce cas, la première étape calcule toutes les probabilités provenant des bases de données de retour d'expérience. Mais, seules quelques probabilités marginales peuvent être calculées de cette façon. En effet, lorsque le nombre de variables est important, il faut une grande quantité de données pour estimer les probabilités conditionnelles.

Règle 4 : Cette règle permet de favoriser les probabilités marginales données par les experts, en considérant que celles-ci sont plus faciles à évaluer, notamment pour les experts non statisticiens. Ainsi, il reste à choisir une des probabilités conditionnelles par nœud parent.

Prenons comme exemple le réseau bayésien avec trois variables présenté figure 2.2. Une

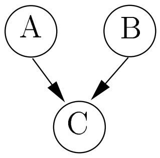


FIG. 2.2 – Un réseau bayésien avec deux parents.

stratégie peut consister à demander en quelles probabilités l'expert a le plus confiance. Cela étant, la plupart du temps, le choix est très restreint. Pour s'en convaincre, supposons que pour le réseau bayésien figure 2.2 avec A une variable à trois modalités, l'expert ait donné comme probabilités celles figurant dans le tableau 2.1.

$P(C) = 0.25$	
$P(C\|A = 0) = 0.05$	$P(A = 0) = 0.33$
$P(C\|A = 1) = 0.25$	$P(A = 1) = 0.66$
$P(C\|A = 2) = 0.30$	$P(A = 2) = 0.01$

TAB. 2.1 – Exemple de probabilités données par l'expert.

D'après le tableau 2.1, il est possible de calculer une valeur de la probabilité de C :

$$P_{calc}(C) = 0.05 * 0.33 + 0.25 * 0.66 + 0.30 * 0.01 = 0.183,$$

qui est différente de la valeur donnée par l'expert. Supposons maintenant que l'expert ait choisi de modifier la dernière probabilité conditionnelle. Alors, il est possible de recalculer cette probabilité :

$$P_{calc}(C|A = 2) = \frac{0.25 - 0.05 * 0.33 - 0.25 * .66}{0.01} = 6.85$$

Cette anomalie est dû au fait que l'expert change la probabilité conditionnelle qui a le poids le plus faible. Ici le poids correspond à la probabilité marginale du nœud parent. De plus, la probabilité conditionnelle avec le plus fort poids ($P(C|A = 1)$ avec un poids de $P(A = 1) = 0.66$) est égale à la probabilité marginale et l'autre probabilité conditionnelle $P(C|A = 0)$ avec un poids important $P(A = 0) = 0.33$ est assez faible. Ainsi, pour compenser la différence entre les deux probabilités marginales (entre 0.183 et 0.25), la probabilité conditionnelle est fortement augmentée et dépasse le seuil de 1.

Règle 5 : Il est préférable pour ces raisons, de changer les probabilités conditionnelles qui ont le plus de poids, si celles-ci sont très différentes de la probabilité marginale.

Dans l'exemple précédent, l'application de la règle 5 donne :

$$P_{calc}(C|A = 1) = \frac{0.25 - 0.05 * 0.33 - 0.30 * 0.01}{0.66} = 0.3492$$

qui remplace l'ancienne valeur de 0.25. Il est à noter que l'ordre des trois probabilités conditionnelles n'est plus respecté. Ce point peut paraître gênant, c'est pourquoi nous proposons plus loin dans ce paragraphe, une procédure prenant en compte les ratios entre les probabilités conditionnelles et donc en gardant l'ordre de ces probabilités. Si l'on avait changé la première probabilité conditionnelle $P(C|A=0)$, on aurait trouvé :

$$P_{calc}(C|A=0) = \frac{0.25 - 0.25 * 0.66 - 0.30 * 0.01}{0.33} = 0.2485,$$

ce qui élimine le facteur multiplicatif qui existait entre les probabilités conditionnelles. c'est pour cette raison que l'on préférera la première solution.
Supposons maintenant que l'expert donne les probabilités comme indiqué dans le tableau 2.2.

$P(C) = 0.05$	
$P(C\|A=0) = 0.01$	$P(A=0) = 0.33$
$P(C\|A=1) = 0.03$	$P(A=1) = 0.66$
$P(C\|A=2) = 0.05$	$P(A=2) = 0.01$
$P(C\|B=0) = 0.10$	$P(B=0) = 0.10$
$P(C\|B=1) = 0.03$	$P(B=1) = 0.90$

TAB. 2.2 – Un autre exemple de probabilités données par l'expert.

La probabilité marginale peut être vue comme une combinaison convexe des probabilités conditionnelles, où les poids sont les probabilités marginales des nœuds parents. Ainsi, dans la quatrième ligne du tableau 2.2, la probabilité conditionnelle est égale à la probabilité marginale, $P(C|A=2) = P(C)$. De plus, les deux autres probabilités conditionnelles ($P(C|A=0)$ et $P(C|A=1)$ sont strictement inférieures à la probabilité marginale. Ainsi, le poids de la probabilité conditionnelle $P(C|A=2)$, à savoir $P(A=2)$ doit être égale à 1. Dans ce cas particulier, nous proposons de changer la probabilité marginale de C, qui plus est, sera proche de la probabilité donnée. On calcule $P(C)$ via les deux nœuds parents :

$$\begin{aligned} P_{calc}^{A}(C) &= 0.029, \\ P_{calc}^{B}(C) &= 0.037. \end{aligned}$$

La première probabilité n'est pas une combinaison convexe des probabilités conditionnelles de C sachant B. Ainsi, la deuxième probabilité calculée est préférée, et nous appliquons les mêmes règles que précédemment afin de changer une des probabilités conditionnelles de C sachant A.

Enfin, nous présentons un dernier cas avec les probabilités suivantes données dans le tableau 2.3.
La probabilité marginale calculée est égale à

$$P_{calc}(C) = 0.05 * 0.33 + 0.175 * 0.66 + 0.25 * 0.01 = 0.1345.$$

$P(C) = 0.233$	
$P(C\|A=0) = 0.05$	$P(A=0) = 0.33$
$P(C\|A=1) = 0.175$	$P(A=1) = 0.66$
$P(C\|A=2) = 0.25$	$P(A=2) = 0.01$

TAB. 2.3 – Exemple de probabilités données par l'expert dans le cas où l'ordre est changé.

Cette probabilité est assez différente de la probabilité donnée, car la probabilité conditionnelle avec un gros poids est assez faible comparée à la probabilité marginale. On calcule respectivement les probabilités conditionnelles en fixant les autres :

$$\begin{aligned} P(C|A=0) &= \frac{0.233 - 0.175 * 0.66 - 0.25 * 0.01}{0.33} = 0.3485, \\ P(C|A=1) &= \frac{0.233 - 0.05 * 0.33 - 0.25 * 0.01}{0.66} = 0.3242, \\ P(C|A=2) &= \frac{0.233 - 0.05 * 0.33 - 0.175 * 0.66}{0.01} = 10.1. \end{aligned}$$

Si l'on veut garder l'ordre des trois probabilités conditionnelles, il est nécessaire de changer deux probabilités conditionnelles en essayant de garder les facteurs multiplicatifs entre ces probabilités. Par exemple, si on garde la première probabilité conditionnelle et que l'on fixe $P(C|A=2) = 0.30$, cela donne :

$$P(C|A=1) = \frac{0.233 - 0.05 * 0.33 - 0.30 * 0.01}{0.66} = 0.2607$$

On peut remarquer que dans ce cas, il n'est pas possible de garder simultanément l'ordre $P(C|A=1) < P(C)$ et la valeur de la probabilité conditionnelle $P(C|A=0)$. Cependant, tous ces cas peuvent être traitée comme suit. Une solution peut consister à résoudre un programme linéaire classique. Dans le cas du dernier exemple donné dans le tableau 2.3, cela donne

$$\min_{0 \leq x \leq 1} |P(C) - \sum_{i=1}^{3} k_i x P(A=i)|$$

où k_i représente les facteurs multiplicatifs donnés par les experts. Ainsi si on choisi $k_0 = 1$, on a $k_1 = 3.5$ et $k_2 = 5$. Ce problème a pour solution :

$$P(C|A=0) = \frac{0.233}{0.33 + 3.5 * 0.66 + 5 * 0.01} = 0.0866,$$

et donc $P(C|A=1) = 0.3032$ et $P(C|A=2) = 0.4331$. Puis les nouvelles probabilités proposées sont soumises aux experts, pour leur accord ou désaccord.
On peut résumer la méthodologie en donnant les règles suivantes :

i) garder toutes les probabilités provenant des bases de données de retour d'expérience,

ii) choisir les experts et préparer un questionnaire afin de réaliser un interview pour chaque expert,
iii) avoir la plus grande confiance pour les probabilités marginales plutôt que les probabilités conditionnelles, ces dernières étant plus difficile à donner,
iv) vérifier si la probabilité marginale calculée est égale à la probabilité donnée et plus précisément vérifier si la probabilité marginale donnée est une combinaison convexe des probabilités conditionnelles,
v) garder l'ordre des probabilités conditionnelles. Pour cela, il existe deux solutions :
  * changer un minimum de probabilité comme dans l'exemple précédent,
  * résoudre un programme linéaire en prenant en compte les facteurs multiplicatifs entre les probabilités conditionnelles.
vi) Et finalement, proposer les solutions aux experts, pour mener à une discussion ouverte dans le but final de valider le modèle.

Mais, tout ce que l'on vient de décrire ne suffit pas pour connaître la probabilité jointe. Ainsi, une hypothèse de simplification doit être faite. Pour cette étude, nous avons opté pour l'utilisation des **modèles log-linéaires** non saturés pour réduire l'inflation des probabilités à donner. Cela permet en particulier d'expliciter les hypothèses faites sur les indépendances conditionnelles et donc de pouvoir les relâcher simplement. En première approximation, seules les interactions d'ordre inférieur à deux ont été considérées. Cette hypothèse revient à supposer des indépendances conditionnelles. Une inférence a été ensuite réalisée avec ces données. Puis, au vu des résultats de l'analyse, certaines de ces hypothèses ont été relâchées. Plus précisément, des termes d'interaction d'ordre 2 ont été ajoutés par les experts, conduisant à un nouveau questionnaire, complétant le précédent. L'intérêt de ce point de vue est de préserver la structure du réseau bayésien et de pouvoir contrôler la nature des hypothèses simplificatrices faites. Nous présentons dans la section suivante les modèles log-linéaires et l'utilisation que nous en avons faites.

2.3.2.2 Approche des réseaux bayésiens par des modèles log-linéaires

Les modèles log-linéaires ont fait l'objet de multiples publications. On peut citer par exemple Goodman [51], Bishop et al. [19], Christensen [31]. Ces modèles sont une généralisation des modèles de Bernoulli. Ils permettent d'analyser les liaisons entre les variables qualitatives à plusieurs modalités. Ces modèles ont pour avantages d'être flexibles, de pouvoir s'associer facilement avec une analyse de variance ou une régression, et surtout d'avoir une interprétation en termes d'indépendance. Les modèles log-linéaires ont une utilisation très simple dans le cas de tableau de contingence. Mais, plus le nombre de facteurs (variables) est grand et plus les modèles log-linéaires sont complexes. En pratique, un nombre de facteur supérieur à trois pose de réelles difficultés. Il existe deux types de modèles log-linéaires : les modèles log-linéaires saturés et non saturés. Les modèles saturés épousent parfaitement les données, mais ne produisent aucune simplification. Les modèles non saturés résultent d'hypothèses d'indépendance et sont donc plus facile à interpréter. Le meilleur modèle est alors le plus petit modèle qui ajuste bien les données.

Prenons comme exemple, un modèle avec quatre variables discrètes binaires, A, B, C et D. On peut écrire le logarithme de la probabilité jointe comme :

$$\begin{aligned} log(p(a,b,c,d)) &= u + u_A(a) + u_B(b) + u_C(c) + u_D(d) \\ &+ u_{AB}(a,b) + u_{AC}(a,c) + u_{AD}(a,d) + u_{BC}(b,c) + u_{BD}(b,d) + u_{CD}(c,d) \\ &+ u_{ABC}(a,b,c) + u_{ABD}(a,b,d) + u_{BCD}(b,c,d) \\ &+ u_{ABCD}(a,b,c,d), \end{aligned} \quad (2.2)$$

où les termes d'interaction u, appelés u-termes, sont des fonctions des modalités des variables. La première ligne de l'équation 2.2 correspond aux termes d'effet principaux. La deuxième correspond aux termes d'interaction d'ordre 1, la troisième aux termes d'interaction d'ordre 2, et la dernière aux termes d'interaction d'ordre 3. Dans ce cas, le modèle log-linéaire est dit saturé. Ce n'est donc pas une approximation. Dans cette écriture, la plupart des termes sont redondants et donc inutiles. Seul le terme d'interaction d'ordre 3 est utile, et pour cette raison, on notera ce modèle $[ABCD]$.
Un modèle log-linéaire non saturé pour cet exemple pourrait être :

$$\begin{aligned} log(p(a,b,c,d)) &= u + u_A(a) + u_B(b) + u_C(c) + u_D(d) \\ &+ u_{AB}(a,b) + u_{AC}(a,c) + u_{BC}(b,c) + u_{ABC}(a,b,c) \\ &+ u_{AD}(a,d) + u_{BD}(b,d). \end{aligned} \quad (2.3)$$

En éliminant les termes redondants, cela donne

$$log(p(a,b,c,d)) = u_{ABC}(a,b,c) + u_{AD}(a,d) + u_{BD}(b,d). \quad (2.4)$$

On notera ce modèle $[ABC][AD][BD]$. Dans ce modèle, C et D sont conditionnellement indépendants sachant A et B. Un modèle log-linéaire non saturé sous-entend des indépendances conditionnelles. Or, nous avons vu dans le premier chapitre que les réseaux bayésiens sont eux aussi des modèles de représentation des indépendances conditionnelles. Nous donnons maintenant la définition d'un modèle log-linéaire graphique (cf. Christensen [31]).

Définition 12 *Un modèle log linéaire peut être représenté par un réseau bayésien si, lorsque le modèle contient tous les termes d'interaction d'ordre 2 générés par un terme d'interaction plus grande, alors le modèle contient cette interaction d'ordre plus élevé.*

Par exemple, le modèle log-linéaire $[ABC][AD][BD]$ peut être représenté par le réseau bayésien de la figure 2.3. Cette figure nous montre l'indépendance conditionnelle de A et B sachant D, mais également l'indépendance conditionnelle de C et D sachant A et B.
Un exemple où le modèle log-linéaire n'est pas représentable par un réseau bayésien est le modèle log-linéaire $[ABC][AD][BD][AB]$. En effet, il manque le terme u_{ABD} traduisant la dépendance conditionnelle entre ces trois variables.

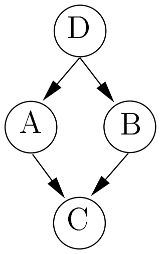


Fig. 2.3 – Représentation graphique du modèle log-linéaire [ABC][AD][BD].

En revanche, la réciproque est toujours vraie : on peut toujours écrire un réseau bayésien comme un modèle log-linéaire. Nous donnons quelques exemples simples de réseaux bayésiens et leurs écriture sous forme de modèle log-linéaires.
Prenons pour exemple, un réseau bayésien simple avec trois variables binaires A, B, C :

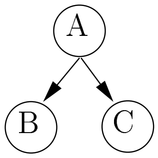


Fig. 2.4 – Un réseau bayésien simple.

La probabilité peut s'écrire

$$\log f_{ABC}(x) = u_\phi + u_A a + u_B b + u_C c + u_{AB} ab + u_{AC} ac + u_{BC} bc + u_{ABC} abc \tag{2.5}$$

où a, b et c prennent comme valeurs 0 ou 1. Si les variables ne sont pas binaires, l'écriture de la probabilité devient

$$\log p_{ABC}(x) = u_\phi + u_A(x) + u_B(x) + u_C(x) + u_{AB}(x) + u_{AC}(x) + u_{BC}(x) + u_{ABC}(x) \tag{2.6}$$

où les termes en u ne dépendent que des variables indiquées en indice. Dans le cas de variables binaires, les probabilités jointes peuvent s'écrire :

$$\begin{cases} \log p(0,0,0) = u_\phi \\ \log p(0,0,1) = u_\phi + u_C \\ \log p(0,1,0) = u_\phi + u_B \\ \log p(0,1,1) = u_\phi + u_B + u_C + u_{BC} \\ \log p(1,0,0) = u_\phi + u_A \\ \log p(1,0,1) = u_\phi + u_A + u_C + u_{AC} \\ \log p(1,1,0) = u_\phi + u_A + u_B + u_{AB} \\ \log p(1,1,1) = u_\phi + u_A + u_B + u_C + u_{AB} + u_{AC} + u_{BC} + u_{ABC}. \end{cases}$$

De la sorte le triplet $(0,0,0)$ joue le rôle de valeur de référence. Dans le réseau bayésien de la figure 2.4, B et C sont conditionnellement indépendants sachant A. Cela implique que $u_{BC} = u_{ABC} = 0$ (voir [98]). Les probabilités jointes peuvent se réécrire

$$\begin{cases} \log p(0,0,0) = u_\phi \\ \log p(0,0,1) = u_\phi + u_C \\ \log p(0,1,0) = u_\phi + u_B \\ \log p(0,1,1) = u_\phi + u_B + u_C \\ \log p(1,0,0) = u_\phi + u_A \\ \log p(1,0,1) = u_\phi + u_A + u_C + u_{AC} \\ \log p(1,1,0) = u_\phi + u_A + u_B + u_{AB} \\ \log p(1,1,1) = u_\phi + u_A + u_B + u_C + u_{AB} + u_{AC}. \end{cases}$$

C'est un système de huit équations à six inconnues, dont une est déterminée par les autres variables (la somme des probabilités étant égale à 1). On peut ainsi déterminer toutes les probabilités a partir de quatre probabilités conditionnelles et d'une probabilité a priori.

$$\begin{cases} p(B=0|A=0) = \dfrac{1}{1+e^{u_B}} \\[2ex] p(C=0|A=0) = \dfrac{1}{1+e^{u_C}} \\[2ex] p(B=0|A=1) = \dfrac{1}{1+e^{u_B}e^{u_{AB}}} \\[2ex] p(C=0|A=1) = \dfrac{1}{1+e^{u_C}e^{u_{AC}}} \\[2ex] p(A=0) = e^{u\phi}\left(1+e^{u_B}\right)\left(1+e^{u_C}\right). \end{cases}$$

On peut donc à partir de ces probabilités, calculer u_B, u_C, u_{AB}, u_{AC}, u_B et $u\phi$. On déduira u_A, par l'équation qui assure que la somme de toutes les probabilités est égale à 1. Dans le cas de variables binaires, on peut également écrire le système précédent de la façon suivante :

$$\begin{aligned} \log p(A=a, B=b, C=c) &= u_\phi + (-1)^a u_A + (-1)^b u_B + (-1)^c u_C \\ &+ (-1)^{a+b} u_{AB} + (-1)^{a+c} u_{AC} + (-1)^{b+c} u_{BC} \\ &+ (-1)^{a+b+c} u_{ABC} \end{aligned}$$

où a, b et c prennent comme valeurs 0 ou 1. Dans cette écriture, u_ϕ représente l'effet moyen et non celui de l'état $(0,0,0)$

$$\begin{aligned} u_\phi &= \frac{1}{8}\sum_{a,b,c}\log\{p(A=a,B=b,C=c)\} \\ &= \frac{1}{8}\log\left\{\prod_{a,b,c}p(A=a,B=b,C=c)\right\}. \end{aligned}$$

Dans la suite, l'écriture avec l'effet moyen a été choisi.

Le nombre de probabilités à évaluer pour une variable est fonction exponentielle du nombre de parents de cette même variable. Ainsi, un nœud avec beaucoup de parents requiert une grande quantité de probabilités à déterminer par expertise. Soit un réseau bayésien avec deux parents (cf. figure 2.2). Supposons que A, B et C soient des variables binaires. On peut écrire de la même façon les probabilités jointes

$$\begin{cases} \log p(0,0,0) = u_\phi \\ \log p(0,0,1) = u_\phi + u_C \\ \log p(0,1,0) = u_\phi + u_B \\ \log p(0,1,1) = u_\phi + u_B + u_C + u_{BC} \\ \log p(1,0,0) = u_\phi + u_A \\ \log p(1,0,1) = u_\phi + u_A + u_C + u_{AC} \\ \log p(1,1,0) = u_\phi + u_A + u_B + u_{AB} \\ \log p(1,1,1) = u_\phi + u_A + u_B + u_C + u_{AB} + u_{AC} + u_{BC} + u_{ABC}. \end{cases}$$

Dans ce cas, on ne peut plus nécessairement dire que A et B sont conditionnellement indépendants sachant C. En revanche, A et B sont indépendants, ce qui implique que $u_{AB} = 0$. Soit

$$\begin{cases} \log p(0,0,0) = u_\phi \\ \log p(0,0,1) = u_\phi + u_C \\ \log p(0,1,0) = u_\phi + u_B \\ \log p(0,1,1) = u_\phi + u_B + u_C + u_{BC} \\ \log p(1,0,0) = u_\phi + u_A \\ \log p(1,0,1) = u_\phi + u_A + u_C + u_{AC} \\ \log p(1,1,0) = u_\phi + u_A + u_B \\ \log p(1,1,1) = u_\phi + u_A + u_B + u_C + u_{AC} + u_{BC} + u_{ABC}. \end{cases}$$

On peut déduire u_A et u_B par :

$$u_A = \log\left[\frac{p(C=0|A=1)p(A=1)}{p(C=0|A=0)p(A=0)}\right] \tag{2.7}$$

et

$$u_B = \log\left[\frac{p(C=0|B=1)p(B=1)}{p(C=0|B=0)p(B=0)}\right]. \tag{2.8}$$

Si l'on suppose que $u_{ABC} = 0$, le système d'équation devient alors inversible. Dans le cas contraire, on peut déduire u_{ABC} par l'équation $\sum_{ABC} P(A, B, C) = 1$.

Soit maintenant un réseau simple avec n parents. Pour ce modèle comportant un noeud, noté B avec n parents, notés $A_1, \ldots, A_n$, il faut définir, dans le cas de variables binaires que nous considérons ici, 2^n probabilités conditionnelles et n probabilités a priori ! Pour $n \geq 4$, on comprend aisément que la tâche de l'expert devient complexe. L'idée est donc de se ramener à un modèle log-linéaire simple. Le modèle log-linéaire saturé conduit à écrire un système de 2^{n+1} équations à 2^{n+1} variables. La somme de ces équations étant égale à 1, il reste $2^{n+1}-1$ variables à déterminer. On fait l'hypothèse simplificatrice que les noeuds parents sont indépendants. Cela entraîne que les termes $u_{A_iA_j}$ où $i = 1, \ldots, n-1$ et $j \neq i$ sont tous nuls. Ainsi, $2^n - n - 1$ variables peuvent être éliminées. Il reste donc $2^{n+1} - 1 - (2^n - n - 1) = 2^n + n$ variables à déterminer. Pour n grand, cela conduit déjà à une réduction notable de complexité. Les experts peuvent donner facilement les n probabilités a priori et les $2n$ probabilités conditionnelles, respectivement $p(A_1), \ldots, p(A_n)$ et $p(B|A_1), p(B|\bar{A}_1)), \ldots, p(B|A_n), p(B|\bar{A}_n))$, où $\bar{A}_i = non(A_i)$. Ce qui donne un système de $3n$ équations à $2^n + n$ variables. Il nous faut donc rajouter $2^n - 2n$ contraintes ou équations. Le tableau 2.4 donne le nombre de contraintes ou équations nécessaires pour la résolution du système d'équations.

n	Nbre de contraintes
2	0
3	2
4	8
5	22
6	52

TAB. 2.4 – Nombre de contraintes en fonction du nombre de parents.

Dans la suite, nous présentons un exemple de prise d'information et la façon de l'intégrer dans le modèle log-linéaire.
Considérons un réseau bayésien, avec trois noeuds parents A, B, C et un noeud fils D. A, B et C sont indépendants, mais non conditionnellement indépendant sachant D. Pour faire de l'inférence sur ce modèle, nous avons besoin de $p(A)$, $p(B)$, $p(C)$ et $p(D|ABC)$, soit 11 probabilités. En écrivant le modèle log-linéaire, on obtient un système de 16 équations à 16 inconnues. Sous l'hypothèse d'indépendance des parents, on a $u_{AB} = u_{AC} = u_{BC} = u_{ABC} = 0$, il reste donc 12 variables. Sachant que la somme des probabilités est égale à 1, on peut déterminer u_{ABCD} en fonction des autres variables. Il reste donc 11 variables. L'expert nous donne $p(A)$, $p(B)$, $p(C)$, $p(D|A)$, $p(D|\bar{A})$, $p(D|B)$, $p(D|\bar{B})$ et $p(D|C)$, $p(D|\bar{C})$, soit 9 probabilités ! Il reste 2 contraintes à mettre en place (voir tableau 2.4 avec $n = 3$). Par exemple, l'expert peut affirmer que la probabilité que l'on observe la variable D et C, connaissant A et B est la même que si l'on observe D et $\neg C$, sachant A et B. En d'autres termes, lorsque l'on observe A et B, la valeur de D ne dépend pas de la valeur de C. L'effet C est négligeable

devant les effets cumulés de A et B. Cela nous donne,

$$\log p(A = 1, B = 1, C = 1, D = 1) = \log p(A = 1, B = 1, C = 0, D = 1).$$

D'où

$$u_{CD} + u_{ACD} + u_{BCD} + u_{ABCD} = 0,$$

Cet exemple illustre la proposition de Whittaker [98] page 207, à savoir que si C et D sont conditionnellement indépendants sachant A et B, alors tous les termes d'interaction avec comme indice les variables C et D sont tous nuls. L'intérêt des modèles log-linéaires est l'équivalence entre les relations d'indépendances et la nullité des termes d'interaction. Ces modèles ont l'avantage d'être hiérarchiques. Il est possible de maîtriser la complexité du modèle. Notre approche consiste à supposer beaucoup d'indépendances conditionnelles, de le transcrire via le modèle log-linéaire, puis quand cela est nécessaire de considérer des interactions d'ordre plus élevés.

2.3.3 Conclusions

Pour résumer la méthodologie proposée, les experts doivent construire la structure du graphe, puis donner toutes les probabilités marginales, et enfin toutes les probabilités conditionnelles du premier ordre. En conséquence, pour obtenir un système d'équation qui soit résolvable, nous ajouterons des contraintes sur les termes d'interaction d'ordre du plus élevé au plus faible. En pratique, il est difficile de tenir compte des interactions supérieures ou égales à 2 (voir [28]). Cependant, si les experts décident de conserver des interactions d'ordre supérieur, celles-ci seront ajoutées au modèle log-linéaire. Puis, les probabilités nécessaires pour la résolution du système d'équations devront être données par les mêmes experts.

2.4 Application

La méthodologie précédemment décrite a été appliquée sur un système de centrale nucléaire.

2.4.1 Construction de la structure du réseau bayésien

2.4.1.1 Choix du système, des experts et du médiateur

Dans notre cas, le médiateur a été à l'origine de l'étude. Ainsi, avec l'accord des statisticiens, le médiateur a choisi le système et les experts. Pour la construction de la structure, le nombre d'experts choisis est de quatre et le système choisi est le joint 1 d'une pompe primaire 900 MW. La pompe primaire possède trois joints successifs. Le joint 1 est le premier et est donc celui qui subit le plus de pression et est essentiel à la survie du système. Un schéma du joint 1 est donné en annexe (cf. figure 6.2).

2.4.1.2 Les variables du modèle

Les premières réunions entre tous les acteurs ont permis d'identifier les variables, ainsi que leur groupe d'appartenance. Les variables d'entrée sont :

- X_1 = **Ad** : âge de la douille (**trois modalités : < 1 an, entre 1 et 6 ans, > 6 ans**),
- X_2 = **Ag** : âge de la glace (**trois modalités : < 1 an, entre 1 et 6 ans, > 6 ans**),
- X_3 = **Ab** : âge de la bague (**deux modalités : < 1 an, entre 1 et 3 ans**),
- X_4 = **PI2** : valeur du débit au démarrage à 25 bars sur la période de montage du joint (**deux modalités : faible, élevé**),
- X_5 = **PI3** : présence d'impuretés dans le circuit RCV sur la période de montage du joint (**deux modalités : oui, non**),
- X_6 = **PI4** : vibrations et déplacement d'arbre sur la période de montage du joint (**deux modalités : oui, non**),
- X_7 = **PI6** : températures du palier, nombre d'excursion autour du niveau moyen avec une amplitude > 10 degrés sur la période de montage du joint (**deux modalités : faible, importante**),
- X_8 = **DI** : débits inverses joint 1 en arrêt de tranche (**deux modalités : oui, non**),
- X_9 = **DJ** : démontage joints pendant l'arrêt de tranche précédant le cycle actuel (**deux modalités : oui, non**).

Les variables intermédiaires sont :

- X_{10} = **M1**$'$: éclats, fissures des faces actives des glaces (**deux modalités : oui, non**),
- X_{11} = **M1**$''$: traces de frottement sur les faces actives des glaces (**deux modalités : oui, non**),
- X_{12} = **M2** : modification des profils des glaces par usure (**deux modalités : oui, non**),
- X_{13} = **M3** : dégradation de la douille logement par usure et/ou rayures (**deux modalités : oui, non**),
- X_{14} = **M4** : dégradation de l'étanchéité secondaire de la bague de glissement par extrusion (**deux modalités : oui, non**),
- X_{15} = **M5** : mauvais positionnement des glaces par rapport à leurs supports en fonctionnement par impuretés ou défauts de surface (**deux modalités : oui, non**),
- X_{16} = **M6** : coulissement difficile de la bague (**deux modalités : oui, non**),
- X_{17} = **O1** : niveau moyen du débit de fuite (**trois modalités : faible, moyen, élevé**),
- X_{18} = **O5** : sensibilité anormale du joint à la mise en pression du circuit primaire (**deux modalités : oui, non**),
- X_{19} = **O2** : allure du débit de fuite au cours des différents cycles de montage du joint (**deux modalités : stable, non stable**),
- X_{21} = **O2**$'$: plage de variation du débit de fuite (**deux modalités : faible, élevée**),
- X_{20} = **O2**$''$: régularité du débit de fuite (**trois modalités : monotone décroissante, monotone croissante, irrégulière**).

La variable de sortie ou variable d'intérêt est :

- X_{22} = **E** : état (**trois modalités : sain, dégradé, défaillant**).

Ce modèle comporte donc 9 nœuds racines (nœuds d'entrée), 12 nœuds intermédiaires et 1 nœud terminal (variable de sortie), soit un total de **22** variables. Toutes les variables sont discrètes, 17 étant binaires, les 5 autres ayant trois modalités.

2.4.1.3 Structure du graphe

Pour déterminer la structure du graphe, les variables ont été divisées en sous-groupes. Les variables $\mathbf{PI2}, \mathbf{PI3}, \mathbf{PI4}, \mathbf{PI6}$ font partie du groupe des "paramètres d'influence". Les variables $\mathbf{Ab}, \mathbf{Ag}, \mathbf{Ad}$ sont dans le groupe "âge des composants". Les variables $\mathbf{DI}, \mathbf{DJ}$ font partie du groupe "paramètres d'exploitation". Les variables $\mathbf{O1}, \mathbf{O5}, \mathbf{O2}, \mathbf{O2}', \mathbf{O2}''$ sont dans le groupes des "paramètres d'observations". Les variables $\mathbf{M1}', \mathbf{M1}'', \mathbf{M2}, \mathbf{M3}, \mathbf{M4}, \mathbf{M5}, \mathbf{M6}$ sont dans le groupe "maladies", et enfin $\mathbf{E}$ est un groupe à lui tout seul, "état du système". Ces groupes sont en relation, comme l'indique la figure 2.5.

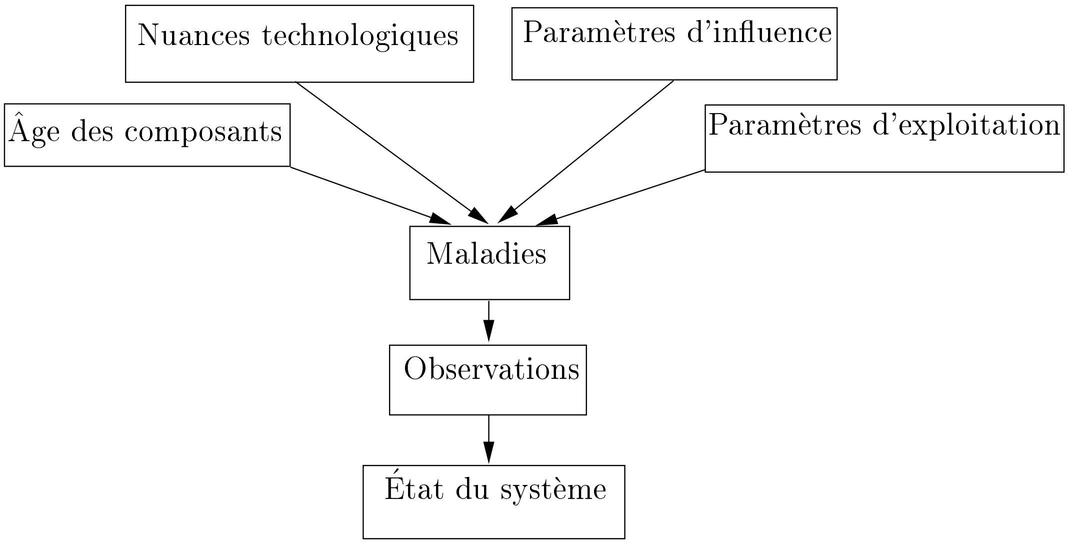


FIG. 2.5 – Regroupement des variables et représentation de la causalité.

2.4.1.4 Le réseau bayésien

Nous présentons dans ce paragraphe le réseau bayésien du joint 1 d'une pompe 900 MW, décrit dans la figure 6.6, accepté lors de la réunion du 15/02/2001 (cf. [25] et l'annexe F pour les multiples étapes de la construction du réseau bayésien).

2.4.2 Évaluation des probabilités

2.4.2.1 Loi jointe

On peut écrire la loi jointe comme le produit des probabilités des variables sachant leurs parents :

$$\begin{aligned} P(U) &= P(Ab)P(Ad)P(Ag)P(PI2)P(PI3)P(PI4)P(PI6)P(DI)P(DJ) \\ &\times P(M1^{'}|Ag,DJ)P(M1^{''}|DJ,PI2)P(M2|Ag,PI3)P(M3|Ad,PI3) \\ &\times P(M4|Ab,PI4,PI6)P(M5|DI,PI3)P(M6|Ad,Ab) \\ &\times P(O1|M1^{''},M4,M5,M6)P(O5|M3,M4,M5,M6)P(O2|M5) \\ &\times p(O2^{''}|M2,M3,M4,M6,O2)P(O2^{'}|M1^{'},M1^{''},M2,M3,M4,M6,O2^{''}) \\ &\times p(E|O1,O5,O2^{'}) \end{aligned}$$

Dans cette équation, afin de connaître la probabilité jointe $P(U)$, il est nécessaire d'évaluer entre autres la probabilité conditionnelle $P(O2^{'}|M1^{'},M1^{''},M2,M3,M4,M6,O2^{''})$, ce qui nécessite de fournir 192 probabilités conditionnelles, et en tout plus de 400 probabilités.

2.4.2.2 Approche par les modèles log-linéaires

Nous considérons ici le modèle log-linéaire saturé. Soit $X_1,\dots,X_J$, les variables du réseau bayésien du joint 1. On note $1,\dots,M_j$ les modalités de la variable X_j, et m_j la valeur prise par X_j. Ces variables sont binaires, à l'exception de $X_1, X_2, X_{17}, X_{20}, X_{22}$. Notons $I = \{1,2,17,21,22\}$, l'ensemble des indices tel que $X_{i\in I}$ ne soit pas une variable binaire. La probabilité jointe, tenant compte de toutes les interactions possibles (modèle saturé) s'écrit (cf. [28]) :

$$\begin{aligned} \ln P_{m_1,\dots,m_J} &= \ln P(X_1 = m_1,\dots,X_J = m_J) \\ &= \alpha \qquad \text{constante} \\ &+ \sum_{j=1}^{J} \alpha_j(m_j) \qquad \text{effets principaux} \\ &+ \sum_{j<l=1}^{J} \alpha_{jl}(m_j,m_l) \quad \text{interactions d'ordre 1} \\ &+ \vdots \\ &+ \alpha_{1\dots J}(m_1,\dots,m_J) \quad \text{interactions d'ordre J-1} \end{aligned}$$

avec les contraintes suivantes sur les paramètres :

$$\sum_{m_j=1}^{M_j} \alpha_j(m_j) = 0 \qquad \text{pour tout } j = 1, \ldots, J$$

$$\left\{ \begin{array}{l} \displaystyle\sum_{m_j=1}^{M_j} \alpha_{jl}(m_j, m_l) = 0 \\ \\ \displaystyle\sum_{m_l=1}^{M_l} \alpha_{jl}(m_j, m_l) = 0 \end{array} \right. \qquad \text{pour tous les termes d'interaction d'ordre 1 } (1 \leq j < l \leq J)$$

$$\vdots$$

$$\sum_{m_j=1}^{M_j} \alpha_{1\ldots J}(m_1, \ldots, m_J) = 0 \qquad \text{pour tout } j = 1, \ldots, J$$

Pour les variables binaires, ces contraintes deviennent :

$$\alpha_j(m_j) = (-1)^{m_j} u_j \quad \text{pour tout } j \in J \setminus I.$$

et

$$\alpha_{jl}(m_j, m_l) = (-1)^{m_j+m_l} u_{jl} \quad \text{pour tout } j, l \in J \setminus I.$$

et ainsi de suite.
Pour des variables non binaires, une seule variable ne suffit pas. Si on considère les interactions d'ordre 1 (α_{jl}) où $j \in I$ et $l \in J \setminus I$. Les contraintes sont

$$\begin{array}{rcl} \alpha_{jl}(0,0) + \alpha_{jl}(0,1) & = & 0 \\ \alpha_{jl}(1,0) + \alpha_{jl}(1,1) & = & 0 \\ \alpha_{jl}(2,0) + \alpha_{jl}(2,1) & = & 0 \\ \alpha_{jl}(0,0) + \alpha_{jl}(1,0) + \alpha_{jl}(2,0) & = & 0 \\ \alpha_{jl}(0,1) + \alpha_{jl}(1,1) + \alpha_{jl}(2,1) & = & 0. \end{array}$$

Soit en posant $\alpha_{jl}(0,0) = u_{jl}$ et $\alpha_{jl}(1,0) = v_{jl}$,

$$\alpha_{jl}(m_j, m_l) = (-1)^{m_l} \left((1 - m_j) u_{jl} + \frac{m_j}{2}(5 - 3m_j) v_{jl} \right).$$

Considérons maintenant le cas où X_l est une variable non binaire et X_j une variable binaire (soit $l \in I$ et $j \in J \setminus I$), on a par symétrie en posant $\alpha_{jl}(0,0) = u_{jl}$ et $\alpha_{jl}(0,1) = v_{jl}$

$$\alpha_{jl}(m_j, m_l) = (-1)^{m_j} \left((1 - m_l) u_{jl} + \frac{m_l}{2}(5 - 3m_l) v_{jl} \right).$$

Considérons maintenant, l'interaction d'ordre 1 (α_{jl}) entre deux variables non binaires (ici 3 modalités), c'est-à-dire où $j, l \in I$. Les contraintes deviennent :

$$\begin{aligned}
\alpha_{jl}(0,0)+\alpha_{jl}(0,1)+\alpha_{jl}(0,2) &= 0\\
\alpha_{jl}(1,0)+\alpha_{jl}(1,1)+\alpha_{jl}(1,2) &= 0\\
\alpha_{jl}(2,0)+\alpha_{jl}(2,1)+\alpha_{jl}(2,2) &= 0\\
\alpha_{jl}(0,0)+\alpha_{jl}(1,0)+\alpha_{jl}(2,0) &= 0\\
\alpha_{jl}(0,1)+\alpha_{jl}(1,1)+\alpha_{jl}(2,1) &= 0\\
\alpha_{jl}(0,2)+\alpha_{jl}(1,2)+\alpha_{jl}(2,2) &= 0.
\end{aligned}$$

Soit en posant $\alpha_{jl}(0,0)=u_{jl}$, $\alpha_{jl}(0,1)=u'_{jl}$, $\alpha_{jl}(1,0)=v_{jl}$ et $\alpha_{jl}(1,1)=v'_{jl}$,

$$\begin{aligned}
\alpha_{jl}(m_j,m_l) &= (1-m_l)\left\{(1-m_j)u_{jl}+\frac{m_j}{2}(5-3m_j)u'_{jl}\right\}\\
&+ \frac{m_l}{2}(5-3m_l)\left\{(1-m_j)v_{jl}+\frac{m_j}{2}(5-3m_j)v'_{jl}\right\}.
\end{aligned}$$

Afin de réduire le nombre de probabilités à donner, nous nous intéressons aux modèles log-linéaires non saturés. D'après le réseau bayésien, certaines variables sont d'une part indépendantes, et d'autre part conditionnellement indépendantes sachant une suite de variables. Ces propriétés sont faciles à déterminer connaissant la structure du réseau bayésien. Pour ce qui concerne le réseau bayésien du joint 1 d'une pompe primaire 900 MW, les variables d'entrée sont toutes mutuellement indépendantes. Ainsi, on peut en déduire (cf. [77])

$$\text{pour toute sous-suite d'indice } K\in[1,9] \text{ tel que card}K\geq 2, \alpha_K=0.$$

Comme le nombre de variables est ici important, nous allons uniquement considérer les interactions d'ordre 1 (cf. [98]). Pour connaître les dépendances conditionnelles entre les variables, il faut construire le graphe moral. Ce graphe s'obtient à partir de la figure 6.6 en reliant les parents de chaque nœud et en abandonnant les directions des arcs. Ainsi, on doit tenir compte des interactions d'ordre 1 pour tous les noeuds connectés entre eux :

- le terme constant u (1 terme),
- les facteurs principaux pour les variables binaires u_j (17 termes),
- les facteurs principaux pour les variables non binaires u_j et u'_j (10 termes),
- les interactions avec X_1 : $u_{1,13}$, $v_{1,13}$, $u_{1,16}$ et $v_{1,16}$ (4 termes),
- les interactions avec X_2 : $u_{2,10}$, $v_{2,10}$, $u_{2,12}$ et $v_{2,12}$ (4 termes),
- les interactions avec X_3 : $u_{3,14}$ et $u_{3,16}$ (2 termes),
- les interactions avec X_4 : $u_{4,11}$ (1 terme),
- les interactions avec X_5 : $u_{5,12}$, $u_{5,13}$ et $u_{5,15}$ (3 termes),
- les interactions avec X_6 : $u_{6,14}$ (1 terme),
- les interactions avec X_7 : $u_{7,14}$ (1 terme),

- les interactions avec X_8 : $u_{8,15}$ (1 terme),
- les interactions avec X_9 : $u_{9,10}$, $u_{9,11}$ (2 termes),
- les interactions avec X_{10} : $u_{10,21}$ (1 terme),
- les interactions avec X_{11} : $u_{11,17}$, $v_{11,17}$ et $u_{11,21}$ (3 termes),
- les interactions avec X_{12} : $u_{12,21}$, $u_{12,20}$ et $v_{12,20}$ (3 termes),
- les interactions avec X_{13} : $u_{13,18}$, $u_{13,21}$, $u_{13,20}$ et $v_{13,20}$ (4 termes),
- les interactions avec X_{14} : $u_{14,17}$, $v_{14,17}$, $u_{14,18}$, $u_{14,21}$, $u_{14,20}$ et $v_{14,20}$ (6 termes),
- les interactions avec X_{15} : $u_{15,17}$, $v_{15,17}$, $u_{15,18}$ et $u_{15,19}$ (4 termes),
- les interactions avec X_{16} : $u_{16,17}$, $v_{16,17}$, $u_{16,18}$, $u_{16,21}$, $u_{16,20}$ et $v_{16,20}$ (6 termes),
- les interactions avec X_{17} : $u_{17,20}$, $u'_{17,20}$, $v_{17,20}$ et $v'_{17,20}$ (4 termes),
- les interactions avec X_{18} : $u_{18,22}$ (1 terme),
- les interactions avec X_{19} : $u_{19,20}$, $v_{19,20}$ (2 termes),
- les interactions avec X_{20} : $u_{20,21}$, $v_{20,21}$ (2 termes),
- les interactions avec X_{21} : $u_{21,22}$ et $v_{21,22}$ (2 termes).

Soit un total de 85 termes. La somme des probabilités étant égale à 1, il reste 84 inconnues. Nous avons vu dans les sections précédentes que le nombre de probabilités nécessaire augmente très vite avec le nombre de parents. Le but est donc d'établir un questionnaire réaliste, par l'ajout de contraintes sur les termes u du modèle log-linéaire associé. Ainsi, le nombre de probabilités à demander aux experts est considérablement réduit.
Notre méthodologie consiste donc à demander aux experts les probabilités marginales des 22 probabilités (dont 9 sont des probabilités d'entrée). Cependant, si une valeur précise paraît difficile à donner pour l'expert, on choisira une échelle comme suit :

– $0,05$ pour quasi-impossible,
– $0,25$ pour peu probable,
– $0,5$ pour probable,
– $0,75$ pour quasi-probable,
– et enfin $0,95$ pour quasi-certain.

Pour connaître la loi jointe sur ce graphe, il nous faut :

– 6 probabilités marginales pour les trois variables d'entrée à trois modalités,
– 6 probabilités marginales pour les six autres variables d'entrée,
– 53 probabilités conditionnelles pour toutes les variables maladies,
– 338 probabilités conditionnelles pour toutes les variables observations,
– 24 probabilités conditionnelles pour la variable de sortie.

Soit un total de 427 probabilités à demander ! Les experts peuvent nous donner facilement **136** probabilités, 27 probabilités marginales, et 109 probabilités conditionnelles. Les 109 probabilités conditionnelles sont dites de premier ordre, c'est-à-dire qu'elles sont conditionnelles sachant un seul nœud parent. Nous appliquons ici un modèle linéaire non saturé en ne tenant compte que des interactions d'ordre inférieur à 2. Ainsi, le système peut être résolu et de plus comporte plus d'équations que d'inconnues. Le nombre plus important d'équations que de variables vient du fait des relations entre les probabilités. Par exemple,

$$p(A) = \sum_b p(A|B = b)p(B = b). \tag{2.9}$$

Ainsi, l'expert donne les probabilités qu'il souhaite, celles qu'il donne avec le plus d'aisance et pour lesquelles il possède plus de connaissances.

En reprenant l'exemple avec les quatre nœuds A, B, C qui pointent vers D, cela revient à demander aux experts les probabilités suivantes :

- les probabilités marginales $P(A)$, $P(B)$, $P(C)$ et $P(D)$,
- les probabilités conditionnelles d'ordre 1, $P(D|A)$, $P(D|\neg A)$, $P(D|B)$, $P(D|\neg B)$, $P(D|C)$, et $P(D|\neg C)$.

Ces probabilités avec l'hypothèse que tous les termes d'interaction d'ordre supérieur à deux sont nuls suffisent à calculer la probabilité jointe.

