Introduction du Retour d'Expérience dans les Réseaux Bayésiens

Franck Corset, Gilles Celeux, André Lannoy

Projet IS2, INRIA Rhône-Alpes / EDF R&D, département MRI
655 av. de l'Europe, Montbonnot/ 6, quai Watier
38334 Saint Ismier cedex, France / 78401 Chatou Cedex, France.
{Franck.Corset,Gilles.Celeux}@inrialpes.fr, Andre.Lannoy@edf.fr
tel : (33)4.76.61.53.25 fax : (33)4.76.61.52.52

Abstract

This paper presents experience feedback integration in Bayesian Networks. Bayesian Networks structure and probabilities are designed with experts judgement only. In this paper, we assume that networks structure are fixed and known. With growing experience, data are registered and must be used to improve the model. In order to model the ignorance of an expert, we choose Jeffreys law. Therefore, we choose a procedure to parameter the confidence in experts judgement in a rapid way. This parameter represents an imaginary sample size. We give a simple example of experience feedback integration and an application on system of a Reactor Coolant Plant.

Ce papier présente l'intégration du retour d'expérience dans les réseaux bayésiens. La structure et les probabilités du réseau bayésien sont évaluées uniquement à partir d'avis d'experts. Dans ce papier, nous supposons que la structure du graphe est fixe et connue. Avec l'expérience grandissante, les données sont répertoriées et doivent contribuer à enrichir le modèle. Dans le but de modéliser l'ignorance d'un expert, nous choisissons la loi de Jeffreys. Nous choisissons pour cela une procédure simple et rapide, permettant de quantifier la confiance que l'on a dans les avis d'experts. Ce paramètre représente une taille d'échantillon fictif. Nous donnons un exemple simple et une application sur un système d'une centrale nucléaire.

Keywords : Bayesian Networks, Bayesian Inference, Expert Judgement, Jeffreys Law, Experience Feedback Integration.
Mots Clés : Réseaux bayésiens, inférence bayésienne, avis d'experts, loi de Jeffreys, intégration du retour d'expérience.

1 Introduction

Notre intérêt porte sur l'intégration des données de retour d'expérience dans un réseau bayésien pour la mise à jour de ses paramètres. Pour notre étude, le processus de vieillissement d'un système mécanique est décrit par un réseau bayésien, via une représentation causale du phénomène étudié. La modélisation par un réseau bayésien offre un cadre d'étude très vaste et de nombreuses possibilités. Pour le domaine qui nous préoccupe, à savoir la sûreté de fonctionnement et l'aide à l'optimisation de la maintenance, l'utilisation des réseaux bayésiens est très bénéfique. En effet, ces modèles nous ont permis de réaliser des analyses de sensibilité et une analyse de données. Ils sont également utilisé comme aide au diagnostic et aide à la décision. Enfin, l'intégration de multitudes de variables comme les coûts, les experts, ou les tâches de maintenance se fait d'une manière simple en ajoutant des nœuds au réseau bayésien.

Dans ce document, la structure du graphe est supposée connue et fixé, 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'expert. 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 classique. Avec les données de retour d'expérience, la probabilité a posteriori est calculée, via la formule de Bayes. 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. 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 non informatives pour des lois binomiales ou multinomiales. On peut citer par exemple le document de Laskey et Mahoney [9] pour l'intégration des données dans les réseaux bayésiens, Bernard [1] pour une description bayésienne des procédures fréquentistes, Walley [14] pour l'étude de modèles de Dirichlet imprécis, Ramoni and Sebiastani [10], pour les problèmes à données manquantes.

L'article est organisé comme suit. Dans un premier temps, nous étudions l'inférence bayésienne classique permettant d'intégrer les données de retour d'expérience. Puis, une brève étude comparative entre les lois a priori non informatives est exposée. Puis, nous étudions des modèles de Dirichlet imprécis où les hyperparamètres varient. Enfin, un exemple d'intégration est donné.

2 Fusion entre connaissance et données dans un réseau bayésien

Soit un réseau bayésien, ou modèle graphique acyclique direct dont la structure est connue, et soit le vecteur de variables aléatoires $X = (X_1, \dots, X_n)$, représentées 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.