2.4.2.3 Mise en place du questionnaire

Un questionnaire de 136 questions a été soumis à deux experts du département SDM de la Division Recherche et Développement de EDF. Ces deux entretiens se sont déroulés sur deux demi-journées en tête-à-tête (le 21 mars 2001 avec Roger Chevalier et le 27 mars avec Benoît Ricard) et de façon indépendante. En effet, le deuxième expert interrogé ne connaissait pas les réponses du premier. Cependant on est en position de penser que les deux experts ont la même connaissance du matériel, et pis encore il se peut qu'un expert ait des connaissances via l'autre expert. Ainsi, il peut exister une corrélation entre les réponses des deux experts. Un troisième expert a été également interrogé. Ce troisième expert est Jean-Paul Miclot, qui a de plus une expérience du terrain, notamment une expérience d'exploitation et de maintenance sur un site de production nucléaire . Il a désiré souligner que ses souvenirs sur le joint ne sont pas récents (période de 1981 à 1985), et que cette période correspondait à une période de démarrage et de rodage, où il y avait donc beaucoup de défauts de jeunesse. La vision d'un expert venu du terrain est évidemment très profitable et très favorable. Dans toute la suite, les experts seront nommés expert RC, expert BR et expert JPM.

2.4.2.4 Retour vers les experts

Après une première inférence avec les hypothèses précédentes, les experts ont analysé les résultats. Les experts se sont aperçus que certaines dépendances manquaient dans le modèle. En effet, il est bien connu que certains effets combinés multiplient par exemple les risques de maladie. On peut citer l'exemple du paragraphe 2.3.2.1 avec la cigarette et l'alcool, multipliant le risque du cancer de la gorge. Ainsi, lorsque ces dépendances sont identifiées, il suffit de relâcher les hypothèses d'indépendances conditionnelles que l'on avait supposées dans les modèles log-linéaire.

Le paragraphe 2.4.2.2 concernait les modèles log-linéaires où tous les termes d'interaction d'ordre plus grand que deux sont égaux à zéro. Cette hypothèse a permis de collecter une information et de pouvoir faire une inférence. Cependant, elle peut sembler restrictive. En certaines circonstances il est nécessaire d'ajouter des interactions d'ordre deux, voire plus. Il est à noter que cette nouvelle hypothèse va demander un nouvel effort aux experts, i.e. de nouvelles probabilités conditionnelles à donner. Dans l'exemple présenté figure 2.2, l'expert peut décider qu'il existe un effet combiné entre A et B sur C. Alors le modèle log-linéaire va contenir le terme u_{ABC}, qui n'est plus égal à zéro. Pour calculer ce terme, et grâce à notre méthodologie, seule une probabilité conditionnelle supplémentaire est nécessaire, à savoir $P(C|a,b)$. On a avec l'indépendance entre A et B :

$$\begin{aligned} P(C|a) &= P(C|a,b)P(b) + P(C|a,\bar{b})P(\bar{b}) \\ P(C|b) &= P(C|a,b)P(a) + P(C|\bar{a},b)P(\bar{a}) \\ P(C|\bar{a}) &= P(C|\bar{a},b)P(b) + P(C|\bar{a},\bar{b})P(\bar{b}) \end{aligned} \tag{2.10}$$

Les experts sont de nouveau questionnés pour donner la probabilité conditionnelle, dans laquelle ils ont le plus confiance, puis on en déduit les autres par les formules (2.10). Une autre méthode consiste à demander les quatre probabilités conditionnelles, puis d'appliquer les règles dans le paragraphe précédent. On remarque toute l'importance d'avoir demandé au préalable les probabilités marginales, puis les probabilités conditionnelles du premier ordre. On peut imaginer une méthode en profondeur, qui prend en compte des interactions d'ordre de plus en plus élevé.

Pour notre étude, neuf interactions d'ordre 2 ont été ajoutées par les experts. Celles-ci concernent les variables données dans la première colonne du tableau 2.5. Dans la deuxième et la troisième colonne sont données les variables qui interagissent et influent sur la variable de la première colonne. Enfin, la quatrième colonne donne la probabilité conditionnelle de la variable sachant une combinaison des variables parents. Les autres probabilités conditionnelles sont calculées par la méthode décrite ci-dessus (voir les équations 2.10).

2.5 Conclusions

Dans ce chapitre, nous modélisons le vieillissement d'un système mécanique à partir d'avis d'experts. La complexité du phénomène étudié et le nombre important de variables mis en jeux nécessitent obligatoirement des avis d'experts. Ainsi, pour modéliser le vieillissement, nous avons choisi les réseaux bayésiens. La capacité des réseaux bayésiens à représenter des probables causalités est précieuse pour une application comme la nôtre. Trois étapes sont nécessaires pour construire de tels modèles : le choix du système et des experts, la construction de la structure du graphe et enfin l'évaluation des probabilités correspondantes.

Variable	Parent 1	Parent 2	Probabilité conditionnelle
Coulissement difficile bague de glissement $M6$	Age de la douille $Ad > 6$ ans	Age de la bague $Ab > 1$ an	0.0297
Mauvais positionnement des glaces $M5$	Débit inverse DI oui	Présence d'impuretés $PI3$	0.8532
Sensibilité anormale du joint $O5$	Dégradation douille $M3$	Dégradation étanchéité $M4$	0.5943
Sensibilité anormale du joint $O5$	Dégradation douille $M3$	Coulissement difficile bague $M6$	0.8211
Plage de variation du débit de fuite élevée $O2^{'}$	Dégradation douille $M3$	Dégradation étanchéité $M4$	0.2190
Plage de variation du débit de fuite élevée $O2^{'}$	Dégradation douille $M3$	Coulissement difficile bague $M6$	0.2190
Irrégularité du débit de fuite $O2^{''}$	Dégradation douille $M3$	Dégradation étanchéité $M4$	0.4816
Irrégularité du débit de fuite élevée $O2^{''}$	Dégradation douille $M3$	Coulissement difficile bague $M6$	0.4619
Joint 1 défaillant E	Niveau moyen $O1$ débit de fuite faible	Plage de variation $O2^{'}$ débit de fuite élevée	0.0025

Tab. 2.5 – Interactions d'ordre 2 pour le réseau bayésien du joint 1 d'une pompe primaire 900 MW.

Le choix du système et des experts doit prendre en compte plusieurs paramètres :

- il doit exister des connaissances sur le système (REX, experts),
- le système a une fonction vitale et est donc bien surveillé,
- dans le cas où il n'y a pas de données, choisir des experts disponibles (moins de dix experts), sachant que l'étude peut être longue,
- choisir si possible des experts avec des formations différentes, n'appartenant pas au même projet, afin d'éviter d'avoir des corrélations trop fortes.

La construction de la structure du graphe nécessite les avis d'experts, même dans le cas où il existe des données de retour d'expérience. Cette étape doit s'appuyer, à travers les avis d'experts, sur les diagrammes de défaillance du système étudié. Les réseaux bayésiens permettent de transcrire facilement les informations de ces diagrammes en représentant les probables causalités entre les variables du modèle.
Après avoir identifié les variables du modèle, nous regroupons ces variables par famille afin de réduire le nombre de connexions possibles. Les familles sont définies généralement par des variables de même nature, par exemple les paramètres de température et de pression, ou par exemple les dégradations des composants. Toutes les familles sont ensuite hiérarchisées, de

telle sorte que des relations de causalité relient les différentes familles. Ainsi, des variables dans une famille donnée ne peuvent influer sur des variables d'une famille plus en amont. De plus, on distingue trois types de variables :
– les variables d'entrée (indépendantes deux à deux), généralement des variables environnementales, des paramètres d'influence,
– les variables intermédiaires, pour notre réseau, les maladies et les observations,
– les variables d'intérêt ou variables de sortie, ici l'état du joint.

Il faut bien avoir conscience que cette phase de construction de structure du graphe, qui est cruciale, est lourde et peut demander beaucoup de temps. Pour en témoigner, nous donnons en annexe (cf. annexe F) quelques structures sur lesquelles notre groupe d'étude, constitué d'experts et de statisticiens, s'était arrêté avant d'aboutir à la structure finale de notre réseau.

La troisième étape concerne l'évaluation des probabilités marginales et conditionnelles. Là encore, les avis d'experts sont indispensables, et ce même quand il existe une base de données. En effet, le nombre de données qui serait nécessaire pour estimer raisonnablement les probabilités n'est jamais, et loin de là, atteint. Une difficulté importante à laquelle nous nous sommes attachés est que le nombre de probabilités à renseigner par les experts devient vite trop important. Pour remédier à ce problème, nous considérons le réseau bayésien comme un modèle log-linéaire où l'on va supposer des indépendances conditionnelles afin de réduire le nombre de probabilités à demander aux experts. Nous avons dans un premier temps supposé que les termes d'interaction d'ordre supérieur à 2 étaient égaux à zéro. Cette réduction drastique de la complexité du réseau fait que le nombre de probabilités à renseigner par les experts est réduit de manière considérable. De plus, nous demandons toutes les probabilités marginales qui à notre sens sont souvent plus faciles à donner par les experts surtout s'ils sont peu familiers avec le calcul des probabilités. Cela entraîne que le nombre de probabilités données par les experts est supérieur à ce qui est devenu nécessaire. Ainsi, il est possible de pointer des incohérences dans les avis d'experts par confrontation des résultats de différentes équations. L'élimination de ces incohérences permet d'obtenir des probabilités plus fiables et plus stables. Pour éliminer ces incohérences et donc éliminer des probabilités, nous donnons quelques règles à suivre par ordre de préférence dans ce souci de fiabilité et de stabilité des probabilités finalement produites :
– garder toutes les probabilités provenant du REX,
– avoir une plus grande confiance en les probabilités marginales,
– vérifier si la probabilité marginale calculée est égale à la probabilité marginale donnée et plus précisément vérifier si cette probabilité est une combinaison convexe des probabilités conditionnelles,
– garder l'ordre des probabilités conditionnelles :
    * changer un minimum de probabilités,
    * résoudre un programme linéaire en prenant en compte les facteurs multiplicatifs entre ces probabilités.

Cependant, restreindre tous les termes d'interaction d'ordre supérieur à 2 égaux à zéro est une hypothèse très forte. Après une inférence statistique faite avec cette hypothèse, les experts ont perçu qu'il manquait des interactions d'ordre supérieur. Ainsi, un retour aux experts pour pointer ces interactions et pour demander les probabilités correspondantes a été nécessaire.

Afin d'éviter ce retour aux experts, nous suggérons donc que ce travail soit effectué en une seule fois. Pour cela, le réseau bayésien est vu comme un modèle log-linéaire. Puis, toutes les probabilités marginales et toutes les probabilités d'ordre 1 sont demandées aux experts. Ensuite, les experts doivent identifier des interactions d'ordre 2 importantes et par conséquent renseigner les probabilités correspondantes. Puis, de façon hiérarchique, des interactions d'ordre 3 peuvent être identifiées. Les probabilités conditionnelles correspondantes doivent être alors renseignées et ainsi de suite. Cette façon de procéder a l'avantage, comme nous l'avons vu, que le nombre de probabilités nécessaires n'explose pas avec l'ordre des interactions considérées. Ainsi, l'évaluation des probabilités d'un réseau bayésien via une approche par des modèles log-linéaires permet de transcrire facilement les avis d'experts, sans que cette tâche ne soit trop complexe. Notons toutefois que notre procédure requiert un travail important de la part du statisticien qui fait appel à un savoir-faire difficile à automatiser.

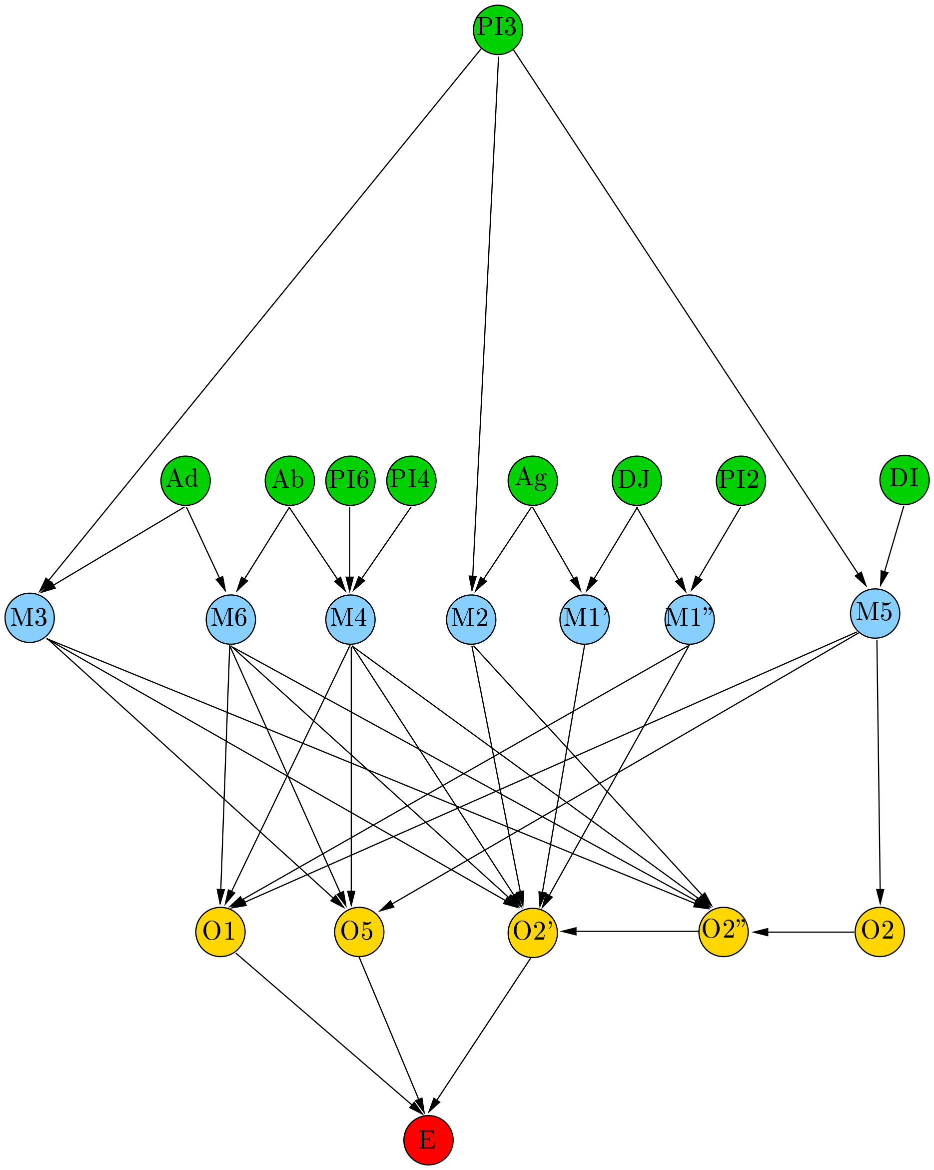


FIG. 2.6 – Réseau bayésien du processus de dégradation du joint 1 d'une pompe primaire.

Chapitre 3

Inférence et Analyse dans le réseau bayésien

Ce chapitre traite de l'inférence et de l'analyse dans un réseau bayésien. Le réseau bayésien considéré ici est celui défini dans le chapitre précédent avec les probabilités données par les trois experts. Dans ce chapitre, nous avons effectué les analyses et les inférences dans le réseau bayésien expert par expert, c'est-à-dire une analyse pour un expert. Mais, comme nous le verrons dans les perspectives, il est possible d'intégrer les experts comme des nouvelles variables du réseau bayésien et de donner via les probabilités conditionnelles les poids correspondants à la confiance donnée aux avis d'experts.

L'inférence classique dans les réseaux bayésiens consiste à calculer des probabilités marginales, détecter les configurations du modèle les plus probables et donc pour notre étude d'identifier les variables contribuant le plus à la dégradation ou à la défaillance du système, et de faire des simulations à but prévisionnel. Grâce à des outils classiques de la statistique, il est possible par exemple de rechercher les variables influant le plus sur la variable d'intérêt, de simuler des scénarios critiques, mais également de faire de l'analyse de sensibilité. Les réseaux bayésiens peuvent également être vus comme un outil d'aide à la décision, où des décisions sont prises lorsque des probabilités dépassent des seuils fixés par les experts. Les décisions peuvent aussi être des actions de maintenance et peuvent être intégrées comme des nouveaux nœuds du réseau bayésien.

Le chapitre est organisé comme suit. Dans la première partie de ce chapitre, nous listons les inférences possibles dans les réseaux bayésiens, comme le calcul des probabilités marginales, les simulations, et la recherche des configurations les plus probables. Nous appliquons ces méthodes sur le réseau bayésien du joint 1 d'une pompe primaire 900 MW. Puis, différentes analyses sont proposées. Nous rappelons tout d'abord l'analyse de sensibilité, où Jensen [59] définit plusieurs indices d'analyse et où Monti et al. [75] préconisent de faire varier les probabilités d'entrée de 0 à 1 et de regarder l'influence sur la variable d'intérêt. Une brève description d'aide à la décision est décrite d'après les travaux de Van Der Gaag et al. [91]. Puis, nous proposons comme outil d'analyse plusieurs indices basés sur la définition d'un score, où l'on mesure l'importance d'une variable sur les variables d'intérêt. Enfin, nous proposons comme outil d'aide à la décision d'intégrer les actions de maintenance comme des nouvelles variables du réseau bayésien. Ces nouvelles variables ont la particularité d'être toujours observées. Une application avec six actions de maintenance est donnée à la fin de ce chapitre.

3.1 Inférence classique

L'inférence classique consiste à estimer des probabilités, à trouver la ou les configurations les plus probables et à simuler divers scénarios.

3.1.1 Calcul de probabilités marginales

Comme nous l'avons vu dans le chapitre 1, l'arbre de jonction permet de calculer les probabilités marginales des cliques. Ainsi, pour le calcul de probabilités marginales, il suffit de marginaliser cette probabilité jointe selon la variable qui nous intéresse. L'intérêt de passer par l'arbre de jonction est de réduire considérablement le nombre de variables à considérer et donc le nombre de sommes à effectuer.

3.1.2 Configuration la plus probable

La recherche de la configuration la plus probable est un problème intéressant. Une manière directe d'aborder ce problème consiste, dans le cas de variables discrètes, à calculer les probabilités de toutes les configurations possibles et de choisir le maximum de toutes ces probabilités. Ceci peut se faire dans le cas de modèles avec peu de variables. Lorsque le nombre de variables est trop important, il est nécessaire d'appliquer une méthode plus appropriée. Cette méthode va être la même que pour le calcul des probabilités marginales. En effet, nous avons vu que la propagation dans l'arbre de jonction consiste à organiser les sommes d'une certaine façon puis de propager des messages entre les cliques. Pour la recherche d'un maximum, nous allons procéder de la même manière, en remplaçant les sommes par des maximums. Prenons comme exemple, un réseau avec trois nœuds A, B et C, comme présenté dans la figure 3.1.

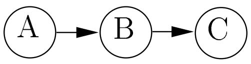


FIG. 3.1 – Configuration la plus probable avec un réseau bayésien à trois variables.

La configuration la plus probable a pour probabilité :

$$\begin{aligned} \alpha = max_{A,B,C}P(A,B,C) &= max_{A,B,C}P(A)P(B|A)P(C|B) \\ &= max_A(P(A)max_B(P(B|A)max_C(P(C|B)))) \end{aligned}$$

Des probabilités fictives sont données dans le tableau 3.1. Ainsi, afin d'éviter le calcul

a	$\bar{a}$
0.4	0.6

	a	$\bar{a}$
b	0.6	0.2
$\bar{b}$	0.4	0.8

	b	$\bar{b}$
c	0.2	0.7
$\bar{c}$	0.8	0.3

$P(A)$ $P(B|A)$ $P(C|B)$

TAB. 3.1 – Exemple fictif de probabilités.

de la probabilité jointe $P(A, B, C)$, il est préférable de calculer en premier lieu le terme $\phi(B) = max_C P(C|B)$, soit $\phi(b) = 0.8$ et $\phi(\bar{b}) = 0.7$. Puis, cette fonction doit être multipliée par la probabilité $P(B|A)$, comme indiqué dans le tableau 3.2. La fonction $\phi(A) =$

	a	$\bar{a}$
b	0.48	0.16
$\bar{b}$	0.28	0.56

TAB. 3.2 – Calcul de la probabilité maximale $P(B|A) max_C P(C|B)$.

$max_B P(B|A)\phi(B)$ prend donc les valeurs $\phi(a) = 0.48$ et $\phi(\bar{a}) = 0.56$. Cette fonction est multipliée par la probabilité a priori $P(A)$, pour donner $(0.192, 0.336)$, et donc $\alpha = 0.336$. la valeur de A, maximisant la probabilité jointe est alors $\bar{a}$. Pour connaître la valeur de B, il suffit de se reporter au tableau 3.2, en sachant que $A = \bar{a}$, c'est-à-dire $B = \bar{b}$. Enfin, la valeur de C est donnée par le tableau 3.1, sachant que $B = \bar{b}$. On trouve donc $C = c$. Ainsi, la configuration la plus probable, ainsi que sa probabilité sont données par cet algorithme. On peut généraliser à un ensemble quelconque de variables, en s'inspirant des algorithmes présentés dans le chapitre 1. Afin d'appliquer les mêmes algorithmes que pour le calcul des probabilités marginales, nous citons le théorème permettant de calculer la probabilité de la configuration la plus probable (cf Jensen [59]).

Théorème 3 *Soit un réseau bayésien avec une probabilité jointe $P(U)$, et soit T, un arbre de jonction associé. Soient $e = (e_1, \ldots, e_m)$ les évidences, après une propagation entière (collecte et distribution des évidences) dans l'arbre de jonction, nous avons :*

(i) pour chaque séparateur S, $max_{U \setminus S} P(U, e)$ est le produit des messages entrants et sortants dans le séparateur S,

(ii) pour chaque noeud V, $max_{U \setminus V} P(U, e)$ est le produit du potentiel associé à V et des messages entrants.

Ce théorème est l'équivalent du théorème 2, en remplaçant les sommes par les maximums.

3.1.3 Simulations

Dans ce paragraphe, nous exposons une méthode pour simuler les états des variables dans un réseau bayésien. L'idée est de parcourir le réseau bayésien dans le sens causal, c'est-à-dire pour notre réseau dans le sens des arcs. En reprenant l'exemple précédent, cela revient tout d'abord à simuler A selon sa probabilité a priori. Un générateur aléatoire est alors utilisé comme suit. On tire un nombre entre 0 et 1. Si ce nombre est inférieur à 0.4 alors $A = a$, sinon $A = \bar{a}$. Supposons que ce nombre soit inférieur à 0.4 et donc que $A = a$. Alors on tire B sachant $A = a$ grâce au générateur aléatoire selon la probabilité conditionnelle correspondante. Ainsi, si le nombre tiré est inférieur à 0.6 alors $B = b$, sinon $B = \bar{b}$. Supposons que ce nombre soit supérieur à 0.6 et donc que $B = \bar{b}$. Enfin C est tiré de la façon suivante. Si le nombre aléatoire est inférieur à 0.7 alors $C = c$ sinon $C = \bar{c}$. Supposons que ce nombre soit inférieur à 0.7 et donc que $C = c$. Finalement, une donnée tirée de la probabilité jointe du réseau bayésien de la figure 3.1 est $(a,\ \bar{b},\ c)$. Pour obtenir un échantillon de taille n, il suffit de répéter cette procédure n fois.

L'intérêt de la simulation apparaît lorsque l'on désire connaître la conséquence de l'occurrence d'une évidence sur le réseau. Une méthode simple est de simuler un échantillon jusqu'à ce que les états simulés soient les mêmes que les états observés. Cependant, cette méthode peut s'avérer très longue si la probabilité des états observés est faible ou si le nombre de variables est important. Une possibilité est d'appliquer l'algorithme de l'échantillonnage de Gibbs ("Gibbs sampling" en anglais, cf. Gilks et al. [48]). Pour cela, une valeur initiale est nécessaire, ce qui correspond ici à une configuration initiale. Cette configuration doit être compatible avec les évidences. Puis les états des variables non observés sont simulés selon l'algorithme décrit ci-dessus. Supposons par exemple que la variable observée soit B, et que sa valeur soit égale à b et que la configuration initiale soit $(\bar{a}, b, c)$. Dans un premier temps, l'état de A est simulé selon la probabilité $P(A|B = b, C = c)$. Cette probabilité est égale d'après le réseau bayésien à $P(A|B = b)$. En appliquant le théorème de Bayes, on trouve $(2/3, 1/3)$. Supposons que le nombre aléatoire généré soit supérieur à $2/3$ et donc que $A = \bar{a}$. Puis, la prochaine variable libre est simulée, ici C selon la probabilité $P(C|A = \bar{a}, B = b) = P(C|B = b)$, soit $(0.2, 0.8)$. Supposons que le nombre aléatoire généré soit supérieur à 0.2 et donc que le nouvel état de C soit égal à $\bar{c}$. Puis l'algorithme est répété avec comme valeur initiale $(\bar{a}, b, \bar{c})$ et ainsi de suite. Généralement, les premières itérations de l'algorithme de l'échantillonnage de Gibbs ne sont pas retenus (période de chauffe ou "burn in" en anglais). Un inconvénient de cet algorithme est la possible stagnation de la chaîne de Markov générée. De plus, la corrélation entre deux états successifs peut être importante. Ainsi, l'état final peut être très dépendant de l'état initial, si l'échantillonnage de Gibbs n'est pas utilisé avec un nombre important d'itérations. Pour toutes ces questions techniques, on pourra se référer au livre de Robert [84].

3.1.4 Applications

Dans ce paragraphe, nous appliquons les méthodes décrites sur le réseau bayésien du joint 1 d'une pompe primaire 900 MW. Nous ne donnons pas d'applications pour le calcul des probabilités marginales, car d'une part celles-ci ont été données par les experts lors de la phase de construction du réseau, et d'autre part les calculs des probabilités marginales sont à la base de toutes les analyses d'un réseau bayésien. Ainsi, dans un premier temps, nous recherchons la configuration la plus probable du réseau bayésien, puis des simulations sont faites dans ce même réseau pour des événements rares. Toutes les applications ont été faites à partir des données d'un expert (expert RC).

3.1.4.1 Configurations les plus probables

Le tableau 3.3 donne la configuration la plus probable de toutes les variables du réseau bayésien du joint 1.
Il est possible d'envisager la configuration la plus probable suite à l'observation d'une ou

Ad	entre 1 et 6 ans	DJ	non	$O1$	moyen
Ag	entre 1 et 6 ans	$M1'$	non	$O5$	non
Ab	supérieur à 1 an	$M1''$	non	$O2$	stable
$PI2$	élevé	$M2$	non	$O2''$	irrégulière
$PI3$	non	$M3$	oui	$O2'$	faible
$PI4$	non	$M4$	non	E	sain
$PI6$	faible	$M5$	non		
DI	non	$M6$	non		

TAB. 3.3 – Configuration la plus probable du réseau bayésien du joint 1 d'une pompe primaire 900 MW.

plusieurs variables. Par exemple, on peut s'intéresser au cas où l'âge de la bague est inférieur à un an, celui de la douille est supérieur à 6 ans, on observe une sensibilité anormale du joint et un faible niveau moyen du débit de fuite, i.e. $Ab = 1$, $Ad = 3$, $O5 = 1$ et $O1 = 1$. Les résultats sont donnés dans le tableau 3.4 avec uniquement les variables qui ont changé de modalités. On peut constater, avec ces observations, l'apparition de la maladie de la bague

$M6$	oui
$O2'$	élevée
E	dégradé

TAB. 3.4 – Configuration la plus probable du réseau bayésien du joint 1 d'une pompe primaire 900 MW après observations.

(coulissement difficile) et la dégradation du joint.

3.1.4.2 Simulations dans le réseau bayésien du joint 1

Dans ce paragraphe, nous effectuons une simulation dans le réseau bayésien du joint 1 d'une pompe primaire 900 MW, en sachant par exemple que la douille et la glace ont un âge supérieur à 6 ans et que la bague a plus d'un an. Pour cela, nous avons effectué 1000 simulations dans le réseau bayésien (via la boite à outils Matlab BNT [102]), et nous n'avons gardé que les scénarios où ces trois modalités étaient vérifiées, c'est-à-dire 9 scénarios dont on détaille les résultats dans le tableau 3.5. Nous nous sommes ici intéressés à des scénarios qui ont une faible probabilité d'apparition, ou autrement dit à des événements rares.

$PI2$	faible (**1**)	élevé (**8**)
$PI3$	oui (**0**)	non (**9**)
$PI4$	oui (**0**)	non (**9**)
$PI6$	faible (**9**)	importante (**0**)
DI	oui (**1**)	non (**8**)
DJ	oui (**2**)	non (**7**)
$M1^{'}$	oui (**0**)	non (**9**)
$M1^{''}$	oui (**0**)	non (**9**)
$M2$	oui (**0**)	non (**9**)
$M3$	oui (**9**)	non (**0**)
$M4$	oui (**0**)	non (**9**)
$M5$	oui (**0**)	non (**9**)
$M6$	oui (**0**)	non (**9**)
$O1$	faible (**1**)	moyen (**8**)
$O5$	oui (**2**)	non (**7**)
$O2$	stable (**7**)	non stable (**2**)
$O2^{''}$	croissante (**2**)	irrégulière (**7**)
$O2^{'}$	faible (**8**)	élevée (**1**)
E	sain (**7**)	dégradé (**2**)

TAB. 3.5 – Simulations dans le réseau bayésien du joint 1 d'une pompe primaire 900 MW.

Pour chaque variable, le tableau donne entre parenthèses le nombre d'apparitions de la modalité correspondante. On peut remarquer qu'avec une douille, une glace et une bague âgées, la maladie la plus fréquente est la dégradation de la douille de logement par usure ou rayures. On peut également remarquer que le joint est déclaré dégradé dans deux cas sur neuf (plus de 20% des cas).

On peut imaginer de nombreuses simulations afin de cerner des scénarios critiques peu probables. Cette méthode est facilement utilisable et peut s'avérer très utile.

3.2 Analyse dans le réseau bayésien

L'analyse dans un réseau bayésien consiste à rechercher les variables les plus discriminantes pour une variable d'intérêt, et à les mettre en cause par une analyse de sensibilité. Nous présentons également dans ce paragraphe, notre méthode consistant à classer les scénarios et à identifier des paramètres jugeant de l'importance d'une variable.

3.2.1 Analyse de sensibilité

L'analyse de sensibilité vise à regarder l'influence d'observations, que l'on notera e, sur des variables d'intérêt, notées $Y = y$. On désignera la modalité de la variable d'intérêt par le terme "hypothèse", noté $h_Y : Y = y$. Plus précisément, l'analyse de sensibilité tend à répondre à plusieurs questions :

- Quelle influence peut avoir l'ajout d'un bruit dans les probabilités ?
- Quelles observations vont favoriser, ou pénaliser, ou n'avoir aucune influence sur la modalité y ?
- Quelles observations vont le mieux discriminer deux modalités de la variable d'intérêt ?

Pour illustrer ce type d'analyse, on reprend l'exemple cité dans le livre de Jensen [59]. M. Holmes quitte sa maison un matin, et s'aperçoit que sa pelouse est humide ($H = y$). Il pense donc qu'il a plu cette nuit ($R = y$) ou qu'il a oublié de débrancher le jet d'eau ($S = y$). Il regarde donc les pelouses de ses deux voisins, celle du Dr. Watson et celle de Mlle Jibbon. Les deux pelouses sont sèches ($W = n$ et $J = n$). Il en conclut qu'il a dû oublier de fermer le robinet. On peut représenter le réseau bayésien par la figure 3.2, avec les probabilités associées dans le tableau 3.6.

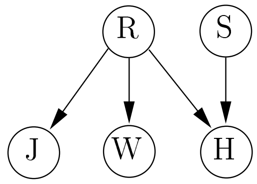


FIG. 3.2 – Exemple d'analyse de sensibilité dans un réseau bayésien.

On s'intéresse ici à la variable d'intérêt $S = y$, à savoir si M. Holmes a oublié de fermer son robinet. Les évidences sont notées e_H : "$H = y$", e_W : "$W = n$" et e_J : "$J = n$". D'après le réseau bayésien S, W et J sont indépendants et donc nous avons $P(S = y|e_W) = P(S = y|e_J) = P(S = y) = 0.1$. En revanche $P(S = y|e_H) = 0.5053$. La probabilité que M. Holmes ait oublié de fermer son robinet, sachant que son gazon est humide et que les pelouses de ses voisins sont sèches, est égale à $P(S = y|e) = \frac{0.65611}{0.65611+0.00000891} = 0.9999$. Ainsi, les évidences e_W et e_J prises seules n'ont aucune influence sur la variable d'intérêt. Aussi, l'évidence e_H ne suffit pas non plus pour la conclusion. Ainsi, la conclusion rapide que e_W et e_J n'ont aucune influence sur la variable d'intérêt est fausse. Pour s'en convaincre,

$R=y$	$R=n$
0.99	0.1

$P(J=y|R)=P(W=y|R)$

	$R=y$	$R=n$
$S=y$	1	0.9
$S=n$	0.99	0

$P(H=y|R,S)$

TAB. 3.6 – Probabilités initiales pour l'exemple du gazon humide avec $P(R=y)=P(S=y)=0.1$.

regardons les probabilités en combinant les évidences. Nous avons $P(S=y|e_W,e_J)=0.1$ et $P(S=y|e_H,e_J)=P(S=y|e_H,e_W)=0.988$. Jensen [59] définit des notions basées sur le rapport de vraisemblance. En effet, afin de quantifier l'impact de ces évidences sur l'hypothèse, on peut diviser ces probabilités par la probabilité marginale de l'hypothèse $P(h_S)$, soit une vraisemblance normalisée $\frac{P(h_S|e')}{P(h_S)}=\frac{P(e'|h_S)}{P(e')}$. Ces rapports de vraisemblances sont données dans le tableau 3.7.

e_W	e_J	e_H	$\frac{P(h_S\|e')}{P(h_S)}$
1	1	1	9.999
1	1	0	1
1	0	1	9.88
1	0	0	1
0	1	1	9.88
0	1	0	1
0	0	1	5.1
0	0	0	1

TAB. 3.7 – Vraisemblance normalisée comme fonction des évidences. Les 1 indiquent la présence ou non de l'évidence correspondante.

Afin de mesurer l'impact des évidences sur l'hypothèse, Jensen [59] définit plusieurs indices, que nous donnons ci-après.

Définition 13 *Soit e une évidence et h une modalité de variable d'intérêt (appelé parfois hypothèse). On souhaite connaître la sensibilité de $P(h|e)$ en fonction de e. L'évidence $e' \subset e$ est dite suffisante si $P(h|e')$ est au moins égale à $P(h|e)$, ou lorsque la différence normalisée entre ces deux probabilités est inférieure à un seuil, noté θ_1.*

- ** l'évidence e' est dite suffisante minimale si elle est suffisante et si aucune sous-suite de e' n'est suffisante,*
- ** l'évidence e' est dite cruciale si elle est incluse dans toutes les suites suffisantes,*

* *l'évidence e' est dite importante si la probabilité de l'hypothèse varie beaucoup sans cette évidence (ou est plus grande qu'un seuil fixé, noté θ_2), i.e. lorsque* $\|\frac{P(h|e\backslash e')}{P(h|e)} - 1\| > \theta_2$.

Ainsi, si on choisit comme seuil $\theta_1 = 0.05$ et $\theta_2 = 0.2$ pour l'exemple précédent, on trouve que (e_H, e_J) et (e_H, e_W) sont suffisantes minimales, que (e_W, e_J) sont importantes et que e_H est cruciale. D'autres mesures peuvent être appliquées ici, comme notamment le facteur de Bayes [62] égal à un rapport de probabilités a posteriori, comparant deux hypothèses possibles, soit $\frac{P(e|h)}{P(e|\bar{h})}$.

Monti and Carenini [75] proposent de faire varier les probabilités des nœuds d'entrée entre 0 et 1, et de voir l'influence sur la probabilité de la variable de sortie ou la variable d'intérêt. Plus précisément, pour notre réseau bayésien, imaginons que l'on souhaite connaître la sensibilité de la modalité $Ab = 0$ (âge de la bague inférieur à un an) sur la défaillance du joint1 d'une pompe primaire 900 MW ($E = 2$). Leur méthode consiste donc à regarder la variation de la probabilité $p(E|Ab = 0)$ en faisant varier la probabilité $p(Ab = 1|E = 2)$ de 0 à 1. Puis, en traçant les trois courbes correspondantes aux trois modalités de la variable "état du joint" (sain, dégradé et défaillant), il est possible de remarquer si une petite variation de la probabilité d'entrée change de façon importante la probabilité de la variable de sortie, et donc change la modalité.

Pradhan et al. [82] se sont intéressés à la sensibilité du réseau bayésien en ajoutant un bruit sur les probabilités. Le bruit est ajouté sur les log-ratios afin de s'affranchir des problèmes aux bords. En effet, une probabilité proche de 1 avec l'ajout d'un bruit peut dépasser la valeur limite. De plus une probabilité de 0.1 que l'on fait varier entre 0 et 0.2 n'a pas le même poids et n'a pas la même signification qu'une probabilité de 0.5 que l'on fait varier de 0.4 à 0.6.

Toutes ces techniques d'analyse de sensibilité permettent de connaître les variables importantes du réseau bayésien et permettent de connaître les efforts à faire pour l'obtention d'une probabilité à partir des avis d'experts. Cependant, nous n'avons pas utilisé ces méthodes. En effet, nous avons proposé, comme nous le verrons dans le paragraphe 3.2.3, quelques méthodes basées sur un score afin de déterminer l'importance d'une variable.

3.2.2 Outil d'aide à la décision

Dans ce paragraphe, nous présentons les travaux de Van Der Gaag et Veerle [91], et de Agre [1], portant sur l'aide à la décision dans les réseaux bayésiens. Pour les premiers auteurs cités, l'outil d'aide à la décision est basé sur l'analyse de sensibilité dans un réseau bayésien. Une variable d'intérêt est désignée ainsi que des seuils, permettant de décider des actions. En règle générale, la probabilité de la variable d'intérêt est une fonction simple de probabilités conditionnelles que l'on souhaite étudier. Ces fonctions sont soit linéaires, soit des fonctions homographiques (rapports de fonctions linéaires), de type $\frac{ax+b}{cx+d}$. De la sorte, il est possible de calculer les seuils correspondants pour les probabilités conditionnelles, ainsi que la robustesse des décisions prises. Nous reprenons l'exemple précédent du paragraphe 3.2.1 où la variable d'intérêt est la variable S. On se fixe un seuil θ tel que si $P(S = y) \geq \theta$, alors M. Holmes doit investir dans un détecteur de fuite. Reprenons pour notre étude les mêmes observations e. On peut écrire $P(S = y) = ax + b$, où x représente $P(S = y|e)$, la probabilité conditionnelle que l'on va faire varier, et où a et b sont des constantes. Ici on trouve $a = 0.0729$ et $b = 0.0344$. Il est alors possible de calculer le seuil correspondant, $x^s = \frac{\theta - b}{a}$. Ainsi, si x est plus grand que ce seuil, la décision d'acheter un détecteur est alors prise. Ce type d'outil est très utilisé pour le diagnostic et l'aide à la décision dans le domaine médical. On pourra également voir les travaux de Agre [1], qui construit un réseau bayésien de diagnostic, et qui utilise des algorithmes d'inférence.

Ces techniques sont facilement envisageables pour notre étude. Pour cela, il suffit que les experts donnent par exemple des seuils pour les probabilités sur les maladies, et qu'ils donnent l'action à entreprendre. Ainsi, à chaque observation, il est possible d'associer une action. Cependant, nous n'avons pas appliqué ces techniques. En effet, pour le réseau bayésien du joint 1, l'aide à la décision qui préoccupait les experts concernait principalement les actions de maintenance. Le but est donc de mesurer quels impacts peuvent avoir les actions de maintenance. Cette approche est décrite dans le paragraphe 3.3, en considérant les tâches de maintenance comme des nouveaux nœuds du réseau bayésien.

3.2.3 Notre approche

Nous proposons pour l'analyse de sensibilité d'un réseau bayésien quelques critères et paramètres simples basés sur un score.

3.2.3.1 Définition d'un score

Nous nous intéressons aux possibles scénarios et à leur influence sur la variable d'intérêt. Supposons que la probabilité jointe du modèle est connue. Dans notre étude, les variables étant discrètes, cette probabilité se résume en un vecteur de longueur $3^5 * 2^{17} = 31850496$, 17 variables ayant 2 modalités et les 5 autres ayant trois modalités. Afin de pouvoir comparer ces probabilités entre experts, la probabilité jointe a été normalisée pour conduire à un score égal à $\dfrac{P_{jointe}}{\max P_{jointe}}$. De plus, pour simplifier l'analyse, les variables ont été divisées en quatre catégories, de la même manière que lors de la conception du réseau bayésien. Les quatre catégories sont les variables d'entrée, les maladies, les variables d'observations et la variable d'intérêt, notée E. Ainsi, nous cherchons à analyser l'influence des variables dans chaque groupe. Par exemple, avec le groupe des variables d'entrée comprenant 9 variables, notés $X_1, \ldots, X_9$, le score est calculé suivant la formule :

$$score = \frac{P(E|X_1, \ldots, X_9)}{\max\limits_{X_1, \ldots, X_9} P(E|X_1, \ldots, X_9)}.$$

Parmi les variables d'entrée, deux variables ont trois modalités, les autres étant binaires, ce vecteur a une longueur égale à $3^2 * 2^7 = 1152$, le nombre de scénarios possibles. L'objectif étant d'identifier les variables importantes sous-jacentes pour la dégradation ou la défaillance du système, deux scores sont ainsi calculés. On notera $score_{def}$ et $score_{deg}$, respectivement le score quand la variable d'intérêt (E) a pour modalités la défaillance et la dégradation (on pourra également s'intéresser au cas où l'état du joint est sain). En triant le score du plus grand au plus petit, il est possible d'effectuer une première analyse. Après avoir identifié le ou les scénarios les plus critiques (avec un score égal à 1), seuls les scénarios dont une nouvelle variable a changé de modalité sont gardés. Par cette méthode il est possible d'identifier les modalités des variables qui contribuent le plus à la dégradation ou à la défaillance. De plus, les grandes variations des scores vont permettre d'isoler des scénarios types, ayant la même contribution à la probabilité maximale.

Nous appliquons cette méthode pour le système étudié avec pour variable d'intérêt (variable de sortie), l'état du système. Nous maximisons la probabilité d'être dégradé sous contraintes des nœuds d'entrée. Pour l'expert RC, les probabilités sont normalisées par la probabilité maximale d'être dégradé, égale à 0.6723. Les scénarios sont classés par ordre décroissant de leurs scores. Dans le tableau 3.8 ci-après sont répertoriés uniquement les scénarios les plus critiques dont au moins une variable a changé de modalités, où Ab est l'*âge de la bague*, Ag l'*âge de la glace*, Ad l'*âge de la douille*, *PI2 la valeur du débit de fuite au démarrage à 25 bars*, *PI3 la présence d'impuretés dans le circuit RCV*, *PI4 vibration et déplacement d'arbre*, *PI6 température du palier, nombre d'excursions autour du niveau moyen avec une amplitude supérieure à 10°C* , *DI* le *débit inverse* et *DJ* le *démontage du joint*. Ainsi, les deux scores des scénarios les plus critiques ont un score égal à 1. Entre ces

Ab	Ag	Ad	$PI2$	$PI3$	$PI4$	$PI6$	DI	DJ	$Score_{deg}$
0	**2**	**2**	**0**	**0**	**1**	**1**	**0**	**1**	1
0	**2**	**2**	**0**	**0**	0	**1**	**0**	**1**	1
0	1	**2**	**0**	**0**	1	**1**	**0**	**1**	0.9834
0	2	1	**0**	**0**	0	**1**	**0**	**1**	0.9736
0	2	2	**0**	**0**	1	0	**0**	**1**	0.9637
0	2	2	1	**0**	0	1	**0**	**1**	0.9538
0	2	2	0	**0**	1	1	1	0	0.9536
1	2	2	0	**0**	1	1	0	1	0.9091
0	2	2	0	1	0	1	0	1	0.8389

TAB. 3.8 – Comparaison de scénarios en fonction du score relatif à la **dégradation** du système.

deux scénarios, seule la variable *vibrations et déplacement d'arbre* ($PI4$) a changé de modalité. Le scénario suivant dans la liste correspond donc au scénario le plus critique dont au moins une variable (autre que $PI4$) a changé de modalité (ici la variable *âge de la glace* Ag), et ainsi de suite.

Dans la figure 3.3, nous comparons les scores avec un deuxième expert. Il y a 1152 ($= 2^7 * 3^2$) scénarios possibles lus sur les axes des abscisses. Dans cette figure, nous avons voulu comparer l'évolution du score des deux experts sans se préoccuper des scénarios. Le score a été trié de manière décroissante.

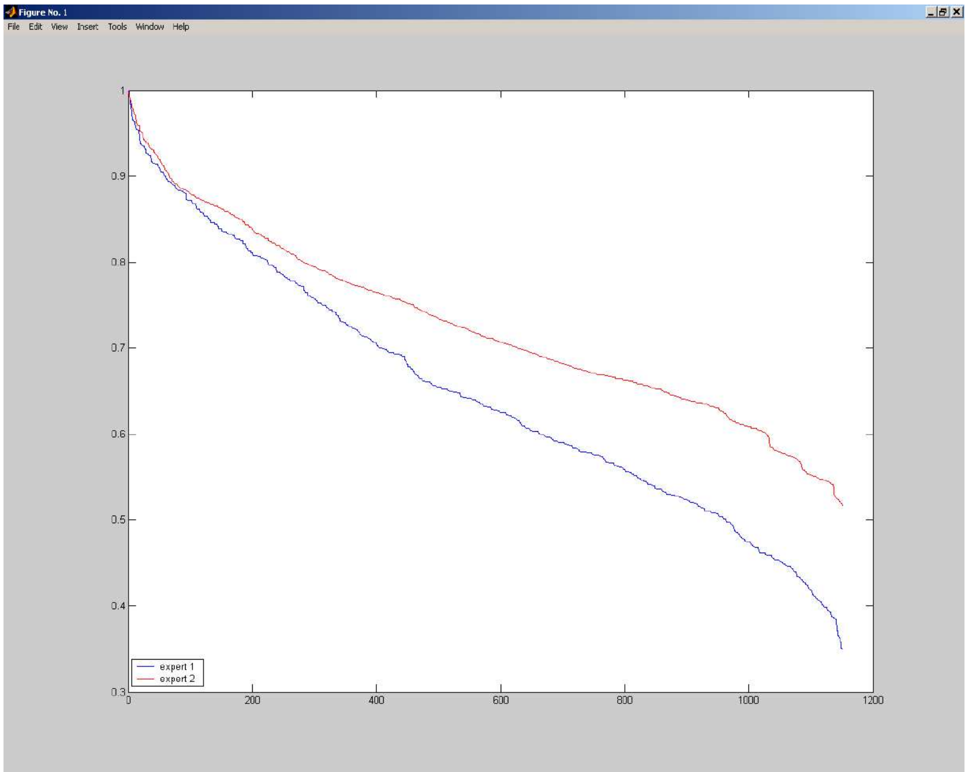


FIG. 3.3 – Comparaison des scores des deux experts avec comme variable d'intérêt la dégradation et comme scénarios les variables d'entrée.

Les variations de ces deux scores semblent être continues sans grande variation, ni saut. La même analyse a été effectué avec l'autre expert et les résultats obtenus sont analogues. Ainsi, pour les deux experts, l'*âge de la bague* (Ab) et la *présence d'impuretés dans le circuit* RCV ($PI3$) sont deux variables très influentes sur la dégradation du joint.

Afin de savoir quelles modalités sont importantes pour les deux experts, il est intéressant de regarder l'évolution du score d'un expert trié de manière décroissante et de comparer avec le score de l'autre expert pour les mêmes scénarios. La figure 3.4 compare les scénarios des deux experts, où l'on a trié le score d'un expert (ici celui de l'expert BR) et calculé le score de l'autre expert pour ces scénarios triés.
On remarque que la tendance entre les deux experts est inversée. En effet, ceci vient du fait que l'expert RC pense que les défauts de jeunesse de la bague sont le plus souvent

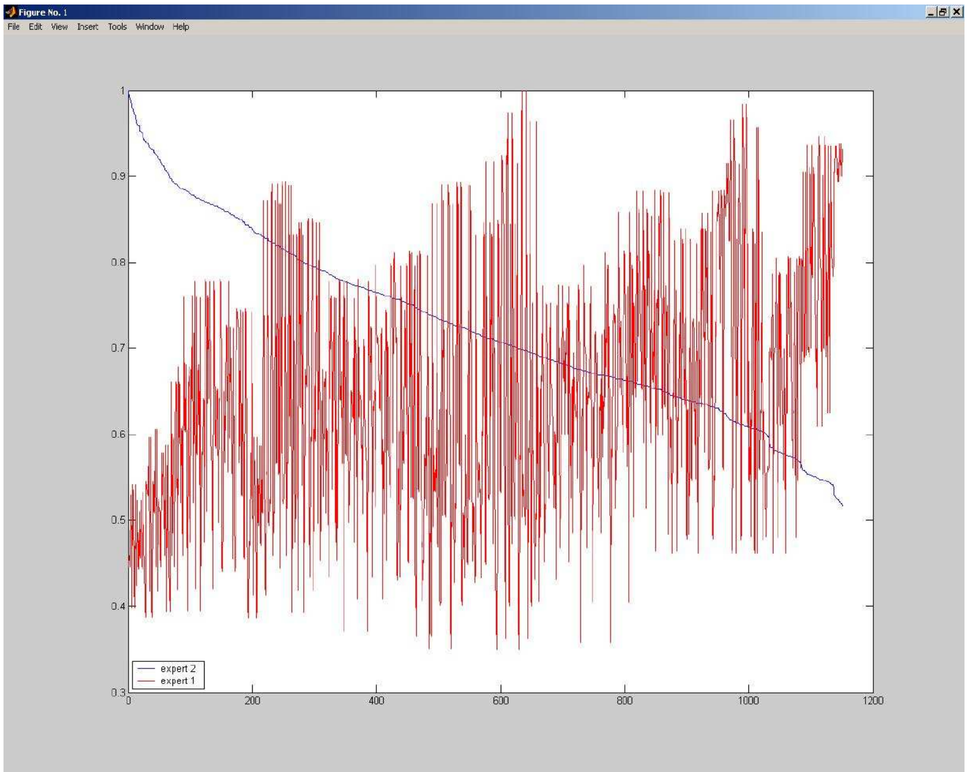


FIG. 3.4 – Comparaison des scores des deux experts dont l'un est trié de manière décroissante, avec comme variable d'intérêt, la dégradation et comme scénarios les variables d'entrée.

responsables de la dégradation du joint, alors que l'expert BR pense que le vieillissement de la bague est responsable de la dégradation. Toutefois, ces résultats doivent être modérés compte tenu des faibles variations des deux scores.

Nous nous sommes également intéressés à la **défaillance** du joint 1 comme modalité de la variable d'intérêt E. Tout d'abord comme précédemment, nous avons classé les scénarios selon leurs scores. Les résultats de l'expert RC sont donnés dans le tableau 3.9. La probabilité maximale que le joint soit défaillant est, pour cet expert, égal à 0.0014.

Dans ce tableau, on remarque que pour l'expert RC, l'*âge de la bague* et l'*âge de la douille* sont des variables contribuant le plus à la *défaillance du joint*. Plus précisément, lorsque l'*âge de la bague* est *inférieur à un an* et lorsque la *douille a un âge supérieur à six ans*, la probabilité de *défaillance* est la plus élevée. Comme pour la dégradation, cet expert considère que ce sont les défauts de jeunesse de la bague qui sont le plus fréquemment responsables de la défaillance du joint 1 d'une pompe primaire 900 MW. On a également comparé les variations des scores des deux experts (cf. figure 3.5).

On peut remarquer que pour l'expert BR, l'influence des variables d'entrée est moindre sur la défaillance du joint 1 que pour l'expert RC (il existe un facteur multiplicatif de 5). Sur cette figure, on voit la grande variation du score qui passe de 0.984 à 0.5003, lorsque l'âge

Ab	Ag	Ad	$PI2$	$PI3$	$PI4$	$PI6$	DI	DJ	$Score_{def}$
0	**2**	**2**	**1**	**0**	**1**	**0**	**0**	**0**	1
0	**2**	**2**	**1**	1	0	**0**	**0**	**0**	1
0	**2**	**2**	**1**	0	1	1	**0**	**0**	0.9995
0	**2**	**2**	0	0	1	0	**0**	**0**	0.9864
0	**2**	**2**	1	0	0	1	1	1	0.9863
0	1	**2**	1	1	0	0	0	0	0.9840
0	2	1	1	0	1	0	0	0	0.5003
1	2	2	1	0	1	0	0	0	0.4506

TAB. 3.9 – Comparaison de scénarios en fonction du score relatif à la **défaillance** du système.

de la douille passe de plus de 6 ans à moins de 6 ans. Nous n'avons pas tracé la figure (dans un souci de lisibilité) comparant les deux scores lorsque celui d'un expert est trié et que l'on calcule les scores correspondants de l'autre expert, mais celle-ci présente encore, comme pour la dégradation une symétrie entre les deux experts. En effet, l'expert BR pense que le vieillissement de la bague est une des causes principales de la défaillance du joint.

3.2.3.2 Dérivée du score

Afin de prendre en compte les grandes variations du score, nous étudions une dérivée de ce score. Ainsi, après avoir trié le vecteur score, il est possible de calculer la dérivée de ce score, c'est-à-dire ici la différence entre deux valeurs consécutives du score. L'objectif de cette analyse est de capter de grands sauts de la fonction score et donc d'identifier des modalités de variables susceptibles de faire varier ce score. Ainsi, par exemple, il est possible après avoir classé les scénarios du plus critique au moins critique, d'établir des classes de scénarios types. Pour cela, un seuil s est fixé de sorte qu'après un saut plus grand que ce seuil, une nouvelle classe de scénarios est définie. Le nombre de classes dépend du seuil choisi et est donc parfaitement arbitraire. L'objectif de cette méthode est de mettre en évidence les modalités communes des variables d'entrée au sein d'une même classe. Prenons comme exemple la dérivée du score de la figure 3.6 avec comme variable d'intérêt la défaillance et avec les données de l'expert RC : Sur cette figure, cinq seuils ont été testés :

- le seuil $s = 0.25$ correspond à deux classes,
- le seuil $s = 0.05$ correspond à trois classes,
- le seuil $s = 0.04$ correspond à quatre classes,
- le seuil $s = 0.03$ correspond à cinq classes,
- le seuil $s = 0.02$ correspond à six classes.

Un premier seuil est associé à deux types de scénarios (deux classes). Les modalités des variables d'entrée communes à une même classe, sont des modalités importantes contribuant beaucoup à une défaillance. Ainsi, pour un seuil égal à 0.25, les 96 scénarios les plus critiques ont comme modalités communes, l'*âge de la bague (inférieur ou égal à 1 an)* et l'*âge de la*

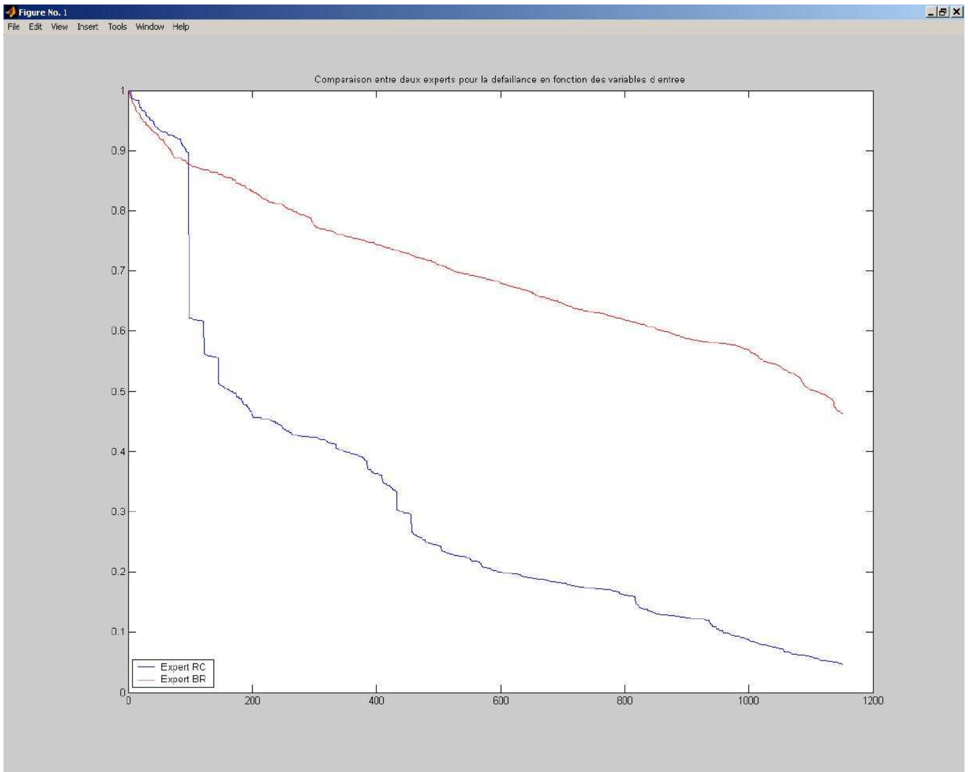


Fig. 3.5 – Comparaison des scores des deux experts avec comme variable d'intérêt, la défaillance et comme scénarios les variables d'entrée.

douille (supérieur à 6 ans). Ce brusque saut est dû à un changement de modalité de l'âge de la glace (de < 1 an à > 6 ans) mais aussi à d'autres changements de modalités d'autres variables. Avec le deuxième seuil (0.05), une troisième classe est formée (du 97ième scénarios le plus critique au 120ième), avec comme modalités communes les mêmes que précédemment. Lors de chaque saut important du score, nous remarquons que, pour ce cas, les modalités des variables supposées importantes restent inchangées.

3.2.3.3 Influence marginale relative d'une variable d'entrée

Dans ce paragraphe, nous étudions l'influence d'une variable d'entrée (parmi les neufs du réseau bayésien) sur *l'état du joint*, lorsque celui est *dégradé* ou *défaillant*. Nous proposons comme indice :

$$\gamma_i = \Delta_i(E) = \frac{\max_j(p(E|X_i = j))}{\min_j(p(E|X_i = j))}.$$

Cette quantité permet de calculer l'influence de la variable d'entrée X_i, sans se préoccuper des autres (contrairement à l'analyse précédente). Elle permet ainsi de classer les modalités d'une variable. Par exemple, si la modalité de la variable d'intérêt est la défaillance, alors on peut s'intéresser à l'influence de l'âge de la bague. Pour cela avec les probabilités données

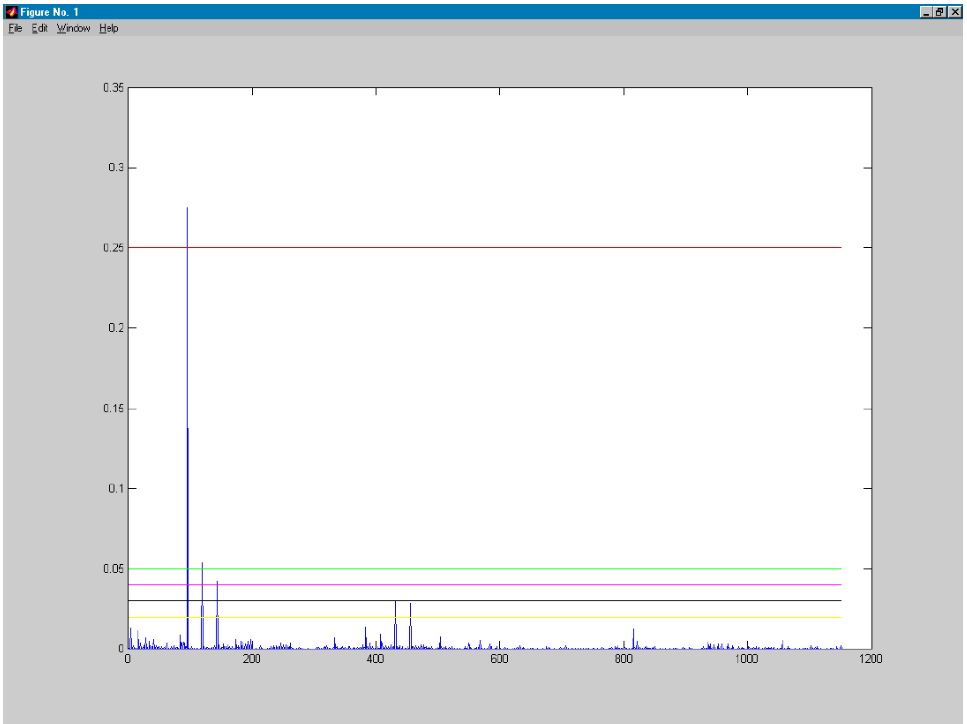


FIG. 3.6 – Dérivée du score et différents seuils pour l'expert RC.

par l'expert RC, on trouve

$$P(E=2|Ab=0) = 5.035.10^{-5} \quad \text{et} \quad P(E=2|Ab=1) = 2.222.10^{-5},$$

et donc $\Delta_i(E=2) = 2.27$. Ainsi, avec une bague jeune (moins de 1 an), la probabilité de défaillance est deux fois plus grande qu'avec une bague âgée. Comme nous l'avons vu précédemment, cet expert pense que les défauts de jeunesse de la bague sont susceptible d'influer grandement sur la défaillance du joint. Toutefois, cette probabilité est de l'ordre de 10^{-5}.

3.2.3.4 Analyse toutes choses égales par ailleurs

Une autre analyse consiste à comparer les scénarios toutes choses égales par ailleurs. La variable d'intérêt reste toujours l'état du joint. On recherche alors l'importance de variables à l'intérieur d'un groupe à partir du calcul du score. Ce groupe peut par exemple être les variables d'entrée. Le but est donc de faire une inférence, avec comme évidence l'observation des variables d'entrée, et comme visée la probabilité que le joint soit dégradé ou défaillant, puis de modifier la modalité d'une variable d'entrée, et de garder les autres. Par cette méthode, il est plus facile d'interpréter le changement de modalités d'une variable. Ainsi, le scénario le plus critique a pour modalité : *âge de la bague inférieur ou égal à 1 an*, *âge de la glace supérieur à 6 ans*, *âge de la douille supérieur à 6 ans*, *valeur du débit* *au démarrage à 25 bars élevée, présence d'impuretés dans le circuit RCV, pas de vibrations ni de déplacement d'arbre, nombre d'excursions autour de la température moyenne avec une amplitude supérieure à 10°C faible, débit inverse et démontage du joint.* Avec le même scénario mais avec une *glace âgée de plus de trois ans*, la probabilité de départ que le joint soit défaillant est divisée par 2 (score passant de 1 à 0.4506). Le même scénario avec une *douille dont l'âge est inférieur à 6 ans*, donne aussi une probabilité de défaillance divisée par 2 (score passant de 1 à 0.5003).

3.2.3.5 Synthèse des résultats

Les variables d'entrée contribuant le plus à la dégradation sont l'âge de la bague et la présence d'impuretés dans le circuit RCV. Pour l'âge de la bague, l'expert BR considère que les effets du vieillissement de la bague sont préjudiciables pour la dégradation du joint, alors que l'expert RC considère que ce sont au contraire les défauts de jeunesse de la bague, ou des problèmes au montage par exemple. Néanmoins, ces résultats doivent être modérés compte tenu des faibles variations des deux scores.

Les solutions peuvent être par exemple :
- d'améliorer la chimie,
- la mise en place d'un filtre,
- rédiger de meilleures procédures.

En ce qui concerne la défaillance du joint 1, il existe une variable d'entrée commune aux deux experts qui contribue le plus à la défaillance, à savoir l'âge de la bague. En revanche, les modalités pour cette même variable sont différentes et même opposées (défauts de jeunesse pour l'expert RC et vieillissement pour l'expert BR). Pour l'expert RC, une autre donnée importante est un âge élevé de la douille (supérieur à six ans). Pour l'expert BR, l'autre variable importante est la présence d'impuretés dans le circuit RCV.

La même démarche a été suivie pour les variables d'observations. Tous les résultats sont donnés dans le rapport [25]. Ainsi, après ces analyses dans le réseau bayésien et la détection des variables importantes, il est possible d'envisager des actions de maintenance visant à éliminer les modalités des variables contribuant le plus à la dégradation ou à la défaillance du joint 1.

3.3 Intégration des actions de maintenance

3.3.1 Objectifs

Dans ce paragraphe, l'objectif est de retenir les variables critiques retenues à la suite de l'analyse faite dans les paragraphes précédents, afin de prévoir des tâches de maintenance sur ces variables du modèle. De plus, les variables retenues doivent faire partie des variables sur lesquelles les agents de maintenance EDF peuvent intervenir. Ces variables ont été choisies par les experts. Ensuite, les experts ont sélectionné des tâches de maintenance sur les variables et donc sur les composants choisis. Ces actions de maintenance peuvent être vues comme des variables du modèle graphique. Ces nœuds se situent en amont des variables sur lesquelles elles agissent. Enfin, des probabilités conditionnelles doivent être également données par les experts ou par le retour d'experience. De plus, des coûts de maintenance ont été déterminés pour chaque tâche de maintenance. Dans un premier temps, ces coûts sont définis comme des variables discrètes. Un des objectifs consiste donc à définir des politiques de maintenance, visant à réduire au maximum la probabilité de dégradation et de défaillance, en tenant compte des contraintes de coûts. Une représentation des actions de maintenance est donnée par les experts en fonction du gain apporté contre la dégradation ou la défaillance et en fonction des coûts de ces opérations de maintenance.

3.3.2 Choix des actions de maintenance

Dans ce paragraphe, nous recensons les variables critiques retenues par l'analyse et validées par les experts comme étant des variables sur lesquelles les ingénieurs de maintenance peuvent agir. Ces variables sont les suivantes :

- *présence d'impuretés* ($PI3$),
- *âge de la douille* (Ad),
- *débit inverse* (DI),
- *démontage du joint* (DJ),
- *dégradation de l'étanchéité secondaire de la bague de glissement* ($M4$),
- *coulissement difficile de la bague* ($M6$),
- *température du palier de la pompe, nombre d'excursions avec une amplitude supérieure à 10°C autour de la valeur moyenne* ($PI6$).

Les actions de maintenance, notées $A_1, \ldots, A_m$, proposées par les experts concernant les variables décrites ci-dessus sont maintenant ici listées. Les coûts de ces tâches de maintenance sont également évalués. Dans un premier temps, ces coûts sont discrets et possèdent trois modalités (faible, moyen et élevé) :

- A_1 : ajout d'un filtre (coût moyen) qui va diminuer la probabilité de *présence d'impuretés dans le circuit RCV* de l'ordre de 30 %.
- A_2 : mise en place de procédures de montage, agissant directement sur *dégradation de l'étanchéité secondaire de la bague de glissement* et sur le *coulissement difficile de la bague*. Cette action de maintenance a un faible coût et permet un gain faible (10 %) sur la première maladie et un gain fort sur la deuxième (probabilité divisée par 2, soit un gain de 50 %).
- A_3 : changement de la douille avant x années d'exploitation (coût élevé). Le gain de cette action de maintenance est calculé au prorata de la répartition des douilles. Par exemple, supposons que la probabilité que *la douille ait moins d'un an* est égale à p_1, et que la probabilité que *la douille ait entre un an et six ans* est égale à p_2 et supposons que la répartition à l'intérieur des classes est uniforme. Ainsi, si l'action de maintenance consiste à changer toutes les douilles avant 6 ans, p_1 est remplacé par

$p_1 + \frac{1-p_1-p_2}{6}$ et p_2 par $p_2 + \frac{5}{6}(1-p_1-p_2)$. Supposons maintenant que l'on change toutes les douilles avant 5 ans d'exploitation, p_1 sera alors remplacé par $p_1 + \frac{1}{5}(1 - p_1 - \frac{4}{5}p_2)$ et p_2 par $\frac{4}{5}p_2 + \frac{4}{5}(1 - p_1 - \frac{4}{5}p_2)$. Enfin si la politique de maintenance consiste à changer les douilles avant 4 ans d'exploitation, p_1 devient $p_1 + \frac{1}{4}(1 - p_1 - \frac{3}{5}p_2)$ et p_2 devient $\frac{3}{5}p_2 + \frac{3}{4}(1 - p_1 - \frac{3}{5}p_2)$.

- A_4 : mise en place de procédures d'exploitation (coût faible-moyen) agissant sur *la température du palier* et permettant un gain de 50 %.
- A_5 : de meilleures procédures de requalification (coût faible) agissant sur le débit inverse et ayant un gain de 50 %.
- A_6 : mise en place d'une maintenance conditionnelle (coût faible) agissant sur le démontage joint et ayant un gain de 33 %.

Ainsi, il est possible de définir $2^6 = 64$ politiques de maintenance, en combinant toutes les actions de maintenance recensées ci-dessus. Dans un premier temps, nous analysons ces actions de maintenance une à une, toutes choses égales par ailleurs.

3.3.3 Intégration des actions de maintenance

Le réseau bayésien du joint 1 de la pompe primaire 900 MW est présenté dans la figure 6.7 avec les actions de maintenance en amont des variables sur lesquelles les ingénieurs de maintenance peuvent intervenir. L'ajout de telles variables a été réalisé par les experts, ainsi que l'évaluation de leurs probables influences sur les variables du modèle. Les variables sur lesquelles les actions de maintenance sont susceptibles d'agir peuvent être soit des variables d'entrée, mais aussi des variables intermédiaires, et même pourquoi pas directement la variable d'intérêt (ce qui est peu fréquent). Une action de maintenance peut également influer sur plusieurs variables simultanément, et pas nécessairement des variables du même type (entrée, intermédiaire ou variable d'intérêt). Une variable peut également être influencée par plusieurs actions de maintenance. Cette dernière situation n'est pas rencontrée dans notre étude, mais peut, par exemple, mettre en compétition deux actions de maintenance agissant sur une même variable (comparaison du gain en probabilité, du gain en terme de coûts, ...). Dans toute la suite, les actions de maintenance seront supposées indépendantes les unes des autres et conditionnellement indépendantes à toutes les autres variables connaissant les variables sur lesquelles elles agissent. De plus, si une action de maintenance agit sur plusieurs variables du modèle, alors ces dernières ne pourront être dépendantes. Si tel est le cas, c'est-à-dire si les experts donnent un tel cas, il faudra choisir une seule variable sur laquelle l'action de maintenance agit. Après avoir introduit la méthodologie dans le paragraphe qui suit, nous traitons dans un deuxième temps le cas où deux (ou plus) actions de maintenance influent sur une même variable puis finissons par le cas rencontré pour notre étude. Toutes ces hypothèses permettent de faciliter les calculer de probabilités conditionnelles, et sont analogues à celles supposées au chapitre 2.

3.3.3.1 La méthodologie

Les variables définissant les actions de maintenance sont décrites comme des variables du réseau bayésien. Ces variables sont toutes des variables d'entrée. En effet, nous considérons que ces variables sont susceptibles d'agir sur les variables du modèle, mais que l'inverse n'est pas possible. De plus, comme c'est le cas dans notre étude, les actions de maintenance influent sur des paramètres d'entrée, mais également sur des variables intermédiaires. Ainsi, les nouvelles probabilités marginales des variables (influencées par ces actions de maintenance) sont remplacées dans ce nouveau réseau bayésien par des probabilités conditionnelles sachant la réalisation de l'action de maintenance ou non. Les nouvelles variables d'entrée définies par les actions de maintenance, sont des variables particulières dans le sens où aucune probabilité a priori ne leur sont attribuées, ou autrement dit on suppose, a priori, que l'on fait ou non l'action de maintenance. Ces variables, notées $A_1, \ldots, A_m$, sont donc dans tous les cas **toutes observées**. Elles sont binaires et ont pour modalités : oui et non (A_j et $\bar{A}_j$ pour $j = 1, \ldots, m$). Les probabilités conditionnelles sont donc définies comme suit :

- elles sont égales aux probabilités marginales s'il n'y a pas d'action de maintenance,
- elles sont calculées à partir des probabilités marginales, grâce au gain estimé par les experts.

Par exemple pour l'action de maintenance A_1, les experts ont donné un gain égal à 30 %, la probabilité de *présence d'impuretés dans le circuit RCV* sachant qu'il y a un filtre pour l'expert BR (la probabilité a priori était égale à $P(PI3 = 0) = 0.1$) devient $P(PI3 = 0|A_1) = 0.07$. La probabilité de *présence d'impuretés dans le circuit RCV* sachant qu'il n'y a pas de filtre est égale à $P(PI3 = 0|\bar{A}_1) = 0.1$.

3.3.3.2 Cas où deux actions influent sur une même variable

Dans le cas où plusieurs actions de maintenance agissent sur une même variable, ce qui n'est pas le cas du joint 1 de la pompe primaire 900 MW, nous supposerons alors l'indépendance conditionnelle des deux (ou plusieurs) actions de maintenance, connaissant la variable sur laquelle ces actions influent. Soit A et B deux actions de maintenance agissant sur une variable X. Soit la probabilité marginale de X, noté p. Soit g_A et g_B, respectivement le gain donné par l'expert si l'action de maintenance A est appliquée et si l'action de maintenance B est appliquée, les nouvelles probabilités conditionnelles sont $P(X|A) = p \times g_A$, $P(X|\bar{A}) = p$ et $P(X|B) = p \times g_B$, $P(X|\bar{B}) = p$. Alors, en supposant l'indépendance entre les actions de maintenance et en supposant l'indépendance conditionnelle entre les actions de maintenance connaissant la variable X (ce qui ne correspond pas à un modèle graphique), on a

$$p(X|A, B) = \frac{p(A, B|X)p(X)}{p(A)p(B)} = \frac{p(X|A)p(X|B)}{p(X)} = p \times g_A \times g_B. \tag{3.1}$$

Si la variable X est une variable d'entrée, la probabilité p est une probabilité a priori et la formule (3.1) s'applique sans difficulté. Si X est une variable intermédiaire, le gain donné par les experts correspond à un gain sur les probabilités marginales. Supposons que le nœud X ait comme nœuds parents $Y_1, \ldots, Y_n$, alors en supposant que les actions de maintenance

A et B sont indépendantes et qu'elles sont conditionnellement indépendantes à toutes les autres variables sachant X, et donc aux parents de X, on a

$$\begin{aligned} p(X|Y_1, \ldots, Y_n, A, B) &= \frac{p(Y_1, \ldots, Y_n, A, B|X)p(X)}{p(Y_1, \ldots, Y_n)p(A)p(B)} \\ &= \frac{p(Y_1, \ldots, Y_n|X)p(A|X)p(B|X)p(X)}{p(Y_1, \ldots, Y_n)p(A)p(B)} \\ &= p(X|Y_1, \ldots, Y_n)\frac{P(X|A)}{p(X)}\frac{P(X|B)}{p(X)}. \end{aligned}$$

Cette formule est analogue à l'équation (3.1), en notant respectivement la nouvelle probabilité conditionnelle de X sachant ses parents et l'ancienne p^{new} et p^{old}, on a

$$p^{new}(X|A, B) = p^{old} \times g_A \times g_B.$$

Ainsi, le gain donné par les experts correspond à un gain sur la probabilité marginale de la variable sur laquelle l'action de maintenance agit.

3.3.3.3 Autre cas

Nous traitons dans ce paragraphe le cas rencontré pour notre étude. Une action de maintenance peut concerner une ou plusieurs variables, mais plusieurs actions de maintenance ne peuvent concerner la même variable. De plus, si une action de maintenance influe sur deux variables, ces dernières doivent être indépendantes. Si tel n'est pas le cas, l'expert devra choisir parmi les deux variables, laquelle est influencée par l'action de maintenance. Avec cette convention, l'action de maintenance est indépendante des ancêtres de la variable sur laquelle elle agit (cf. Lauritzen [67], et Whittaker [98] pour les propriétés d'indépendance et d'indépendance conditionnelle dans les modèles graphiques). De plus, l'indépendance conditionnelle entre l'action de maintenance et les parents de la variable (sur laquelle influe l'action) sachant cette variable est supposée. Ainsi en notant $pa(X)$ les nœuds parents de X et A l'action de maintenance influant la variable X, on a

$$p(X|pa(X), A) = p(X|pa(X))\frac{p(X|A)}{p(X)}.$$

Soit g, la variable qui prend comme valeur le gain donné par l'expert si l'action de maintenance A est menée, et vaut 1 sinon. En notant $p^{new} = p(X|pa(X), A)$ la nouvelle probabilité conditionnelle de X et $p^{old} = p(X|pa(X))$ l'ancienne, on a

$$p^{new} = g \times p^{old}.$$

Ainsi, en supposant l'indépendance conditionnelle entre l'action de maintenance et les parents de la variable concernée, le gain donné par l'expert sur les probabilités marginales est le même pour les probabilités conditionnelles.

3.3.4 Applications

Le nouveau réseau bayésien avec les actions de maintenance est donné dans la figure 6.7. Dans ce paragraphe, une nouvelle inférence est faite avec les actions de maintenance. Ces analyses sont faites action de maintenance par action de maintenance, mais il est possible de combiner deux ou plusieurs actions de maintenance.

3.3.4.1 Mise en place d'un filtre

Cette action permet de diminuer la probabilité de *présence d'impuretés dans le circuit RCV* de 30 %. Ainsi, cette probabilité passe de 0.01 à 0.007 pour l'expert RC, et de 0.10 à 0.07 pour l'expert BR. On mesure le gain de cette action de maintenance en terme de probabilité de dégradation et de défaillance du joint 1. Les résultats sont donnés dans le tableau 3.10.

Probabilité de dégradation			
Expert RC		Expert BR	
Sans Filtre	Avec filtre	Sans Filtre	Avec filtre
0.2970	0.2968	0.2341	0.2326

Probabilité de défaillance			
Expert RC		Expert BR	
Sans Filtre	Avec filtre	Sans Filtre	Avec filtre
0.0003	0.0003	0.0102	0.0101

TAB. 3.10 – Comparaison entre experts des probabilités de dégradation et de défaillance du joint 1 en mettant en place un filtre.

On peut remarquer que cette action de maintenance a peu d'influence sur la probabilité de dégradation et de défaillance.

3.3.4.2 Mise en place de procédures de montage

Cette action permet de diminuer la probabilité de 10 % de *dégradation de l'étanchéité secondaire de la bague* et de 50 % le *coulissement difficile de la bague*. Le tableau 3.11 donne les gains obtenus sur les probabilités de dégradation et de défaillance du joint1.

Là encore cette action de maintenance a peu d'influence sur les probabilités de dégradation et de défaillance.

3.3.4.3 Changement de la douille

La même procédure est effectuée pour différentes politiques de changement de la douille (cf. tableau 3.12).

Probabilité de dégradation			
Expert RC		Expert BR	
Sans procédure	Avec procédure	Sans procédure	Avec procédure
0.2970	0.2917	0.2341	0.2244

Probabilité de défaillance			
Expert RC		Expert BR	
Sans procédure	Avec procédure	Sans procédure	Avec procédure
0.0003	0.0002	0.0102	0.0098

Tab. 3.11 – Comparaison entre experts des probabilités de dégradation et de défaillance du joint 1 en mettant en place une procédure de montage.

Probabilité de dégradation			
Expert RC		Expert BR	
Sans procédure	0.2970	Sans procédure	0.2341
Avant 6 ans	0.2961	Avant 6 ans	0.2341
Avant 5 ans	0.2948	Avant 5 ans	0.2327
Avant 4 ans	0.2931	Avant 4 ans	0.2307

Probabilité de défaillance			
Expert RC		Expert BR	
Sans procédure	0.0003	Sans procédure	0.0102
Avant 6 ans	0.0003	Avant 6 ans	0.0102
Avant 5 ans	0.0003	Avant 5 ans	0.0101
Avant 4 ans	0.0003	Avant 4 ans	0.0100

Tab. 3.12 – Comparaison entre experts des probabilités de dégradation et de défaillance du joint 1 en changeant les douilles plus souvent.

On peut en conclure que quelle que soit la politique, les probabilités de dégradation et de défaillance du joint varient très peu.

3.3.4.4 Mise en place de procédures d'exploitation

Cette action permet selon les avis d'experts de diminuer de 50 % la probabilité que la température du palier pompe varie beaucoup. Nous donnons les résultats dans le tableau 3.13.

Probabilité de dégradation			
Expert RC		Expert BR	
Sans procédure	Avec procédure	Sans procédure	Avec procédure
0.2970	0.2965	0.2341	0.2339

Probabilité de défaillance			
Expert RC		Expert BR	
Sans procédure	Avec procédure	Sans procédure	Avec procédure
0.0003	0.0003	0.0102	0.0102

Tab. 3.13 – Comparaison entre experts des probabilités de dégradation et de défaillance du joint 1 en mettant en place des procédures d'exploitation.

D'après ces résultats, on peut en conclure que la mise en place de procédures d'exploitation n'apporte pas de gain en termes de probabilités de dégradation et de défaillance du joint 1.

3.3.4.5 Meilleures procédures de requalification

Cette action permet selon les avis d'experts de diminuer de 50 % la probabilité qu'il y ait un débit inverse. Le tableau 3.14 donne les gains en probabilité.

Probabilité de dégradation			
Expert RC		Expert BR	
Sans procédure	Avec procédure	Sans procédure	Avec procédure
0.2970	0.2938	0.2341	0.2325

Probabilité de défaillance			
Expert RC		Expert BR	
Sans procédure	Avec procédure	Sans procédure	Avec procédure
0.0003	0.0003	0.0102	0.0101

Tab. 3.14 – Comparaison entre experts des probabilités de dégradation et de défaillance du joint 1 en mettant en place des meilleures procédures de requalification.

Là encore, même si l'influence de cette action de maintenance semble plus forte que les autres actions, le gain en probabilité reste faible.

3.3.4.6 Mise en place d'une maintenance conditionnelle

Cette action permet selon les avis d'experts de diminuer de 33 % la probabilité qu'il y ait un démontage joint. Nous donnons dans le tableau 3.15 les gains obtenus sur les probabilités de dégradation et de défaillance du joint 1 pour cette action de maintenance.

Probabilité de dégradation			
Expert RC		Expert BR	
Sans maintenance	Avec maintenance	Sans maintenance	Avec maintenance
0.2970	0.2943	0.2341	0.2338

Probabilité de défaillance			
Expert RC		Expert BR	
Sans maintenance	Avec maintenance	Sans maintenance	Avec maintenance
0.0003	0.0003	0.0102	0.0102

Tab. 3.15 – Comparaison entre experts des probabilités de dégradation et de défaillance du joint 1 en mettant en place une maintenance conditionnelle.

On peut conclure que cette dernière action de maintenance n'apporte pas de gain en termes de probabilités de dégradation et de défaillance du joint 1.

3.3.4.7 Avec toutes les actions de maintenance

Au vu des résultats décevants que l'on a obtenus, nous effectuons, dans ce paragraphe, une inférence en supposant que toutes les actions de maintenance sont réalisées. Les résultats sont données dans le tableau 3.16. On peut remarquer que toutes les actions de maintenance

Probabilité de dégradation			
Expert RC		Expert BR	
Sans maintenance	Avec maintenance	Sans maintenance	Avec maintenance
0.2970	0.2812	0.2341	0.2176

Probabilité de défaillance			
Expert RC		Expert BR	
Sans maintenance	Avec maintenance	Sans maintenance	Avec maintenance
0.0003	0.0002	0.0102	0.0095

Tab. 3.16 – Comparaison entre experts des probabilités de dégradation et de défaillance du joint 1 en effectuant toutes les actions de maintenance.

ont peu, voire pas d'influence sur la probabilité de dégradation et surtout de défaillance. Ceci est probablement du à la profondeur du graphe, c'est-à-dire au nombre important de variables intermédiaires entre les variables d'entrée et la variable de sortie. Ainsi, nous avons appliqué la même méthode pour une maladie, la dégradation de la douille de logement par usure (maladie la plus fréquente).

3.3.4.8 Influence des actions de maintenance sur une maladie

On choisit de faire toutes les actions de maintenance données dans le paragraphe précédent, et on regarde l'influence sur la probabilité que la douille soit dégradée ($M3 = oui$). Les résultats sont donnés dans le tableau 3.17.

Probabilité de dégradation			
Expert RC		Expert BR	
Sans maintenance	Avec maintenance	Sans maintenance	Avec maintenance
0.7	0.65	0.25	0.22

TAB. 3.17 – Comparaison entre experts des probabilités de dégradation de la douille de logement en effectuant toutes les actions de maintenance.

La maintenance permet, comme attendu, de diminuer la probabilité de la maladie de la douille, plus fortement qu'elle n'avait pu le faire pour la probabilité de dégradation ou de défaillance du joint 1. Cependant, les gains apportés, au vu des coûts de maintenance restent faibles.

3.4 Conclusions

Nous avons présenté plusieurs méthodologies d'inférence et d'analyse dans un réseau bayésien. La plupart de ces méthodes sont basées sur l'analyse de sensibilité. Nous proposons des indices simples basés sur un rapport de vraisemblance. Ce score nous permet de classer les scénarios possibles, suivant leur influence sur la variable d'intérêt, ici la dégradation et la défaillance du joint 1 d'une pompe primaire. Une fois ces variables importantes identifiées, l'aide à la décision prend part via l'intégration des actions de maintenance sur ces variables jugées importantes. Pour cela, nous considérons les tâches de maintenance comme de nouveaux nœuds du réseau bayésien. Cela permet notamment de prendre en compte le gain apporté par cette action de maintenance grâce à de nouvelles probabilités conditionnelles. Ces actions de maintenance ainsi que leurs gains et leurs coûts sont donnés par les experts. Puis, une nouvelle inférence est faite avec la prise en compte des actions de maintenance. Les résultats en termes de gain sur la probabilité de dégradation ou de défaillance sont plutôt décevants. En effet, l'inférence montre que le gain apporté par les actions de maintenance est négligeable dans la plupart des cas. Il existe à notre avis deux raisons à ces résultats.

D'une part, le nombre de variables intermédiaires entre les variables d'entrée et la variable de sortie peut être important, ce qui atténue l'influence des actions de maintenance. On a pu notamment remarquer que le gain était plus important lorsque la variable d'intérêt est une maladie (exemple du paragraphe 3.3.4.8). Par conséquent, pour réduire ce nombre, il serait nécessaire d'effectuer la même étude avec toutes les maladies, comme variables d'intérêt. D'autre part, les gains donnés par les experts, même s'ils semblent importants, n'agissent que sur des événements rares. Par exemple, un gain de 50% sur une probabilité de 0.01 est moins significatif qu'un gain de 10% sur une probabilité de 0.5. Ainsi, nous préconisons plutôt que de demander aux experts des pourcentages de gain, de leur demander directement les probabilités conditionnelles des variables sachant que l'on effectue l'action de maintenance afin que l'ordre de grandeur de l'ancienne probabilité soit prise en compte par l'expert. Enfin, s'il s'avère que les actions aient plus d'influence sur la dégradation du joint, des diagrammes de gain en probabilité et en coût peuvent être donnés. En effet, les décideurs pourront juger de l'utilité d'une action de maintenance en fonction du gain qu'elle apporte sur la probabilité de la variable de sortie, mais aussi du coût qu'elle engendre.

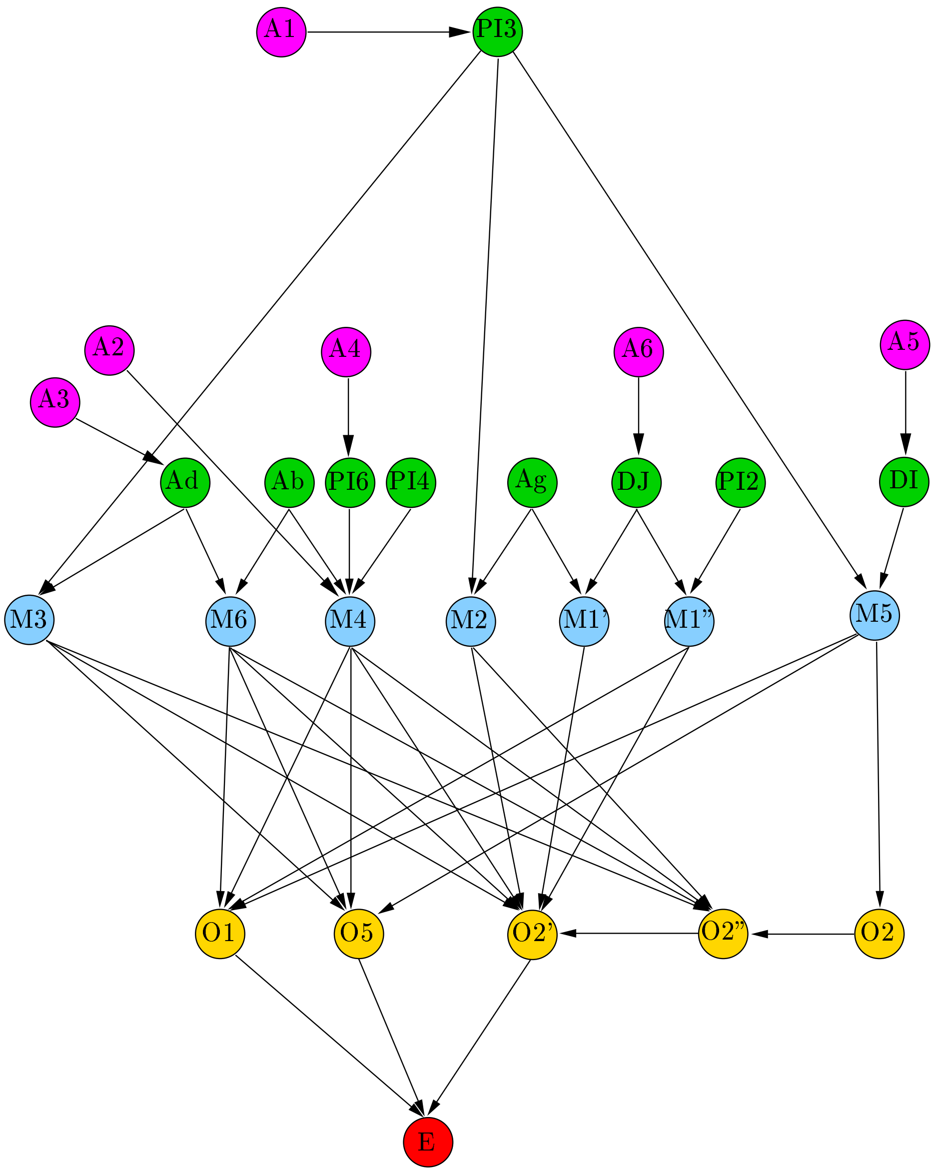


FIG. 3.7 – Réseau bayésien pour le joint 1 de la pompe primaire 900 MW avec intégration des actions de maintenance.

Chapitre 4

Intégration du retour d'expérience dans les réseaux bayésiens

Dans ce chapitre, notre intérêt porte sur l'intégration des données de retour d'expérience dans un réseau bayésien pour la mise à jour de ses paramètres. Dans toute la suite, la structure du graphe est supposée connue et fixée, les paramètres se résumant aux probabilités marginales et conditionnelles. Pour notre étude, la structure et les probabilités servant à l'inférence du modèle proviennent des avis d'experts. Ceci est justifié par le fait que les réseaux bayésiens sont généralement utilisés comme des systèmes experts capable de capitaliser et de modéliser la connaissance. Puis au fur et à mesure que le temps passe et que l'expérience grandit, des données de retour d'expérience sont répertoriées. Ces données doivent obligatoirement enrichir le réseau bayésien. Ainsi, les avis d'experts qui ont servi à évaluer les probabilités du graphe, sont dans cette optique considérées comme les probabilités a priori dans un cadre probabiliste bayésien habituel. Avec les données de retour d'expérience, la probabilité a posteriori est calculée, via la formule de Bayes. Ce chapitre s'attache également à paramétrer la confiance que l'on peut avoir dans les avis d'experts. Les modèles de Dirichlet imprécis [94] permettent une telle paramétrisation. Nous proposons pour modéliser cette confiance d'utiliser une loi a priori non informative, la loi de Jeffreys [57]. Puis, les paramètres de cette loi sont calibrés de telle sorte que la loi a posteriori soit la loi a priori correspondant aux avis d'experts. Autrement dit, les avis d'experts sont vus comme un échantillon fictif où la taille de l'échantillon va donner la confiance que l'on a dans les avis d'experts.

Dans notre cas, les variables étudiées sont toutes discrètes, et pour la plupart binaires, les lois des variables sont donc multinomiales, et pour la plupart binomiales. Dans un cadre bayésien, les lois a priori choisies sont en règle générale des lois conjuguées afin de faciliter les calculs. Dans le cas où les lois a priori sont non conjuguées, des simulations de Monte-Carlo sont en général nécessaires afin de calculer, par exemple, le maximum a posteriori. Ici les lois a priori conjuguées choisies sont des lois de Dirichlet dans le cas multinomial et des lois Beta dans le cas binomial. Beaucoup d'articles traitent du problème de choix de lois a priori pour des variables binomiales ou multinomiales. On peut citer par exemple le document de Laskey et Mahoney [65] pour l'intégration des données dans les réseaux bayésiens, Bernard [15] pour une description bayésienne des procédures fréquentistes, Walley [94] pour l'étude de modèles de Dirichlet imprécis, Ramoni and Sebiastani [83], pour les problèmes à données manquantes.

Le présent chapitre est organisé comme suit. Dans un premier temps, nous introduisons l'inférence bayésienne, puis exposons pour notre étude le moyen de "fusionner" les avis d'experts avec les données de retour d'expérience. Puis, une brève étude comparative entre les lois a priori est exposée. Puis, nous proposons une procédure simple afin de modéliser la confiance que l'on a dans les avis d'experts. Enfin, nous appliquons ces méthodes dans des cas simples, avec notamment une application sur le joint 1 d'une pompe 1300 MW.