2.1 Variables binomiales

Supposons que toutes les variables du réseau bayésien sont binomiales, noté $\mathcal{B}(n, \theta)$. 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'index de l'état $x_{i,j}$ de X_i et c_i l'indice des configurations des nœuds parents du nœud X_i. Soit $\theta_{ijc} = p(X_i = x_{ij} | X_{pa(i)} = x_{ic_i})$, la distribution jointe peut s'écrire sous la forme

$$P(x_{1j_1}, \dots, x_{nj_n}) = \prod_{i=1}^{n} \theta_{ij_ic_i}.$$

Notons que pour les nœuds racines, $c = \emptyset$, 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} 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 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 suppport $[0, 1]$:

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

où B est la fonction beta qui vaut $\frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$ 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. La loi a posteriori est donc une loi Beta $\mathcal{BE}(\alpha_1 + a, \alpha_2 + n - a)$. Supposons 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 $\hat{\theta} = \frac{a + 1}{n + 2}$. Cette valeur diffère de l'estimateur empirique $\frac{a}{n}$. Ainsi, comme l'ont déjà fait remarquer Haldane [6] et Jeffreys [8], la loi a priori uniforme, contrairement aux idées reçues, introduit un biais dans l'estimation. De plus, la loi uniforme n'existent que sur des espaces bornés. Il est donc préférable d'utiliser des lois gén'eralisées. 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 [6] suggère de prendre comme loi a priori la loi limite $\mathcal{BE}(0,0)$ ayant pour densité :

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

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 ($\hat{\theta} = \dfrac{a}{n}$).

Jeffreys [7] 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

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

où $I(\theta) = \mathbb{E}_\theta\left[(\dfrac{\partial log f(X;\theta)}{\partial\theta})^2\right]$ 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 [3] 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. La loi de Jeffreys est généralement une loi impropre. Dans notre exemple, la loi de Jeffreys correspondantes 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 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.

2.2 Variables multinomiales

On suppose maintenant que les variables $X_1, \ldots, X_n$ 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

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

Dans l'article de Laskey et Mahoney [9] et dans bon nombre d'articles, la loi a priori choisie pour θ_{ic_i} est une loi de Dirichlet (la généralisation de la loi Beta dans le cas multinomial) avec pour hyperparamètres $\alpha_{ic_i} = (\alpha_{i1c_i}, \ldots, \alpha_{ik_ic_i})$.

La densité de la loi de Dirichlet est la suivante avec $\displaystyle\sum_{j=1}^{k_i} \theta_{ijc_i} = 1$:

$$f(\theta_{i1c_i}, \ldots, \theta_{ik_ic_i}) = \frac{\Gamma(\sum_j \alpha_{ijc_i})}{\Gamma(\alpha_{i1c_i}) \ldots \Gamma(\alpha_{ik_ic_i})} \theta_{i1c_i}^{\alpha_{i1c_i}-1} \ldots \theta_{ik_ic_i}^{\alpha_{ik_ic_i}-1}.$$

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

$$\mathbb{E}\left[\theta_{ijc}\right] = \frac{\alpha_{ijc}}{\sum_{j'} \alpha_{ij'c}}.$$

Dans un tel contexte et par analogie au cas binomial, on montre que la loi a posteriori est une loi de Dirichlet avec la propriété d'une loi conjuguée et a pour paramètres $(\alpha_{i1c_i} + n_{i1c_i}, \ldots, \alpha_{ik_ic_i} + n_{ik_ic_i})$, où $n_{ic_i} = (n_{i1c_i}, \ldots, n_{ik_ic_i})$ représentent les observations de l'échantillon, distribué suivant une loi multinomiale de paramètre $(\theta_{i1c_i}, \ldots, \theta_{ik_ic_i})$. En effet, la vraisemblance du paramètre d'une loi multinomiale s'écrit avec $n = \sum_{j=1}^{k_i} n_{ijc_i}$ (cf. Saporta [12])

$$\mathcal{L}(n_{ic_i}|\theta_{ic_i}) = \frac{n!}{n_{i1c_i}! \ldots n_{ik_ic_i}!} \theta_{i1c_i}^{n_{i1c_i}} \ldots \theta_{ik_ic_i}^{n_{ik_ic_i}},$$

et donc la densité a posteriori s'écrit alors

$$f(\theta_{ic_i}|n_{ic_i}) = \frac{\Gamma(\sum_j \alpha_{ijc_i} + n_{ijc_i})}{\Gamma(\alpha_{i1c_i} + n_{i1c_i}) \ldots \Gamma(\alpha_{ik_ic_i} + n_{ik_ic_i})} \theta_{i1c_i}^{\alpha_{i1c_i}+n_{i1c_i}-1} \ldots \theta_{ik_ic_i}^{\alpha_{ik_ic_i}+n_{ik_ic_i}-1}.$$

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

$$\hat{\theta}_{ijc_i} = \frac{\alpha_{ijc_i} + n_{ijc_i}}{\displaystyle\sum_{j'} \alpha_{ij'c_i} + \sum_{j'} n_{ij'c_i}}.$$

Dans cette dernière formule, $\sum_j \alpha_{ijc_i}$ est une quantité à fixer. En effet, l'expert donne des proportions, et non des effectifs. Il est à noter que $\sum_j \alpha_{ijc_i}$ représente la taille de l'échantillon virtuel, donné par l'expert. Comme on le détaillera en 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.