4.1 Introduction à l'inférence bayésienne

Dans un premier temps, nous rappelons le cadre bayésien habituel. En pratique, on dispose de quelques informations sur l'évènement que l'on étudie. L'approche bayésienne fournit une méthode pour inclure ces informations dans notre modèle. Soit $(X_1, \ldots, X_n)$ un échantillon de variables indépendantes et identiquement distribuées, ayant pour densité de probabilité $f(x;\theta)$, où θ est un élément de Θ, espace des paramètres. L'approche bayésienne consiste à supposer que les paramètres d'intérêt θ sont aléatoires, de loi de densité $g(\theta)$, dite loi *a priori*. Ainsi, la densité marginale de X, appelé densité prédictive, est alors :

$$f^*(x) = \int_\Theta f(x;\theta) g(\theta)\, d\theta.$$

La densité conditionnelle de θ sachant les données observées, appelée loi *a posteriori*, est la densité de θ sachant $X = x$, c'est-à-dire d'après le théorème de Bayes :

$$g(\theta/x_1, \ldots, x_n) = \frac{g(x;\theta)}{f^*(x)} = \frac{\prod_{k=1}^{n} f(x_k;\theta) g(\theta)}{\int_\Theta \prod_{k=1}^{n} f(x_k;\theta) g(\theta)\, d\theta} \tag{4.1}$$

Il nous faut maintenant définir une fonction de perte $L(\hat{\theta}, \theta)$ représentant le coût de décider $\hat{\theta}$ alors que la vraie valeur du paramètre est θ. Dans la majorité des cas, on choisit une fonction de perte comme une distance entre l'estimateur et la vraie valeur :

$$L(\hat{\theta}, \theta) = D\left(\hat{\theta} - \theta\right)$$

En prenant $D(x) = x^2$ ou $D(x) = |x|$, on parle respectivement d'erreur quadratique ou d'erreur absolue.

Définition 14 *On définit le risque a posteriori comme la perte moyenne par rapport à la distribution a posteriori, i.e*

$$R(\hat{\theta}, x_1, \ldots, x_n) = \int_{\Theta} L(\hat{\theta}, \theta) g(\theta / x_1, \ldots, x_n)\, d\theta.$$

L'estimateur de Bayes, associé à la fonction de perte $L(\hat{\theta}, \theta)$, est alors la valeur $\hat{\theta}_{Bayes}$ qui minimise le risque a posteriori sachant $x_1, \ldots, x_n$.

Si on utilise la fonction de perte quadratique, l'estimateur bayésien correspond à la moyenne a posteriori. Nous nous placerons toujours dans ce cas là.

4.2 Fusion entre connaissance et données dans un réseau bayésien

Soit un réseau bayésien, ou modèle graphique acyclique orienté dont la structure est connue, et soit le vecteur de variables aléatoires $X = (X_1, \ldots, X_v)$, représentés par les nœuds du graphe. Les variables du modèle sont toutes supposées discrètes. Nous étudions dans un premier temps le cas de variables binomiales pour ensuite généraliser l'étude au cas multinomial. La probabilité jointe s'écrit selon la factorisation récursive, et donc nous raisonnerons dans toute la suite sur un nœud du graphe et ses parents.

4.2.1 Variables binomiales

Supposons que toutes les variables du réseau bayésien suivent une loi de Bernouilli, notée $\mathcal{B}(\theta)$. On observe ici un échantillon de taille n pour cette variable. La loi de la somme est donc une loi binomiale, notée $\mathcal{B}(n, \theta)$. La loi a priori conjuguée de la loi binomiale est la loi Beta, noté $\mathcal{BE}(\alpha_1, \alpha_2)$, ayant comme moyenne

$$\mathbb{E}[\theta] = \frac{\alpha_1}{\alpha_1 + \alpha_2}$$

et comme variance

$$\mathbb{V}[\theta] = \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2 (\alpha_1 + \alpha_2 + 1)}.$$

On désigne par $X_{pa(i)}$, la suite des variables aléatoires représentant les parents de X_i. On note j l'indice de l'état $x_{i,j}$ de X_i et c_i l'indice des configurations des nœuds parents du nœud X_i. Soit $\theta_{ijc_i} = p(X_i = x_{ij} | X_{pa(i)} = x_{ic_i})$, la distribution jointe (avec v sommets) du réseau bayésien peut s'écrire sous la forme

$$P(x_{1j_1}, \ldots, x_{vj_v}) = \prod_{i=1}^{v} \theta_{ij_ic_i}.$$

Notons que pour les nœuds racines, $c_i = \emptyset$, et que la probabilité correspondante est une probabilité marginale. Dans notre étude, ces probabilités sont données par l'expert et constituent donc dans la théorie bayésienne les probabilités a priori, mais qui, rappelons-le, sont pour la plupart des probabilités conditionnelles. Dans un cadre bayésien, les paramètres θ_{ic_i} suivent une loi a priori (ici donnée par les experts), fonction d'hyperparamètres. Le choix de la loi a priori est d'une importance capitale. En règle générale et dans un souci de simplification, les lois a priori choisies sont des lois conjuguées, de telle sorte que la loi a posteriori soit de la même forme. La loi a priori Beta des paramètres θ a pour densité sur le support $[0, 1]$:

$$f(\alpha_1, \alpha_2) = \frac{1}{B(\alpha_1, \alpha_2)}(\theta)^{\alpha_1 - 1}(1-\theta)^{\alpha_2 - 1}$$

où B est la fonction Beta qui vaut $\dfrac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$ avec $\Gamma(x) = \int_0^{+\infty} e^{-t}t^{x-1}dt$ désignant la fonction Gamma.
Supposons que pour une variable, on observe a cas favorables sur n données, alors la vraisemblance du paramètre θ associé à cette variable s'écrit :

$$C_n^a \theta^a (1-\theta)^{n-a}.$$

La loi a posteriori est donc égale à

$$\frac{\theta^{\alpha_1+a-1}(1-\theta)^{n-a+\alpha_2-1}}{\int_0^1 \theta^{\alpha_1+a-1}(1-\theta)^{n-a+\alpha_2-1}d\theta} = \frac{\Gamma(\alpha_1+\alpha_2+n)}{\Gamma(\alpha_1+a)\Gamma(\alpha_2+n-a)}\theta^{\alpha_1+a-1}(1-\theta)^{\alpha_2+n-a-1},$$

i.e. une loi Beta $\mathcal{BE}(\alpha_1 + a, \alpha_2 + n - a)$.
Supposons que l'on ait aucune connaissance sur ce phénomène et que l'on choisisse comme loi a priori non informative la loi Beta $\mathcal{BE}(1, 1)$, égale à la loi uniforme, ce qui semble assez naturel. Alors la moyenne a posteriori, qui est l'estimateur bayésien pour une fonction de perte quadratique, vaut $\hat{\theta} = \dfrac{a+1}{n+2}$. Cette valeur diffère de l'estimateur empirique $\dfrac{a}{n}$. Ainsi, comme l'ont déjà fait remarquer Haldane [53] et Jeffreys [58], la loi a priori uniforme, contrairement aux idées reçues, introduit un biais dans l'estimation. De plus, la loi uniforme n'existe que sur des espaces bornés. Il est donc préférable d'utiliser des lois généralisées. Une autre critique est plus fondamentale. Elle concerne le problème de l'*invariance par reparamétrisation*. Ce principe dit que si on passe d'un paramètre θ à un autre paramètre $\eta = g(\theta)$ par une transformation bijective g, l'information a priori ne doit pas être modifiée. Pour la loi uniforme, ce principe n'est pas vérifié.

Haldane [53] suggère de prendre comme loi a priori la loi limite $\mathcal{BE}(0, 0)$ ayant pour densité :

$$f(\theta) = \frac{1}{\theta(1-\theta)},$$

ce qu'il justifie par le fait que pour les événements rares, ce n'est plus θ qui suit une loi uniforme, mais seules les petites valeurs sont équiprobables. Dans ce cas la loi a posteriori est une loi Beta $\mathcal{BE}(a, n-a)$, et donc l'estimateur bayésien (moyenne a posteriori) est sans biais ($\hat{\theta} = \frac{a}{n}$).

Jeffreys [57] pour sa part s'est attaqué au principe d'*invariance par reparamétrisation* et a généralisé ce principe en se basant sur l'information de Fisher. La loi a priori associée est définie par

$$\pi(\theta) = I^{1/2}(\theta), \tag{4.2}$$

où $I(\theta) = \mathbb{E}_\theta\left[(\frac{\partial log f(X;\theta)}{\partial\theta})^2\right]$ est l'information de Fisher dans le cas unidimensionnel. Ce principe de maximisation de l'information de Fisher a été repris et généralisé par Bernardo [17] pour construire des lois a priori de référence. Le principe est de séparer les paramètres en deux types : les paramètres d'intérêt et les paramètres de nuisance, afin d'éviter le paradoxe de marginalisation. Afin de comprendre ce paradoxe, supposons que l'on ait un échantillon $x = (x_1, \ldots, x_n)$ avec un paramètre d'intérêt η. Supposons également que la loi a posteriori de η, notée $\pi(\eta|x)$ ne dépende que d'une variable z, fonction de x. Le paradoxe vient du fait qu'il arrive que l'on ne puisse écrire $\pi(\eta|x)$ en fonction de la loi a posteriori de z, notée $\pi(z|x)$ et ce pour toute loi a priori $\pi(\eta)$.

La loi de Jeffreys qui est donc la loi de référence est généralement une loi impropre. Dans notre exemple, la loi de Jeffreys correspondante est la loi a priori Beta, $\mathcal{BE}(1/2, 1/2)$. Ainsi, l'estimateur bayésien vaut $\frac{a+\frac{1}{2}}{n+1}$. Cet estimateur est également biaisé, mais cependant avec un biais plus faible que dans le cas de la loi a priori uniforme.

Dans un cas pratique, l'expert est capable de fournir la valeur $\frac{\alpha_1}{\alpha_1+\alpha_2}$ qui s'interprète comme une proportion. Comme on le verra dans le paragraphe 4.4, la constante de normalisation $\alpha_1 + \alpha_2$ permettra de régler la confiance sur les avis d'experts. On peut maintenant examiner la situations pour des variables multinomiales.

4.2.2 Variables multinomiales

On suppose maintenant que les variables $X_1, \ldots, X_v$ sont multinomiales. Pour chaque variable X_i, les modalités sont $1, \ldots, k_i$. La probabilité jointe du réseau bayésien garde la même forme et s'écrit

$$P_{jointe}(x_{1j_1}, \ldots, x_{vj_n}) = \prod_{i=1}^{v} \theta_{ij_ic_i}.$$

Dans toute la suite du chapitre, nous raisonnerons sur un seul nœud et une seule configuration de ses parents. Ainsi, il est possible d'abandonner les indices afin d'alléger la notation. Le paramètre est donc noté $\theta = (\theta_1, \ldots, \theta_j, \ldots, \theta_k)$. Dans l'article de Laskey et

Mahoney [65] et dans bon nombre d'articles, la loi a priori choisie pour θ est une loi de Dirichlet (la généralisation de la loi Beta dans le cas multinomial) avec pour hyperparamètres $\alpha = (\alpha_1, \dots, \alpha_k)$. La densité de la loi de Dirichlet est la suivante avec $\displaystyle\sum_{j=1}^{k} \theta_j = 1$:

$$f(\theta_1, \dots, \theta_k) = \frac{\Gamma(\sum_j \alpha_j)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)} \theta_1^{\alpha_1 - 1} \dots \theta_k^{\alpha_k - 1}.$$

L'espérance mathématique d'une loi de Dirichlet est égale à

$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum_{j'} \alpha_{j'}}.$$

Dans un tel contexte et par analogie au cas binomial, on montre que la loi a posteriori est une loi de Dirichlet de paramètres $(\alpha_1 + n_1, \dots, \alpha_k + n_k)$, où $n = (n_1, \dots, n_k)$ représentent les observations de l'échantillon, distribué suivant une loi multinomiale de paramètre $(\theta_1, \dots, \theta_k)$. Autrement dit, la loi de Dirichlet est une loi conjuguée pour le cas multinomial. En effet, la vraisemblance du paramètre d'une loi multinomiale s'écrit avec $n = \sum_{j=1}^{k} n_j$ (cf. Saporta [86])

$$\mathcal{L}(n|\theta) = \frac{n!}{n_1! \dots n_k!} \theta_1^{n_1} \dots \theta_k^{n_k},$$

et donc la loi jointe

$$\mathcal{L}(n, \theta) = \frac{n!}{n_1! \dots n_k!} \frac{\Gamma(\sum_j \alpha_j)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)} \theta_1^{n_1 + \alpha_1 - 1} \dots \theta_k^{n_k + \alpha_k - 1}.$$

La densité a posteriori s'écrit alors

$$f(\theta|n) = \frac{\Gamma(\sum_j \alpha_j + n_j)}{\Gamma(\alpha_1 + n_1) \dots \Gamma(\alpha_k + n_k)} \theta_1^{\alpha_1 + n_1 - 1} \dots \theta_k^{\alpha_k + n_k - 1}.$$

Si la fonction de perte choisie est quadratique, l'estimateur bayésien correspond à la moyenne de la distribution a posteriori :

$$\hat{\theta}_j = \frac{\alpha_j + n_j}{\displaystyle\sum_{j'} \alpha_{j'} + \sum_{j'} n_{j'}}.$$

Dans cette dernière formule, $\sum_j \alpha_j$ est une quantité à fixer. En effet, l'expert donne des proportions, et non des effectifs. Il est à noter que $\sum_j \alpha_j$ représente la taille de l'échantillon virtuel, donné par l'expert. Comme on le détaillera en 4.4.2, cette taille virtuelle peut s'interpréter comme un niveau de confiance sur les dires de l'expert. Si cette taille est élevée, le poids donné aux avis d'expert va être important comparé aux données. En revanche, si elle est très faible, l'estimateur bayésien prendra peu en compte les avis d'expert.

4.3 Quelle loi a priori ?

Dans tout ce paragraphe, nous étudions les différentes possibilités de lois a priori dans le cas de variables multinomiales (le cas binomial étant un cas particulier du cas multinomial). On peut distinguer pour ce problème deux cas : soit les experts ont un avis sur le phénomène étudié et l'on appliquera une loi de Dirichlet tenant compte de ces avis avec la nécessité de quantifier la confiance en eux. Soit les experts n'ont pas d'avis et on appliquera une loi non informative. Dans ce dernier cas, il existe plusieurs choix de lois non informatives, que l'on décrit ci-après. Parmi ces lois a priori non informatives, certaines sont dites impropres, c'est-à-dire qu'elles ne définissent pas de lois de probabilité (cf. Robert [84]).

Définition 15 *Une loi a priori $\pi(\theta)$ d'un paramètre θ appartenant au support Θ est impropre si*

$$\int_{\Theta} \pi(\theta) d\theta = +\infty.$$

4.3.1 La loi uniforme de Bayes-Laplace

Elle correspond à la loi de Dirichlet $\mathcal{D}(1, \ldots, 1)$ sur le simplex $(\{x_i \geq 0 \textit{ tel que } \sum_i x_i \leq 1\})$. Dans ce cas, la loi a posteriori est une loi de Dirichlet $\mathcal{D}(n_1 + 1, \ldots, n_{k_i} + 1)$, soit un estimateur bayésien pour un coût quadratique égal à $\hat{\theta}_j = \dfrac{n_j + 1}{n + k_i}$, qui induit un biais qui disparaît asymptotiquement (lorsque $n \to +\infty$). L'inconvénient de cette loi est son support borné et qu'elle ne vérifie pas le principe d'invariance par paramétrisation.

4.3.2 Le cas limite de Haldane [53]

Ce cas limite correspond à une loi a priori de Dirichlet $\mathcal{D}(0, \ldots, 0)$, qui est une loi impropre et menant à une loi a posteriori de Dirichlet $\mathcal{D}(n_1, \ldots, n_{k_i})$, soit un estimateur bayésien $\hat{\theta}_j = \dfrac{n_j}{n}$ (estimateur sans biais) qui vérifie le principe de compatibilité de Villegas [92]. Ce cas limite pose de gros problèmes notamment lorsqu'un évènement n'est pas réalisé (il existe j tel que $n_j = 0$). En effet, supposons que l'on réalise n expériences, et que l'on observe 0 succès. La loi a posteriori est alors une loi impropre. Elle s'écrit

$$\frac{\theta^{-1}(1-\theta)^{n-1}}{\int_0^1 \theta^{-1}(1-\theta)^{n-1}\, d\theta}.$$

Dans le cas d'un réseau bayésien comme le nôtre, le nombre de données est particulièrement faible. Ainsi, pour un nœud donné dans le réseau avec un nombre de parents important, la probabilité d'avoir dans les bases de données des cases vides est importante. Cet inconvénient est majeur, car on ne peut alors effectuer d'inférence bayésienne, et l'on va voir que cet handicap disparaît avec la loi de Jeffreys.

4.3.3 La loi de Jeffreys

Cette loi a priori impropre correspond à une loi a priori de Dirichlet $\mathcal{D}(1/2, \ldots, 1/2)$ et conduit à une loi a posteriori de Dirichlet $\mathcal{D}(n_1 + 1/2, \ldots, n_{k_i} + 1/2)$, soit un estimateur bayésien $\hat{\theta}_j = \dfrac{n_j + 1/2}{n + k_i/2}$. Cet estimateur présente un faible biais (plus faible que pour la loi uniforme) qui disparaît rapidement avec les données (lorsque $n \to +\infty$). De plus, cette loi est la loi de référence au sens de Bernardo (cf. page 119 dans [17]). Ainsi, si le paramètre d'intérêt est une fonction bijective de θ, alors la loi a priori sur le nouveau paramètre reste inchangée. Cette propriété se nomme l'invariance par reparamétrisation. Enfin, si l'on observe 0 succès sur n expériences, la loi a posteriori est une loi Beta $\mathcal{BE}(1/2, n + 1/2)$ et s'écrit :

$$\frac{\theta^{-1/2}((1-\theta)^{n-1/2})}{\int_0^1 \theta^{-1/2}(1-\theta)^{n-1/2}\, d\theta},$$

qui n'est pas une loi impropre. Ainsi, une inférence bayésienne peut être faite. C'est la loi que nous préconisons.

4.3.4 La loi de Perks [81]

Cette loi impropre correspond à une loi a priori de Dirichlet $\mathcal{D}(1/k_i, \ldots, 1/k_i)$ et conduit à une loi a posteriori de Dirichlet $\mathcal{D}(n_1 + 1/k_i, \ldots, n_{k_i} + 1/k_i)$, soit un estimateur bayésien $\hat{\theta}_j = \dfrac{n_j + 1/k_i}{n + 1}$ et donc qui conduit à un faible biais. Remarquons que cette loi correspond à la loi de Jeffreys dans le cas binomial.

4.3.5 Le modèle de Dirichlet imprécis

Ce modèle consiste à faire dépendre la loi a priori de Dirichlet d'un paramètre et de faire varier ce paramètre. Cette approche a été proposée par Walley [94]. Soit une distribution a priori de Dirichlet paramétrée comme suit $\mathcal{D}(s, \mathbf{t})$, où s est un réel strictement positif et $\mathbf{t} = (t_1, \ldots, t_{k_i})$, avec $0 < t_j < 1$ pour tout j et $\sum_{j=1}^{k_i} t_j = 1$. Pour reprendre les notations précédentes, on a $\alpha_j = st_j$. La moyenne a priori de θ_j est donc égale à t_j. Ce paramètre s comme nous le verrons dans le paragraphe suivant va nous permettre, soit de calibrer le degré d'imprécision en cas d'ignorance des experts, soit de calibrer la confiance données aux avis d'experts. Ainsi, la moyenne a posteriori est égale à $\dfrac{n_j + st_j}{n + s}$. Il est possible de faire varier t_j de 0 à 1. On obtient alors deux bornes pour la moyenne a posteriori, une borne supérieure

$$\hat{\theta}_j^{sup} = \frac{n_j + s}{n + s},$$

et une borne inférieure

$$\hat{\theta}_j^{inf} = \frac{n_j}{n + s}.$$

On peut interpréter s, comme un nombre de données non observées. La borne supérieure correspond à ce que toutes les données non observées aient comme valeur x_j. La borne inférieure correspond à ce que toutes les données non observées aient pris des valeurs différentes de x_j. Walley [94] définit également un degré d'imprécision sur cet évènement égal à

$$\hat{\theta}_j^{sup} - \hat{\theta}_j^{inf} = \frac{s}{n+s} \longrightarrow 0 \quad \text{quand } n \to +\infty.$$

Ce degré d'imprécision tend vers 0 lorsque le nombre de données augmente. Si aucune donnée n'est disponible alors $\hat{\theta}_j^{sup} = 1$ et $\hat{\theta}_j^{inf} = 0$, ce qui correspond à une imprécision maximale égale à 1. De plus, la quantité s va permettre de calibrer la confiance en les avis d'experts. En effet, plus s est grand et plus on accorde de confiance à l'opinion d'experts.

4.3.6 Conclusions sur les différentes lois a priori

Dans ce paragraphe, nous nous limitons au cas où les avis d'experts n'existent pas, le cas contraire étant étudié dans le paragraphe suivant. Dans un tel cas, il paraît légitime de choisir une loi a priori de Dirichlet symétrique, c'est-à-dire pour des variables multinomiales à k_i modalités, une loi de Dirichlet $\mathcal{D}(s/k_i, \ldots, s/k_i)$, ne privilégiant ainsi aucune modalité. Dans ce cas, il est possible de retrouver les lois a priori citées précédemment, $s = k_i$ pour la loi uniforme, $s \to 0$ pour le cas limite de Haldane, $s = k_i/2$ pour la loi non informative de Jeffreys et $s = 1$ pour la loi de Perks. Pour une discussion plus approfondie sur ces valeurs on pourra regarder les travaux de Bernard [16] et Good [50]. En ce qui nous concerne, nous privilégions la loi non informative de Jeffreys, son faible biais, et pour sa stabilité numérique (au cas où une modalité ne soit pas observée) et surtout pour son invariance par reparamétrisation.

4.4 Quelle confiance attribuer aux avis d'experts ?

Afin de quantifier la confiance que l'on peut attribuer aux avis d'experts, nous allons procéder de la manière suivante. Nous effectuons une analyse bayésienne, où l'a priori sera non informatif, considérant qu'aucune information n'existe sur le phénomène. Puis, nous allons estimer les hyperparamètres de façon à ce que la loi a posteriori modélise les avis d'experts. Autrement dit, les avis d'experts seront représentés par un échantillon fictif, où la taille de cet échantillon va modéliser la confiance attribuée aux avis d'experts.

4.4.1 Dans le cas binomial

Dans ce cas, supposons tout d'abord que l'avis d'experts soit inexistant. Si on choisit comme loi a priori non informative la loi uniforme sur $[0, 1]$ pour le paramètre θ, cela correspond pour $(\theta, 1-\theta)$ à une loi beta $\mathcal{B}(1, 1)$. Supposons maintenant que l'on soit en présence d'une échantillon virtuel avec comme données, $\alpha_1 - 1$ pour $X_i = x_i^1$ et $\alpha_2 - 1$ pour $X_i = x_i^2$. On note $s = \alpha_1 + \alpha_2$. Ainsi, la loi a posteriori est une loi Beta $\mathcal{B}(\alpha_1, \alpha_2)$. On a donc directement la taille de l'échantillon fictif associé à une loi a priori $\mathcal{B}(\alpha_1, \alpha_2)$, elle est de $\alpha_1+\alpha_2-2$. Autrement dit, une loi a priori $\mathcal{B}(\alpha_1, \alpha_2)$ modélisant les avis d'experts correspond à un échantillon fictif de taille $s-2$, c'est-à-dire, la confiance donné à l'expert équivaut à un échantillon de taille $s-2$.

Si l'on suppose que la loi de référence pour décrire l'absence d'information sur un phénomène est la loi de Jeffreys, alors afin d'obtenir comme loi a priori, une loi beta $\mathcal{B}(\alpha_1, \alpha_2)$, cela revient à un échantillon fictif, où $\alpha_1 - 1/2$ données ont pour modalité x_1 et $\alpha_2 - 1/2$ données ont pour modalité x_2. La taille de l'échantillon fictif est donc de $s-1$. Généralement les experts donnent des proportions, par exemple dans le cas binomiale, p pour la modalité x_1 et $1-p$ pour x_2. Ainsi, la loi a priori sera une loi beta $\mathcal{B}(sp, s(1-p))$, où $s-1$ représente la taille de l'échantillon fictif, correspondant à la confiance dans les dires de l'expert.

4.4.2 Dans le cas multinomial

Dans le cas d'une loi multinomiale, le même raisonnement peut être fait. Une variable X_i prend des valeurs $(x_i^1, \ldots, x_i^{k_i})$, avec des probabilités $(\theta_{i1}, \ldots, \theta_{ik_i})$. Supposons qu'il n'y ait aucun avis d'expert et donc, dans un cadre bayésien, que les paramètres suivent une loi de Dirichlet $\mathcal{D}(1, \ldots, 1)$. Si on observe un échantillon fictif avec comme observations $(\alpha_{i1}-1, \ldots, \alpha_{ik_i}-1)$, alors la loi a posteriori est une Dirichlet $\mathcal{D}(\alpha_{i1}, \ldots, \alpha_{ik_i})$. Ainsi, un avis d'expert modélisé par une loi a priori de Dirichlet avec de tels paramètres peut s'interpréter comme la réalisation d'un échantillon fictif, de taille $\sum_{j=1}^{k_i} \alpha_{ij} - k_i$, en partant d'une loi a priori non informative. Remarquons que les termes α_{ij} doivent être supérieurs ou égaux à 1. Dans le cas où tous ces termes sont égaux à 1, cela correspond au cas où il n'existe pas d'avis d'expert. En effet l'échantillon fictif est de taille nulle avec aucune observation. La loi a priori est alors une Dirichlet $\mathcal{D}(1, \ldots, 1)$. Pour qu'il y ait un avis d'expert, il faut que $\sum_{j=1}^{k_i} \alpha_{ij} \geq k_i + 1$, ce qui correspond à une taille d'échantillon fictif au moins égale à 1. Ce raisonnement est valable uniquement en considérant que la loi non informative classique est la loi uniforme.

Si l'on modélise la non information par la loi de Jeffreys alors pour obtenir une loi a priori informative $\mathcal{D}(\alpha_{i1}, \ldots, \alpha_{ik_i})$, il est équivalent de considérer les avis d'experts comme la réalisation d'un échantillon fictif. Cet échantillon serait composé de $s - k_i/2$ données, dont $\alpha_1 - 1/2$ pour la première modalité, $\alpha_2 - 1/2$ pour la deuxième, etc, et $\alpha_{k_i} - 1/2$ pour la dernière. Ainsi, quand l'expert donne des proportions $(p_1, \ldots, p_{k_i})$, la loi a priori correspondante est $\mathcal{D}(sp_1, \ldots, sp_{k_i})$, où $s-k_i/2$ représente le nombre de données fictives, qui est équivalent à la confiance pour cet avis. Là encore, nos préférences vont pour la loi non informative de Jeffreys et donc pour le cas que l'on vient de décrire.

Cette traduction en termes d'échantillon fictif d'un avis d'expert va permettre de calibrer facilement l'importance que l'on veut lui donner vis-à-vis des données de retour d'expérience. Cette importance est mesurée en taille d'échantillon virtuel comparée à la taille de l'échantillon réel du retour d'expérience. En pratique, il sera assez simple par exemple de convertir un nombre d'années d'expérience de l'expert sur le phénomène étudié en taille d'échantillon fictif qu'il aurait rencontré au cours de sa carrière. Une fois cette taille déterminée, il suffit de rapporter les proportions données par l'expert pour calculer tous les hyperparamètres. De plus, chaque paramètre s peut être différent pour un même expert pour différents nœuds du graphe.

4.4.3 Exemple d'application

Nous appliquons la méthodologie présentée dans ce chapitre à un exemple fictif, et à un composant d'une pompe primaire 1300 MW. Pour cela, nous avons repris le même réseau bayésien que pour la pompe 900 MW. En effet, par manque de données sur le joint 1 d'une pompe primaire 900 MW, nous n'avons pas pu appliquer la méthode sur le réseau bayésien de ce travail de thèse.

4.4.3.1 Taille fictive en fonction de l'expérience de l'expert

Rappelons que l'on quantifie la confiance que l'on attribue dans un avis d'expert par un paramètre s, représentant une taille d'échantillon fictif. Ainsi, supposons que l'expert ait dix ans d'expérience sur un phénomène analogue et que par conséquent on considère que cela correspond à m données, avec m assez grand. On choisira alors comme paramètre de confiance $s = m + 1$.

On souhaitait illustrer cette méthodologie sur le réseau bayésien du joint 1 d'une pompe primaire 900 MW. Cependant, EDF ne possédait pas de données réelles disponibles pour cette pompe. Ainsi, nous allons appliquer la méthode sur une pompe primaire 1300 MW, où l'on supposera que le réseau bayésien est identique à celui d'une pompe primaire 900 MW (cf. figure 6.7). Les données de retour d'expérience disponibles concernent la dégradation de l'étanchéité secondaire de la bague de glissement ($M4$), l'âge de la bague (Ab) et la température du palier pompe ($PI6$). Toutes ces variables sont binaires. Le tableau 4.1 donne les probabilités données par les experts.

$p(Ab = 1)$	0.5
$p(PI6 = 1)$	0.6
$p(M4 = 1 \vert Ab = 1, PI6 = 1)$	0.0833
$p(M4 = 1 \vert Ab = 1, PI6 = 2)$	0.125
$p(M4 = 1 \vert Ab = 2, PI6 = 1)$	0.0833
$p(M4 = 1 \vert Ab = 2, PI6 = 2)$	0.125

TAB. 4.1 – Probabilités données par les experts pour le réseau bayésien d'une pompe primaire 1300 MW.

Ces avis d'experts correspondent à une probabilité que l'étanchéité secondaire soit dégradée ($M4 = 1$) égale à 0.1. On va s'intéresser par exemple à la probabilité conditionnelle $p(M4 = 1|Ab = 2, PI6 = 2)$. La confiance dans les avis d'experts est mesuré par le paramètre $s - 1$, taille d'un échantillon fictif. Prenons ici $s = 2$, c'est-à-dire une taille d'échantillon fictif égale à 1. Ainsi, la probabilité a priori correspondant aux avis d'experts est une loi Beta, $\mathcal{BE}(0.25, 1.75)$.

Des données de retour d'expérience ont été récupérées pour la centrale nucléaire de Belleville. Elles sont au nombre de 14 et sont données dans le tableau 4.2.

		Âge de la bague				
		≤ 1 an		> 1 an		
	PI6	faible	importante	faible	importante	total
Dégradation de l'étanchéité secondaire	oui	0	0	0	2	2
	non	0	0	7	5	12
	total	0	0	7	7	14

TAB. 4.2 – Données de REX pour le joint 1 d'une pompe primaire 1300 MW.

La loi a posteriori de la probabilité conditionnelle $p(M4 = 1|Ab = 2, PI6 = 2)$ est donc une loi Beta, $\mathcal{BE}(0.25 + 2, 1.75 + 5)$. L'estimateur bayésien de la probabilité conditionnelle, $p^{new}(M4 = 1|Ab = 2, PI6 = 2)$, sera la moyenne a posteriori

$$p^{new}(M4 = 1|Ab = 2, PI6 = 2) = \frac{2.25}{2.25 + 6.75} = \frac{1}{4} = 0.25,$$

ce qui diffère de l'ancienne valeur de cette probabilité, qui était de 0.125 (exactement le double). Celle donnée par le retour d'expérience est égale à $\frac{2}{7} = 0.2857$. Ainsi, on peut remarquer que le poids donné aux avis d'experts est assez faible. Afin de rendre ce poids plus fort, on peut pour cette probabilité conditionnelle prendre une taille d'échantillon fictif égale au nombre de données (ici 7), c'est-à-dire $s = 8$. La loi a priori devient alors une loi Beta, $\mathcal{BE}(1, 7)$. La loi a posteriori de la probabilité conditionnelle $p(M4 = 1|Ab = 2, PI6 = 2)$ est donc une loi Beta, $\mathcal{BE}(1+2, 7+5)$. L'estimateur bayésien de la probabilité conditionnelle, $p^{new}(M4 = 1|Ab = 2, PI6 = 2)$, sera la moyenne a posteriori

$$p^{new}(M4 = 1|Ab = 2, PI6 = 2) = \frac{3}{3 + 12} = \frac{1}{5} = 0.2.$$

On remarque dans ce cas, que la moyenne a posteriori est égale au barycentre de la moyenne a priori (0.125) et de celle donnée par le REX (0.2857), ce qui est cohérent avec une confiance dans les avis d'experts égale à la taille des données de REX.

La même démarche peut être effectuée facilement pour les autres probabilités conditionnelles et pour les probabilités a priori.

4.4.3.2 Sur un réseau bayésien fictif

Soit le réseau bayésien concernant l'exemple avec la variable "fumeur" (S), "alcool" (A), et "cancer de la gorge" (C). Supposons que ce réseau bayésien ait été construit uniquement par des avis d'experts. Les probabilités données par l'expert sont données dans le tableau 4.3. Au fil des années, les données de retour d'expérience sont répertoriées. On suppose que

$p(S)$	0.5
$p(A)$	0.01
$p(C\|\bar{s}, \bar{a})$	0.00001
$p(C\|\bar{s}, a)$	0.0001
$p(C\|s, \bar{a})$	0.01
$p(C\|s, a)$	0.6

TAB. 4.3 – Exemple fictif d'avis d'experts.

ces données, pour notre exemple fictif ont été tirées selon les probabilités conditionnelles données dans le tableau 4.4. L'échantillon est répertorié dans le tableau 4.5.

$p(S)$	0.6
$p(A)$	0.05
$p(C\|\bar{s}, \bar{a})$	0.0001
$p(C\|\bar{s}, a)$	0.0005
$p(C\|s, \bar{a})$	0.05
$p(C\|s, a)$	0.8

TAB. 4.4 – Exemple fictif de probabilités.

Dans cette exemple, nous avons considéré que les experts sous-estimaient la probabilité d'avoir le cancer. L'objectif est donc de montrer l'évolution des probabilités et de son biais en fonction du nombre de données et de la pondération choisie pour l'avis d'expert.

Pour prendre en compte le retour d'expérience, nous fixons la confiance pour les avis d'experts, $s-1$ à l'unité, c'est-à-dire $s=2$. Ainsi la loi a priori pour la probabilité conditionnelle d'avoir le cancer de la gorge sachant que l'on est fumeur et alcoolique est une loi Beta $\mathcal{B}(1.2, 0.8)$. La probabilité a posteriori correspondante est également une loi Beta, $\mathcal{B}(3.2, 0.8)$. L'estimateur bayésien est égal à

$$\hat{P}(C = oui|f, a) = 0.8.$$

En prenant une taille fictive de 10, i.e. en ayant un grande confiance en l'expert, la loi a posteriori devient $\mathcal{B}(8.6, 4.4)$ et l'estimateur bayésien $\hat{P}(C = oui|f, a) = 0.6615$.

		Fumeur				
		Oui		Non		
	Alcool	oui	non	oui	non	total
Cancer de la gorge	oui	2	0	0	0	2
	non	0	6	0	2	8
	total	2	6	0	2	10

TAB. 4.5 – Exemple de données de REX fictif avec un échantillon de taille 10, conditionnellement aux variables F et A .

Supposons maintenant que 90 données de REX soit disponible. Ces données sont répertoriées dans le tableau 4.6. En choisissant une confiance pour l'expert équivalent à une donnée, la loi a posteriori devient une loi beta, $\mathcal{B}(7.2, 1.8)$, soit un estimateur bayésien pour cette probabilité égal à 0.8.

En prenant une taille de 10 données, correspondant à la confiance aux avis d'experts, la loi a posteriori devient une loi Beta $\mathcal{B}(12.6, 5.4)$, soit un estimateur bayésien pour cette probabilité égal à 0.7.

		Fumeur				
		Oui		Non		
	Alcool	oui	non	oui	non	total
Cancer de la gorge	oui	6	4	0	0	10
	non	1	52	4	33	90
	total	7	56	4	33	10

TAB. 4.6 – Exemple de données de REX fictif avec un échantillon de taille 100, conditionnellement aux variables F et A.

4.5 Conclusions

Les probabilités d'un réseau bayésien sont données par les experts. Au fil du temps, des données de REX vont servir régulièrement à mettre à jour les probabilités, grâce à la théorie bayésienne. Dans ce cadre, les avis d'experts sont modélisés par des lois a priori. Pour les probabilités d'une variable multinomiale, les lois a priori conjuguées sont les lois de Dirichlet, et des lois beta pour les variables binomiales. En modélisant une absence d'information par une loi de Jeffreys, nous établissons une procédure simple pour paramétrer la confiance dans les avis d'experts. Cette confiance se mesure via une taille d'échantillon fictif correspondant aux avis d'experts. Ce paramètre de confiance peut varier pour un même expert selon les probabilités. Nous préconisons comme taille d'échantillon fictif l'unité. En effet, pour notre part le nombre de données de REX est de quelques dizaines. Les avis d'experts seront donc de moins en moins prédominants au fur et à mesure que les données de REX augmenteront. Cette procédure doit être effectuée sur chaque nœud du réseau bayésien et sur ses parents. On peut imaginer que la confiance attribuée aux avis d'experts diffère selon le nœud du graphe, notamment pour les variables d'entrée qui n'ont pas de parents (la probabilité donnée par l'expert est une probabilité marginale).

Une autre approche consisterait à demander un intervalle de confiance pour chaque probabilité, ou de façon équivalente une moyenne et une variance. La variance fait alors office de paramètre jugeant la confiance dans l'avis d'expert. Ainsi, les paramètres de la loi Beta sont alors calibrés par les dires des experts. Mais la calibration des avis d'experts par une taille d'échantillon fictif traduisant leur avis nous semble plus naturelle et plus simple à manipuler.

En tout cas, il nous semble clair que la prise en compte de données de REX dans les réseaux bayésiens construits à partir d'avis d'experts doit se faire dans le cadre de l'inférence bayésienne, où elle ne pose d'ailleurs aucune difficulté particulière, du moins si les nœuds du réseau sont des variables discrètes.

Deuxième partie

Fiabilité dans un contexte doublement censuré et prise en compte d'un changement de comportement de maintenance

Introduction

Dans cette partie, nous étudions un composant mécanique très sensible et très coûteux, pour lequel aucune défaillance n'est observée. En effet, le composant est régulièrement maintenu, de telle sorte que lors d'une inspection il est déclaré rebuté (dégradé) ou non dégradé. Ainsi, la date d'apparition de la dégradation qui cause son rebut n'est pas connue. Les données sont alors dites doublement censurées. Elles sont censurées à gauche si le composant est déclaré rebuté lors de l'inspection et censurées à droite si la dégradation n'est pas encore survenue lors de l'inspection de maintenance. Les deux chapitres de cette partie traitent de données doublement censurées. Beaucoup d'articles portent sur des données doublement censurées [3], [7], [89] pour des méthodes non paramétriques et [11], [69], [90] pour des inférences paramétriques. Pour la majorité des articles cités, le terme "doublement censuré" signifie qu'il existe des censures à gauche et à droite, mais également des observations de défaillance, ce qui n'est pas notre cas. Notre cas peut être modélisé par des variables de Bernoulli indépendantes et non identiquement distribuées [11], [55]. Dans les articles cités traitant d'un modèle paramétrique, la principale loi utilisée est la loi exponentielle, loi usuelle en fiabilité mais qui n'est pas en mesure de modéliser le possible vieillissement du composant. En outre, la loi de Weibull est capable de modéliser le vieillissement d'un composant via son paramètre de forme. L'ajout pour cette loi d'un paramètre de forme rend les calculs théoriques et numériques complexes. Ainsi, le premier objectif de cette partie était donc de proposer des solutions pour l'estimation des paramètres de la loi de Weibull pour des variables doublement censurées (comme nous l'entendons). L'autre objectif de cette partie a été de modéliser, toujours dans le même cadre, des facteurs non observés par des variables cachées.

Plan de la deuxième partie

Le **chapitre 5** propose une modélisation paramétrique pour des données doublement censurées. Nous considérons dans un premier temps la loi exponentielle, puis la loi de Weibull. Pour la loi exponentielle, nous étudions l'estimateur du maximum de vraisemblance où des théorèmes asymptotiques (consistence et normalité asymptotique), que nous rappelons, existent déjà (voir principalement [11]). Puis, nous proposons une approche bayésienne, qu'il existe ou non des avis d'expert. En effet, même en l'absence d'avis d'experts, des lois a priori non informatives peuvent permettre de "stabiliser" les estimateurs. Puis, nos efforts se sont concentrés sur la loi de Weibull. Pour cette loi, la difficulté principale est la dimension 2 de l'espace des paramètres. La consistence de cet estimateur n'est pas démontrée dans ce chapitre, mais la preuve peut s'inspirer de [11], [93], [55]. Nous démontrons la normalité asymptotique de l'estimateur du maximum de vraisemblance en vérifiant les conditions [11]. Nous proposons également une approche bayésienne avec des lois a priori informatives et non informatives, et semi informatives (où l'on impose un vieillissement). Enfin, toutes les méthodes proposées sont testées sur des composants de centrales nucléaires. Des simulations sont également données à la fin du chapitre.

Le **chapitre 6** traite d'une modélisation d'un changement de comportement de maintenance pour deux problèmes distincts. Le premier concerne la modélisation d'un facteur humain et le deuxième modélise une période de garantie. Ces deux modèles ont été regroupés dans le même chapitre car la méthodologie est similaire dans les deux cas.

Le premier modèle de ce chapitre reprend les mêmes données que le chapitre 5, en y ajoutant une variable cachée, modélisant le changement d'un comportement de maintenance. Suite à l'étude du chapitre 5, les experts ont été surpris des faibles moyennes des durées de vie des composants. Par une étude descriptive des données, nous nous sommes aperçus que les données pouvaient être séparées en deux, une partie correspondant à des inspections anciennes (avant une date butoir) et une autre correspondant aux inspections les plus récentes. Avec cette découverte, les experts nous ont alors reporté que les responsables de maintenance étaient très préventifs avec des composants qu'ils ne connaissaient pas, et qu'avec l'expérience cette appréhension disparaissait. Ceci se traduit par des remplacements de composants qui n'avaient pas lieu d'être avant une date d_0, considérée comme la date à partir de laquelle le composant est supposé bien connu par les ingénieurs de maintenance. Ces remplacements inopportuns, impliquant une sous-estimation de la moyenne des durées de vie, sont modélisés par des variables cachées. Nous proposons dans ce chapitre d'estimer la durée de vie moyenne en utilisant un modèle de données incomplètes et de l'identifier en appliquant l'algorithme EM ([73] et voir annexe A). Une approche bayésienne est également proposée via l'algorithme d'échantillonnage de Gibbs (cf. [84]). Une application et des simulations ont également été faites. Ce travail a fait l'objet d'une publication (voir [26]).

Le deuxième modèle du **chapitre 6** a pour objectif de construire un modèle similaire au précédent, afin de prendre en compte une période de garantie d'un composant (voir [36]). Cette période de garantie est caractérisée par un comportement de maintenance différent. En effet, lors de la période de garantie, les responsables de maintenance rebutent tous les composants qui présentent la moindre dégradation, ce qui n'est plus le cas lorsque le composant n'est plus garanti. Ainsi, lorsqu'un composant est rebuté lors de la période de garantie, on ne sait pas si cela est dû à un défaut de jeunesse par exemple (donc à une vraie défaillance) ou à un remplacement non justifié. Les données sont également doublement censurées. Nous modélisons ce changement de comportement par une variable cachée. Nous proposons comme solution au problème des données incomplètes, l'algorithme EM. Des simulations sont données à la fin du chapitre. Pour l'heure, aucune application sur des données réelles n'a pu être effectuée. En effet, la modélisation d'un changement de comportement en maintenance du chapitre 6 nous a amenés naturellement vers une prise en compte d'une période de garantie des composants. Cependant, cette période de garantie n'existe pas pour les composants des centrales nucléaires, EDF étant à la fois fournisseur et exploitant. Un travail futur sera d'appliquer ce modèle sur des données réelles, par exemple pour l'industrie automobile.

Chapitre 5

Fiabilité dans un contexte doublement censuré

Dans ce chapitre, nous nous intéressons aux modèles paramétriques de fiabilité pour des données doublement censurées. Ces données proviennent du département Surveillance, Diagnostique, Maintenance (SDM) de la Division de Recherche et Développement d'Électricité de France. La fiabilité des composants étudiés étant très grande, la politique de maintenance d'EDF consistait jusqu'à présent en des inspections régulières (environ tous les 6 ans, durée bien inférieure à la durée de vie moyenne des composants). De plus, ces composants étant intégrés dans un système jouant un rôle important dans la sécurité de l'installation, ils sont rebutés dès qu'une dégradation majeure est présente. Ainsi, les données sont doublement censurées, sans jamais observer de défaillances. L'objectif de ce chapitre est donc par des modèles paramétriques de modéliser la durée de vie sans dégradation des composants, considérés comme des matériels non réparables. Nous étudions deux lois paramétriques usuelles en fiabilité, la loi exponentielle et la loi de Weibull. Des résultats théoriques ont été établis pour l'estimateur du maximum de vraisemblance, et de nombreux résultats numériques sont donnés, notamment par des approches bayésiennes.

Le chapitre est organisé comme suit. Tout d'abord, une présentation des données est faite et leur caractère doublement censuré est exhibé. Puis, après un rappel de quelques notions en fiabilité, nous introduisons différents modèles paramétriques pour des données censurées à gauche et à droite, principalement pour deux lois usuelles, la loi exponentielle et la loi de Weibull. Pour la loi exponentielle, nous rappelons les résultats théoriques de l'estimateur du maximum de vraisemblance (consistence et normalité asymptotique). Pour cette même loi, nous proposons une alternative bayésienne. Pour la loi de Weibull, nous établissons la normalité asymptotique de cet estimateur. Une approche bayésienne est également proposée pour la loi de Weibull, avec notamment un a priori que l'on qualifie de semi-informatif (où un rajeunissement est proscrit). Enfin, nous avons appliqué ces méthodes sur des données réelles, puis sur des données simulées, afin d'étudier la stabilité, et la validité de nos modèles.

5.1 Présentation des données doublement censurées

Les données nous ont été fournies par le département Surveillance, Diagnostic et Maintenance d'EDF. Ce département souhaite modéliser statistiquement la durée de vie des principaux composants qui constituent les pompes primaires 900 MW. Ces composants sont les suivants :

* la glace tournante du joint 2 (noté c_1),
* la glace flottante du joint 2 (noté c_2),
* les douilles flottantes du joint 2 (noté c_3),
* la glace tournante du joint 3 (noté c_4),
* la glace flottante du joint 3 (noté c_5),
* les douilles flottantes du joint 3 (noté c_6),
* la cartouche (noté c_7).

Lors des précédentes inspections sur ces composants, une importante quantité d'information a été enregistrée dans des bases de données (fichiers Excel). Les variables étaient au nombre de 9 et se présentaient sous la forme suivante :

* la provenance des pompes primaires (le site de la centrale nucléaire),
* la tranche nucléaire (numérotée de 1 à 4 pour la plupart des centrales),
* la date d'arrivée sur le site des pompes primaires,
* la date d'expertise,
* le nombre d'heures de fonctionnement,
* le composant,
* les observations lors de l'expertise (décollement, fissures, griffures, ...),
* les propositions pour éventuellement remettre en état le composant (pierrer les griffes, adoucir, remplacer,...),
* l'état du composant (accepté ou rebuté).

Lors d'une inspection, tous les composants cités sont inspectés et donc acceptés ou rebutés. Il est à noter que pour un composant donné une seule inspection de ce composant est disponible. En effet, aucune traçabilité des composants n'était encore possible. Ainsi, pour un composant donné, les n données de l'échantillon (même composant observé sur des sites différents) sont le couple $(t_i, \delta_i)_{i=1,\dots,n}$, où les t_i sont les dates d'inspections (en heures) et les δ_i sont les variables binaires définies par la décision de l'ingénieur de maintenance. Les temps d'inspections t_i sont supposés ne pas tous être égaux et sont déterministes. Les variables indicatrices sont définies comme suit

$$\delta_i = \left[\begin{array}{ll} 0 & \text{s'il y a une censure à droite} \\ 1 & \text{s'il y a une censure à gauche} \end{array} \right.$$

Si l'on note X_i les durées de vie sans dégradation des composants, alors on a la relation $\delta_i = \mathbb{I}_{\{X_i \leq t_i\}}$. Aucune défaillance n'étant observée sur ces composants, c'est l'apparition d'une dégradation majeure (ici X_i qui n'est pas observé) que l'on souhaite modéliser ici. Ou si l'on préfère, un état de dégradation majeure est considéré comme une défaillance. Ainsi, une censure à gauche correspond à une dégradation détectée lors de la visite (c'est-à-dire survenue avant l'inspection, i.e. $X_i \leq t_i$). Une censure à droite correspond, quant à elle, à un matériel non dégradé (mais susceptible de se dégrader ultérieurement, i.e. $X_i > t_i$). Les composants sont identiques et les durées de vie sans dégradation sont supposées indépendantes et identiquement distribuées. Dans un cadre paramétrique, si on note F la fonction de répartition de la loi de X, alors δ_i est une variable de Bernoulli avec comme paramètre $F(t_i)$, où t_i est la date (en heures) d'inspection du composant i. Remarquons que les δ_i sont alors des variables aléatoires indépendantes, mais non identiquement distribuées. En effet, les temps d'inspections t_i ne sont pas tous égaux.

Le tableau 5.1 représente les caractéristiques des composants étudiés. Les différentes tailles d'échantillons s'expliquent par les valeurs manquantes, assez importantes dans certains cas (comme pour le composant c6 et le composant c7). L'information pertinente dans

Composant	Nom du composant	taux de censures à gauche	taille de l'échantillon
c1	Glace tournante du joint 2	40,5 %	n=168
c2	Glace flottante du joint 2	26,4 %	n=72
c3	Douille flottante du joint 2	1,4 %	n=71
c4	Glace tournante du joint3	59,8 %	n=72
c5	Glace flottante du joint 3	44,6 %	n=74
c6	Douille flottante du joint 3	0 %	n=39
c7	Cartouche	17,8 %	n=28

Tab. 5.1 – Composants et caractéristiques des échantillons.

ces données est exclusivement contenue dans les censures à gauche. Pour le composant c6, on remarque qu'il n'y a que des censures à droite, ce qui n'apporte pas d'information sur la dégradation (sauf que celle-ci est plus grande que toutes les observations). Dans toute la suite, on supposera qu'il existe au moins une censure à gauche et au moins une censure à droite. On verra notamment que pour le composant c6, les modèles utilisés ne sont pas en mesure d'estimer la moyenne de la durée de vie (sauf dans le cas d'une inférence bayésienne, où l'estimateur bayésien sera la moyenne *a priori*).

Dans toute la suite du chapitre, les données de nos échantillons sont considérées doublement censurées.

5.2 Rappel de fiabilité

Nous effectuons dans ce paragraphe un rappel des principales définitions et des lois utilisées en fiabilité (cf. [40], [68]).

5.2.1 Définitions

Dans toute la suite $(\Omega, \mathcal{F}, \mathbb{P})$ désigne un espace probabilisé, $\theta \in \Theta$ le paramètre à estimer.

On note $\mathcal{L}_n$ la vraisemblance du modèle et ℓ_n son logarithme népérien. $\dot{\ell}_n$ et $\ddot{\ell}_n$ désignent respectivement la dérivée première et seconde de la log-vraisemblance.

On note $(X_1, \ldots, X_n)$ les durées de vie, que l'on suppose indépendantes et identiquement distribuées, de fonction de répartition F et de densité f par rapport à la mesure de Lebesgue, et $(T_1, \ldots, T_n)$ les temps d'observations de distribution G et de densité g. On suppose que ces deux variables aléatoires sont indépendantes entre elles. Dans toute la suite du chapitre, nous considérons des matériels non réparables.

Définition 1 *On appelle fiabilité ou fonction de survie la fonction $R(x)$ telle que :*

$$\begin{aligned} R(x) = 1 - F(x) &= P(X > x) \\ &= P(\textit{le composant ait fonctionné sur } [0, x]). \end{aligned}$$

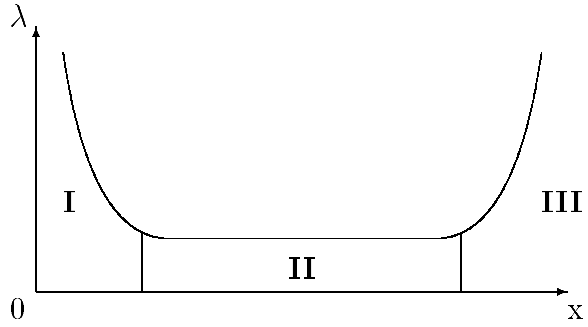
Définition 2 *Dans le cas où X est de loi absolument continue on peut définir le taux de panne instantané λ par*

$$\lambda(x) = \lim_{h \to 0^+} \frac{1}{h} P\left(X \in]x, x+h] / X > x\right) = \lim_{h \to 0^+} \frac{F(x+h) - F(x)}{hR(x)} = \frac{f(x)}{R(x)}$$

et le taux de panne cumulé Λ par

$$\Lambda(x) = \int_0^x \lambda(s)\, ds.$$

Exemple : Le taux de panne instantané au cours de la vie d'un matériel est décrit par la courbe en baignoire.



Dans ce graphe la période **I** correspond à la jeunesse du composant pendant laquelle le taux de panne instantané décroît, **II** correspond à la période de vie utile où λ est constant, et **III** à la période de vieillesse ou d'usure (la probabilité de tomber en panne augmente avec le temps).
Parmi les lois usuelles en fiabilité, la loi de Weibull permet de caractériser une de ces trois périodes grâce à son paramètre de forme. La loi exponentielle est un cas particulier de la loi de Weibull et caractérise la période de vie utile. Nous présentons ces deux lois ci-après.

5.2.2 Lois usuelles

5.2.2.1 Loi exponentielle $\mathcal{E}(\lambda)$

On suppose que la distribution des durées de vie du composant est exponentielle de paramètre λ, appelé intensité ou taux de défaillance. La fonction de répartition de support $\mathbb{R}_+$ s'écrit

$$F(t) = 1 - \exp(-\lambda t). \tag{5.1}$$

La densité de probabilité est égale à

$$f(t) = \lambda \exp(-\lambda t). \tag{5.2}$$

Il existe une autre paramétrisation de la loi exponentielle, avec comme paramètre la moyenne de la durée de vie, notée $\eta = \frac{1}{\lambda}$. Le taux de défaillance pour cette loi est constant et égale au paramètre λ. Ainsi, cette loi suppose qu'il n'existe pas de vieillissement pour le composant et que les défaillances sont donc dues à des accidents. Cette loi est largement utilisée en électronique, et en mécanique.

5.2.2.2 Loi de Weibull $\mathcal{W}(\eta, \beta)$

Si l'on suppose que les durées de vie sont distribuées selon une loi de Weibull, notée $\mathcal{W}(\eta, \beta)$, alors la fonction de répartition de support $\mathbb{R}_+$ s'écrit

$$F(t) = 1 - \exp\left\{-\left(\frac{t}{\eta}\right)^{\beta}\right\}. \tag{5.3}$$

La densité de probabilité est égale à

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left\{-\left(\frac{t}{\eta}\right)^{\beta}\right\}. \tag{5.4}$$

On peut donc calculer le taux de défaillance instantané :

$$\lambda(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}. \tag{5.5}$$

Le paramètre de forme β de la loi de Weibull modélise le vieillissement ou non des composants. En effet, un paramètre de forme inférieur à 1 traduit une période où le composant se "bonifie" avec le temps, et donc modélise une période de défauts de jeunesse. Un paramètre de forme supérieur à 1 traduit un vieillissement du composant étudié. On peut distinguer trois cas :

- $1 < \beta < 2$: le taux de défaillance est croissant et concave,
- $\beta = 2$: le taux de défaillance est croissant et linéaire (avec un pente égale à $\frac{2}{\eta^2}$),
- $\beta > 2$: le taux de défaillance est croissant et convexe.

Enfin, lorsque $\beta = 1$, cette loi équivaut à une loi exponentielle, avec un taux de défaillance égale à $\lambda = \frac{1}{\eta}$.

5.2.3 Données censurées

Dans la majorité des cas, les données observées sont censurées, c'est-à-dire que l'on n'observe pas directement la durée de vie du composant. Nous recensons ici le cas des censures à droite (le plus souvent rencontré) et le cas qui nous intéresse (doublement censuré).

5.2.3.1 Rappel : censures à droite

Dans ce paragraphe, nous introduisons la notion de censure qui est très fréquente en fiabilité, notamment les censures à droite. Il vise à introduire le paragraphe suivant qui nous intéresse, à savoir les données doublement censurées.

Il existe plusieurs types de censures à droite. Nous considérons des *censures progressives à droite de type I*, i.e. elles sont déterministes et connues mais peuvent être différentes suivant les observations. On note $(T_1, \ldots, T_n)$, les n dates d'observations du composant. Les durées de vie sont notées X_i, de densité f, la variable indicatrice δ sur la présence d'une défaillance, et les censures (à droite) de densité g. La fonction de répartition des durées de vie est notée F, celle des censures est notée G. On observe donc les couples

$$(T_i, \delta_i) \text{ avec } \delta_i = \mathbb{I}_{\{X_i \leq T_i\}}, \text{ pour } i = 1, \ldots, n$$

$$\delta_i = \left[\begin{array}{ll} 0 & \text{s'il y a une censure} \\ 1 & \text{s'il y a une panne} \end{array}\right.$$

On désire obtenir les densités correspondant à l'observation d'une panne et à l'observation d'une censure à droite. Pour cela, on calcule la probabilité de constater une panne en inspectant le composant au temps t :

$$\begin{aligned} P(X \leq t, \delta = 1) &= \int_{x \leq t, x \leq c} f(x) g(c)\, dx dc \\ &= \int_{x \leq t} f(x) \left(\int_x^{\infty} g(c)\, dc \right) dx \\ &= \int_{x \leq t} f(x) \bar{G}(x)\, dx \end{aligned}$$

pour tout $t \geq 0$, et où $\bar{G} = 1 - G$. La densité correspondant à l'observation d'une panne est alors $f(x)\bar{G}(x)$. De même la probabilité d'obtenir une censure à droite est donnée par :

$$\begin{aligned} P(X \leq t, \delta = 0) &= \int_{c \leq x, x \leq t} f(x) g(c)\, dx dc \\ &= \int_{c \leq t} g(c) \left(\int_c^{\infty} f(x)\, dx \right) dc \\ &= \int_{c \leq t} g(c) R(c)\, dc \end{aligned}$$

Ainsi, la densité correspondant à l'observation d'une censure est alors $g(x)R(x)$. On peut donc déterminer la densité conjointe du couple (T, δ) par rapport à la mesure $\lambda \bigotimes (\delta_0 + \delta_1)$, où λ est la mesure de Lebesgue sur $\mathbb{R}^+$ et δ_x la mesure de Dirac :

$$[f(x)\bar{G}(x)]^{\delta} [R(x) g(x)]^{1-\delta} \mathbb{I}_{\{x \in \mathbb{R}^+, \delta_x \in \{0,1\}\}}$$

La vraisemblance des observations est alors :

$$\mathcal{L}_n((t_1, \delta_1), \ldots, (t_n, \delta_n); \theta) = \prod_{i=1}^{n} [f_\theta(t_i) \bar{G}(t_i)]^{\delta_i} [R_\theta(t_i) g(t_i)]^{1-\delta_i}$$

De plus, on suppose que la loi de la censure à droite ne dépend pas du paramètre d'intérêt de la loi de durée de vie ; on peut donc travailler uniquement sur la vraisemblance informative ou pertinente [76], i.e.

$$\mathcal{L}_n((t_1, \delta_1), \ldots, (t_n, \delta_n); \theta) = \prod_{i=1}^{n} [f_\theta(t_i)]^{\delta_i} [R_\theta(t_i)]^{1-\delta_i}$$

Ainsi, le paramètre d'intérêt peut être estimé par le maximum de vraisemblance.

Remarque : Il est possible de considérer notre échantillon comme un ensemble de données censurées à droite. Pour cela, il suffit de considérer les censures à gauche comme des temps de défaillance. Cette approximation est grossière, mais permet de donner une borne supérieure à la moyenne des durées de vie (surestimation de celle-ci). Des méthodes dites actuarielles sont également envisageables, où les temps de défaillances sont tirés selon une loi uniforme (cf. [35]).

5.2.3.2 Censures à gauche et à droite

Ce paragraphe est consacré à l'étude de données doublement censurées, ce qui correspond à la caractéristique de nos données. De nombreux articles traitent de telles données, voir Turnbull [89], Ayer et al. [7], Andersen et al. [3] pour les méthodes non paramétriques et Leemis et al. [69], Basu et Ghosh [11] pour les méthodes paramétriques. Dans la plupart des cas, le terme "doublement censuré" signifie qu'il existe trois possibilités : censure à gauche, défaillance, et censure à droite. Ce type de données est souvent rencontré en médecine où les patients sont observés sur une longue période (par exemple un mois) et donc où l'on est susceptible d'observer des défaillances. Un exemple célèbre est l'apprentissage de la lecture par un groupe d'enfants. La variable aléatoire est l'âge des enfants à partir duquel ils savent lire. Ce groupe d'enfants est en observation pendant un mois où ils apprennent à lire. Les censures à gauche correspondent à des enfants qui savent déjà lire. Les censures à droite correspondent à des enfants qui après un mois d'observation ne savent toujours pas lire. Enfin les données non censurées correspondent aux enfants qui ont appris à lire pendant le stage. Dans notre étude, on peut considérer que la période d'observation est réduite à néant, et il n'existe donc que deux cas possibles (voir [11]). En effet, lorsque les composants sont inspectés, le système est arrêté, et une dégradation pendant l'inspection est alors impossible. On observe donc, comme nous l'avons vu, les couples

$$(t_i, \delta_i) \text{ avec } \delta_i = \mathbb{I}_{\{X_i \leq t_i\}}, \text{ pour } i = 1, \ldots, n$$

$$\delta_i = \left[\begin{array}{ll} 0 & \text{s'il y a censure à droite (le composant ne présente pas de défaut)} \\ 1 & \text{s'il y a censure à gauche (le composant présente un défaut)} \end{array} \right.$$

Les variables δ_i sont donc des variables de Bernoulli, avec comme paramètre $F(t_i) = P(X_i \leq t_i)$. Cet échantillon peut donc être vu comme indépendant mais non identiquement distribué (dans le cas où tous les t_i ne sont pas égaux). Ainsi, en gardant les mêmes notations que précédemment, on peut écrire la vraisemblance [3] :

$$\mathcal{L}_n = \prod_{i=1}^{n} \left[F(t_i)\right]^{\delta_i} \left[1 - F(t_i)\right]^{1-\delta_i}. \tag{5.6}$$

La log-vraisemblance que l'on cherche à maximiser est de la forme :

$$l_n = \sum_{i=1}^{n} \delta_i \log\left[F(t_i)\right] + (1 - \delta_i) \log\left[1 - F(t_i)\right]. \tag{5.7}$$

Dans toute la suite, nous allons étudier à partir de l'équation 5.7 l'estimateur du maximum de vraisemblance pour deux lois, la loi exponentielle et la loi de Weibull.

5.3 Modélisation paramétrique

Dans ce paragraphe, nous étudions tout d'abord la loi exponentielle. Un rappel sur les propriétés asymptotiques de l'estimateur du maximum de vraisemblance est fait. Puis, une approche bayésienne est envisagée avec des lois a priori informatives et non informatives. La même démarche est faite pour la loi de Weibull, où principalement nous démontrons la normalité asymptotique de l'estimateur du maximum de vraisemblance. Dans toute la suite, la vraie valeur du paramètre est noté θ_0. Dans un premier temps, nous donnons la définition des informations de Fisher intervenant dans les théorèmes asymptotiques. On note $\mathcal{F}_{i-1} = \sigma(X_1, \ldots, X_{i-1})$, la tribu engendrée par le passé du processus étudié, si i représente le présent.

Définition 3 *On appelle :*

- **information incrémentale**, *la variation quadratique du score notée*

$$J_n(\theta) = \sum_{i=1}^{n} \left(\dot{\ell}_i(\theta) - \dot{\ell}_{i-1}(\theta)\right) \left(\dot{\ell}_i(\theta) - \dot{\ell}_{i-1}(\theta)\right)' ;$$

- **information de Fisher conditionnelle**, *la variation quadratique prévisible du score notée*

$$I_n(\theta) = \sum_{i=1}^{n} \mathbb{E}_\theta \left[\left(\dot{\ell}_i(\theta) - \dot{\ell}_{i-1}(\theta)\right) \left(\dot{\ell}_i(\theta) - \dot{\ell}_{i-1}(\theta)\right)' |\mathcal{F}_{i-1}\right] ;$$

- **information de Fisher moyenne**, *la matrice semi-définie positive notée*

$$i_n(\theta) = \mathbb{E}_\theta \left[I_n(\theta)\right] = \mathbb{E}_\theta \left[J_n(\theta)\right] ;$$

- **information de Fisher observée**, *la matrice notée*

$$j_n(\theta) = -\ddot{\ell}_n.$$

5.3.1 Loi exponentielle

Dans le cas d'une loi exponentielle, la vraisemblance du modèle pour le paramètre $\theta = \lambda$ s'écrit

$$\mathcal{L}_n = \prod_{i=1}^{n} \left[1 - \exp\left(-\theta t_i\right)\right]^{\mathbb{I}_{\{x_i \le t_i\}}} \left[\exp\left(-\theta t_i\right)\right]^{\mathbb{I}_{\{x_i > t_i\}}} , \tag{5.8}$$

où $\theta \in \Theta = \mathbb{R}_+^*$. Dans toute la suite, on suppose que la vraie valeur du paramètre est à l'intérieur de Θ, sous-ensemble fermé de $\mathbb{R}$. De plus, on suppose que les temps d'inspections ne sont pas tous égaux et sont bornés.

5.3.1.1 Propriétés de l'estimateur du maximum de vraisemblance

Nous rappelons ici les propriétés asymptotiques de l'estimateur du maximum de vraisemblance pour la loi exponentielle.

Théorème 4 *Sous les hypothèses précédentes, l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ est consistent et de plus on a*

$$\sqrt{i_n(\theta_0)}\left(\hat{\theta}_n - \theta_0\right) \longrightarrow \mathcal{N}(0,1). \tag{5.9}$$

Preuve : Pour la consistence p.s. et la normalité asymptotique de l'estimateur, voir les travaux de Basu and Ghosh [11] qui se servent du théorème d'Andrews [4] sur la convergence uniforme (voir annexe B).
En pratique, les intervalles de confiance seront données avec $i_n(\hat{\theta}_n)$.

5.3.1.2 Approche bayésienne

En pratique, on dispose de quelques informations sur le comportement des matériaux, sur les pannes, ou sur les premières expérimentations. L'approche bayésienne fournit une méthode pour inclure ces informations dans notre modèle (cf. chapitre 4). Si on utilise la fonction de perte quadratique, l'estimateur bayésien correspond à la moyenne a posteriori. Nous nous placerons toujours dans ce cas là.

Pour la loi exponentielle, deux lois a priori ont été considérées, une loi a priori informative, modélisée par une loi inverse Gamma, et une loi a priori non informative, modélisée par la loi de Jeffreys.

Dans un premier temps, nous nous intéressons au cas d'une loi a priori informative, c'est-à-dire lorsqu'ils existent des avis d'experts. En estimation bayésienne, il est commun de prendre des lois a priori qui facilitent le calcul de la loi a posteriori. C'est le cas des lois dites conjuguées d'un modèle pour lequel les lois a priori et a posteriori sont de même nature. Citons, par exemple, les lois gamma et beta pour la plupart des modèles classiques. Cependant, dans notre contexte (données doublement censurées), ces lois ne sont plus conjuguées. En revanche, elles ont l'avantage de présenter différentes formes suivant le choix des paramètres. Si on choisit comme loi *a priori* une loi inverse gamma de paramètre (r,s), notée $\mathcal{IG}(r,s)$ [84]. La densité de θ est alors :

$$g(\theta) = \frac{1}{s^r\Gamma(r)\theta^{r+1}} \exp\left(-\frac{1}{s\theta}\right) \qquad \theta > 0.$$

avec

$$\Gamma(r) \quad = \quad \int_0^{+\infty} \theta^{r-1} \exp\left(-\theta\right)\, d\theta,$$

et on a :

$$\begin{aligned}
\mathbb{E}[\theta] &= \frac{1}{s(r-1)} \\
\mathbb{V}[\theta] &= 1/s^2\,(r-1)^2\,(r-2)
\end{aligned}$$

Ainsi, la densité *a posteriori* peut se mettre sous la forme :

$$\mathcal{L}_n = \frac{\exp\left(-\dfrac{1}{s\theta}\right)}{\theta^{r+1}K} \prod_{k=1}^{l} \left[1 - \exp\left(-\theta t_k\right)\right] \exp\left(-\theta \sum_{k=l+1}^{n} t_k\right).$$

En choisissant la fonction de perte quadratique, on peut calculer la moyenne de cette loi a posteriori, ainsi que sa variance, par des méthodes numériques de type Monte-Carlo. En effet, il n'existe pas d'expression analytique explicite pour la loi a posteriori et donc pour l'estimateur bayésien. La proposition suivante fournit malgré tout une bonne approximation de l'estimateur bayésien :

Proposition 2 : *Soit $(X_1, \ldots, X_n)$ les durées de vie, suivant une loi exponentielle de paramètre θ, doublement censurées, telles que l'on observe $(t_1, \ldots, t_l)$ correspondant à des censures à gauche et $(t_{l+1}, \ldots, t_n)$ correspondant à des censures à droite.*

On note ℓ_n la log-vraisemblance de ce modèle. On suppose comme loi a priori sur θ une loi inverse gamma notée $\mathcal{IG}(r, s)$. Si on simule $(\theta_1, \ldots, \theta_N) \rightsquigarrow \mathcal{IG}(r-1, s)$ et $(\tilde{\theta}_1, \ldots, \tilde{\theta}_N) \rightsquigarrow \mathcal{IG}(r, s)$ et si on choisit la fonction de perte quadratique alors on a :

$$\frac{1}{s\,(r-1)} \frac{\sum_{i=1}^{N} \ell_n(\theta_i)}{\sum_{i=1}^{N} \ell_n(\tilde{\theta}_i)} \overset{N\to+\infty}{\longrightarrow} \hat{\theta}_{Bayes} \qquad p.s.$$

Remarque : Le terme $\dfrac{1}{s\,(r-1)}$ correspond à la moyenne de la loi a priori.

Preuve : On peut écrire :

$$\hat{\theta}_{Bayes} = \frac{\displaystyle\int_0^{+\infty} \frac{\exp\left(-\dfrac{1}{s\theta}\right)}{\theta^{r}} \prod_{k=1}^{l} \left[1 - \exp\left(-\theta t_k\right)\right] \exp\left(-\sum_{k=l+1}^{n} \theta t_k\right) d\theta}{\displaystyle\int_0^{+\infty} \frac{\exp\left(-\dfrac{1}{s\theta}\right)}{\theta^{r+1}} \prod_{k=1}^{l} \left[1 - \exp\left(-\theta t_k\right)\right] \exp\left(-\sum_{k=l+1}^{n} \theta t_k\right) d\theta}.$$

On peut donc calculer la moyenne a posteriori par une simulation de type Monte-Carlo. Soit $\theta_1, \ldots, \theta_n$ suivant une loi inverse gamma $\mathcal{IG}(r-1, s)$ de densité $f_{r-1,s}$, alors d'après la loi des grands nombres :

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{i=1}^{N} \ell_n(\theta_i) = s^{r-1}\Gamma(r-1) \int_0^{+\infty} f_{r-1,s}(\theta)\ell_n(\theta)\, d\theta$$

où le membre de droite correspond au numérateur de la moyenne a posteriori.

Pour calculer le dénominateur de $\hat{\theta}_{Bayes}$, on va procéder de la même manière avec des $\tilde{\theta}_1, \ldots, \tilde{\theta}_n$ suivant une loi de gamma inverse $\mathcal{IG}(r,s)$ de densité $f_{r,s}$:

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{i=1}^{N} \ell_n(\tilde{\theta}_i) = s^r \Gamma(r) \int_0^{+\infty} f_{r,s}(\theta) \ell_n(\theta)\, d\theta$$

D'où le résultat.

■

Remarque : Il est aisé de calculer la variance de la loi a posteriori par la même méthode.

Supposons maintenant qu'il n'existe pas de connaissance a priori. Une approche bayésienne peut, tout de même être retenue, en implémentant une loi a priori non informative. Dans ce paragraphe, nous allons donc nous intéresser à une famille de loi *a priori* non informative introduite par Martz et Waller (1982) [72] :

$$g(\theta) = \frac{1}{\theta^c} \qquad c > 0.$$

Il s'agit d'un *a priori* impropre (i.e. $\displaystyle\int_0^{+\infty} \frac{1}{\theta^c}\, d\theta = +\infty$) et qui correspond à une loi de Jeffreys si $c = 1$. La loi a priori choisie est la suivante :

$$g(\theta) = \frac{1}{\theta}.$$

D'où

$$\hat{\theta}_{Bayes} = \frac{\displaystyle\int_0^{+\infty} \mathcal{L}_n\, d\theta}{\displaystyle\int_0^{+\infty} \frac{1}{\theta} \mathcal{L}_n\, d\theta}.$$

En pratique, nous avons simulé une loi inverse gamma pour le paramètre θ, puis nous avons calculé ces intégrales par simulation de Monte-Carlo. En effet si θ suit une loi inverse gamma, $\mathcal{IG}(r,s)$ de densité $g(\theta)$ alors si on tire $\tilde{\theta}_1, \ldots, \tilde{\theta}_N$ selon cette loi, on a

$$\frac{\displaystyle\frac{1}{N}\sum_{i=1}^{N} \frac{\mathcal{L}_n(\tilde{\theta}_i)}{g(\tilde{\theta}_i)}}{\displaystyle\frac{1}{N}\sum_{i=1}^{N} \frac{\mathcal{L}_n(\tilde{\theta}_i)}{\tilde{\theta}_i g(\tilde{\theta}_i)}} \overset{N\to+\infty}{\longrightarrow} \hat{\theta}_{Bayes} \tag{5.10}$$

5.3.2 Loi de Weibull

Dans ce paragraphe, nous étudions tout d'abord les propriétés asymptotiques de l'estimateur du maximum de vraisemblance. Plus précisément, nous établissons un théorème central limite en vérifiant les conditions de Basu and Ghosh [11]. Une approche bayésienne est également proposée.

Les paramètres de la loi de Weibull sont notés $\theta = (\eta, \beta)$. On note les vraies valeurs des paramètres $\theta_0 = (\eta_0, \beta_0)$. La vraisemblance est alors donnée par :

$$\mathcal{L}_n = \prod_{i=1}^{n} \left\{ 1 - \exp\left[-\left(\frac{t_i}{\eta}\right)^{\beta} \right] \right\}^{\delta_i} \left\{ \exp\left[-\left(\frac{t_i}{\eta}\right)^{\beta} \right] \right\}^{1-\delta_i}$$

D'où la log-vraisemblance :

$$\ell_n = \sum_{i=1}^{n} \left[\mathbb{I}_{\{x_i \leq t_i\}} \log \left\{ 1 - \exp\left[-\left(\frac{t_i}{\eta}\right)^{\beta} \right] \right\} - \mathbb{I}_{\{x_i > t_i\}} \left(\frac{t_i}{\eta}\right)^{\beta} \right] \tag{5.11}$$

5.3.2.1 Propriétés de l'estimateur du maximum de vraisemblance

Comme pour la loi exponentielle, nous supposerons que les temps d'observations sont bornés et que les vraies valeurs des paramètres appartiennent à l'intérieur de Θ, c'est-à-dire $\eta_0 > 0$ et $\beta_0 > 0$. Dans ces conditions, l'estimateur du maximum de vraisemblance est consistent (voir pour la preuve Hoadley [55] qui généralise les conditions de Wald [93] pour les cas indépendants et non identiquement distribués). Nous nous intéressons ici à la normalité asymptotique.

Nous supposons que les temps d'observations ne sont pas tous égaux et qu'ils appartiennent à l'intervalle $[t_0, t_f]$. Dans le cas contraire, les variables sont indépendantes et identiquement distribuées. Soit γ_i la quantité définie par

$$\gamma_i = \mathbb{I}_{\{X_i \leq t_i\}} \frac{\exp\left[-\left(\frac{t_i}{\eta}\right)^{\beta} \right]}{1 - \exp\left[-\left(\frac{t_i}{\eta}\right)^{\beta} \right]} - \mathbb{I}_{\{X_i > t_i\}}.$$

Pour la suite, nous avons besoin du lemme suivant.

Lemme 1 $\mathbb{E}[\gamma_i] = 0$ *et* $\mathbb{E}[\gamma_i^2] = \dfrac{\exp\left[-\left(\frac{t_i}{\eta}\right)^{\beta} \right]}{1 - \exp\left[-\left(\frac{t_i}{\eta}\right)^{\beta} \right]}.$

Preuve :
Nous avons

$$\mathbb{E}\left[\mathbb{I}_{\{X_i \leq t_i\}}\right] = P\left(X_i \leq t_i\right) = F(t_i) = 1 - \exp\left[-\left(\frac{t_i}{\eta}\right)^\beta\right],$$

et

$$\mathbb{E}\left[\mathbb{I}_{\{X_i > t_i\}}\right] = 1 - \mathbb{E}\left[\mathbb{I}_{\{X_i \leq t_i\}}\right] = R(t_i) = \exp\left[-\left(\frac{t_i}{\eta}\right)^\beta\right],$$

ce qui conclut la première partie du lemme.
De plus, comme $\mathbb{I}^2_{\{X_i > t_i\}} = \mathbb{I}_{\{X_i > t_i\}}$, $\mathbb{I}^2_{\{X_i \leq t_i\}} = \mathbb{I}_{\{X_i \leq t_i\}}$ et $\mathbb{I}_{\{X_i > t_i\}}\mathbb{I}_{\{X_i \leq t_i\}} = 0$, on peut en déduire la deuxième partie, ce qui conclut la preuve.

Soit maintenant $u_i = \begin{bmatrix} -\beta/\eta \\ \log \frac{t_i}{\eta} \end{bmatrix}$ et $\xi_i = \left(\frac{t_i}{\eta}\right)^{2\beta} \mathbb{E}[\gamma_i^2]$, le gradient de la log-vraisemblance peut s'écrire $\nabla \ell_k = \sum_{i=1}^k \psi_i(X_i, \theta)$ où

$$\psi_i = \gamma_i \left(\frac{t_i}{\eta}\right)^\beta u_i.$$

Soit I^k l'information conditionnelle de Fisher donnée par

$$I^k(\theta) = \sum_{i=1}^k \psi_i \psi_i^\mathsf{T} = \sum_{i=1}^k \gamma_i^2 \left(\frac{t_i}{\eta}\right)^{2\beta} u_i u_i^\mathsf{T}. \tag{5.12}$$

Soit i^k l'information moyenne de Fisher, donnée par

$$\begin{aligned} i^k &\triangleq \mathbb{E}\left[I^k\right] \\ &= \sum_{i=1}^k \xi_i u_i u_i^\mathsf{T}. \end{aligned} \tag{5.13}$$

Soit J^k l'information observée de Fisher, donnée par

$$J^k(\theta) = \begin{bmatrix} \frac{\partial^2 \ell_k}{\partial \eta^2} & \frac{\partial^2 \ell_k}{\partial \beta \partial \eta} \\ \\ \frac{\partial^2 \ell_k}{\partial \eta \partial \beta} & \frac{\partial^2 \ell_k}{\partial \beta^2} \end{bmatrix}. \tag{5.14}$$

Les lemmes suivants sont nécessaires pour la suite.

Lemme 2 *Soit $h_{\eta_0,\beta_0}(\cdot)$ une fonction continue sur le compact $[t_0, t_f]$, et ν_i des variables aléatoires indépendantes avec $\mathbb{E}[\nu_i] = 0$ et $\sup_{i \in \mathbb{N}} \mathbb{E}[\nu_i^2] < +\infty$ alors*

$$\lim_{k \to +\infty} \frac{1}{k} \sum_{i=1}^k h_{\eta_0,\beta_0}(t_i) \nu_i = 0$$

Preuve

Comme h_{η_0,β_0} est continue sur $[t_0, t_f]$, on peut choisir deux constantes réelles K_1, K_2 tel que

$$K_1 \frac{1}{k} \sum_{i=1}^{k} \nu_i \leq \frac{1}{k} \sum_{i=1}^{k} h_{\eta_0,\beta_0}(t_i)\nu_i \leq K_2 \frac{1}{k} \sum_{i=1}^{k} \nu_i$$

Maintenant, comme $\sup_{i\in\mathbb{N}} \mathbb{E}[\nu_i^2] < +\infty$ et comme $\mathbb{E}[\nu_i] = 0$, on peut appliquer la loi des grands nombres pour des variables indépendantes (cf. [44] chapitre 17 par exemple) pour obtenir

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \nu_i = 0. \tag{5.15}$$

ce qui conclut la preuve.

Lemme 3 *Nous avons* $i^k \triangleq \mathbb{E}\left[I^k\right] = -\mathbb{E}\left[J^k\right]$.

Preuve :

Voir annexe D.

Lemme 4 *La matrice* $i^k(\theta_0)$ *est définie positive et* $\lim_{k\to\infty} \|i^k(\theta_0)\| = +\infty$.

Preuve :
On déduit de l'expression 5.13 que i^k est semi définie positive. Soit x tel que $x^\mathsf{T} \sum_{i=1}^{k} \xi_i u_i u_i^\mathsf{T} x = 0$. Ainsi, $\sum_{i=1}^{k} \xi_i \|x^\mathsf{T} u_i\|^2 = 0$. Or $\xi_i > 0$ pour tout i, on peut conclure que $x^\mathsf{T} u_i = 0$ pour tout i. Maintenant, comme nous avons supposé que $t_i \neq t_j$ pour au moins un couple (i, j), l'espace engendré par $(u_1, \ldots, u_n)$ est égal à $\mathbb{R}^2$. Ainsi, nous obtenons que $x = 0$ et donc que i^k est définie positive.

Supposons maintenant que $\|i^k\| \leq L$, pour un réel L. Soit $a = (1, 0)^\mathsf{T}$, calculons

$$a^\mathsf{T} i^k a = \sum_{i=1}^{k} \xi_i \left(\frac{\beta_0}{\eta_0}\right)^2. \tag{5.16}$$

Or $\|i^k\| \leq L$, la série dans la précédente équation est convergente. Alors, $\lim_{i\to\infty} \xi_i = 0$, ce qui implique que t_i tend vers $+\infty$ ce qui est contraire à l'hypothèse que les t_i sont bornés.

Nous sommes maintenant en mesure d'établir le théorème central limite pour l'estimateur du maximum de vraisemblance pour la loi de Weibull dans un contexte doublement censuré.

Théorème 5 *Soit* $\hat{\theta}^k$ *l'estimateur du maximum de vraisemblance. La variable* $(i^k)^{1/2}\left(\hat{\theta}^k - \theta_0\right)$ *converge en distribution vers une loi normale centrée réduite.*

Preuve :

Nous avons $i^k(\theta_0) \triangleq E_{\theta_0}[I^k(\theta_0)] = E_{\theta_0}[J^k(\theta_0)]$ comme le montre le lemme 3 et avec le lemme 4 nous concluons que $i^k(\theta_0) > 0$ et $i^k(\theta_0) \to +\infty$. Pour $i = 1, \ldots, n$, il existe une constante $K(t_0, t_f, \eta_0, \beta_0)$ tel que $\mathbb{E}[|\psi_i|^3] \leq K$, où ψ_i a été défini précédemment. De plus, $i^k(\theta_0) \to +\infty$, ce qui implique que $\psi_i(\theta_0)$ satisfont la condition de Lindeberg (cf. chapitre 18 dans [44]), sous θ_0.

Finalement, il reste à montrer que $||i^k + J^k|| \to 0$ quand $k \to +\infty$. Avec l'expression de J^k (voir annexe D), nous devons montrer que les composantes de la matrice de J^k tendent vers les composantes correspondantes de la matrice $-i^k$. Pour rendre plus lisible cette démonstration, nous ne montrons cette convergence uniquement pour la première composante. La démonstration pour les trois autres composantes est identique et sans difficulté. Soit $\nu_i = \gamma_i$ et $h_{\eta_0,\beta_0}(t_i) = \left(\frac{t_i}{\eta}\right)^\beta \frac{\beta(\beta+1)}{\eta^2}$ et appliquons le lemme 2, on obtient

$$\lim_{k\to+\infty} \frac{1}{k} \sum_{i=1}^{k} \left(\frac{t_i}{\eta}\right)^\beta \frac{\beta\,(\beta+1)}{\eta^2}\gamma_i = 0.$$

Prenons maintenant $\nu_i = \mathbb{I}_{\{Y_i \leq t_i\}} - (1 - e^{-(\frac{t_i}{\eta})^\beta})$ et $h_{\eta_0,\beta_0}(t_i) = \left(\frac{t_i}{\eta}\right)^{2\beta} \left(\frac{\beta}{\eta}\right)^2 \frac{e^{-(\frac{t_i}{\eta})^\beta}}{\left(1-e^{-(\frac{t_i}{\eta})^\beta}\right)^2}$, nous avons

$$\lim_{k\to+\infty} \frac{1}{k} \sum_{i=1}^{k} \left(\frac{t_i}{\eta}\right)^{2\beta} \left(\frac{\beta}{\eta}\right)^2 \mathbb{I}_{\{Y_i \leq t_i\}} \frac{e^{-(\frac{t_i}{\eta})^\beta}}{\left(1 - e^{-(\frac{t_i}{\eta})^\beta}\right)^2} =$$

$$\lim_{k\to+\infty} \frac{1}{k} \sum_{i=1}^{k} \left(\frac{t_i}{\eta}\right)^{2\beta} \left(\frac{\beta}{\eta}\right)^2 \frac{e^{-(\frac{t_i}{\eta})^\beta}}{1 - e^{-(\frac{t_i}{\eta})^\beta}}.$$

Ainsi, $\lim_{k\to+\infty}(i^k_{(1,1)} - J^k_{(1,1)}) = 0$ et finalement $\lim_{k\to+\infty} ||i^k - J^k|| = 0$. Les conditions dans [11] sont donc satisfaites et la normalité asymptotique de l'estimateur du maximum de vraisemblance pour la loi de Weibull est donc démontrée.

5.3.2.2 Approche bayésienne

Nous appliquons l'inférence bayésienne pour la loi de Weibull. Dans un premier temps, nous considérons des lois a priori informatives. Nous avons choisi, pour β, la loi Beta de paramètres p et q et de support $[\beta_g, \beta_d]$, notée $\mathcal{B}\,(p, q, [\beta_g, \beta_d])$. La densité est alors :

$$f(\beta) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} * \frac{(\beta - \beta_g)^{p-1}\,(\beta_d - \beta)^{q-1}}{(\beta_d - \beta_g)^{p+q-1}}$$

Ses caractéristiques sont :

$$\begin{aligned}\mathbb{E}[\beta] &= \beta_g + \frac{(p-1)(\beta_d-\beta_g)}{p+q} \\ \mathbb{V}[\beta] &= \frac{pq(\beta_d-\beta_g)^2}{(p+q+1)(p+q)^2}\end{aligned}$$

Nous avons choisi $\beta_g = 0.5$, $\beta_d = 4$, $p = 3.6$ et $q = 14.6$ (ce qui correspond à une moyenne a priori égale à 1, et à une variance a priori égale à 0.1)(cf. [8]). Il est à noter que si $x \rightsquigarrow \mathcal{B}(p,q,[0,1])$, alors $(\beta_d-\beta_g)x+\beta_g \rightsquigarrow \mathcal{B}(p,q,[\beta_g,\beta_d])$.

La loi sur η est une inverse gamma $\mathcal{IG}(r,s)$ (comme dans le cas exponentiel)

$$g(\eta) = \frac{1}{s^r\Gamma(r)\eta^{r+1}}\exp\left(-\frac{1}{s\eta}\right) \qquad \eta > 0.$$

Ainsi, les deux estimateurs bayésiens sont

$$\hat{\beta}_{Bayes} = \frac{\displaystyle\int_{\beta_g}^{\beta_d}\int_0^{+\infty} \beta f(\beta)g(\eta)\mathcal{L}_n\,d\eta\,d\beta}{K}$$

et

$$\hat{\eta}_{Bayes} = \frac{\displaystyle\int_0^{+\infty}\int_{\beta_g}^{\beta_d} \eta f(\beta)g(\eta)\mathcal{L}_n\,d\beta\,d\eta}{K}$$

où $K = \displaystyle\int_{\beta_g}^{\beta_d}\int_0^{+\infty} f(\beta)g(\eta)\mathcal{L}_n\,d\eta\,d\beta$.

En pratique, on procèdera de la façon suivante

* On simule $\beta \rightsquigarrow \mathcal{B}(p,q,[\beta_g,\beta_d])$
* et $\eta \rightsquigarrow \mathcal{IG}(r,s)$

Alors

$$\frac{\displaystyle\sum_{i=1}^N\sum_{j=1}^N \eta_i\mathcal{L}_n(\eta_i,\beta_j)}{\displaystyle\sum_{i=1}^N\sum_{j=1}^N \mathcal{L}_n(\eta_i,\beta_j)} \overset{N\to+\infty}{\longrightarrow} \hat{\eta}_{Bayes} \quad et \quad \frac{\displaystyle\sum_{i=1}^N\sum_{j=1}^N \beta_j\mathcal{L}_n(\eta_i,\beta_j)}{\displaystyle\sum_{i=1}^N\sum_{j=1}^N \mathcal{L}_n(\eta_i,\beta_j)} \overset{N\to+\infty}{\longrightarrow} \hat{\beta}_{Bayes}.$$

Supposons maintenant qu'il n'existe aucune information a priori.

Avec un a priori non informatif

La loi a priori de référence [87] est

$$g(\eta, \beta) = \frac{1}{\eta\beta}$$

En pratique, nous avons simulé deux lois inverse gamma pour η et β. Les paramètres de ces lois ont été ajustés suivant l'ordre de grandeur de η et β. La technique est la même que dans l'équation 5.10 pour le cas exponentiel, excepté que l'on calcule la vraisemblance sur une matrice (indice i pour η et indice j pour β) et que les sommes sont donc doubles.

Avec un a priori "semi-informatif"

On va supposer dans ce paragraphe que les durées de vie des composants $(X_1, \dots, X_n)$ sont régies par une loi de Weibull, notée $W(\eta, \beta)$. On observe toujours $((t_1, \delta_1), \dots, (t_n, \delta_n))$ où $\delta_i = \mathbb{I}_{\{X_i \leq T_i\}}$. De plus, on supposera que le paramètre d'échelle de la loi de Weibull, η suit une loi de Jeffreys, que le paramètre de forme β suit une loi uniforme sur [1,4] et que ces deux paramètres sont indépendants. Par cette méthode, nous imposons que le paramètre de forme soit plus grand que 1, c'est-à-dire que le matériel vieillisse, ou plus précisément qu'il ne rajeunisse pas. Ce sont ces deux a priori conjugués que l'on appelle a priori "semi-informatif" : nous n'apportons aucune information sur le paramètre η, mais nous imposons au paramètre β d'appartenir à $[1, 4]$. Ceci peut se justifier par le fait que l'on essaye de modéliser le comportement d'un composant, subissant des contraintes mécaniques et thermiques. Il y a alors tout lieu de penser que ce composant vieillit, et que sa probabilité de tomber en panne, ou de présenter une dégradation, augmente avec le temps. D'autre part, dans ce contexte mécanique, les paramètres de forme de la loi de Weibull sont toujours plus petits que 4 (cf. [9]).

Cependant la loi uniforme est très peu informative. En effet, celle-ci impose "uniquement" que β appartienne à l'intervalle $[1, 4]$, mais laisse celui-ci parfaitement libre à l'intérieur de cet intervalle.

Ainsi, la loi *a priori* est la suivante :

$$g(\eta, \beta) = \frac{1}{3\eta} \mathbb{I}_{\{\beta \in [1,4]\}}$$

En remplaçant dans la formule de vraisemblance (5.6), on trouve :

$$\mathcal{L}_n(\eta, \beta / (t_1, \delta_1), \dots, (t_n, \delta_n)) =$$

$$\frac{\prod_{k=1}^{l} \left\{ 1 - \exp\left[- \left(\frac{t_k}{\eta} \right)^{\beta} \right] \right\} \prod_{k=l+1}^{n} \exp\left[- \left(\frac{t_k}{\eta} \right)^{\beta} \right]}{K * \eta} \mathbb{I}_{\{\beta \in]1,4[\}} \tag{5.17}$$

Proposition 3 : *Si on suppose qu'il existe au moins une censure à gauche et une censure à droite ($l \geq 1$ et $l < n$), alors la distribution a posteriori (5.17) de (η, β) en utilisant la loi a priori de Jeffreys pour η et une loi uniforme sur* $[1,4]$ *n'est pas une loi impropre.*

Preuve : On cherche à savoir si cette double intégrale est finie :

$$K = \int_1^4 \int_0^{+\infty} \frac{1}{\eta} \prod_{k=1}^{l} \left\{ 1 - \exp\left[- \left(\frac{t_k}{\eta} \right)^{\beta} \right] \right\} \prod_{k=l+1}^{n} \exp\left[- \left(\frac{t_k}{\eta} \right)^{\beta} \right] \, d\eta \, d\beta$$

La fonction à intégrer est continue sur le domaine $]0, +\infty[\times[1,4]$. Il faut donc s'intéresser à l'intégrale impropre en η. On a :

$$\lim_{\eta \to 0} \mathcal{L}_n \left(\eta, \beta / (t_1, \delta_1), \ldots, (t_n, \delta_n) \right) = 0$$

Le premier produit étant équivalent à 1 et le deuxième tendant vers 0, grace à la prédominance de la fonction exponentielle sur la fonction puissance, $\mathcal{L}_n$ est prolongeable par continuité en 0. Le comportement asymptotique de la distribution a posteriori est :

$$\mathcal{L}_n \left(\eta, \beta / (t_1, \delta_1), \ldots, (t_n, \delta_n) \right) \sim \frac{\prod_{k=1}^{l} t_k^{\beta}}{\eta^{\beta l + 1}}$$

Or $\beta l > 1$, l'intégrale est donc convergente, et par conséquent la loi a posteriori n'est pas impropre.