3 Quelle loi a priori ?

Dans tout ce paragraphe, nous étudions les différentes possibilités de la loi a priori dans le cas de variables multinomiales. Parmi ces lois, certaines sont

dites impropres, c'est-à-dire qu'elles ne définissent pas de lois de probabilité (cf. Robert [11]).
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 loi non informatives que l'on décrit ci-après.

3.1 La loi uniforme de Bayes-Laplace

Elle correspond à la loi de Dirichlet $\mathcal{D}(1,\ldots,1)$ sur le simplexe. 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. L'inconvénient de cette loi est son support borné et qu'elle ne vérifie pas le principe d'invariance par paramétrisation.

3.2 Le cas limite de Haldane [6]

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 [13]. Ce cas limite pose de gros problèmes notamment lorsque 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

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

Cette inconvénient est majeur, car on ne peut effectuer d'inférence bayésienne, et l'on va voir que ce handicap disparaît avec la loi de Jeffreys.

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. De plus, cette loi est la loi de référence au sens de Bernardo (cf. page 119 dans [3]). 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

expérience, la loi a posteriori est une loi Beta $\mathbb{BE}(1/2, n+1/2)$ 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.

3.4 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. Ces travaux ont été proposés par Walley [14]. Soit une distribution a priori de Dirichlet paramétré comme suit $\mathcal{D}(s,t)$, où s est un réel strictement positif et $t = (t_1, \ldots, t_{k_i})$, 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 : $\hat{\theta}_j^{sup} = \frac{n_j+s}{n+s}$ et $\hat{\theta}_j^{inf} = \frac{n_j}{n+s}$. On peut interpréter s, comme un nombre de données non observées. La borne supérieur correspond èe que toutes les données non observées aient comme valeur x_j. La borne inférieur correspond à ce que toutes les données non observées aient pris des valeurs différentes de x_j. Walley définit également un degré d'imprécision sur cet évènement égal à

$$\hat{\theta}_j^{sup} - \hat{\theta}_j^{inf} = \frac{s}{n+s} \longrightarrow 0 \tag{2}$$

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.

3.5 Conclusions sur les différentes lois a priori

Afin de modéliser une non information, il paraît légitime de choisir une loi a priori de Dirichlet symétrique, c'est-à-dire dans un cas multinomiale à k_i modalités, une Dirichlet $\mathcal{D}(s/k_i, \ldots, s/k_i)$, ne privilégiant ainsi aucune modalité. Dans cas 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. Pour une discussion plus approfondie sur ces valeurs on pourra regarder les travaux de Bernard [2] et Good [5]. En ce qui nous concerne, nous privilégions la loi non informative de Jeffreys, pour son invariance par reparamétrisation, son faible biais, et pour sa robustesse (au cas où une modalité ne soit pas observée).

4 Quelle confiance en les avis d'experts ?

Dans ce paragraphe, nous supposerons que les avis d'experts existent et que par conséquent la loi de Dirichlet n'a aucune raison d'être symétrique. L'objectif de ce paragraphe est de permettre de choisir la taille de l'échantillon fictif qui traduit la confiance en l'avis d'experts. Ce paramètre est égal à $\sum_j \alpha_j = \sum_j st_j = s$, avec les notations utilisées précédemment.

4.1 Dans le cas binomiale

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, ce qui 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 $X_i = x_i^2$. 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 une loi a priori $\mathcal{B}(\alpha_1, \alpha_2)$, elle est de $s = \alpha_1 + \alpha_2 - 2$. 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 une non-information sur un phénomène est la loi de Jeffreys. 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 sont égales à x_1 et $\alpha_2 - 1/2$ données sont égales à 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. Dans ce cas, la loi a priori est 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.

4.2 Dans le cas multinomiale

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 supérieur 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 informatives 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 comparé à la taille de léchantillon réel du retour d'expérience. En pratique, il faudra par exemple convertir un nombre d'année d'expérience de l'expert sur le phénomène étudié en taille d'échantillon fictif. 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, suivant les probabilités.

4.3 Exemple d'application

Nous appliquons la méthodologie présentée dans ce papierà un exemple fictif, et à un composant d'une pompe primaire 1300 MW. Pour cela, nous avons utilisé un réseau bayésien, construit lors d'une étude précédente sur une pompe primaire 900 MW (voir [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 souhaite illustrer cette méthodologie sur le réseau bayésien du joint 1 d'une pompe primaire 900 MW. Cependant, EDF ne posséde pas encore de données réelles disponibles pour cette pompe. Ainsi, nous appliquons la

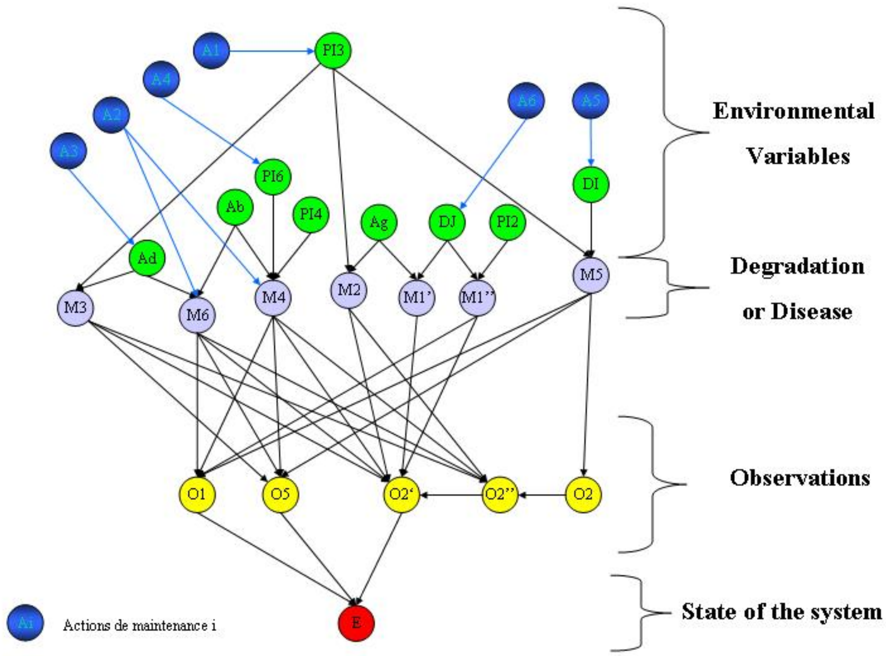

Figure 1: Réseau bayésien pour le joint 1 de la pompe primaire 900 MW.

méthode sur une pompe primaire 1300 MW, où l'on suppose que le réseau bayésien est identique à celui d'une pompe primaire 900 MW (cf. figure 1). 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 1 donne les probabilités données par les experts.

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

Table 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 $p(M4 = 1 \mid Ab = 2, PI6 = 2)$. 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 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

Table 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.

4.3.2 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 3. Au fil des années, les données de retour

$p(S)$	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

Table 3: Exemple fictif d'avis d'experts.

d'expérience sont répertoriés. On suppose que ces données, pour notre exemple fictif ont été tirées selon les probabilités conditionnelles données dans le tableau 4. L'échantillon est répertorié dans le tableau 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

Table 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.

		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

Table 5: Exemple de données de REX fictif avec un échantillon de taille 10, conditionnellement aux variables F et A .

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 $\hat{P}(C = oui|f, a) = 0.6615$.

Supposons maintenant que 90 données de REX soit disponible. Ces données sont répertoriées dans le tableau 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

Table 6: Exemple de données de REX fictif avec un échantillon de taille 100, conditionnellement aux variables F et A.

5 Conclusions

Le réseau bayésien, et notamment les probabilités, 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 conjugués sont les lois de Dirichlet, et des lois de beta pour les variables binomiales. En modélisant une non 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, 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. Cependant, il est envisageable de prendre différentes valeurs de s selon les probabilités conditionnelles.

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.

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.

References

[1] J.-M. Bernard. Bayesian interpretation of frequentist procedures for a bernouilli process. *The American Statistician*, 50(1):7–13, February 1996.

[2] 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.

[3] J.M. Bernardo. Reference posterior distributions for bayesian inference. *J. R. Statist. Soc. B*, 41(2):113–147, 1979.

[4] 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.

[5] 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.

[6] J.B.S. Haldane. The precision of observed values of small frequencies. *Biometrika*, 35:297–300, 1948.

[7] 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.

[8] H. Jeffreys. *Theory of Probability*. Oxford University Press, 1961.

[9] 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.

[10] M. Ramoni and P. Sebastiani. Robust learning with missing data. *Machine Learning*, 2001.

[11] C. P. Robert. *The Bayesian Choice*. Springer Texts in Statistics, Springer-Verlag, 1994.

[12] G. Saporta. *Probabilités, Analyse des données et Statistique*. PARIS TECHNIP, 1996.

[13] C. Villegas. On the representation of ignorance. *Journal of the American Statistical*, 1977.

[14] P. Walley. Inferences from multinomial data: Learning about a bag of marbles. *J.R.Statist. Soc. B*, 58(1):3–57, 1996.