■

On est maintenant en mesure de calculer les lois marginales a posteriori, et donc de déterminer les estimateurs de Bayes.

$$\hat{\eta}_{bayes} = \frac{\displaystyle\int_1^4 \int_0^{+\infty} \prod_{k=1}^{l} F(t_k) \prod_{k=l+1}^{n} R(t_k) \, d\eta \, d\beta}{\displaystyle\int_1^4 \int_0^{+\infty} \frac{1}{\eta} \prod_{k=1}^{l} F(t_k) \prod_{k=l+1}^{n} R(t_k) \, d\eta \, d\beta},$$

$$\hat{\beta}_{bayes} = \frac{\displaystyle\int_1^4 \int_0^{+\infty} \frac{\beta}{\eta} \prod_{k=1}^{l} F(t_k) \prod_{k=l+1}^{n} R(t_k) \, d\eta \, d\beta}{\displaystyle\int_1^4 \int_0^{+\infty} \frac{1}{\eta} \prod_{k=1}^{l} F(t_k) \prod_{k=l+1}^{n} R(t_k) \, d\eta \, d\beta}.$$

Ces intégrales seront aussi calculées par simulation de Monte-Carlo, comme dans le cas non informatif.

5.4 Résultats numériques

Les méthodes présentées précédemment ont été appliquées pour des données simulées, pour se donner une idée de la qualité de ces estimateurs. Puis, nous les avons appliquées sur les composants des pompes primaires. Les programmes d'optimisation ont été effectués sous Matlab, ainsi que le calcul des informations de Fisher et des intervalles de confiance. Dans toute la suite, τ désigne le pourcentage de censures à gauche présentes dans l'échantillon.

5.4.1 Simulations

Afin d'être dans le même cadre que nos données réelles, nous nous sommes servis des temps d'observations $t_1, \dots, t_n$ du composant 1, $n = 168$, et du composant 7, $n = 28$.

5.4.1.1 Pour la loi exponentielle

Nous avons simulé plusieurs lois exponentielles, afin de vérifier la validité de nos méthodes d'estimation. Les étapes sont les suivantes :

* On simule $X_1, \dots, X_n$, représentant les durées de vie suivant une loi exponentielle de paramètre θ_0 connu,
* on compare les x_i et les t_i pour tout $i \in < 1, n >$,
* si $x_i \leq t_i$ alors $\delta_i = 1$, sinon $\delta_i = 0$.

Les observations sont $(t_i, \delta_i)_{1 \leq i \leq n}$. Nous sommes alors dans les mêmes conditions que les données réelles. Le tableau 5.2 donne les estimateurs du maximum de vraisemblance pour différentes lois exponentielles simulées, ainsi que les intervalles de confiance. L'estimation

Simulations	EMV	Intervalle de confiance	τ
$Exp(10^{-5})$	1.10×10^{-5}	$[0.75 \times 10^{-5}; 1.44 \times 10^{-5}]$	24.4 %
$Exp(1.4.10^{-5})$	1.21×10^{-5}	$[0.85 \times 10^{-5}; 1.56 \times 10^{-5}]$	26.8 %
$Exp(2.10^{-5})$	1.90×10^{-5}	$[1.42 \times 10^{-5}; 2.39 \times 10^{-5}]$	38.1 %
$Exp(2.6.10^{-5})$	1.96×10^{-5}	$[1.46 \times 10^{-5}; 2.45 \times 10^{-5}]$	38.7 %
$Exp(3.4.10^{-5})$	3.54×10^{-5}	$[2.79 \times 10^{-5}; 4.29 \times 10^{-5}]$	57.7 %

TAB. 5.2 – EMV et Intervalles de confiance pour des lois exponentielles simulées avec n=168.

ponctuelle du paramètre semble correcte dans la plupart des cas, excepté pour le quatrième modèle ($Exp(2, 6.10^{-5})$), pour lequel l'intervalle de confiance ne contient pas la vraie valeur du paramètre. Les intervalles de confiance pour ces estimateurs sont dans la plupart des cas assez précis. Ceci étant, si l'on s'intéresse au temps moyen de bon fonctionnement $\eta = \frac{1}{\lambda}$, il est nécessaire pour avoir un bon intervalle de confiance pour ce paramètre, d'avoir des bornes très proches pour le taux de défaillance. Par exemple, pour le dernier modèle simulé, l'intervalle de confiance du MTBF (Mean Time Before Failure) est $[22000 \text{ heures}, 40000 \text{ heures}]$

pour un seuil de confiance à 95 % , malgré une taille d'échantillon honorable ($n = 168$).

Remarque : La variance asymptotique théorique de l'estimateur est croissante avec la valeur de l'estimateur. Ainsi, plus le taux de censure à gauche est important et plus la variance de l'estimateur du maximum de vraisemblance est grande.

Pour vérifier la stabilité de l'estimateur du maximum de vraisemblance pour des données doublement censurées, nous avons itéré 100 fois le procédé. Ainsi, ce bootstrap paramétrique nous donne un écart type empirique que l'on pourra comparer à l'écart type asymptotique théorique. On peut remarquer d'après le tableau 5.3 que l'estimateur du maximum de vraisemblance possède de bonnes propriétés, avec une taille d'échantillon honorable ($n = 168$). L'écart type calculé à partir des estimations est très proche de l'écart type théorique. Ce qui nous fait penser que nous pouvons considérer que les calculs asymptotiques sont valables pour cette valeur de n.

Simulations	EMV moyen	Écart type asymptotique	τ moyen
$Exp(10^{-5})$	1.03×10^{-5} *(1.63×10^{-6})*	1.78×10^{-6}	23.3 %
$Exp(1.4.10^{-5})$	1.39×10^{-5} *(1.80×10^{-6})*	1.79×10^{-6}	29.9 %
$Exp(2.10^{-5})$	2.00×10^{-5} *(2.40×10^{-6})*	2.45×10^{-6}	39.5 %
$Exp(2.6.10^{-5})$	2.63×10^{-5} *(2.84×10^{-6})*	2.55×10^{-6}	47.8 %
$Exp(3.4.10^{-5})$	3.43×10^{-5} *(4.01×10^{-6})*	3.83×10^{-6}	57.7 %

TAB. 5.3 – Simulations et écarts types asymptotiques de lois exponentielles pour $n = 168$.

Nous allons maintenant étudier le comportement de l'estimateur du maximum de vraisemblance en fonction de la taille de l'échantillon. Les résultats numériques pour une taille d'échantillon égale à 28 sont donnés dans le tableau 5.4.

Avec la baisse de la taille de l'échantillon ($n = 28$), on remarque que l'estimateur du maximum de vraisemblance est toujours performant. Ceci étant dit, les écarts types empiriques et théoriques sont bien supérieurs au cas où $n = 168$ (environ cinq fois plus grand). Les intervalles de confiance sont donc peu précis et ne peuvent donc être pris en compte dans le cas $n = 28$ (nous ne sommes en aucun cas dans des conditions asymptotiques !).

L'estimation ponctuelle semble donc, pour ce modèle, assez stable, et peu perturbée par une faible taille d'échantillon. En revanche, l'écart type asymptotique empirique ou théorique peut être assez important, soit pour des valeurs du paramètre importantes, soit pour une taille d'échantillon faible. Ce modèle est donc satisfaisant pour des échantillons de grande

Simulations	EMV moyen	Écart type asymptotique	τ moyen
$Exp(10^{-5})$	1.02×10^{-5} (5.08×10^{-6})	3.22×10^{-6}	17.9 %
$Exp(1.4.10^{-5})$	1.40×10^{-5} (5.47×10^{-6})	3.66×10^{-6}	23.7 %
$Exp(2.10^{-5})$	2.03×10^{-5} (6.79×10^{-6})	1.22×10^{-5}	32.4 %
$Exp(2.6.10^{-5})$	2.69×10^{-5} (7.97×10^{-6})	8.04×10^{-6}	40.2 %
$Exp(3.4.10^{-5})$	3.41×10^{-5} (0.90×10^{-5})	1.29×10^{-5}	47.7 %

TAB. 5.4 – Simulations et écarts types asymptotiques de lois exponentielles pour $n = 28$.

taille. Pour des échantillons de petite taille, une approche bayésienne peut être envisagée.

Estimateurs bayésiens

Ces calculs ont été effectués par des simulations de Monte-Carlo avec $N = 2000$. Nous comparons deux estimateurs bayésiens, l'un informatif avec une loi a priori Gamma, et l'autre non informatif avec une loi a priori de Jeffreys. Tous les résultats sont reportés dans le tableau 5.5. On peut remarquer tout d'abord la cohérence des résultats. En effet, l'estimateur

Simulations	m et σ a priori	Bayes informatif	Bayes non informatif	tau
$Exp(10^{-5})$	$m = 1.6 \times 10^{-5}$ $\sigma = 0.4 \times 10^{-5}$	1.19×10^{-5} (3.94×10^{-6})	1.47×10^{-5} (1.71×10^{-6})	22.6 %
$Exp(1.4.10^{-5})$	$m = 1.6 \times 10^{-5}$ $\sigma = 0.4 \times 10^{-5}$	1.49×10^{-5} (3.24×10^{-6})	1.92×10^{-5} (2.16×10^{-6})	32.1 %
$Exp(2.10^{-5})$	$m = 1.6 \times 10^{-5}$ $\sigma = 0.4 \times 10^{-5}$	1.78×10^{-5} (5.61×10^{-7})	2.22×10^{-5} (2.35×10^{-6})	37.5 %
$Exp(2.6.10^{-5})$	$m = 1.6 \times 10^{-5}$ $\sigma = 0.4 \times 10^{-5}$	2.05×10^{-5} (9.70×10^{-6})	2.81×10^{-5} (3.02×10^{-6})	46.4 %
$Exp(3.4.10^{-5})$	$m = 1.6 \times 10^{-5}$ $\sigma = 0.4 \times 10^{-5}$	3.35×10^{-5} (6.08×10^{-6})	3.16×10^{-5} (1.45×10^{-6})	58.3 %

TAB. 5.5 – Estimateurs bayésiens pour une loi exponentielle avec $n = 168$.

de Bayes informatif est toujours compris entre l'estimateur du maximum de vraisemblance et la moyenne a priori donnée par l'expert. L'estimateur de Bayes non informatif a tendance a surestimer le paramètre d'intérêt. Pour ces deux estimateurs l'écart type de la loi a posteriori, donné entre parenthèses, semble convenable pour ce modèle. Néanmoins, comme l'estimateur du maximum de vraisemblance possède de bonnes propriétés pour ce modèle, l'intérêt de l'analyse bayésienne ne subsiste que dans l'apport d'information d'experts.

En fiabilité industrielle, l'objectif est de mesurer le vieillissement d'un composant, ce que le modèle exponentiel ne prend pas en compte. Ainsi, nous poursuivons naturellement les simulations pour la loi de Weibull.

5.4.1.2 Pour la loi de Weibull

Nous avons simulé plusieurs lois de Weibull à deux paramètres, et procédé de la même façon que pour le modèle exponentiel. Les résultats sont donnés dans le tableau 5.6. Les paramètres ont été choisies de façon à ce que la durée moyenne de survie soit à peu près constante.
Il existe une difficulté pour cet estimateur, à bien estimer le paramètre de forme β, lorsque

Simulations	EMV	Intervalle de confiance	τ
$\eta_0 = 80000$ $\beta_0 = 0.6$	$\hat{\eta} = 217767$ $\hat{\beta} = 0.35$	$IC = [61145; 374389]$ $IC = [0.26; 0.43]$	37.5 %
$\eta_0 = 60000$ $\beta_0 = 0.8$	$\hat{\eta} = 77461$ $\hat{\beta} = 0.63$	$IC = [44352; 110571]$ $IC = [0.55; 0.71]$	38.7 %
$\eta_0 = 40000$ $\beta_0 = 1$	$\hat{\eta} = 37663$ $\hat{\beta} = 1.16$	$IC = [27623; 47703]$ $IC = [1.12; 1.19]$	46.4 %
$\eta_0 = 30000$ $\beta_0 = 1.5$	$\hat{\eta} = 33376$ $\hat{\beta} = 1.19$	$IC = [26380; 40374]$ $IC = [0.99; 1.38]$	50.6 %
$\eta_0 = 20000$ $\beta_0 = 2$	$\hat{\eta} = 19808$ $\hat{\beta} = 2.26$	$IC = [18462; 21155]$ $IC = [1.53; 2.98]$	72 %

TAB. 5.6 – EMV et intervalle de confiance au risque $\alpha = 0.05$ pour la loi de Weibull avec n=168.

celui-ci est inférieur à 1. Lorsque $\beta > 1$, l'estimateur du maximum de vraisemblance a un bon comportement.

Le modèle exponentiel étant un cas particulier d'un modèle de Weibull ($\beta = 1$ et $\eta = 1/\theta$), il est possible de faire un parallèle. Ainsi, la variance asymptotique sur η, comme dans le cas exponentiel, croît avec la valeur de celui-ci. Pour $\eta = 20000$, l'intervalle de confiance est bon, alors que pour $\eta = 80000$, il devient catastrophique. Pour β, sachant qu'il est très important pour nous de connaître la position de β par rapport à 1, les intervalles de confiance nous donnent une réponse correcte dans les 5 cas simulés.

Nous étudions la variance asymptotique empirique de l'estimateur du maximum de vraisemblance pour le modèle de Weibull (voir tableau 5.7). Il existe une grande difficulté à

Simulations	EMV moyen	Écart type asymptotique	τ moyen
$\eta_0 = 80000$ $\beta_0 = 0.6$	n'existe pas $\hat{\beta} = 0.57$ $(\mathit{0.27})$	$+\infty$ 0.07	38.5 %
$\eta_0 = 60000$ $\beta_0 = 0.8$	$\hat{\eta} = 99148$ $(\mathit{175443})$ $\hat{\beta} = 0.81$ $(\mathit{0.31})$	17027 0.14	39.4 %
$\eta_0 = 40000$ $\beta_0 = 1$	$\hat{\eta} = 45000$ $(\mathit{24179})$ $\hat{\beta} = 0.81$ $(\mathit{0.28})$	6660 0.20	46.5 %
$\eta_0 = 30000$ $\beta_0 = 1.5$	$\hat{\eta} = 30267$ $(\mathit{2764})$ $\hat{\beta} = 1.48$ $(\mathit{0.27})$	3867 0.27	52.8 %
$\eta_0 = 20000$ $\beta_0 = 2$	$\hat{\eta} = 19955$ $(\mathit{1208})$ $\hat{\beta} = 2.13$ $(\mathit{0.45})$	931 0.28	71.3 %

TAB. 5.7 – Simulations et écarts types asymptotiques pour la loi de Weibull pour $n = 168$.

estimer correctement les paramètres lorsque $\beta < 1$. En effet, il existe des cas où la méthode d'estimation donne $\eta = +\infty$ et $\beta = 0$ (4 % des cas pour la première loi simulée). Ceci affecte peu la moyenne des estimateurs de β. En revanche, la moyenne de η est très surestimée. Il est possible d'envisager d'exclure les estimateurs du couple (η, β) quand celui-ci est trop proche de $(0, +\infty)$, qui n'a aucun sens physique. Pour $\beta > 1$, l'estimation des deux paramètres avec une taille d'échantillon égale à 28 est correct. Pour étudier l'effet de la taille de l'échantillon, nous nous sommes placés dans le même cas que le composant c7, à savoir n=28 (cf. tableau 5.8).

Simulations	EMV moyen	Écart type asymptotique	τ moyen
$\eta_0 = 80000$ $\beta_0 = 0.6$	n'existe pas $\hat{\beta} = 0.78$ $(\mathit{0.96})$	515218 0.07	34 %
$\eta_0 = 60000$ $\beta_0 = 0.8$	n'existe pas $\hat{\beta} = 1.32$ $(\mathit{1.29})$	7953 0.31	32.. %
$\eta_0 = 40000$ $\beta_0 = 1$	n'existe pas $\hat{\beta} = 1.24$ $(\mathit{1.14})$	31449 0.42	38.25 %
$\eta_0 = 30000$ $\beta_0 = 1.5$	n'existe pas $\hat{\beta} = 2.40$ $(\mathit{4.24})$	118788 0.53	41.8 %
$\eta_0 = 20000$ $\beta_0 = 2$	$\hat{\eta} = 19810$ $(\mathit{3182})$ $\hat{\beta} = 2.39$ $(\mathit{1.38})$	3582 1.02	60.8 %

TAB. 5.8 – Simulations et écarts types asymptotiques pour la loi de Weibull avec $n = 28$.

À n=28, l'estimation du paramètre d'échelle devient quasiment impossible pour le maximum de vraisemblance (excepté pour des faibles valeurs de η et un β plus grand que 1). β est légèrement surestimé, mais ne peut constituer, eu égard à son écart type, une estimation valable (par exemple pour $W(30000, 1.5)$, l'écart type sur β est de 4.24).

L'estimateur du maximum de vraisemblance montre plusieurs difficultés à bien estimer les paramètres, principalement η. L'estimation sur β est meilleure quand celui-ci est plus grand que 1. Les estimateurs du maximum de vraisemblance sont moins stables que pour le modèle exponentiel, quand la taille de l'échantillon diminue. Cependant, le point positif est que ce modèle nous permet a priori de déterminer correctement si le composant vieillit ou pas.

Estimateurs bayésiens

Nous étudions maintenant l'estimateur bayésien. Nous avons simulé plusieurs lois de Weibull à deux paramètres, et procédé de la même façon que pour le modèle exponentiel. Les moyennes a priori pour l'estimation dite informative sont de 30000, pour η, et de 1,18 pour β. On n'impose donc très peu de contraintes sur le paramètre d'échelle. Le paramètre de forme possède une moyenne a priori supérieure à 1, ce qui induit un léger vieillissement a priori. Les résultats sont donnés dans le tableau 5.9. Pour $n = 168$, on remarque là où

Simulation	Bayes non informatif	Bayes semi informatif	Bayes informatif	τ
$\eta_0 = 80000$ $\beta_0 = 0.6$	$\hat{\eta} = 52996$ (*30528*) $\hat{\beta} = 0.78$ (*0.13*)	$\hat{\eta} = 42782$ (*14909*) $\hat{\beta} = 1.12$ (*0.10*)	$\hat{\eta} = 47094$ (*28989*) $\hat{\beta} = 0.88$ (*0.16*)	44 %
$\eta_0 = 40000$ $\beta_0 = 1$	$\hat{\eta} = 38637$ (*14988*) $\hat{\beta} = 1.25$ (*0.30*)	$\hat{\eta} = 36746$ (*38907*) $\hat{\beta} = 1.40$ (*0.24*)	$\hat{\eta} = 37269$ (*14387*) $\hat{\beta} = 1.26$ (*0.22*)	45,8 %
$\eta_0 = 20000$ $\beta_0 = 2$	$\hat{\eta} = 19410$ (*34148*) $\hat{\beta} = 1.98$ (*0.35*)	$\hat{\eta} = 19527$ (*51281*) $\hat{\beta} = 2.06$ (*0.35*)	$\hat{\eta} = 18955$ (*43536*) $\hat{\beta} = 1.65$ (*0.27*)	34 %

TAB. 5.9 – Estimateurs bayésiens pour des lois simulées de Weibull avec $n = 168$.

l'estimateur du maximum de vraisemblance ne donnait pas de résultats (pour η lorsque β est inférieur à 1), l'estimateur bayésien dans les trois cas donne une estimation de η (sous-estimation). Cependant, l'estimation de β semble bonne, excepté le cas où l'on impose $\beta > 1$ (bayésien semi informatif). L'estimateur de Bayes non informatif semble avoir le meilleur comportement. Ces trois estimateurs ne semblent donc pas être meilleurs que l'estimateur du maximum de vraisemblance, l'estimateur bayésien ne subsistera alors que dans l'apport d'information a priori. On va donc tester l'inférence bayésienne avec une taille d'échantillon très faible (cf. tableau 5.10). Pour $n = 28$, l'estimateur bayésien du paramètre d'échelle, même si l'écart type est très grand, a un excellent comportement, et même avec β plus grand que 1. L'estimateur non informatif semble avoir le meilleur comportement de ces trois

Simulation	Bayes non informatif	Bayes semi informatif	Bayes informatif	τ
$\eta_0 = 80000$	$\hat{\eta} = 52021$ $(> 10^9)$	$\hat{\eta} = 34034$ $(> 10^9)$	$\hat{\eta} = 35799$ $(> 10^9)$	44 %
$\beta_0 = 0.6$	$\hat{\beta} = 0.78$ (0.26)	$\hat{\beta} = 1.36$ (0.33)	$\hat{\beta} = 1.03$ (0.25)	
$\eta_0 = 40000$	$\hat{\eta} = 39696$ $(> 10^9)$	$\hat{\eta} = 30928$ $(> 10^9)$	$\hat{\eta} = 34758$ $(> 10^9)$	45,8 %
$\beta_0 = 1$	$\hat{\beta} = 1.54$ (0.89)	$\hat{\beta} = 2.08$ (0.78)	$\hat{\beta} = 1.23$ (0.3)	
$\eta_0 = 20000$	$\hat{\eta} = 18357$ $(> 10^9)$	$\hat{\eta} = 18165$ $(> 10^9)$	$\hat{\eta} = 17383$ $(> 10^9)$	34 %
$\beta_0 = 2$	$\hat{\beta} = 2.19$ (0.72)	$\hat{\beta} = 2.20$ (0.73)	$\hat{\beta} = 1.27$ (0.32)	

TAB. 5.10 – Estimateurs bayésiens pour des lois simulées de Weibull avec $n = 28$].

estimateurs. Ainsi, nous préconiserons d'utiliser celui-ci et seulement les autres quand les avis d'experts sont forts.

5.4.2 Applications sur des données réelles

5.4.2.1 Pour le modèle exponentiel

Des applications numériques ont été effectuées sur les composants des pompes primaires. Les résultats sont donnés dans le tableau 5.11, avec notamment les intervalles de confiance pour l'estimateur du maximum de vraisemblance dans le cas doublement censuré, pour un niveau de confiance à 95 %. La dernière colonne du tableau 5.11 donne τ, le taux de censures à gauche dans l'échantillon. On peut remarquer que l'EMV croît avec le pourcentage de cen-

Composant	*EMV*	Intervalle de confiance	τ
c1	2.03×10^{-5}	$[1.54 \times 10^{-5}; 2.52 \times 10^{-5}]$	40.5 %
c2	1.44×10^{-5}	$[0.79 \times 10^{-5}; 2.09 \times 10^{-5}]$	26.4 %
c3	6.61×10^{-7}	$[0; 19.56 \times 10^{-7}]$	1.4 %
c4	4.40×10^{-5}	$[3.02 \times 10^{-5}; 5.78 \times 10^{-5}]$	59.8 %
c5	2.75×10^{-5}	$[1.80 \times 10^{-5}; 3.69 \times 10^{-5}]$	44.6 %
c6	NaN	NaN	0 %
c7	9.93×10^{-6}	$[1.19 \times 10^{-6}; 18.66 \times 10^{-6}]$	17.8 %

TAB. 5.11 – Intervalle de confiance pour l'EMV pour la loi exponentielle.

sures à gauche. En effet, l'estimateur représente, pour ce modèle, le taux de défaillance, qui est grand quand le nombre de dégradations présentes dans l'échantillon est élevé, et est non défini (composant c6) quand le taux est nul. Les intervalles de confiance sont relativement bons lorsque le taux de censures à gauche est assez important, c'est-à-dire lorsque l'information apporté par les observations est suffisante. Dans l'hypothèse d'un non vieillissement, le modèle exponentiel semble très satisfaisant. Néanmoins, une inférence bayésienne est effectuée.

Estimateurs bayésiens

On veut comparer l'estimateur du maximum de vraisemblance, l'estimateur de Bayes pour un a priori informatif, et l'estimateur de Bayes non informatif. Les lois a priori informatives sont des inverses gamma dont on définit dans la deuxième colonne du tableau 5.12 la moyenne et l'écart type entre parenthèse. Pour le calcul des estimateurs de Bayes, nous avons pris $N = 2000$ simulations.
La première remarque sera la cohérence des résultats présentés. En effet, l'estimateur de

	moyenne a priori	Bayes informatif	Bayes non informatif	EMV	τ
c1	1.6×10^{-5} (0.4×10^{-5})	1.82×10^{-5} (5.62×10^{-6})	2.38×10^{-5} (2.69×10^{-6})	2.03×10^{-5}	40.5 %
c2	1.6×10^{-5} (0.4×10^{-5})	1.48×10^{-5} (3.52×10^{-6})	2.37×10^{-5} (3.80×10^{-6})	1.44×10^{-5}	26.4 %
c3	0.1×10^{-5} (0.05×10^{-5})	8.88×10^{-7} (3.48×10^{-7})	1.39×10^{-5} (2.13×10^{-6})	6.61×10^{-7}	1.4 %
c4	1.6×10^{-5} (0.4×10^{-5})	5.48×10^{-5} (3.14×10^{-5})	3.72×10^{-5} (1.13×10^{-6})	4.40×10^{-5}	59.8 %
c5	1.6×10^{-5} (0.4×10^{-5})	2.28×10^{-5} (6.34×10^{-6})	3.25×10^{-5} (4.09×10^{-6})	2.75×10^{-5}	44.6 %
c6	0.1×10^{-5} (0.05×10^{-5})	0.1×10^{-5} (0.05×10^{-5})	3.77×10^{-5} (2.19×10^{-6})	NaN	0 %
c7	1.6×10^{-5} (0.4×10^{-5})	1.37×10^{-5} (4.14×10^{-6})	2.79×10^{-5} (4.30×10^{-6})	9.93×10^{-6}	17.8 %

TAB. 5.12 – Comparaisons EMV et estimateurs bayésiens pour la loi exponentielle.

Bayes informatif est toujours compris entre l'estimateur du maximum de vraisemblance et la moyenne a priori (excepté pour le composant c4 où la variance de l'estimateur bayésien est très grande). De plus, pour le composant c6 (qui ne présente aucune défaillance !) l'estimateur de Bayes est égal à la moyenne a priori. Ainsi, comme aucune information n'est apportée pour ce composant, l'estimateur "fait une entière confiance" à l'a priori que l'expert lui donne. On peut aussi remarquer que pour le composant 3 (1,4 % de censures à gauche, donc peu d'information sur le comportement du composant), il est très proche de la moyenne a priori donnée par l'expert.

L'estimateur de Bayes non informatif semble uniformiser les résultats. Cet estimateur est plus ou moins croissant avec le pourcentage de censures à gauche, c'est-à-dire avec le taux de défaillance. Le résultat pour le composant c6 semble toutefois curieux. Nous testons pour ces mêmes composants la loi de Weibull pour se renseigner sur leur possible vieillissement.

5.4.2.2 Pour le modèle de Weibull

Nous présentons dans ce paragraphe les résultats des mêmes méthodes pour la loi de Weibull. Nous gardons ici les même notations. Comme pour la loi exponentielle, nous donnons les estimateurs du maximum de vraisemblance ainsi que les intervalles de confiance au seuil 95 %. Tous les résultats sont donnés dans le tableau 5.13.

Composant	EMV	Intervalle de confiance	τ
c1	$\hat{\eta} = 62067$	$IC = [36690; 87444]$	40.5 %
	$\hat{\beta} = 0.72$	$IC = [0.67; 0.77]$	
c2	$\hat{\eta} = 98226$	$IC = [46229; 150223]$	26.4 %
	$\hat{\beta} = 0.77$	$IC = [0.65; 0.88]$	
c3	$\hat{\eta} \geq 10^{16}$	$IC = [0; +\infty]$	1.4 %
	$\hat{\beta} = 0.15$	$IC = [0.08; 0.21]$	
c4	$\hat{\eta} = 22759$	$IC = [15322; 30196]$	59.8 %
	$\hat{\beta} = 0.97$	$IC = [0.07; 1.86]$	
c5	$\hat{\eta} = 237314$	$IC = [0; 611938]$	44.6 %
	$\hat{\beta} = 0.22$	$IC = [0.09; 1.33]$	
c6	n'existe pas	n'existe pas	0 %
	n'existe pas	n'existe pas	
c7	$\hat{\eta} = 211782$	$IC = [0; 437385]$	17.8 %
	$\hat{\beta} = 0.68$	$IC = [0.45; 0.91]$	

TAB. 5.13 – Intervalle de confiance pour l'EMV pour la loi de Weibull.

On peut remarquer que, comme dans le cas exponentiel, l'estimation semble plus réaliste, lorsque le taux de censures à gauche est important. Il apparaît que pour la loi de Weibull, la difficulté à estimer le paramètre de forme β est très importante. Dans tous les cas présentés ci-dessus, l'EMV nous affirme que le paramètre de forme est inférieur à 1. D'après ces résultats, un modèle exponentiel semble être la meilleure modélisation.

Ces résultats ne sont dans l'ensemble pas acceptables. En effet, l'estimateur du maximum de vraisemblance est biaisé pour des faibles tailles d'échantillons, ou pour des échantillons comportant peu de censures à gauche. Ainsi, les intervalles de confiance, calculés à partir de cet estimateur, ne peuvent être corrects. De plus, on remarque la difficulté de cette méthode à bien estimer les deux paramètres simultanément. Une bonne estimation de l'un d'entre eux se fait apparemment au détriment de l'autre (composant c4). Cependant, on constate, comme dans le cas exponentiel, que l'estimation semble stable quand le nombre de censures à gauche est important, comme par exemple pour le composant 1. Enfin, pour tous les composants, l'estimation du paramètre de forme β semble se faire avec très peu de liberté. Il serait peut être bon d'effectuer une optimisation composante par composante, c'est-à-dire d'optimiser d'abord suivant β, puis suivant η, par exemple par une grille sur $\mathbb{R}^2$, où l'on impose une borne inférieure (strictement positive) pour β.

Estimateurs bayésiens

Le tableau 5.14 représente les trois estimateurs bayésiens et l'estimateur du maximum de vraisemblance. La moyenne a priori sur η a été prise à 30000 pour tous les composants. La moyenne a priori sur β a, quant à elle, été prise à 1.22 pour tous les composants pour l'estimateur de Bayes informatif.

	Bayes non informatif	Bayes semi informatif	Bayes informatif	τ
c1	η = 59354 (*32207*) $\beta = 0.82$ (*0.17*)	η = 47175 (*27283*) $\beta = 1.14$ (*0.12*)	η = 51480 (*19261*) $\beta = 0.95$ (*0.96*)	40.5 %
c2	η = 96886 (*90854*) $\beta = 0.98$ (*0.39*)	η = 52135 (*12698*) $\beta = 1.53$ (*0.44*)	η = 60713 (*41322*) $\beta = 1.18$ (*1.21*)	26.4 %
c3	η = 181181 (*177240*) $\beta = 2.71$ (*1.54*)	η = 184670 (*177589*) $\beta = 2.45$ (*0.77*)	η = 213279 (*37019*) $\beta = 1.61$ (*1.63*)	1.4 %
c4	η = 23819 (*33595*) $\beta = 0.93$ (*0.35*)	η = 22864 (*46060*) $\beta = 1.42$ (*0.34*)	η = 22927 (*32291*) $\beta = 1.07$ (*1.10*)	59.8 %
c5	η = 42018 (*19630*) $\beta = 0.88$ (*0.22*)	η = 34417 (*33133*) $\beta = 1.26$ (*0.23*)	η = 37163 (*37034*) $\beta = 0.97$ (*1.00*)	44.6 %
c6	η = 73425 (*61440*) $\beta = 3.48$ (*2.74*)	η = 256663 (*250301*) $\beta = 2.50$ (*0.85*)	η = 29704 (*28962*) $\beta = 1.19$ (*0.85*)	0 %
c7	η = 281261 (*278922*) $\beta = 1.08$ (*0.66*)	η = 63701 (*37807*) $\beta = 1.88$ (*0.72*)	η = 67994 (*38880*) $\beta = 1.29$ (*0.31*)	17.8 %

TAB. 5.14 – Estimateurs bayésiens pour la loi de Weibull.

On peut remarquer que l'estimateur de Bayes semi informatif (on impose le $\beta \geq 1$, donc un “non rajeunissement” du composant) est assez stable pour les composants présentant un taux de censure à gauche non négligeable. En effet l'écart type de la loi a posteriori du β (le chiffre entre parenthèse) est faible. De plus comme pour le modèle exponentiel le paramètre de forme du composant c6 (ne présentant aucune défaillance) est égal à la moyenne a priori (la moyenne d'une loi uniforme sur $[1, 4]$ étant égale à 2.50), et est très proche (2.45) pour le composant c3 qui présente un faible taux de censures à gauche. On peut d'ailleurs remarquer que pour ce composant, la variance de l'estimateur du paramètre de forme est très grande. Eu égard au nombre de censures à gauche (1.4 %) et à l'estimateur du maximum de vraisemblance ($\hat{\eta} \geq 10^{16}$), nous privilégierons pour ce composant une loi exponentielle.

L'estimateur de Bayes non informatif semble tout d'abord biaisé pour des composants ayant des faibles taux de censures à gauche. Pour les autres, le paramètre de forme est la plupart du temps très proche de 1. Il apparaît donc que l'estimateur du maximum de vraisemblance sera retenu, et que la valeur de β sera égale à 1. On retiendra donc un modèle exponentiel, excepté pour le composant c_3. En effet, les conclusions concernant ce composant doivent être prudentes, eu égard au nombre très faible de censures à gauche (2 sur 71 données). De nouvelles connaissances a priori peuvent dans ce cas précis nous permettre d'utiliser les avis d'experts et donc de choisir l'estimateur de Bayes informatif.

5.5 Conclusions

Dans ce chapitre, une modélisation paramétrique est proposée dans un cadre de données doublement censurées, en considérant la loi exponentielle et la loi de Weibull. Nous avons établi un théorème central limite pour la loi de Weibull. Ce point est très important pour les industriels notamment, car il permet de calculer des intervalles de confiance sur les estimateurs. Néanmoins, nous avons mis en évidence, pour des tailles d'échantillons faibles, les difficultés d'estimation dans un contexte doublement censuré en fiabilité industrielle, et notamment pour le modèle de Weibull. Cependant, l'estimation des paramètres semble bonne lorsque β est plus grand que 1. Ainsi, le modèle de la loi de Weibull doit être utilisé si l'on souhaite se renseigner sur le vieillissement d'un composant. Une alternative a été proposée par Celeux et al. [27] en modélisant le vieillissement par un processus de Poisson afin de tester le vieillissement ou non du composant. Nous avons également proposé une estimation bayésienne des paramètres de la loi exponentielle et de la loi de Weibull, avec des lois a priori informative et non informative. Cette approche bayésienne même en l'absence de connaissance a priori a permis dans beaucoup de cas d'améliorer les résultats et de rendre plus stable les estimateurs. Côté application, les résultats n'ont pas apporté d'information décisive sur le vieillissement des composants. En effet, les paramètres de forme obtenus étaient proches de 1. Le modèle exponentiel nous permettra donc de décrire de manière satisfaisante le comportement de ces composants.

Les perspectives de ce travail sont principalement de proposer dans le cadre de la loi de Weibull, des tests statistiques permettant de tester la valeur du paramètre de forme et donc de tester le vieillissement du composant. Par exemple, on pourrait proposer comme test statistique l'hypothèse nulle $\mathcal{H}_0$: "$\beta = 1$" contre l'hypothèse alternative $\mathcal{H}_1$: "$\beta > 1$". Cette question est là encore cruciale pour les industriels. Il serait également intéressant de reproduire la même démarche pour un système constitué de composants montés en série par exemple. En effet, notre étude porte sur un composant, ou comme nous l'avons vu peut se généraliser à un système que l'on remet à neuf à chaque dégradation. Cette approximation peut s'avérer grossière dans beaucoup de cas. Une généralisation de ce type permettrait de pouvoir appliquer ces modèles à la plupart des cas. Mais, il faudra aussi être capable de tracer les limites d'une inférence à partir de données censurées à droite et à gauche pour des modèles complexes.

Chapitre 6

Modélisation d'un changement de comportement de maintenance

Ce chapitre présente deux modèles d'un changement de politique de maintenance, l'un modélisant un comportement préventif des ingénieurs de maintenance et l'autre modélisant une période de garantie d'un composant.

Le premier modèle fait suite au chapitre précédent, car les données sont du même type, à savoir censurées à gauche et à droite. Cette étude vient d'une collaboration entre l'Inria et l'Électricité De France, ayant fait l'objet d'un stage de DEA [22]. Une publication de ce travail est disponible [26]. Les composants étudiés font partie de pompes primaires 900MW et 1300MW. Après des études exploratoires sur la durée de vie de ces composants, nous avons pu constater que la proportion des composants rebutés était plus importante dans les premières années d'exploitation que dans les années suivantes. Par la suite, avec l'expérience, cette proportion diminue sans que la technologie et le mode de fonctionnement du composant soient modifiés. Ainsi, nous avons supposé que cette différence venait d'un comportement différent des ingénieurs de maintenance. En effet, lors des premières années d'exploitation du composant, les ingénieurs de maintenance ont tendance à rebuter le matériel alors que celui-ci n'est pas dégradé. Ce comportement préventif traduit le manque de connaissance sur le processus de vieillissement du matériel et par conséquent une peur de continuer à faire fonctionner le matériel. Puis après un certain nombre d'années de fonctionnement, les ingénieurs de maintenance ne font plus cette erreur. L'objectif est de modéliser via une variable cachée le comportement préventif des ingénieurs de maintenance. La loi des durées de vie est supposée exponentielle. Nous appliquons l'algorithme EM [73] pour résoudre le problème de données incomplètes. Nous proposons également à partir de l'écriture de l'algorithme EM une estimation du maximum de vraisemblance. Puis, une approche bayésienne avec l'échantillonnage de Gibbs est également envisagée. Enfin une application et des simulations grâce à un modèle de chocs [45, 85] sont données.

Le deuxième modèle de ce chapitre a les mêmes caractéristiques que précédemment. Les données de survie sont donc doublement censurées et nous considérons dans cette partie que le matériel étudié possède une période de garantie. L'objectif est de modéliser un changement de comportement courant lors des deux périodes (de garantie et de non garantie). Nous supposons que pendant la période de garantie, les responsables de la maintenance renvoient le matériel au fournisseur dès que le matériel présente le moindre incident. Une fois cette période passée, les responsables de maintenance décident soit de pratiquer, lors des opérations de maintenance, des réparations minimales s'ils jugent que le matériel est peu dégradé, soit de changer le matériel si celui-ci est considéré comme trop dégradé, soit de ne pas intervenir si le matériel est jugé sain. Nous supposons que les réparations minimales faites sur ces matériels ne sont généralement pas signalées. L'hypothèse de travail est donc pour cette dernière période aussi mauvais que vieux ("as bad as old") lors de réparation minimale, et aussi bon que neuf ("as good as new") lorsque le matériel est changé ou considéré comme remis à neuf. Pour l'étude de la durée de vie sans dégradation du matériel, nous modélisons ce changement de comportement par une variable cachée, et proposons une méthode d'estimation via l'algorithme EM de cette loi de durée de vie sans dégradation.

6.1 Modélisation d'un facteur humain

6.1.1 Motivations

Le composant étudié dans cette section fait partie de ceux étudiés dans le chapitre précédent. Ce composant est intégré dans un système très coûteux, très sûr, et qui est une pièce majeure pour la sûreté et la fiabilité de l'équipement. La politique de maintenance avait été jusque là de surveiller ce système plus ou moins périodiquement avec une période faible pour ne prendre aucun risque ; si bien qu'aucune défaillance n'a été observée. À chaque inspection, les ingénieurs de maintenance décident de rebuter ou non le matériel, selon la présence ou non de dégradation. La variable aléatoire à étudier est dans ce cas la durée de vie du composant sans dégradation, ou vu sous un autre angle, la date d'apparition de la dégradation. Ainsi, les données sont doublement censurées, voir l'exemple 1.7 de Meeker et Escobar [74]. En effet, la donnée est censurée à gauche, si le composant est dégradé lors de l'inspection de maintenance, et est censurée à droite lorsque le composant n'est pas dégradé. La politique de maintenance est donc essentiellement basée sur la fiabilité de ce composant, et plus précisément basée sur la durée de vie moyenne du composant. Nous supposons par la suite que le système est non réparable et est remplacé dès qu'il est rebuté lors d'une inspection, ou qu'il est remis à neuf (hypothèse "aussi bon que neuf").

Afin de modéliser le facteur humain que l'on souhaite prendre en compte, nous définissons une variable cachée Z_i. Ces variables modélisent les erreurs que peuvent faire les ingénieurs de maintenance dans les premières années d'exploitation du composant. Ainsi, un ingénieur de maintenance peut décider de rebuter un composant non dégradé. Ces erreurs lors de l'inspection de maintenance peuvent être nombreuses. Dans ce cas, une analyse statistique des durées de vie serait biaisée par des informations erronées dans la base de données de retour d'expérience. Sans la prise en compte de ce phénomène, la vie moyenne de durée de vie du composant peut être grandement sous-estimée. L'objectif de ce chapitre est de prendre en compte ces erreurs et de réduire ce biais.

Les données de durées de vie sont doublement censurées (à gauche et à droite) et nous ne connaissons jamais la date exacte de dégradation. De plus, nous supposons qu'avant une date calendaire d_0, les ingénieurs de maintenance sont susceptibles de rebuter un composant à tort et donc de le remplacer. Puis après cette date d_0, les ingénieurs de maintenance ne commettront plus jamais cette erreur. Ceci crée un biais dans l'estimation du maximum de vraisemblance du taux de défaillance. Ce biais correspond à une vue pessimiste, et diminue l'estimation de la durée de vie moyenne. Ce modèle a été envisagé suite à une première étude où l'erreur humaine n'était pas prise en compte. Les estimations des durées de vie moyennes paraissaient faibles aux experts. Une étude qualitative en choisissant correctement une date d_0 a permis de relever cette anomalie. La figure 6.1 donne un exemple d'un tel changement de comportement (ici la date calendaire d_0 est 1992). Cette figure représente les pourcentages

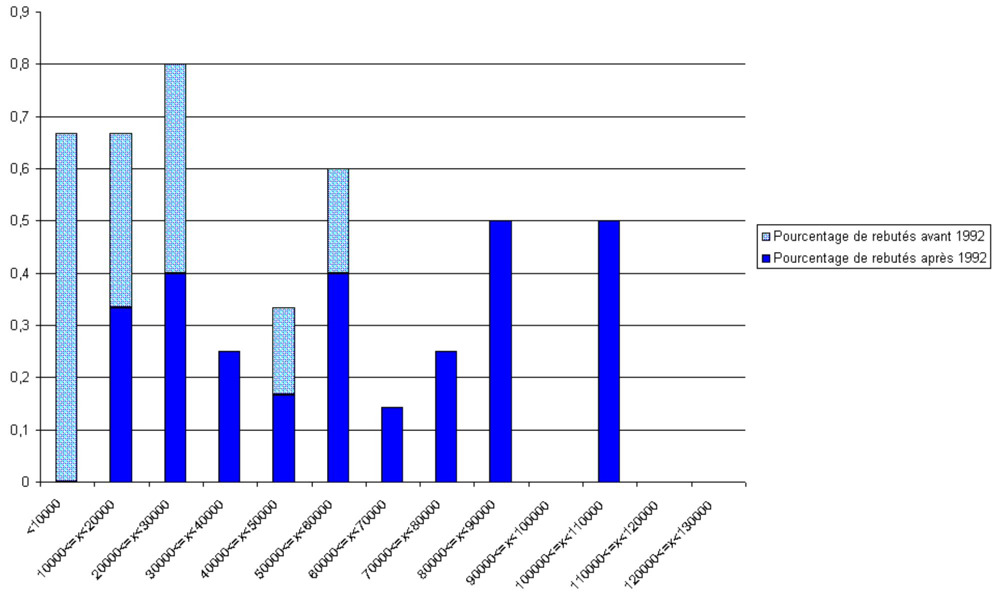


FIG. 6.1 – Pourcentage des rebutés / censurés au cours du temps pour la glace flottante du joint 2 d'une pompe primaire 900 MW

des éléments rebutés en fonction du nombre d'heures de fonctionnement. On a distingué les expertises avant et après une date butoir $d_0 = 1992$. Cette date a été choisie par un expert. Cette figure montre que le pourcentage de composants rebutés lors de leur "jeunesse" avant 1992 est beaucoup plus important que ceux rebutés pendant leur "jeunesse" après 1992. Pour cela, il est intéressant de séparer l'échantillon en prenant en compte cette date. Le tableau 6.1 montre une grande différence dans l'estimation des moyennes de durées de vie.

	Etude sur toutes les données	Données après 1992
MTBF	168002	256909
Taille totale de l'échantillon	68	60
Pourcentage de censures à gauche	32 %	23 %

TAB. 6.1 – Comparaison des moyennes de durées de vie avant et après 1992.

Ces estimations ont été effectuées selon les méthodes décrites dans le chapitre précédent. La taille de l'échantillon totale est de 68, dont 8 inspections ont été effectués avant 1992. Sur ces 8 inspections, les 8 composants ont tous été déclarés rebutés. De plus, la figure 6.1 montre que les durées de vie de ces 8 composants (barres en clair dans l'histogramme de la figure 6.1) ne sont pas particulièrement élevées, ce qui aurait pu justifier les 100% de composants rebutés.

Ces résultats nous ont donc amené à prendre en compte ce changement de comportement par des variables cachées. Notre approche pour traiter cette difficulté consiste à regarder le problème comme un modèle à données incomplètes. Les données manquantes que l'on considère sont des indicatrices traduisant qu'un remplacement a été effectué par précaution. Une approche classique pour traiter un problème de données incomplètes est l'algorithme EM [73] (voir annexe A). Dans toute la suite, on considère la date butoir fixée et connue.

6.1.2 Présentation du modèle

Soit X la durée de vie sans dégradation du matériel. On suppose que cette variable aléatoire est distribuée selon une loi exponentielle, notée $\mathcal{E}(\eta)$ avec pour moyenne η et comme fonction de répartition F

$$\forall t \in \mathbb{R}_+,\ F(t) = 1 - \exp\left(-\frac{t}{\eta}\right), \tag{6.1}$$

et comme fonction de survie R

$$R(t) = \mathbb{P}(X \geq t) = 1 - F(t) = \exp\left(-\frac{t}{\eta}\right). \tag{6.2}$$

On suppose que les composants observés sont indépendants (généralement observés sur différents sites nucléaires) et sont inspectés à des dates déterministes et différentes selon les composants. Après chaque inspection, si le composant est jugé dégradé, le système est alors remplacé et considéré comme aussi bon que neuf. Soit K le nombre total de composants inspectés. Pour le composant i, on note $t_{i,j}$ les temps (en heures) inter-inspections et J_i le nombre d'inspections sur le composant i. Pour ce temps inter-inspection $t_{i,j}$, on note $\delta_{i,j}^{obs}$ l'indicatrice de la décision de l'ingénieur de maintenance. Cette indicatrice prend la valeur 1 si le composant est déclaré rebuté et est donc remplacé, et prend la valeur 0 sinon. L'hypothèse aussi bon que neuf, et la modélisation par la loi exponentielle permet de nous affranchir de l'indice j représentant les inspections. En effet, la loi exponentielle possède la propriété d'absence de mémoire, i.e. $P(X > x + y|X > y) = P(X > x)$. Chaque inspection peut donc être vue comme un nouveau composant. Dans toute la suite, les indices en j seront donc abandonnés et le nombre de composants sera donc égal au nombre d'inspections que l'on notera n. On note $\delta_i = \mathbb{I}_{\{X_i \leq t_i\}}$ indiquant si le matériel est dégradé au temps d'inspection t_i. Pour modéliser le changement de comportement, une date butoir est fixée et notée d_0, de telle sorte qu'avant cette date les ingénieurs de maintenance pouvaient changer les composants par précaution, et qu'après cette date ils ne feront plus jamais cette erreur. Ainsi, après la date d_0, $\delta_i = \delta_i^{obs}$ et donc $P(\delta_i^{obs} = 1) = F(t_i)$. En revanche, avant la date d_0, on a $P(\delta_i^{obs} = 1) \neq F(t_i)$. Nous définissons des variables cachées Z_i pour les composants inspectés avant d_0 lorsque le composant est rebuté ($\delta_i^{obs} = 1$), traduisant le fait que l'ingénieur fasse un changement par précaution. Cette variable prend deux valeurs possibles :

$$z_i = \begin{cases} 1 & \text{si l'ingénieur de maintenance a rebuté à tort le matériel} \\ 0 & \text{si l'ingénieur de maintenannce a rebuté à raison le matériel} \end{cases}$$

On peut résumer toutes les situations par :
- avant d_0
  - * si ($\delta_i = 1$ et $Z_i = 0$) ou ($\delta_i = 0$ et $Z_i = 1$) alors $\delta_i^{obs} = 1$,
  - * si $\delta_i = 0$ alors $\delta_i^{obs} = 0$,
- après d_0 dans tous les cas $\delta_i = \delta_i^{obs}$.

On note α la probabilité de rebut à tord, soit $p(Z_i = 1)$. Finalement pour la période avant d_0, on peut calculer la probabilité d'observer un rebut et que l'ingénieur se soit trompé :

$$P(\delta_i^{obs}, Z_i = 1) = \alpha R(t_i). \tag{6.3}$$

La probabilité d'observer un rebut et que l'ingénieur est rebuté le matériel à raison est égale à :

$$P(\delta_i^{obs}, Z_i = 0) = (1 - \alpha)F(t_i). \tag{6.4}$$

Ainsi la structure des données incomplètes est la suivante :
- Les données observées sont les temps d'inspection et les indicateurs du remplacement d'un composant δ_i^{obs} pour $i = 1, \ldots, n$, n étant le nombre total d'inspections.
- Les données manquantes ou cachées qui sont les indicateurs z_i des remplacements par précaution pour $i = 1, \ldots, n_0$, où n_0 $(1 < n_0 < n)$ est le nombre d'inspections avant d_0.

La vraisemblance du modèle va donc comprendre deux parties, l'une correspondant aux inspections avant d_0 et une autre correspondant aux inspections après d_0. En notant $\mathbf{c} = ((t_i, \delta_i, z_i), i = 1, \ldots, n))$ les données complètes, la vraisemblance des données complétées s'écrit

$$\begin{aligned}\mathcal{L}(\eta,\alpha;\mathbf{c}) &= \prod_{i=1}^{n_0}\left\{(\alpha R(t_i\mid\eta))^{z_i}((1-\alpha)F(t_i\mid\eta))^{1-z_i}\right\}^{\delta_i^{obs}} R(t_i\mid\eta)^{1-\delta_i^{obs}} \\ &\quad \prod_{i=n_0+1}^{n} F(t_i\mid\eta)^{\delta_i^{obs}} R(t_i\mid\eta)^{1-\delta_i^{obs}}. \end{aligned} \tag{6.5}$$

La deuxième partie de cette vraisemblance correspondant aux inspections après d_0 est la même que dans le chapitre précédent. Comme nous allons le voir, l'algorithme EM (cf. annexe A) n'apporte rien de plus que la maximisation directe de la vraisemblance observée. En effet, bien que ce modèle soit à structure incomplète, l'étape M est d'une complexité analogue que le calcul direct de l'estimateur du maximum de vraisemblance. Cependant, nous décrivons ci-après l'algorithme EM pour une possible et future généralisation de ce modèle à la loi de Weibull.

L'étape E de l'algorithme nécessite le calcul de l'espérance de la log-vraisemblance des données complètes. Cette espérance conditionnelle va s'écrire en fonction de l'espérance conditionnelle de la variable cachée sachant les données observées et la valeur courante des paramètres η^r et α^r, où r est l'itération de l'algorithme, c'est-à-dire

$$\mathbb{E}[Z_i|\delta_i^{obs}=1,t_i,\eta^r,\alpha^r] = \frac{\alpha^r R(t_i|\eta^r)}{\alpha^r R(t_i|\eta^r)+(1-\alpha^r)F(t_i|\eta^r)}. \tag{6.6}$$

On notera par la suite cette quantité κ_i^r. L'étape E de l'algorithme revient donc à écrire l'espérance conditionnelle de la log-vraisemblance, notée $\mathcal{Q}(\eta,\alpha|\eta^r,\alpha^r)$:

$$\begin{aligned} Q(\eta,\alpha\mid\eta^r,\alpha) &= \sum_{1\le i\le n_0} \delta_i^{obs}\{\kappa_i^r\log(\alpha R(t_i|\eta)) + (1-\kappa_i^r)\log(1-\alpha)F(t_i|\eta)\} \\ &+ (1-\delta_i^{obs})\log(R(t_i|\eta)) \\ &+ \sum_{n_0<i\le n} \delta_i^{obs}\log(F(t_i|\eta)) + (1-\delta_i^{obs})\log(R(t_i|\eta)) \end{aligned} \tag{6.7}$$

L'étape M consiste à maximiser cette fonction en η et en α , où les maximiseurs seront notés η^{r+1} et α^{r+1}. La maximisation selon le paramètre α peut être facilement réalisée. En effet, cela revient à maximiser, en α, l'expression suivante

$$\sum_{i=1}^{n_0}\delta_i^{obs}\kappa_i^r\log\alpha + \sum_{i=1}^{n_0}\delta_i^{obs}(1-\kappa_i^r)\log(1-\alpha). \tag{6.8}$$

On a donc

$$\alpha^{r+1} = \frac{\sum_{i=1}^{n_0}\delta_i^{obs}\kappa_i^r}{\sum_{i=1}^{n_0}\delta_i^{obs}} = \frac{\sum_{i=1}^{n_0}\delta_i^{obs}\dfrac{\alpha^r R(t_i|\eta^r)}{\alpha^r R(t_i|\eta^r)+(1-\alpha^r)F(t_i|\eta^r)}}{\sum_{i=1}^{n_0}\delta_i^{obs}} \tag{6.9}$$

De la même façon, on peut maximiser en η la fonction

$$\begin{aligned} &\sum_{1\leq i\leq n_0} \delta_i^{obs}\{\kappa_i^r \log(R(t_i|\eta)) + (1-\kappa_i^r)\log(F(t_i|\eta))\} + (1-\delta_i^{obs})\log(R(t_i|\eta)) \\ + &\sum_{n_0<i\leq n} \delta_i^{obs}\log(F(t_i|\eta)) + (1-\delta_i^{obs})\log(R(t_i|\eta)) \end{aligned} \quad (6.10)$$

L'algorithme est donc itératif. Le critère d'arrêt est fonction de la différence entre deux estimations successives.

Pour notre étude, la probabilité de rebut par précaution sera constante et estimée par

$$\alpha = \frac{\sum_{i=1}^{n_0}\delta_i^{obs}}{n_0} - \frac{\sum_{i=n_0+1}^{n}\delta_i^{obs}}{n-n_0}.$$

L'algorithme EM servira donc à mettre à jour uniquement le paramètre d'intérêt, à savoir la durée de vie du matériel.

Une étape importante dans l'algorithme EM est l'initialisation des paramètres. En effet, cet algorithme est sensible au choix des valeurs initiales au risque de converger très lentement. Ainsi, nous proposons deux valeurs initiales, l'une optimiste qui consiste à remplacer toutes les censures à gauche avant d_0 par des censures à droite, et l'autre pessimiste qui consiste à penser que les ingénieurs de maintenance ne remplacent jamais à tort le matériel, que l'on note respectivement η_{opt}^0 et η_{pes}^0. On a

$$\eta_{opt}^0 = \frac{\sum_{i=1}^{n} t_i}{\sum_{i=n_0+1}^{n} \delta_i}, \quad (6.11)$$

et

$$\eta_{pes}^0 = \frac{\sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} \delta_i}. \quad (6.12)$$

En outre, si le nombre de données après d_0 est suffisant, on préconisera alors comme valeur initiale :

$$\eta^0 = \frac{\sum_{n_0+1\leq i\leq n} t_i}{\sum_{n_0+1\leq i\leq n} \delta_i}. \quad (6.13)$$

On peut également calculer directement la vraisemblance observée et donc en déduire l'estimateur du maximum de vraisemblance par une maximisation directe. Pour cela, il suffit, à partir de l'équation 6.5, de sommer sur toutes les valeurs possibles des variables cachées. En notant $\mathbf{o} = ((t_i, \delta_i), i = 1, \ldots, n))$ les données observées, la vraisemblance des données observées est égale à

$$\begin{aligned}\mathcal{L}(\eta; \mathbf{o}) &= \prod_{i=1}^{n_0} R(t_i \mid \eta)^{1-\delta_i^{obs}} \left((1-\alpha)F(t_i \mid \eta) + \alpha R(t_i \mid \eta)\right)^{\delta_i^{obs}} \\ &\times \prod_{i=n_0+1}^{n} (F(t_i \mid \eta))^{\delta_i^{obs}} (R(t_i \mid \eta))^{(1-\delta_i^{obs})}. \end{aligned} \tag{6.14}$$

On pourra choisir comme α la différence entre le pourcentage de rebuts avant d_0 et le pourcentage de rebuts après d_0 :

$$\alpha = \frac{\sum_{i=1}^{n_0} \delta_i^{obs}}{n_0} - \frac{\sum_{i=n_0+1}^{n} \delta_i^{obs}}{n - n_0}.$$

L'estimateur du maximum de vraisemblance est la solution du problème d'optimisation suivant

$$\eta^{ML} = argmax_{\eta \in \mathbb{R}_+} \log \mathcal{L}(\eta; \mathbf{o}). \tag{6.15}$$

Ce problème d'optimisation peut être résolu facilement par un algorithme du gradient qui n'est pas plus complexe que l'étape M de l'algorithme EM. L'algorithme EM et la maximisation de la vraisemblance observée ont été appliqués sur des composants de centrales nucléaires.

6.1.3 Approche bayésienne

En présence d'échantillons de petites tailles et avec peu de censures à gauche, l'estimateur du maximum de vraisemblance peut donner des résultats médiocres (par exemple lorsque n est inférieur à 20). Dans de tels cas, une approche bayésienne peut s'avérer intéressante pour régulariser les estimateurs. Dans ce paragraphe, afin de rendre plus lisible les équations, le paramètre d'intérêt est le taux de défaillance instantané $\lambda = \frac{1}{\eta}$. Ici, la probabilité que l'ingénieur de maintenance fasse une erreur sera estimée par la quantité :

$$\alpha = \frac{\displaystyle\sum_{i=1}^{n_0} \delta_i^{obs}}{n_0} - \frac{\displaystyle\sum_{i=n_0+1}^{n} \delta_i^{obs}}{n - n_0}.$$

Dans un cadre bayésien, λ est une variable aléatoire de loi a priori $\pi(\lambda)$. L'inférence bayésienne consiste à calculer les estimateurs issus de la loi a posteriori $\pi(\lambda|\mathbf{o})$ définie comme

$$\pi(\lambda|\mathbf{o}) = \frac{\pi(\lambda)P(\mathbf{o}|\lambda)}{\int \pi(\lambda)P(\mathbf{o}|\lambda)d\lambda},$$

où $P(\mathbf{o}|\lambda)$ est la vraisemblance du paramètre λ pour les données $\mathbf{o} = (d_0, (t_i, \delta_i), i = 1, \ldots, n))$. Alors, l'estimateur bayésien de λ pour une fonction de perte quadratique est l'espérance a posteriori de λ (cf. [84])

$$E(\lambda|\mathbf{o}) = \frac{\int \lambda\pi(\lambda)P(\mathbf{o}|\lambda)d\lambda}{\int \pi(\lambda)P(\mathbf{o}|\lambda)d\lambda}.$$

Dans un contexte doublement censuré, cette formule n'est pas exploitable. Les méthodes MCMC (Markov Chain Monte Carlo) ont pour but d'évaluer l'espérance a posteriori par l'intégration de Monte Carlo (voir annexe C) en générant des chaînes de Markov (cf. [23], [49]). Plus précisément, nous utilisons l'échantillonnage de Gibbs pour approximer la distribution a posteriori de λ. L'échantillonnage de Gibbs consiste à simuler un échantillon à partir des distributions conditionnelles afin de générer une chaîne de Markov dont la distribution stationnaire correspond à la distribution a posteriori désirée. On décrit cette algorithme ci-dessous.

6.1.3.1 Choix de la loi a priori

La loi a priori choisie pour le paramètre λ est une loi Gamma de paramètres a et b et notée $\mathcal{G}(a, b)$ et a pour densité

$$\pi(\lambda) = \frac{b^a \lambda^{a-1}}{\Gamma(a)} exp(-\lambda b)\mathbb{I}_{R_+}, \tag{6.16}$$

où $\Gamma(n) = \displaystyle\int_{R_+} x^{n-1}e^{-x}dx$ est la fonction Gamma. L'espérance de cette loi est

$$\mathbb{E}[\lambda] = \frac{a}{b}, \tag{6.17}$$

et la variance est

$$Var(\lambda) = \frac{a}{b^2}. \tag{6.18}$$

Plusieurs raisons nous incitent à choisir cette loi comme a priori. Premièrement, grâce à ses deux paramètres, les formes possibles de cette distribution sont nombreuses. De plus, cette loi est une loi conjuguée, c'est-à-dire que la loi a posteriori est également une loi Gamma. En effet, si X suit une loi exponentielle de paramètre $\eta = \frac{1}{\lambda}$, la loi a posteriori est une loi Gamma $\mathcal{G}(a + 1, x + b)$.

6.1.3.2 Description de l'algorithme de Gibbs

On tire λ selon une loi a priori $\mathcal{G}(a, b)$.

◇ **Répéter G fois :**

- **Pour chaque i dans $\{1, n\}$**

  Si Z est la variable aléatoire du leurre, tirer z_i tel que :

∘ *Si* $i \leq n_0$

$$* \text{ Si } \delta_i = 1, \begin{cases} z_i = 1 & \text{avec la probabilité} \quad \alpha, \\ z_i = 0 & \text{avec la probabilité} \quad 1-\alpha. \end{cases}$$

∗ Si $\delta_i = 0$, $z_i = 0$.

∘ *Si* $i > n_0$, $z_i = 0$.

On tire $\tilde{t}_i$ suivant une loi $\mathcal{E}(\eta = \frac{1}{\lambda})$, et :

∗ Si $\delta_i = 1$ et $z_i = 0$, on répète le tirage de $\tilde{t}_i$ jusqu'à ce que $\tilde{t}_i < t_i$ (censure à gauche due à un rebut justifié)

∗ Si $\delta_i = 0$ ou si $\delta_i = 1$ et $z_i = 1$, on répète le tirage de $\tilde{t}_i$ jusqu'à ce que $\tilde{t}_i > t_i$ (censure à droite due à un rebut par précaution ou composant accepté en l'état après expertise)

On a alors : $(\lambda \mid (t_i)_{1 \leq i \leq n}) \sim \mathcal{G}(a+n, \sum_i \tilde{t}_i + b)$

On tire λ suivant cette loi, et on met à jour les paramètres.

Fin Répéter.

Calcul de $\pi^g = \dfrac{\sum_{i=1}^{n_0} z_i^{(g)}}{\sum_{i=1}^{n_0} \delta_i}$ où g est une itération de l'algorithme ($g \in \{1, G\}$)

Fin Répéter.

Remarque : La propriété d'absence de mémoire est utilisée pour le tirage de la loi exponentielle. En effet, il n'est pas nécessaire de répéter les tirages jusqu'à ce que la valeur soit supérieur à $\tilde{t}_i$, il suffit pour cela de tirer selon une loi exponentielle décalée avec comme paramètre de décalage $\tilde{t}_i$. Cela permet d'alléger le nombre de tirages.

À la fin de la procédure, la suite $\{\lambda^g, g = g_0, \ldots, G\}$ peut être vue comme un échantillon de la distribution a posteriori de λ. L'entier g_0 définit la longueur de la période de chauffe. Celui-ci doit être choisi assez grand de telle sorte que la valeur initiale soit "oubliée". L'entier G a été choisi également assez grand afin que la loi a posteriori de λ soit bien approximée. Le choix de g_0 et de G reste un problème difficile et ouvert (voir Robert [84]). Dans les applications numériques, nous avons choisi ces deux entiers de façon empirique. Il s'avère que $g_0 = 1000$ et $G = 10000$ donnent de bons estimateurs de η par la formule

$$\hat{\eta} = \frac{1}{G - g_0} \sum_{g=g_0+1}^{G} \eta^g. \tag{6.19}$$

On peut également estimer le pourcentage de rebuts par précaution :

$$\hat{\pi} = \frac{1}{G} \sum_g \pi^g. \tag{6.20}$$

Par ailleurs, les estimations successives de η peuvent être fortement corrélées. Pour corriger cette corrélation, il est possible de ne pas tenir compte de toutes les itérations. Une solution consiste à garder les itérations à pas constant. Par exemple on choisit un pas de 50, et on ne garde que les estimateurs $\eta_{50}, \eta_{100}, \eta_{150}, \ldots$. L'inconvénient est que pour obtenir un nombre suffisant d'estimateurs, il faut un grand nombre d'itérations de l'algorithme d'échantillonnage de Gibbs. Tout ceci peut rendre cet algorithme très lent.

6.1.4 Résultats numériques

Nous présentons dans ce paragraphe les résultats numériques obtenus sur un composant du joint de la pompe primaire 900 MW. Puis des simulations via un modèle de chocs ont permis de tester la robustesse des estimateurs et d'illustrer l'intérêt de cette approche dans des cas où le manque d'information est important.

6.1.4.1 Applications

Les résultats des données réelles correspondant à la figure 6.1 sont données dans le tableau 6.2. Pour notre exemple, la taille de l'échantillon est de 68, dont 8 ont été inspectés avant 1992. Sur ces 8 composants, tous ont été rebutés. Dans ce tableau, η_m représente le point de vue pessimiste, où l'on suppose que tous les remplacements avant 1992 sont justifiés. η_M représente l'estimateur du maximum de vraisemblance dans un cas optimiste où l'on suppose que tous les remplacements avant 1992 sont considérés comme des rebuts par précaution et sont donc non justifiés. Enfin, η_{EM} est l'estimateur du maximum de vraisemblance, calculé à partir de l'algorithme EM.
L'estimateur du maximum de vraisemblance permet de réduire le biais dû à des remplace-

d_0	% repl. before d_0	% repl. after d_0	η_m	η_M	η_{EM}
1992	100	20	168002	272916	267544

TAB. 6.2 – Estimation de η pour les données correspondantes à la figure 6.1.

ments à tort. En effet, la différence entre η_m et η_{ml} est de 100000 heures. Ici, la probabilité de rebut par précaution (α) a été estimée à 1. Ceci s'explique par le faible nombre d'inspections avant 1992 (8 sur 68 inspections au total) et que sur ces 8 données, toutes sont des défaillances avec des durées de vie assez faibles. Mais ne connaissant pas la vraie valeur du paramètre, il est difficile d'étudier la réduction du biais sur un cas réel. Nous proposons des expérimentations numériques sur des simulations.

6.1.4.2 Simulations

Pour les simulations, un modèle de chocs est défini et simulé selon notre modèle de changement de comportement de maintenance. Supposons que le système reçoive des chocs aléatoirement. Des articles récents traite les modèles de chocs (voir par exemple [45]). Supposons que la distribution du temps entre deux chocs suit une loi Gamma $\mathcal{G}(1/s, \eta)$ et que les chocs sont des événements indépendants. On suppose qu'après s chocs le système est dégradé et que l'on procède à son remplacement. Ainsi, la distribution entre deux remplacements est une loi exponentielle $\mathcal{E}(\eta)$. Autrement dit, la loi des durées de vie du système est une loi exponentielle de paramètre η. Nous supposons également qu'un comportement préventif existe pour les inspections ayant eu lieu avant une date d_0. Pour ces n_0 inspections, les ingénieurs de maintenance décident de remplacer le système dès qu'il a reçu $s' < s$ chocs. Mais, pour les $n - n_0$ inspections restantes, un remplacement est décidé après s chocs. Le problème statistique revient à estimer le paramètre η, durée de vie moyenne du composant. Finalement, ce modèle de chocs dépend des quantités suivantes

- n est le nombre total des inspections,
- n_0 est le nombre d'inspections pour lesquelles un remplacement inopportun est possible (il est équivalent de se donner n_0 ou se donner une date butoir d_0),
- s le nombre de chocs produisant un remplacement justifié (ou si l'on préfère une défaillance),
- s' le nombre de chocs produisant un remplacement non justifié,
- η le paramètre à estimer de la loi exponentielle.

Nous avons simulé ce modèle de chocs pour différentes valeurs des paramètres s', n et n_0. La vraie valeur du paramètre η est $\eta = 20000$. On pose $s = 10$. Les résultats sont données dans le tableau 6.3. Chaque simulation a été répétée 100 fois, puis les estimateurs ont été moyennés. Le tableau donne également entre parenthèses l'écart type des estimateurs. Les résultats sont satisfaisants. Notre procédure a donné des estimateurs raisonnables de η et le biais pessimiste a été en grande partie éliminé. Comme attendu, les résultats sont plus variables pour des tailles d'échantillons faibles ($n = 20$). L'estimateur calculé à partir de l'algorithme EM η_{EM} semble moins fiable quand $n = 20$ et $s' = 5$ et surestime la durée de vie moyenne du composant. Le paramètre s' mesure l'importance du facteur humain. En effet, une valeur proche de $s = 10$ signifie que les ingénieurs de maintenance ne font pas beaucoup de remplacements par précaution et donc que le nombre de censures à gauche est faible. Une faible valeur de s' signifie que beaucoup de remplacements par précaution sont effectués, donc que le nombre de censures à gauche est important. L'estimateur du maximum de vraisemblance arrive à corriger la sur-estimation de η quand s' est grand et la sous-estimation de η quand s' est petit.

Ainsi, pour le cas particulier ($n = 20, n_0 = 10, s = 10, s' = 7$), nous avons proposé une estimation bayésienne par l'algorithme d'échantillonnage de Gibbs avec comme loi a priori, une loi Gamma $G(1/10, 200000)$. Cela correspond à une bonne information a priori, car la moyenne a priori est 20 000 (la vraie valeur) et la variance a priori est de 4×10^9. On peut remarquer que l'estimateur bayésien est plus stable que le maximum de vraisemblance notamment pour des échantillons de petites tailles. Dans cet exemple, nous n'avons pas fait 10 000 itérations pour l'échantillonnage de Gibbs (ceci s'avère trop long), mais nous l'avons fait fonctionner pour 10 échantillons différents. D'après ces résultats, on peut affirmer que si le nombre de censures à gauche est suffisant, alors il y a peu de différence entre l'estimateur du maximum de vraisemblance et l'estimateur bayésien. En revanche, si le nombre de censures à gauche est faible (dans notre exemple moins de 10 remplacements sur un échantillon de

n	n_0	s'	η_M	η_m	η_{EM}	α
100	50	5	57823 (*6551*)	14999 (*1920*)	18116 (*3174*)	0.22 (*0.08*)
100	30	5	36907 (*4310*)	17170 (*2152*)	19302 (*3116*)	0.21 (*0.09*)
100	70	5	46110 (*26200*)	15947 (*2352*)	18016 (*3102*)	0.21 (*0.09*)
20	10	5	70271 (*32973*)	16259 (*4135*)	28680 (*28518*)	0.27 (*0.18*)
20	10	3	67009 (*23237*)	14333 (*4298*)	23869 (*18183*)	0.33 (*0.15*)
20	10	7	62505 (*24912*)	18261 (*6037*)	23359 (*13687*)	0.16 (*0.17*)
100	50	7	56859 (*8220*)	16801 (*2318*)	18673 (*3450*)	0.13 (*0.08*)

TAB. 6.3 – Estimation de η pour des données simulées à partir d'un modèle de chocs.

taille 20), alors l'estimateur bayésien est bien meilleur que l'estimateur du maximum de vraisemblance comme illustré dans le tableau 6.4. Dans ce tableau, l'estimateur bayésien est noté η_{ba}.

num. of repl.	η_m	η_M	η_{ml}	η_{ba}
7	23214	44814	28999	18801
9	16727	34761	22407	18169
9	16727	44814	23536	18423

TAB. 6.4 – EMV et estimateurs bayésiens de η pour trois échantillons de faible taille avec peu de censures à gauche.

6.1.5 Conclusions

Nous avons proposé de modéliser un facteur humain et de la prendre en compte par un modèle de données incomplètes. L'écriture de l'algorithme EM nous a également permis d'écrire simplement la vraisemblance des données observées. Pour ce modèle, nous préconisons donc si la taille de l'échantillon est suffisante d'effectuer un algorithme afin de résoudre ce problème de données incomplètes. Cette prise en compte du facteur humain a permis de réduire considérablement le biais dans l'estimation de la moyenne de durées de vie du composant. Les applications numériques pour la loi exponentielle nous ont montré que notre procédure est facilement implémentable. De plus, même si l'algorithme de l'échantillonnage de Gibbs est plus lourd, l'inférence bayésienne est également facilement envisageable et efficace lorsque la taille de l'échantillon est faible ou lorsque le nombre de censures à gauche est faible.

Une perspective naturelle de ce travail est la prise en compte d'un possible vieillissement modélisé par une loi de Weibull. On peut craindre comme nous l'avons vu dans le chapitre précédent que l'estimateur du maximum de vraisemblance pour cette loi soit beaucoup moins stable. Ainsi, l'algorithme EM permettra certainement dans des travaux futurs de transcrire facilement toutes les équations pour la loi de Weibull et de résoudre certains de ces problèmes. De plus, l'approche bayésienne proposée ici permettra d'établir une solution pour la loi de Weibull (voir [9]).

6.2 Un modèle de garantie

La prise en compte d'un changement de comportement de maintenance dans l'étude précédente nous a conduit à imaginer la modélisation d'une période de garantie. En effet, lorsqu'un matériel est encore sous garantie, il est naturel, avant la fin de la période de garantie, de noter la moindre dégradation du matériel et de renvoyer ce matériel au fournisseur. Puis, lorsque le matériel n'est plus sous garantie, le comportement des ingénieurs de maintenance n'est plus le même. Si le matériel possède une dégradation mineure (ou plus exactement que les ingénieurs de maintenance jugent mineure), alors les ingénieurs de maintenance effectuent une réparation minimale du matériel sans noter nécessairement l'anomalie. De la sorte le matériel est jugé aussi mauvais que vieux. De plus, le fait que la réparation effectuée n'est pas forcément enregistrée dans les bases de données, entraîne un manque d'information que l'on va modéliser par une variable cachée. Ainsi, pendant la période de non garantie, certaines censures à gauche (dégradations) ne sont pas observées.

Dans ce paragraphe, nous traitons d'un matériel dégradable sujet à une période de garantie et où les données sont doublement censurées. La loi des durées de vie sans dégradation est supposée exponentielle. La modélisation est donc analogue à l'étude précédente, avec cette fois-ci une observation du vrai processus avant une date d_0 et un manque d'information après cette date. Ici, la date d_0 n'est pas une date calendaire (contrairement à l'étude précédente), mais un nombre d'heures de fonctionnement. Cependant, cette différence n'entraîne pas de modification dans les équations.

6.2.1 Notations et hypothèses

Soit X_i les durées de vie sans dégradation du matériel i, ou de manière équivalente, les dates d'arrivée de la première dégradation. On suppose, comme dans le modèle précédent, que la loi des durées de vie est une loi exponentielle, notée $\mathcal{E}(\eta)$, où η représente la durée moyenne de vie sans dégradation du matériel, ou de façon équivalente la durée moyenne d'apparition de la dégradation. Cette hypothèse nous permet de traiter les deux périodes (de garantie et de non garantie) séparément et de considérer grâce à la propriété d'absence de mémoire qu'un matériel n'est inspecté qu'une seule fois. Lors d'une inspection, le matériel est déclaré dégradé ou non, mais nous supposons comme dans l'étude précédente que la dégradation ne survient pas pendant la visite du matériel. Cette hypothèse paraît vraisemblable dans la mesure où le matériel n'est en général pas en fonctionnement lors de l'inspection. On note d_0 la durée de la période de garantie pour ce matériel et δ_i^{obs}, la variable indicatrice de la décision de l'opérateur de maintenance au temps inter-inspection t_i. On note

$$\delta_i = I_{X_i \leq t_i},$$

la variable indicatrice de la dégradation du matériel i. Ainsi, δ_i représente le vrai processus non observé, et δ_i^{obs} représente les données observées. La caractérisation du problème que nous voulons traiter est que pendant la période de garantie, ces deux quantités sont égales, c'est-à-dire que l'on observe le vrai processus de dégradation. Pendant la période de garantie, à chaque opération de maintenance effectuée au temps inter-inspection $t_{i,j}$, pour chaque dégradation, le matériel est envoyé au fournisseur et un matériel neuf le remplace. Cette hypothèse correspond au cas aussi bon que neuf, ou à un processus de renouvellement.

6.2.2 Présentation du modèle

Comme nous l'avons vu dans l'étude précédente, l'hypothèse de la loi exponentielle et l'hypothèse "aussi bon que neuf" vont nous permettre de trier les données, de telle sorte que l'on puisse traiter de façon indépendante les données correspondants à la période de garantie et celles de non garantie. Pour la période de garantie, où l'on observe le vrai processus, la vraisemblance s'écrit de façon analogue à celle du chapitre 5. Après la période de garantie (nombre d'heures de fonctionnement supérieur à d_0), à chaque opération de maintenance, une autre politique est adoptée. En effet, si le matériel possède une dégradation jugée mineure, celle-ci n'est pas forcément déclarée et est réparée de façon minimale, i.e. le composant est considéré comme aussi mauvais que vieux. Cette manière d'opérer introduit des censures à gauche non observées dans la durée de vie du matériel qui, si elles ne sont pas prises en compte produisent un biais dans l'estimation du paramètre η, et plus précisément une surestimation de η. Si le matériel est considéré trop dégradé, alors ce matériel est remplacé par un composant neuf, ou réparé et considéré comme neuf (hypothèse "aussi bon que neuf"). Enfin, si celui-ci ne présente pas de dégradation, la donnée correspondante est une censure à droite. Afin de modéliser ce manque d'information, une variable cachée est utile pour décrire le phénomène. On définit, conditionnellement à l'évènement $\delta_i^{obs} = 0$, Z_i, la variable indiquant si une réparation minimale non déclarée a été effectuée pendant l'opération de maintenance. On note $\alpha = P(Z_i = 1)$. L'ensemble des situations possibles pour cette période peuvent être résumé comme suit :

- $\delta_i^{obs} = 0$ et $\delta_i = 0$, i.e. $Z_i = 0$ (censure à droite),
- $\delta_i^{obs} = 0$ et $\delta_i = 1$, i.e. $Z_i = 1$ (une réparation minimale a été effectuée ou censure à gauche non observée),
- $\delta_i^{obs} = 1$ et $\delta_i = 1$ (censure à gauche observée).

Lorsque une censure à gauche est observée, le matériel i est changé ou remis à neuf, et un nouveau composant est considéré (l'indice i est alors incrémenté).

On peut donc écrire la vraisemblance des données complétées

$$\mathcal{L}(\eta,\alpha;\mathbf{c}) = \prod_{i=1}^{n_0} R(t_i)^{1-\delta_i^{obs}} F(t_i)^{\delta_i^{obs}} \prod_{i=n_0+1}^{n} F(t_i)^{\delta_i^{obs}} [(\alpha F(t_i))^{z_i}((1-\alpha)R(t_i))^{(1-z_i)}]^{1-\delta_i^{obs}}. \tag{6.21}$$

Pour l'application de l'algorithme EM, il est nécessaire de calculer la probabilité conditionnelle d'une réparation minimale non déclarée sachant que le matériel n'a pas été rebuté

$$P(Z_i = 1|\delta_i^{obs} = 1, \eta^r, \alpha^r) = \frac{\alpha^r F(t_i|\eta^r)}{\alpha^r F(t_i|\eta^r) + (1-\alpha^r)R(t_i|\eta^r)} \tag{6.22}$$

Pour cette étude, comme dans la précédente, nous avons calculé, à partir de la vraisemblance des données complètes (équation (6.21)), la vraisemblance des données observées en sommant sur toutes les valeurs possibles des variables cachées Z_i :

$$\mathcal{L}(\eta,\alpha;\mathbf{o}) = \prod_{i=1}^{n_0} R(t_i)^{1-\delta_i^{obs}} F(t_i)^{\delta_i^{obs}} \prod_{i=n_0+1}^{n} F(t_i)^{\delta_i^{obs}} \left[(\alpha F(t_i))^{1-\delta_i^{obs}}((1-\alpha)R(t_i))^{1-\delta_i^{obs}}\right]. \tag{6.23}$$

On pourra choisir pour *alpha* la différence entre le pourcentage de censures à droite après la période de garantie et le pourcentage de censures à droite avant la période de garantie :

$$\alpha = \frac{\sum_{i=n_0+1}^{n}(1-\delta_i^{obs})}{n-n_0} - \frac{\sum_{i=1}^{n_0}(1-\delta_i^{obs})}{n_0} = \frac{\sum_{i=1}^{n_0}\delta_i^{obs}}{n_0} - \frac{\sum_{i=n_0+1}^{n}\delta_i^{obs}}{n-n_0}.$$

Ici encore, l'estimateur du maximum de vraisemblance issu de la maximisation de cet équation possède, comme nous le verrons dans les simulations, un bon comportement. Là encore, il est envisageable d'appliquer l'algorithme EM par exemple dans le cadre d'une généralisation par une loi de Weibull. L'écriture de cet algorithme se fait de manière analogue à l'étude précédente et n'est pas rappelé ici.

6.2.3 Simulations

On reprend ici la même méthodologie que le modèle précédent. Afin de modéliser ce type de données, nous appliquons un modèle de chocs (voir par exemple [45] et [85]). Supposons que le matériel subissent des chocs de façon aléatoire. Supposons que la distribution des temps entre les chocs est une loi gamma $\mathcal{G}(1/s_1, \eta)$ et que les chocs sont des événements indépendants entre eux. Supposons également que si le composant reçoit entre s_1 et s_2 chocs (avec $s_1 \leq s_2$) alors les ingénieurs de maintenance considèrent que le composant présente une dégradation mineure, et que si le nombre de chocs est supérieur ou égale à s_2, les mêmes ingénieurs décident que le composant est trop dégradé et décident alors de le remplacer ou de le remettre à neuf. On suppose que la dégradation arrive lorsque le nombre de chocs atteint s_1. Ainsi la loi des durées de vie sans dégradation est la somme de s_1 lois Gamma indépendantes, égale à une loi Gamma $\mathcal{G}(1,\eta) = \mathcal{E}(\eta)$, i.e. une loi exponentielle de moyenne η. Les résultats sont donnés dans le tableau 6.5 avec 100 simulations, $s_1 = 10$, avec 3 temps inter-inspections $t_1 = 1000$, $t_2 = 2000$, 3000, et avec une période de garantie égale aux 1100 premières heures, et avec $\eta = 3500$. Ainsi, tous les composants sont inspectés au bout de 1000, 3000 et 6000 heures de fonctionnement. Nous donnons la moyenne des 100 estimateurs ainsi que leurs écarts types entre parenthèses. Le paramètre s_2 mesure l'importance du changement de comportement de maintenance. Ainsi, lorsque celui-ci est égal à 10, aucune erreur n'est faite et plus ce paramètre est grand, plus les ingénieurs de maintenance font des réparations non déclarées. Dans ce tableau, $\hat{\eta}_{NC}$ représente l'estimateur du maximum de

n	s_2	$\hat{\eta}_{NC}$	$\hat{\eta}_{EM}$
50	20	8912 (*2164*)	4257 (*1294*)
20	20	9850 (*6072*)	5092 (*3240*)
20	10	4569 (*980*)	5876 (*4511*)
20	30	19600 (*9215*)	4857 (*2470*)

TAB. 6.5 – Différence entre deux estimateurs, $\hat{\eta}_{ML}$ prenant en compte le changement de comportement et $\hat{\eta}_{NC}$.

vraisemblance si l'on ne tient pas compte du changement de comportement de maintenance (cf. chapitre 5 pour l'estimation). L'estimateur $\hat{\eta}_{EM}$ est calculé via l'algorithme EM. On peut remarquer le biais produit par un changement de comportement de maintenance peut être très important. L'estimateur du maximum de vraisemblance sur-estime quelque peu le paramètre, mais possède néanmoins un écart type assez faible. L'estimateur du maximum de vraisemblance reste une bonne estimation de la moyenne des durées de vie en présence d'échantillons de petites tailles.

On peut remarquer que dans le cas où il n'existe pas de changement de comportement de maintenance ($s_2 = s_1 = 10$), l'estimateur du maximum de vraisemblance subodorant un changement de comportement reste raisonnable. Notre modèle sous-entend qu'il existe un changement de comportement et induit donc un biais dans l'estimation des durées de vie moyennes, qui, comme le montre le tableau 6.5, reste heureusement faible. De plus, lorsque le changement de comportement est important ($s_2 = 30$), l'estimateur que l'on propose permet une bonne estimation des durées de vie moyennes, alors que l'estimateur issu du modèle simple (sans tenir compte du changement de comportement) est fortement biaisé (19600 au lieu de 3500).

6.2.4 Conclusions

Nous avons proposé de prendre en compte un changement de comportement de maintenance pour des matériels possédant une période de garantie par un modèle de données incomplètes. Pour ce modèle, comme pour le précédent, nous proposons de le résoudre directement par une maximisation de la vraisemblance observée. Cet estimateur permet de réduire considérablement le biais dans l'estimation de la moyenne des durées de vie du composant. Les simulations avec la loi exponentielle nous ont montré que notre procédure est facilement implémentable sans avoir recours à l'algorithme EM bien qu'il s'agisse d'un modèle à structure de données incomplètes.

La perspective immédiate de ce travail est d'appliquer ce modèle sur des données réelles, par exemple pour l'industrie automobile. Une inférence bayésienne est là encore facilement envisageable. Une extension de ce modèle avec une loi de Weibull (généralisée [100], ou un modèle où le vieillissement n'apparaît qu'à une certaine date, voir [18]) est également envisagée. De plus, il serait intéressant de modéliser la période de garantie par une loi de Weibull où le paramètre de forme est inférieur à 1 (correspondant à des défauts de jeunesse) et de considérer pour la période de non garantie une loi de Weibull, où le paramètre de forme est supérieur à 1 (correspondant à un vieillissement). Enfin toujours pour la loi de Weibull, l'hypothèse de réparation minimale pourrait être remplacée par une hypothèse moins forte, entre "aussi bon que neuf" et "aussi mauvais que vieux" (voir la notion d'âge virtuel dans [42]).

Conclusion générale

Bilan et perspectives de la première partie

Dans la première partie de cette thèse, nous avons proposé une démarche complète pour modéliser le processus de dégradation d'un système mécanique d'une installation nucléaire. Les objectifs de ce travail étaient ambitieux, eu égard à la complexité du phénomène étudié, et du faible nombre de données que l'on avait à disposition. Pour modéliser le processus de dégradation, nous avons choisi d'utiliser les réseaux bayésiens. Ce choix est principalement dû à leur capacité à représenter des causalités entre des variables. Ceci est idéal pour reproduire les mécanismes mis en jeux dans les phénomènes mécaniques, physiques et chimiques. La possibilité et les divers champs d'applications des réseaux bayésiens nous ont également séduit. De plus, ces modèles constituent un domaine de recherche très actif et encore peu utilisés en fiabilité et en sûreté de fonctionnement.

Nous nous sommes principalement attachés à la construction des réseaux bayésiens et à leurs analyses dans ce contexte de sûreté de fonctionnement. La complexité algorithmique que peut engendrer une inférence statistique dans ces modèles n'a pas été l'objet de recherche dans ce travail. Nous nous sommes attachés essentiellement à contribuer à la résolution des problèmes apparaissant lors de la construction du réseau, puis à proposer des outils simples d'exploitation d'un tel réseau. En premier lieu, nous avons voulu réduire la complexité de la construction des réseaux bayésiens à partir d'avis d'experts. Tout d'abord, la construction de la structure du graphe a été effectuée lors de nombreuses réunions entre les experts, en s'appuyant sur les diagrammes de défaillance. Pour simplifier la construction du réseau, nous avons regroupé les variables de même type ou de même nature en familles homogènes. Ces familles sont ensuite reliées entre elles par des relations de causalité. Ceci permet de réduire considérablement le nombre d'arcs possibles dans le réseau. Une fois la structure du graphe fixé, il est nécessaire d'évaluer les probabilités correspondantes à ce graphe. Le nombre de probabilités à renseigner est, même dans le cas d'un graphe simple comme le nôtre, très vite gigantesque et rend la mission des experts impossible. Ainsi, nous avons opté pour des modèles log-linéaires afin de représenter le réseau bayésien. Puis, afin d'alléger le nombre de probabilités à donner par expertise, des hypothèses d'indépendances conditionnelles ont été faites en contraignant des termes d'interaction à être nuls. Nous avons proposé une méthode simple permettant d'exploiter pleinement les avis d'experts. Ainsi, les interactions sont identifiées par les experts qui doivent ensuite évaluer les probabilités correspondantes. Il est évident que plus l'ordre des interactions est élevé et plus les probabilités conditionnelles à renseigner sont complexes, notamment avec de multiples combinaisons possibles entre variables. Pour évaluer les probabilités, nous avons questionné les experts indépendamment et nous leurs avons demandé les probabilités les plus simples à donner, à savoir toutes les probabilités marginales et les probabilités conditionnelles sachant une ou deux variables (et pas plus). Cette méthode diffère de la méthode classique qui consiste à ne demander que les probabilités marginales des variables d'entrée et de demander toutes les probabilités conditionnelles sachant toutes les variables parents du graphe. Cette stratégie nous a permis grâce à des probabilités redondantes de vérifier la cohérence dans les avis d'experts. En cas d'incohérence, nous avons établi des règles permettant de choisir quelles probabilités éliminer, ce qui a pour effet de produire des probabilités plus fiables et plus stables.

Pour l'analyse des réseaux bayésiens, nous avons exposé les nombreuses méthodes d'analyse basées principalement sur l'analyse de sensibilité. Nous avons proposé d'identifier les variables importantes en classant les scénarios les plus critiques, où la probabilité de dégradation ou de défaillance du système est importante. D'autre part, nous avons intégré les actions de maintenance comme variables du réseau bayésien. Ainsi, nous avons pu mesurer grâce à une inférence statistique, l'influence des actions de maintenance sur les probabilités de dégradation ou de défaillance du système.

Enfin, nous avons proposé d'intégrer grâce à une inférence bayésienne utilisant la loi de Dirichlet les données de retour d'expérience dans le réseau bayésien. Dans ce cadre, nous proposons de quantifier simplement la confiance que l'on attribue dans les avis d'experts par un paramètre de la loi a priori qui peut être traduit en termes de données de retour d'expérience qu'un expert aurait intégrées dans sa connaissance. Cette mise en correspondance est très adéquate pour une prise en compte raisonnable des opinions d'expert.

Les perspectives de ce travail sont tout d'abord de proposer d'autres méthodes d'analyse d'un réseau bayésien. En effet, peu de travaux, mis à part ceux de Jensen [59], traitent de ces problèmes et nous avons commencé dans ce travail de thèse à proposer des paramètres susceptibles d'analyser le réseau bayésien. En quelque sorte, nous pensons qu'il est souhaitable de définir des outils spécifiques pour la fouille de données (data mining) provenant de réseaux bayésiens. Dans cette thèse, nous avons exploré quelques pistes.
De plus, il est souhaitable d'enrichir le réseau bayésien en y incluant de nouvelles variables. Ces variables peuvent être de différentes natures, comme la motivation du personnel, la conjecture économique, les experts. Il est également primordial que les coûts des actions de maintenance, de réparations, d'indisponibilité soient pris en compte dans l'optimisation de la politique de maintenance. Ces coûts pourront par exemple pondérer les arêtes du graphe. Concernant la confiance dans les dires des experts, il est possible de ne plus demander des probabilités ponctuelles, mais des intervalles de confiance. La variabilité donnée par les experts pourrait ensuite être propagée dans le réseau bayésien.

Enfin, d'un point de vue industriel, il est envisageable d'entreprendre ce même travail pour des systèmes qui ne sont pas encore en fonctionnement. La construction du réseau se ferait donc par les avis d'experts suite aux périodes de conception et de déverminage. Pour finir une perspective naturelle de ce travail est d'effectuer une démarche analogue pour d'autres systèmes, et pas nécessairement dans l'industrie nucléaire.

Bilan et perspectives de la seconde partie

Dans la deuxième partie de cette thèse, nous avons proposé une étude complète de fiabilité pour des données doublement censurées. Pour notre étude, le terme doublement censuré signifie que les données sont censurées à droite et à gauche, mais que l'on n'observe jamais de défaillance. Dans un tel contexte, nous établissons un théorème central limite pour l'estimateur du maximum de vraisemblance pour la loi de Weibull. Nous proposons également une inférence bayésienne qui semble stabiliser les estimateurs dans le cas de fortes censures ou pour des échantillons de petites tailles. Puis, nous avons proposé, pour ce type de données suivant une loi exponentielle, de modéliser différents facteurs humains susceptibles d'entacher l'estimation de la loi de durée de vie par un modèle de données incomplètes. Contrairement à ce que l'on attendait, l'algorithme EM n'apporte pas, dans ce cas précis, plus que l'estimation directe du maximum de vraisemblance. Une approche bayésienne par l'algorithme de Gibbs a également pu être effectuée, et pourra se révéler intéressante en cas d'avis d'experts et de données peu informatives.

Les perspectives de ce travail sont l'étude de tests paramétriques de vieillissement par la loi de Weibull pour des données doublement censurées. En effet, pour les industriels, le diagnostic d'un vieillissement d'un composant est une question cruciale. De plus, une généralisation de ces modèles pour des systèmes, et non plus des composants, est envisageable.

Concernant la modélisation d'un facteur humain, il est dans un premier temps primordial d'appliquer le modèle de garantie sur des données réelles. De plus, il est nécessaire de prendre en compte le possible vieillissement par une loi de Weibull par exemple. Il serait également intéressant de supposer, pour le modèle de garantie, que la réparation d'un composant n'est ni parfaite, ni minimale, mais entre ces deux extrêmes.

Annexes

Annexe A : L'algorithme EM

Algorithme 1 *L'algorithme EM (Expectation Maximisation) vise à rechercher le maximum de vraisemblance avec des données incomplètes. On considère un modèle statistique paramétrique, dont le paramètre est noté θ. On note x l'échantillon complet, défini sur un espace C, qui se décompose en données effectivement observées y, définies sur un espace O, et en données manquantes z, définies sur M. Ainsi, on a $x = (y, z)$ et $C = (O \times M)$. X suit de loi à densité $f(x|\theta)$, Y suit une loi à densité $g(y|\theta)$ avec*

$$g(y|\theta) = \int f(y, z|\theta)\, dz$$

La densité des valeurs manquantes s'écrit alors

$$h(z|y, \theta) = f(x|\theta)/g(y|\theta)$$

Ainsi, les log-vraisemblances sont :

$$\ell(\theta, x) = \ell(\theta, y) + \log h(z|y, \theta) \tag{6.24}$$

où $\ell(\theta, x)$ et $\ell(\theta, y)$ représente respectivement la log-vraisemblance de l'échantillon complet et des données observées. Or, on veut maximiser la vraisemblance observée. L'algorithme EM consiste alors à rechercher la valeur de θ, qui maximise l'information manquante en considérant l'espérance conditionnelle de la vraisemblance de l'échantillon complet $\ell(\theta, x)$ sachant les données observées.

Soit θ^r, un estimateur du paramètre. L'espérance conditionnelle de l'équation (6.24), pour la loi conditionnelle $h(z|y, \theta))$, c'est-à-dire sachant les données observées, est

$$\ell(\theta|y) = Q(\theta|\theta^r) - H(\theta|\theta^r)$$

avec

$$Q(\theta|\theta^r) = \int h(z|y, \theta^r)\ell(\theta, y, z)\, dz$$

et

$$H(\theta|\theta^r) = \int h(z|y, \theta^r) \log h(z|y, \theta)\, dz$$

D'après la théorie de l'information, on a

$$H(\theta|\theta^r) \leq H(\theta^r|\theta^r)$$

Ainsi, si $Q(\theta|\theta^r) \geq Q(\theta^r|\theta^r)$*, on aura* $\ell(\theta, y) \geq \ell(\theta^r, y)$*.*

L'algorithme EM se décompose alors en deux étapes :

* **Etape E :** *Calcul de l'espérance conditionnelle* $Q(\theta|\theta^r)$*.*
* **Etape M :** *Calcul de* θ^{r+1} *qui maximise* $Q(\theta|\theta^r)$*.*

Ainsi, toute suite $(\theta^r)_r$ *engendrée par EM vérifie* $\ell(\theta^{r+1}, y) \geq \ell(\theta^r, y)$*.*

Annexe B : Théorème générique d'Andrews

Proposition 4 *On suppose que*

(i) Θ *est un ensemble relativement compact (de fermeture compact) dans un espace métrique,*

(ii) $\frac{1}{n}\sum_{i=1}^{n}\left(\psi_i\left(X_i,\theta\right) - \mathbb{E}[\psi_i\left(X_i,\theta\right)]\right) \xrightarrow{p.s.} 0,\ \forall\theta \in \Theta,$

(iii) $\forall\theta, \theta' \in \Theta,\ \left|\psi_i\left(X_i,\theta\right) - \psi_i\left(X_i,\theta'\right)\right| \leq B_i(X_i)h\left(d(\theta,\theta')\right)$*, où* h *est une fonction réelle satisfaisant* $lim_{x\to 0}h(x) = 0$*, où* d *est la distance dans* Θ *et où* $(B_n)_{n\geq 1}$ *est une suite de fonctions réelles mesurables,*

(iv) $\sup_{n\geq 1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[B_i(X_i)\right] < +\infty,$

(v) $\frac{1}{n}\sum_{i=1}^{n}\left(B_i(X_i) - \mathbb{E}\left[B_i(X_i)\right]\right) \xrightarrow{p.s.} 0,$

Alors

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\psi_i(X_i,\theta) - \mathbb{E}\left[\psi_i(X_i,\theta)\right]\right)\right| \xrightarrow{p.s.} 0 \ \text{ quand } n \to +\infty$$

et $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\psi_i(X_i,\theta)\right]$ *est continue en* θ

Preuve : cf. Andrews [4].

Annexe C : Simulation de Monte-Carlo

Nous présentons la simulation de Monte-Carlo, usuel en estimation bayésienne. Dans un cadre bayésien, il nous faut déterminer la loi a posteriori. Cette loi étant implicite avec nos types de données, une approximation de la moyenne peut etre faite à l'aide de simulation de Monte-Carlo, dont nous présentons la méthode ci-dessous.

Pour déterminer l'estimateur de Bayes, dans le cas où la fonction de perte est quadratique, il faut calculer la moyenne de la loi marginale a posteriori. Si la dimension de Θ est égale à 1, on trouve :

$$\hat{\theta} = \mathbb{E}[\theta/t_1, \ldots, t_n] = \int_\Theta \theta g(\theta/t_1, \ldots, t_n)\, d\theta$$

Il en va de même si la dimension est strictement supérieure à 1. Par exemple pour la loi de Weibull à deux paramètres, on a :

$$\begin{aligned} \hat{\beta}_{Bayes} &= \mathbb{E}_\beta\left[\beta/t_1, \ldots, t_n\right] = \int_a^b \beta g_\beta(\beta/t_1, \ldots, t_n)\, d\beta \\ \hat{\eta}_{Bayes} &= \mathbb{E}_\eta\left[\eta/t_1, \ldots, t_n\right] = \int_0^{+\infty} \eta g_\eta(\eta/t_1, \ldots, t_n)\, d\eta \end{aligned}$$

où

$$\begin{aligned} g_\beta(\beta/t_1, \ldots, t_n) &= \int_0^{+\infty} g(\eta, \beta/t_1, \ldots, t_n)\, d\eta \\ g_\eta(\eta/t_1, \ldots, t_n) &= \int_a^b g(\eta, \beta/t_1, \ldots, t_n)\, d\beta \end{aligned}$$

et donc,

$$\begin{aligned} \hat{\beta}_{Bayes} &= \int_a^b \int_0^{+\infty} \beta g(\eta, \beta/t_1, \ldots, t_n)\, d\beta\, d\eta \\ \hat{\eta}_{Bayes} &= \int_0^{+\infty} \int_a^b \eta g(\eta, \beta/t_1, \ldots, t_n)\, d\eta\, d\beta \end{aligned}$$

Pour calculer les intégrales de la forme :

$$\int_\Theta g(\theta) f(\theta)\, d\theta,$$

on va simuler des données $X_1, \ldots, X_n$ suite i.i.d. de loi à densité $f(x)$ par rapport à la mesure de Lebesgues.

Ainsi en calculant $g(X_i)$ et grâce à la loi des grands nombres, il vient :

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^n g(X_i) = \mathbb{E}[g(X)] = \int_\Theta g(x) f(x)\, dx.$$

Annexe D : Matrices d'information de Fisher

Les matrices d'information de Fisher correspondent au cas doublement censuré pour la loi de Weibull. Elles servent notamment à établir la normalité asymptotique de l'estimateur du maximum de vraisemblance. Nous calculons les matrices d'information de Fisher I^k, J^k et i^k. La fonction score a l'expression suivante.

$$\left\{\begin{array}{l} \dot{\ell}_k^\eta = \frac{\partial \ell_k}{\partial \eta} = -\sum_{i=1}^k \left(\frac{t_i}{\eta}\right)^\beta \frac{\beta}{\eta}\gamma_i \\ \\ \dot{\ell}_k^\beta = \frac{\partial \ell_k}{\partial \beta} = \sum_{i=1}^k \left(\log \frac{t_i}{\eta}\right)\left(\frac{t_i}{\eta}\right)^\beta \gamma_i \end{array}\right.$$

$$\begin{aligned} I^k(\theta) &= \sum_{i=1}^k \psi_i \psi_i^\mathsf{T} \\ &= \sum_{i=1}^k \left(\mathbb{I}_{\{Y_i \le t_i\}} \frac{e^{-2\left(\frac{t_i}{\eta}\right)^\beta}}{\left(1-e^{-\left(\frac{t_i}{\eta}\right)^\beta}\right)^2} + \mathbb{I}_{\{Y_i > t_i\}} \right) \left(\frac{t_i}{\eta}\right)^{2\beta} \begin{bmatrix} \beta^2/\eta^2 & -\beta/\eta \log t_i/\eta \\ -\beta/\eta \log t_i/\eta & (\log(t_i/\eta))^2 \end{bmatrix}, \end{aligned}$$

$$J^k(\theta) = \begin{bmatrix} \frac{\partial^2 \ell_k}{\partial \eta^2} & \frac{\partial^2 \ell_k}{\partial \beta \partial \eta} \\ \\ \frac{\partial^2 \ell_k}{\partial \eta \partial \beta} & \frac{\partial^2 \ell_k}{\partial \beta^2} \end{bmatrix},$$

avec

$$\frac{\partial^2 \ell_k}{\partial \eta^2} = -\left(\sum_{i=1}^k \left(\frac{t_i}{\eta}\right)^\beta \frac{\beta(\beta+1)}{\eta^2}\gamma_i + \left(\frac{t_i}{\eta}\right)^{2\beta} \left(\frac{\beta}{\eta}\right)^2 \mathbb{I}_{\{Y_i \le t_i\}} \frac{e^{-(\frac{t_i}{\eta})^\beta}}{\left(1-e^{-(\frac{t_i}{\eta})^\beta}\right)^2} \right)$$

$$\frac{\partial^2 \ell_k}{\partial \beta^2} = \sum_{i=1}^k \left(\log \frac{t_i}{\eta}\right)^2 \left(\frac{t_i}{\eta}\right)^\beta \gamma_i - \left(\frac{t_i}{\eta}\right)^{2\beta} \left(\log \frac{t_i}{\eta}\right)^2 \mathbb{I}_{\{Y_i \le t_i\}} \frac{e^{-(\frac{t_i}{\eta})^\beta}}{\left(1-e^{-(\frac{t_i}{\eta})^\beta}\right)^2}$$

$$\frac{\partial^2 \ell_k}{\partial \eta \partial \beta} = \frac{\partial^2 \ell_k}{\partial \beta \partial \eta} = \sum_{i=1}^k - \left(\frac{t_i}{\eta}\right)^\beta \frac{1}{\eta}\left(\beta \log \frac{t_i}{\eta} + 1\right)\gamma_i + \left(\frac{t_i}{\eta}\right)^{2\beta} \frac{\beta}{\eta} \log \frac{t_i}{\eta} \mathbb{I}_{\{Y_i \le t_i\}} \frac{e^{-(\frac{t_i}{\eta})^\beta}}{\left(1-e^{-(\frac{t_i}{\eta})^\beta}\right)^2}.$$

Or $\mathbb{E}\left[\mathbb{I}_{\{Y_i \le t_i\}}\right] = 1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}$ et avec le lemme 1 du chapitre 5, nous avons $\mathbb{E}[\gamma_i] = 0$ et ainsi

$$\begin{aligned} i^k(\theta) &= \mathbb{E}\left[I^k(\theta)\right] = -\mathbb{E}\left[J^k(\theta)\right] \\ &= \sum_{i=1}^k \left(\frac{t_i}{\eta}\right)^{2\beta} \frac{e^{-\left(\frac{t_i}{\eta}\right)^\beta}}{1-e^{-\left(\frac{t_i}{\eta}\right)^\beta}} \begin{bmatrix} \beta^2/\eta^2 & -\beta/\eta \log t_i/\eta \\ -\beta/\eta \log t_i/\eta & (\log(t_i/\eta))^2 \end{bmatrix}. \end{aligned}$$

Annexe E : Schéma fonctionnel du joint 1 d'une pompe primaire 900 MW

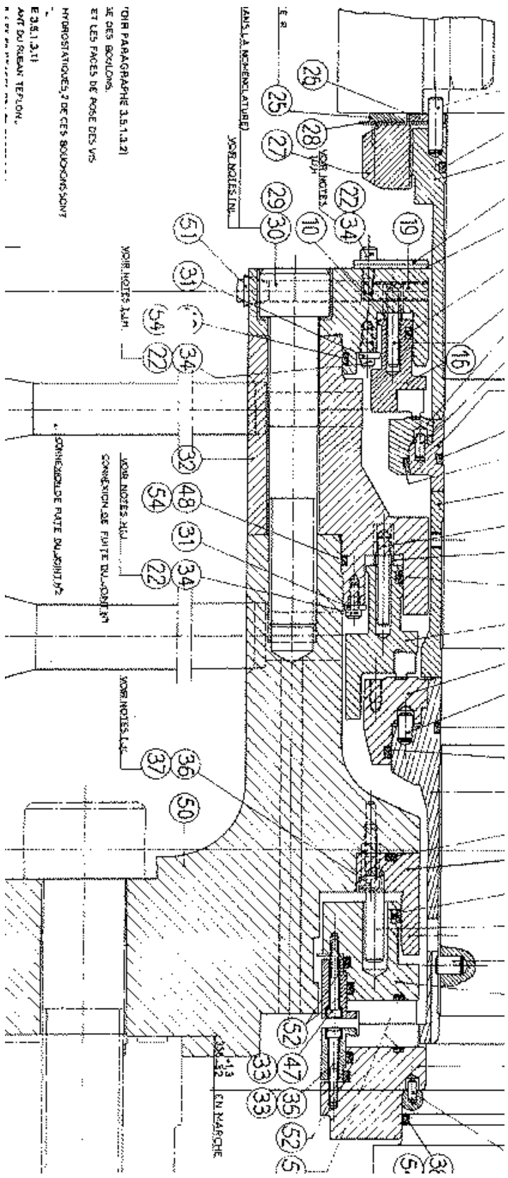


FIG. 6.2 – Schéma fonctionnel d'un joint 1 d'une pompe primaire 900 MW.

Annexe F : Quelques étapes de la construction du réseau bayésien

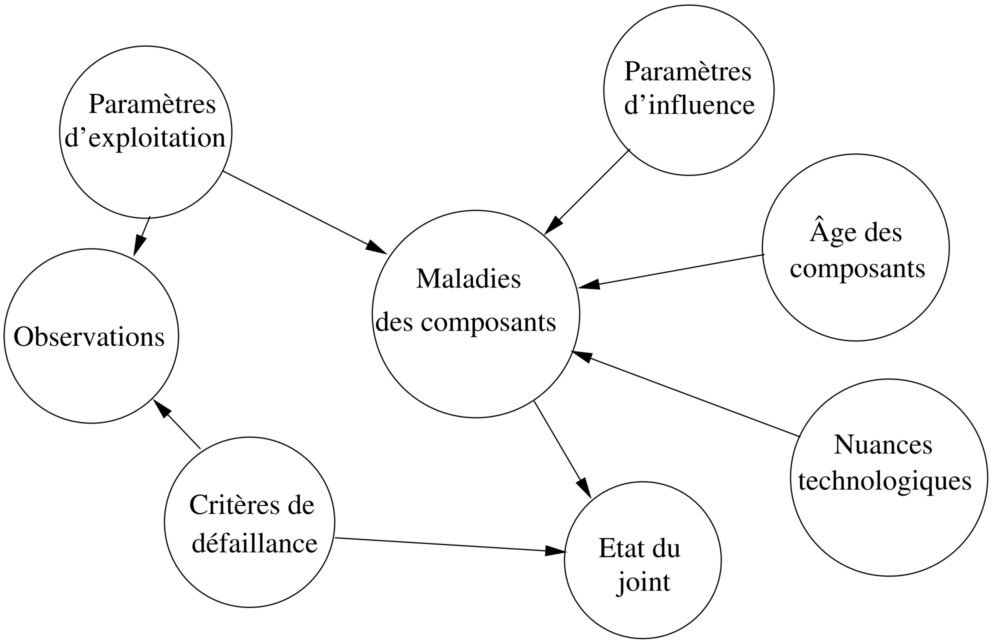


FIG. 6.3 – Première proposition de graphe pour le joint 1 d'une pompe 900 MW.

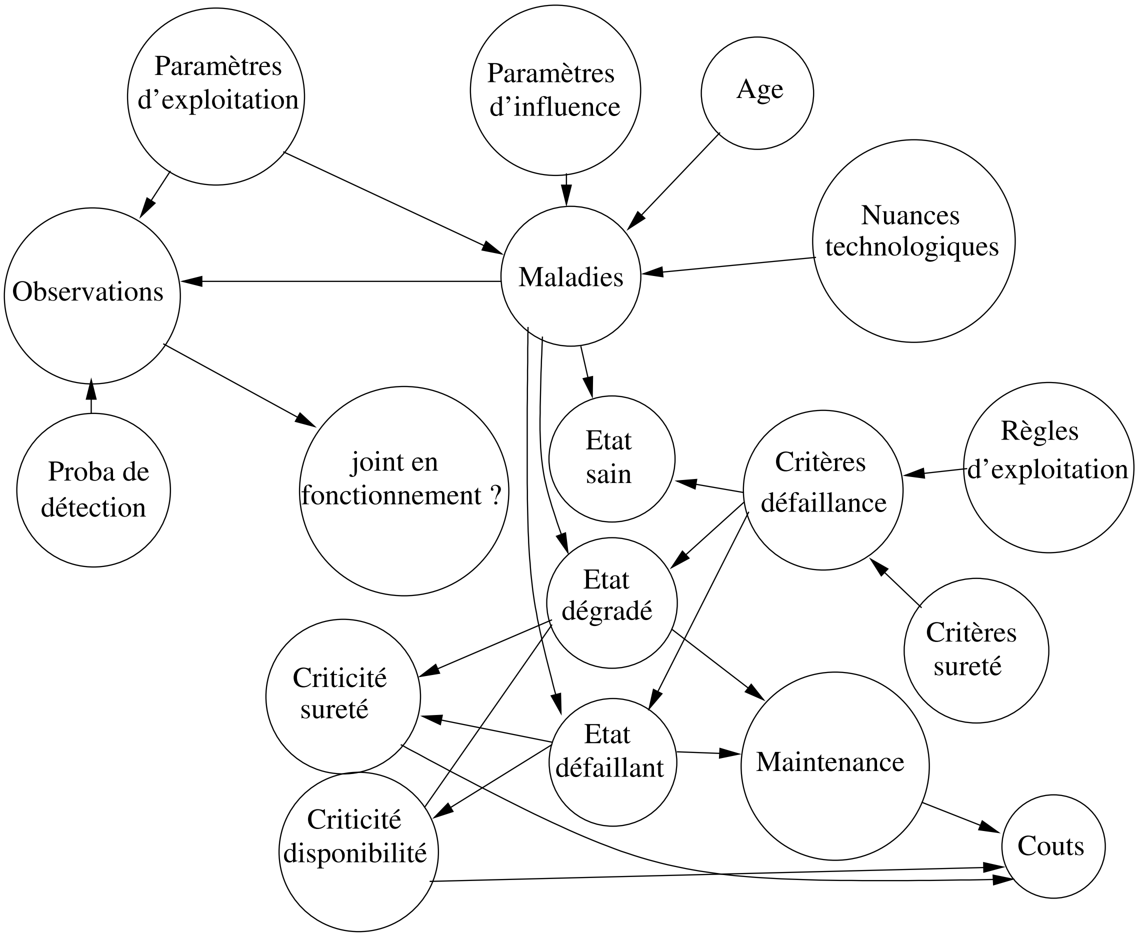


Fig. 6.4 – Seconde proposition de graphe pour le joint 1 d'une pompe 900 MW.

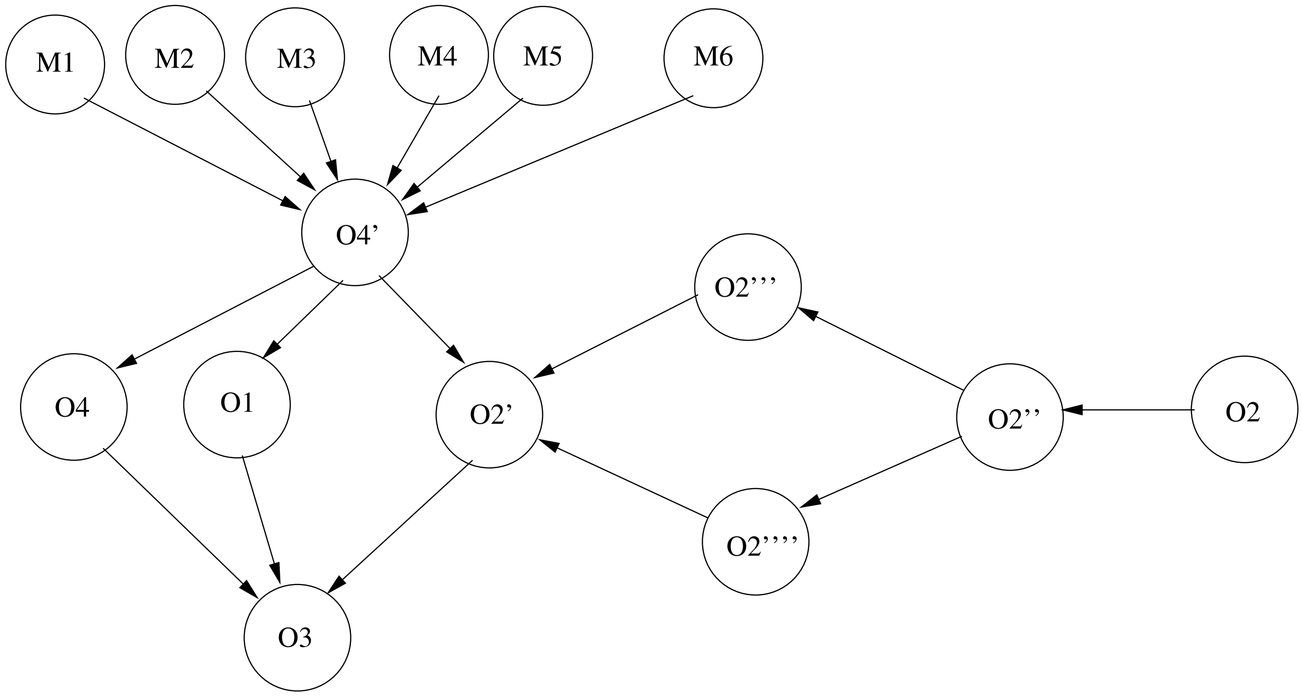


FIG. 6.5 – Naissance des variables observations.

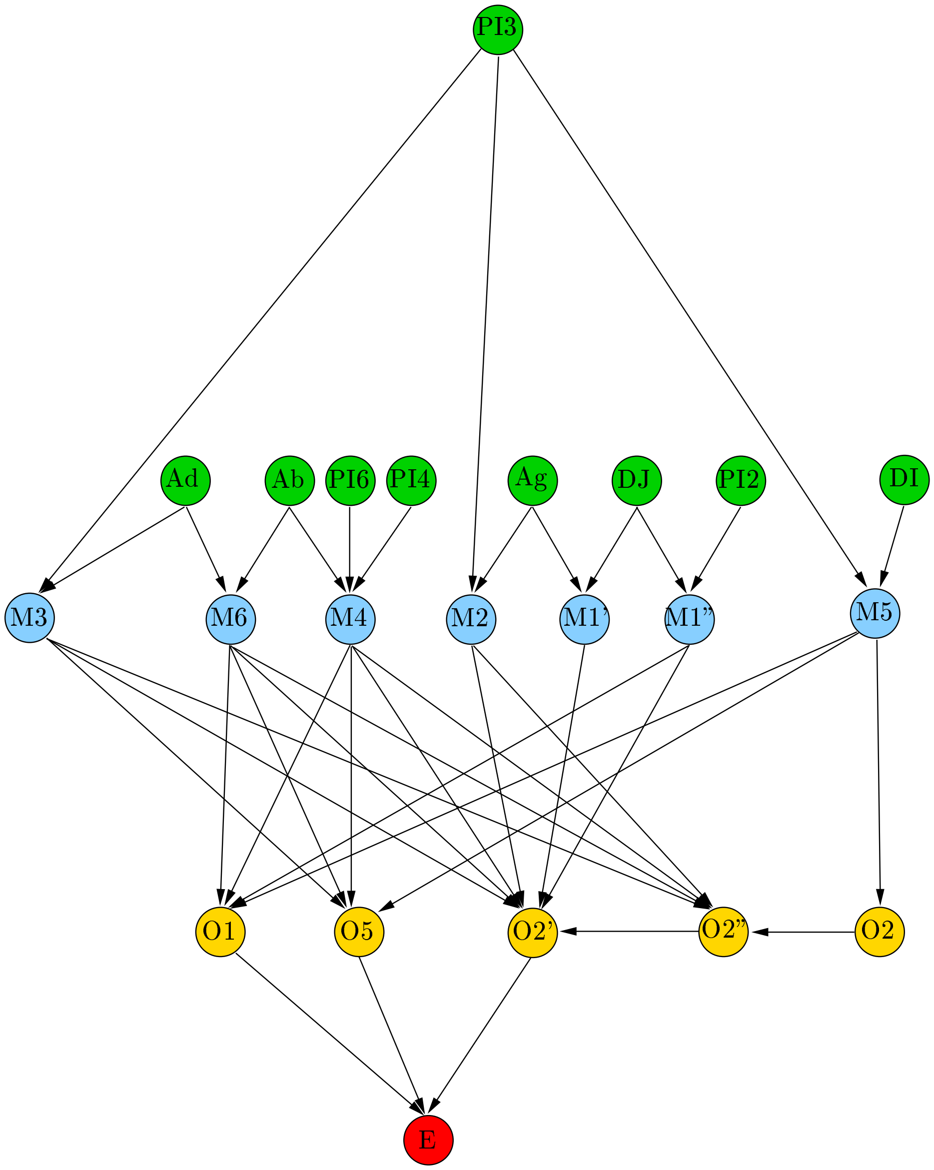


FIG. 6.6 – Réseau bayésien final du joint 1 d'une pompe primaire.

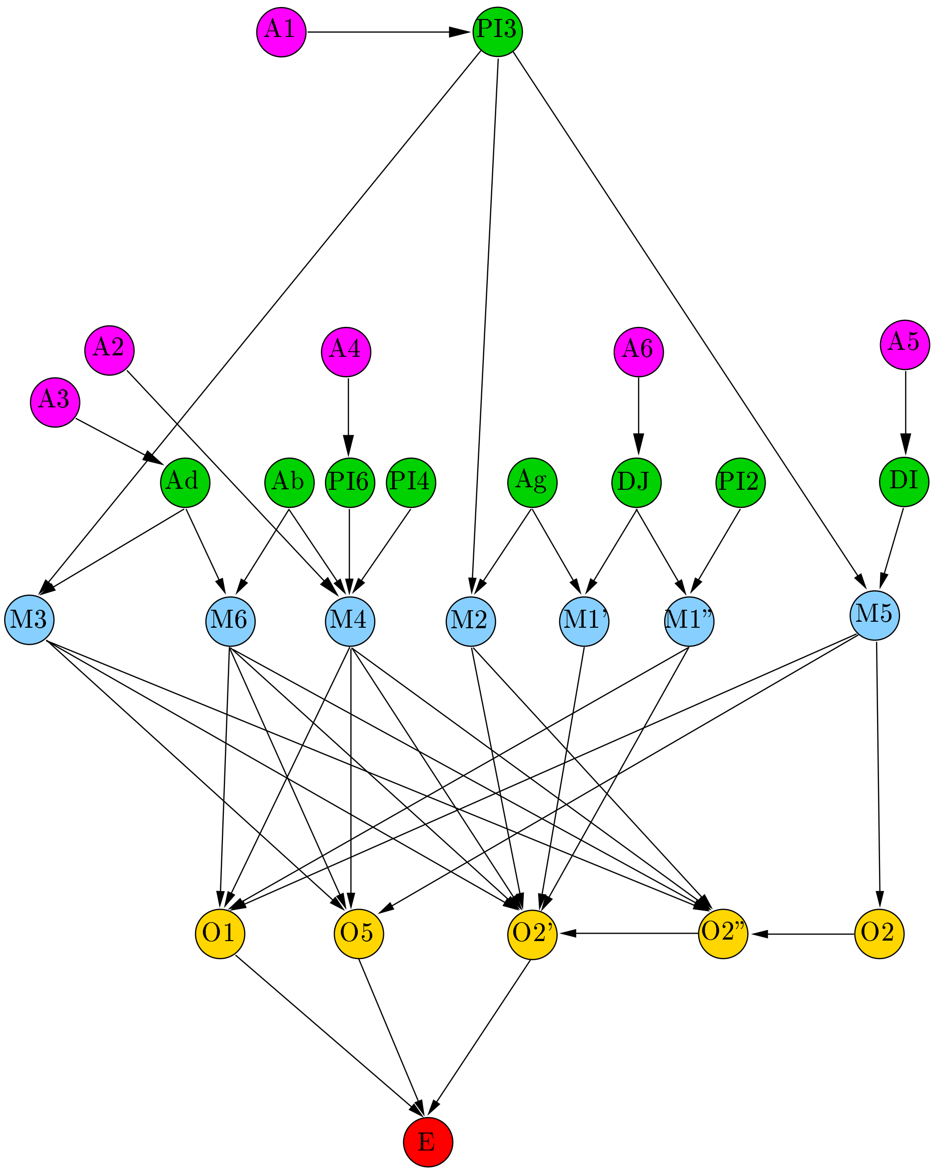


FIG. 6.7 – Réseau bayésien pour le joint 1 de la pompe primaire 900 MW avec intégration des actions de maintenance.

Bibliographie

[1] G. Agre. Diagnostic bayesian networks. *Computers and Artificial Intelligence*, 16(1) :47–67, 1997.

[2] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Autom. Control*, 19 :716–723, 1974.

[3] K. Andersen and B. Rønn. A nonparametric test for comparing two samples where all observations are either left- or right-censored. *Biometrics*, 51(1) :323–329, March 1995.

[4] D.W.K. Andrews. Generic uniform convergence. *Econometric Theory*, 8, 1982.

[5] H. Ascher and H. Feingold. *Repairable System Reliability.* Lecture Notes in Statistics, 1984.

[6] T. Aven. Determination/estimation of an optimal replacement interval under minimal repair. *Optimization*, 16 :743–754, 1985.

[7] M. Ayer, H.D Brunk, G.M Ewing, W.T Reid, and E. Silverman. An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, pages 641–647, 1955.

[8] M. Bacha. *Inférence statistique pour des modèles de durées de vie et applications.* PhD thesis, Université de Rouen, Mars 1996.

[9] M. Bacha, G. Celeux, E. Idée, A. Lannoy, and D. Vasseur. *Estimation de modèles de durées de vie fortement censurées.* Collection de la Direction des Études er Recherches d'Électricité de France, Eyrolles, 1999.

[10] R. E. Barlow. Geometry of the total time on test transform. *Nav. Res. Logis.*, Q.26 :393–402, 1979.

[11] A.P. Basu and J.K. Ghosh. Asymptotic properties of a solution to the likelihood equation with life-testing applications. *Journal of the American Statistical Association*, 75(370) :410–414, June 1980.

[12] A. Becker and P. Naïm. *Les réseaux bayésiens.* Eyrolles, 1999.

[13] M.P. Berg. The marginal cost analysis and its application to repair and replacement policies. *European Journal of Operational Research*, 82 :214–224, 1995.

[14] B. Bergman and B. Klefsjö. A graphical method applicable to age-replacement problems. *IEEE Transactions on Reliability*, R-31(5) :478–481, December 1982.

[15] J.-M. Bernard. Bayesian interpretation of frequentist procedures for a bernouilli process. *The American Statistician*, 50(1) :7–13, February 1996.

[16] J.-M. Bernard. Non-parametric inference about an unknown mean using the imprecise dirichlet model. In *2nd International Symposium on Imprecise Probabilities and Theirs Applications*. Ithaca, New York, 2001.

[17] J.M. Bernardo. Reference posterior distributions for bayesian inference. *J. R. Statist. Soc. B*, 41(2) :113–147, 1979.

[18] H. Bertholon. *Un modéle de vieillissement*. PhD thesis, Université Joseph Fourier, Grenoble, 2001.

[19] Y.M. Bishop, S.E. Fienberg, and P.W. Holland. *Discrete multivariate analysis*. Cambridge, MA : MIT press, 1975.

[20] M. Bouissou, F. Martin, and A. Ourghanlian. Assessment of a safety-critical system including software : A bayesian belief network for evidence sources. In *RAMS'99 Reliability and Maintainability Symposium*, Washington, 1999.

[21] J. S. Breese, E. Horvitz, M. A. Peot, R. Gay, and G. H. Quentin. Automated decision-analytic diagnosis of thermal performance in gas turbines. In *Proceedings of the ASME International Gas Turbine and Aeroengine Congress and Exposition*, Cologne, Germany, 1992.

[22] C. Breuils. Analyse statistique de durée de vie pour les pompes primaires 900 et 1300 mw. Master's thesis, Université de Paris, 2000.

[23] G. Casella and E. I. George. An introduction to Gibbs sampling. *The American Statistician*, 46 :167–174, 1992.

[24] B. Castanier. *Modélisation stochastique et optimisation de la maintenance conditionnelle des systèmes à dégradation graduelle*. PhD thesis, Université de Technologie de Troyes, 2001.

[25] G. Celeux, F. Corset, F. Billy, R. Chevalier, M-A. Garnero, A. Lannoy, and Ricard B. Description d'un modèle graphique pour la caractérisation des maladies du joint 1 de la pompe primaire 900 mw. Technical report, Rapport de collaboration EDF-INRIA, Février 2001.

[26] G. Celeux, F. Corset, M.-A. Garnero, and C. Breuils. Accounting for inspection errors and change in maintenance behaviour. *IMA Journal of Management Mathematics*, 13(1) :51–59, 2002.

[27] G. Celeux, C. Lavergne, and Y. Vernaz. Assessing material aging from doubly censored data : Weibull distribution vs. poisson process. Technical Report rapport de recherche RR-3857, INRIA, janvier 2000.

[28] G. Celeux and J-P. Nakache, editors. *Analyse discriminante sur variables qualitatives*. Polytechnica, 1994.

[29] C. Chatelain, C. Carron, and A. M. Chemali. Guide de la prospective opérationnelle. Technical Report rapport interne EDF-HE-51/98/023A-21/01/1999, EDF, 1999.

[30] C. Chatelain and A. Lannoy. Une application de la technique du réseau bayésien à l'évolution des coûts de maintenance. Technical Report HP-20/00/020/A, EDF, Juillet 2000.

[31] R. Christensen. *Log-Linear Models*. Springer Texts in Statistics. Springer-Verlag, New York, 1990.

[32] C. Cocozza-Thivent. Un modèle de maintenance préventive. Technical report, Equipe d'Analyse et de Mathématiques Appliquées, Université de Marne-La-Vallée, 1998.

[33] C. Cocozza-Thivent, editor. *Un modèle de maintenance conditionnelle et son exploitation*. $\lambda\mu$12, 2000.

[34] G. F. Cooper. *NESTOR : A Computer-Based Medical Diagnosis Aid that Integrates Causal and Probabilistic Knowledge*. Report hpp-84-48, Medical Computer Science Group, Stanford University, November 1984.

[35] F. Corset. Analyse de fiabilité à partir de données censurées à droite et à gauche. Master's thesis, Université de Rennes I, 1999.

[36] F. Corset. Modelling a guarantee period and a change of maintenance behaviour. In *ESReDA 22nd Seminar on maintenance management and optimisation*, Madrid, Spain, May 27-28 2002.

[37] F. Corset, G. Celeux, A. Lannoy, and B. Ricard. Designing a graphical model for preventive maintenance from expert opinions in a rapid and reliable way. In *ESREL and* $\lambda\mu$13, Lyon, France, 2002.

[38] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, and D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Statistics for Engineering and Information Science. Springer, 1999.

[39] D. R. Cox. Regression models and life tables. *J. R. Stat. Soc. Ser. B*, 34 :187–220, 1972.

[40] J.-Y. Dauxois. Fiabilité. Cours de DEA, Université de Rennes I, 1999.

[41] A.P. Dawid. Conditional independance in statistical theory (with discussion). *Journal of the Royal Statistical Society, Series B*, 41 :pp. 1–31, 1979.

[42] L. Doyen and O. Gaudoin. Models for assessing maintenance efficiency. In *3rd International Conference on Mathematical Methods in Reliability*, Trondheim, June 2002.

[43] J.-B. Durand. *Modèles à structure cachée pour l'étude de processus aléatoire*. PhD thesis, Université Joseph Fourier, Grenoble, décembre 2002.

[44] D. Foata and A. Fuchs. *Calcul des probabilités*. Science Sup, Dunod, 1999.

[45] O. Gaudoin and J.L. Soler. Failure rate behavior of components subjected to random stresses. *Reliability Engineering and System Safety*, 58 :19–30, 1997.

[46] D. Geiger and J. Pearl. Logical and algorithmic properties of independence and their application to in bayesian networks. *Annals of Mathematics and Artificial Intelligence*, 2 :165–178, 1990.

[47] D. Geiger, T. Verma, and J. Pearl. Identifying independence in bayesian networks. *Networks*, 20 :507–534, 1990.

[48] W. Gilks, A. Thomas, and D. Spiegelhalter. A language and a program for complex bayesian modelling. *The Statistician*, 43 :169–178, 1994.

[49] W. R. Gilks, S. Richardson, and D. I. Spiegelhalter. *Markov Chains Monte Carlo in Practice*. Chapman and Hall, 1996.

[50] I.J. Good. *The Estimation of Probabilities : An Essay on Modern Bayesian Methods*. The M.I.T. press, Cambridge, Massachusetts, research monograph no. 30 edition, 1965.

[51] L.A. Goodman. The multivariate analysis of qualitative data : interaction among multiple classifications. *J. Amer. Statist. Assoc.*, 65 :226–256, 1970.

[52] C. Gourieroux and A. Monfort. *Statistique et modèles économétriques*, volume 2. Economica, 1996.

[53] J.B.S. Haldane. The precision of observed values of small frequencies. *Biometrika*, 35 :297–300, 1948.

[54] D. Heckerman, A. Mamdani, and M. P. Wellman. Real-world applications of bayesian networks. *Communications of the ACM*, 38(3) :24–26, 1995.

[55] B. Hoadley. Asymptotic properties of maximum likelihood estimators for the independent not-identically-distributed case. *Annals of Mathematical Statistics*, 42 :1977–1991, 1995.

[56] S. Højsgaard. Learning structures from data and experts. *Mathematics and Computers in Simulation*, 42 :143–152, 1996.

[57] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London, Ser. A*, 186 :453–461, 1946.

[58] H. Jeffreys. *Theory of Probability*. Oxford University Press, 1961.

[59] F.V. Jensen. *Bayesian Networks and Decision Graphs*. Statistics for Engineering and Information Science. Springer, 2001.

[60] M. I. Jordan, editor. *Learning in Graphical Models*. The MIT press, 1999.

[61] C.W. Kang and M.W. Golay. A bayesian belief networks-based advisory system for operational availability focused diagnosis of complex nuclear power systems. *Expert systems with applications*, 1(17) :21–32, 1999.

[62] R.E. Kass and A.E. Raftery. Bayes factors. *Journal of the American Statistical association*, 90(440) :773–793, June 1995.

[63] H. Koshimae, T. Dohi, N. Kaio, and S. Osaki. Graphical/statistical approach to repair limit replacement problem. *Journal of the Operations Research Society of Japan*, 39(2) :230–246, June 1996.

[64] A. Lannoy and H. Procaccia. *L'utilisation du jugement d'expert en sûreté de fonctionnement*. Editions TEC et DOC, 2001.

[65] K.B. Laskey and S.M Mahoney. Knowledge and data fusion in probabilistic networks. *submitted to Machine Learning special issue on knowledge and data fusion*, A paraître.

[66] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *J. R. Statist. Soc. B*, 50(2) :157–224, 1988.

[67] S.L. Lauritzen. *Graphical Models.* Oxford Science Publications, 1996.

[68] J. F. Lawless. *Statistical models and methods for lifetime data.* Wiley Series in Probability and Mathematical Statistics, 1982.

[69] L. M. Leemis and L. H. Shih. Exponential parameter estimation for data sets containing left and right censored observations. *Commun. stat., Simulation Comput.*, 18(3) :1077–1085, 1989.

[70] G. Letac and H. Massam. Introduction aux modèles graphiques. Cours de DEA, 2002.

[71] H. A. Linstone and M. Turoff. *The Delphi method. Techniques and applications. With a foreword by Olaf Helmer.* Addison-Wesley Publishing Company.Advanced Book Program, 1975.

[72] H. F. Martz and R. A. Waller. *Bayesian Reliability Analysis.* Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley, 1982.

[73] G. McLachlan and T. Krishnam. *The EM algorithm and extensions.* Wiley, 1997.

[74] W.Q. Meeker and L.A. Escobar. *Statistical Methods for Reliability Data.* Wiley, 1998.

[75] S. Monti and G. Carenini. Dealing with the expert inconsistency in probability elicitation. *IEEE Transactions on Knowledge and Data Engineering*, 12(4) :499–508, 2000.

[76] W. Nelson. *Statistical Applied Life Data Analysis.* Wiley, 1982.

[77] M. Nerlove and S.J. Press. Multivariate log-linear probability models for the analysis of qualitative data. *Bull. Int. Stat. Inst.*, 48(2) :3–20, 1979.

[78] J. Pearl. Reverend bayes on inference engines : a distributed hierachical approach. In *Proceedings of American Association for Artificial Intellingence National Conference on AI, Pittsburg, Pennsylvania*, pages pp. 133–136, 1982.

[79] J. Pearl. *Probabilistic Inference in Intelligent Systems.* Morgan Kaufmann, San Matteo, California, 1988.

[80] C. Pellegrin. A graphical procedure for an on-condition maintenance policy : imperfect-inspection model and interpretation. *IMA Journal of Mathematics Applied in Business and industry*, 3 :177–191, 1992.

[81] W. Perks. Some observations on inverse probability including a new indifference rule. *Journal of the Institute Actuaries*, 73 :285–334, 1947.

[82] M. Pradhan, M. Henrion, G. Provan, B. D. Favero, and K. Huang. The sensitivity of belief networks to imprecise probabilities : an experimental investigation. *Artificial Intelligence*, 85 :363–397, 1996.

[83] M. Ramoni and P. Sebastiani. Robust learning with missing data. *Machine Learning*, 2001.

[84] C. P. Robert. *The Bayesian Choice.* Springer Texts in Statistics, Springer-Verlag, 1994.

[85] A. Rodionov and G. Celeux. A schock model for assessing component aging reliability. In *22nd European Safety, Reliability and Data Association (ESReDA) Seminar*, Madrid, Espagne, mai 2002.

[86] G. Saporta. *Probabilités, Analyse des données et Statistique.* PARIS TECHNIP, 1996.

[87] D. Sun. A note on non informative priors for weibull distributions. *Journal of Statistical Planning and Inference*, 61 :319–338, 1997.

[88] R. E. Tarjan and M. Yannakakis. Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM J. Comput.*, 13 :566–579, 1984.

[89] B.W. Turnbull. Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, 69 :169–173, 1974.

[90] S. K. Upadhyay, V. Shastri, and U. Singh. Bayes estimators of exponential parameters when the observations are left and right censored. *Metron*, 50(3-4) :185–199, 1992.

[91] L.C. Van der Gaag and V.M.H. Coupé. Sensitivity analysis for threshold decision making with bayesian belief networks. *Lecture Notes in Computer Science*, 1792 :37–48, 2000.

[92] C. Villegas. On the representation of ignorance. *Journal of the American Statistical*, 1977.

[93] A. Wald. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist*, 20 :595–601, 1949.

[94] P. Walley. Inferences from multinomial data : Learning about a bag of marbles. *J.R.Statist. Soc. B*, 58(1) :3–57, 1996.

[95] H.R. Warner, A.F. Toronto, L.G. Veasey, and R. Stephenson. A mathematical approach to medical diagnosis -applications to congenital heart disease. *Journal of the American Medical Association*, 177 :177–184, 1961.

[96] P. Weber, M.C. Suhner, and B. Lung. System approach-based bayesian networks to aid maintenance of manufacturing process. In *6th IFAC Symposium on Cost Oriented Automation, Low Cost Automation, Berlin, Germany*, October, 8-9, 2001.

[97] U. Westberg and B. Klefsjö. Applications of the piecewise exponential estimator for the maintenance policy block replacement with minimal repair. *IAPQR Trans*, 20 :197–210, 1995.

[98] J. Whittaker. *Graphical Models in Applied Multivariate Statistics.* Wiley, 1990.

[99] S. Wright. Correlation and causation. *Journal of Agricultural Research*, 20 :557–585, 1921.

[100] M. Xie, T.N. Goh, and Y. Tang. A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering and Systems Safety*, 76(2) :279–285, 2002.

[101] B. Yildiz and M. W. Golay. Development of expert system with bayesian networks for application in nuclear power plants. In *EPRI International Maintenance Conference. Maintaining Reliable Electric Generation*, August 2001.

[102] `http ://www.cs.berkeley.edu/~murphyk/Bayes/bnt.html`.

[103] `http ://bnt.insa-rouen.fr/`.

[104] `http ://www.norsys.com/download.html`.

[105] `http ://www.hugin.com/Products_Services/`.

[106] `http ://www.cs.cmu.edu/ javabayes/HOME`.

Résumé :
La première partie traite de l'application des réseaux bayésiens en maintenance et propose une méthodologie de construction à partir d'avis d'experts. Pour évaluer les probabilités, le réseau est décrit par un modèle log-linéaire saturé. Des indices permettant l'analyse du réseau et l'identification des variables critiques sont donnés. Les actions de maintenance sont intégrées comme nouveaux nœuds du graphe. Une intégration du retour d'expérience est proposée par une inférence bayésienne, en quantifiant la confiance attribuée aux avis d'experts. La deuxième partie traite de la fiabilité dans un contexte doublement censuré. Nous étudions les propriétés asymptotiques des estimateurs du maximum de vraisemblance pour la loi de Weibull. Une inférence bayésienne est proposée avec des lois a priori informatives et non informatives. De plus, nous modélisons par des variables cachées un facteur humain, représentant des manques d'information lors d'opérations de maintenance et résolvons ce problème par maximum de vraisemblance et par une inférence bayésienne.

Mots-clés :
Réseaux bayésiens, Optimisation de la maintenance, Modèle log-linéaire, Intégration du retour d'expérience, Fiabilité, Doublement censuré, Algorithme EM, Échantillonage de Gibbs.

Help to the Maintenance Optimization from Bayesian Networks and Reliability with Doubly Censored Data

Abstract :
The first part deals with application of Bayesian networks in maintenance and proposes a methodology to build a network from expert judgment. In order to evaluate probabilities, networks is seen as a saturated log-linear model. Some tools are defined in order to identify the most critical variables. Some maintenance tasks are integrated as new nodes of the Bayesian Network. An integration of experience feedback data is proposed by a Bayesian inference, by quantifying confidence given to experts judgment. The second part deals with reliability with doubly censored data. We study the asymptotic properties of the maximum likelihood estimators for the law of Weibull. A Bayesian inference is also proposed with informative and non informative priors. Moreover, human factors are described by hidden variables, which represent lack of information during maintenance operations. We solve this problem by maximum likelihood and by a Bayesian inference.

Keywords :
Bayesian Networks, Maintenance Optimization, Log-Linear Model, Experience Feedback Integration, Reliability, Doubly Censored, EM Algorithm, Gibbs Sampling.