

UNIVERSITÉ DE MONTRÉAL

CONTRIBUTION DES CONNAISSANCES A PRIORI DANS LE PRONOSTIC DU
RISQUE DE DÉFAILLANCE DES SYSTÈMES EN EXPLOITATION

AMIRA BERRAIS
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INDUSTRIEL)
MAI 2017

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

CONTRIBUTION DES CONNAISSANCES A PRIORI DANS LE PRONOSTIC DU
RISQUE DE DÉFAILLANCE DES SYSTÈMES EN EXPLOITATION

présenté par : BERRAIS Amira

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. ADJENGUE Luc-Désiré, Ph. D., président

M. OUALI Mohamed-Salah, Doctorat., membre et directeur de recherche

M. CLÉMENT Bernard, Ph. D., membre et codirecteur de recherche

M. BASSETTO Samuel-Jean, Doctorat., membre

DÉDICACE

*À ma mère qui m'aime d'un amour inconditionnel,
À mon père qui m'a toujours soutenue et encouragée,
À mes frères,
À ma belle sœur,
À mes amis,
À tous ceux qui m'ont aidée,
Et à tous ceux que j'aime,
Je vous dédie ce travail comme simple geste de reconnaissance*

Amira

REMERCIEMENTS

Je tiens à exprimer ma plus vive reconnaissance à mon directeur de recherche, Monsieur Mohamed Salah OUALI pour sa disponibilité, son attention et ses conseils judicieux.

Toute ma gratitude va aussi à mon co-directeur de recherche, Monsieur Bernard CLÉMENT pour son importante contribution à l'élaboration de ce travail et ses qualités d'encadrement.

Je remercie également le Ministère de l'Enseignement Supérieur et de la Recherche Scientifique de la Tunisie, la Mission Universitaire de Tunisie en Amérique du Nord pour toutes les formes de support qu'ils m'ont accordées.

De même, j'aimerais remercier les membres du jury d'avoir accepté d'évaluer le présent travail, Monsieur Luc-Désiré Adjengue en tant que président du jury et Monsieur Samuel-Jean Bassetto en tant que membre de jury.

Je voudrais finalement remercier ma famille, mes amis et tous ceux qui m'ont aidée à parvenir à réaliser ce travail.

RÉSUMÉ

La problématique de recherche concerne le pronostic du risque de défaillance des systèmes en exploitation hautement fiables. Ces systèmes sont soumis à des programmes d'inspection et d'intervention réglementaires qui assurent leur sécurité durant leurs périodes de vie utile. Ces systèmes sont assujettis à très peu d'évènements de défaillances et de défauts de fonctionnement. Le défi des experts en maintenance est de pouvoir capitaliser sur les connaissances acquises suite aux diagnostics des causes à l'origine des défaillances et des défauts de fonctionnement de ces systèmes d'une période à l'autre afin d'établir un pronostic valable de leurs risques de défaillance.

Pour ce faire, trois scénarios de modélisation des avis a priori sur le risque de défaillance sont proposés et analysés. Scénario 1 : une distribution a priori constante d'une période d'exploitation à une autre, ne tenant pas compte des évènements récents ; Scénario 2 : une distribution a priori à mémoire courte, tenant compte des évènements récents au cours de la dernière période d'observation ; Scénario 3 : une distribution à mémoire cumulative qui tient compte de tous les évènements durant toutes les périodes d'observation antérieures. À l'aide de l'inférence bayésienne, ces avis a priori sont conjugués avec les données de défaillance afin d'établir le risque de défaillance a posteriori. Ce risque permet ensuite d'établir un pronostic du risque de défaillance tout au long de la durée de vie d'un système.

Une étude comparative de ces différents scénarios est réalisée afin d'évaluer la contribution de chacun des a priori dans la précision de pronostic du risque de défaillance. Deux modèles de risque de défaillance sont étudiés : le modèle binomial qui caractérise la probabilité de défaillance d'un système et le modèle de Poisson qui caractérise le taux de défaillance. Le Deviation Information Criterion (DIC), le p-value Bayésien et le test de Friedman sont utilisés comme critères de validation et d'analyse des scénarios proposés. Deux échantillons de données de défaillance et de défauts de fonctionnement se rapportant à l'exploitation d'une flotte de moteurs d'aéronefs et une autre de turbines à gaz sont utilisés en guise d'illustration.

ABSTRACT

The research problem is about dependable operating systems failure risk prognosis. These systems are monitored and inspected, on a regular basis, in order to insure their safety during their useful life. They are also subject to rare failure and fault events. This is challenging for the maintenance experts, since they have to be able to capitalize on the acquired knowledge during the diagnosis of failure and faults causes from one period of time to another. As a consequence, a valid failure risk prognosis is determined for each period of time.

For this purpose, three scenarios for modeling a priori opinions on the risk of failure are proposed and analyzed. The first a priori remains the same for all the operating periods of time, thus, not taking into account recent events. The second a priori is a short memory a priori, that takes into account the last period events. Finally, the third a priori is cumulative and considers all the events that took place during all previous periods. Thanks to the Bayesian inference, these a priori opinions are conjugated with failure data so as to establish an a posteriori failure risk, which enables us to establish a prognosis of failure risk throughout the useful life of an equipment.

A comparative study of these different scenarios is carried out in order to evaluate the contribution of each of the a priori in the precision of the prognosis of a system failure risk. Two models of failure risk are considered: the binomial model which characterizes the probability of failure of a system, and the Poisson model which characterizes the rate of failures. The Deviation Information Criterion (DIC), the Bayesian p-value and the Friedman test are used as techniques to validate and analyze the proposed scenarios. Two databases relating to the operation of a fleet of aircraft engines and gas turbines are used for illustration purposes.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	ix
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xii
CHAPITRE 1 INTRODUCTION	1
1.1 Contexte général	1
1.2 Problématique	2
1.3 Objectifs de recherche	4
1.4 Hypothèses de travail	5
1.5 Organisation du mémoire	5
CHAPITRE 2 SYSTÈMES À L'ÉTUDE ET REVUE DE LITTÉRATURE	6
2.1 Systèmes à l'étude : moteurs d'aéronefs et turbines à gaz	6
2.2 Traitement et organisation des données	7
2.3 Revue des modèles de pronostic du risque de défaillance	9
2.4 Modélisation du risque de défaillance	11
2.5 Estimation du risque de défaillance de systèmes hautement fiables	13
2.5.1 Stochastic Expectation Maximization (SEM)	14
2.5.2 Bayesian Restoration Maximization (BRM)	14
2.6 Technique d'inférence bayésienne	15
2.6.1 Fonction de vraisemblance	16
2.6.2 Distribution a priori	17
2.6.3 Lois conjuguées	18
2.6.4 Techniques de simulation par chaînes de Markov	19

2.6.5	Critères de sélection des modèles	21
CHAPITRE 3 SCÉNARIOS ET MODÉLISATION DES AVIS D'EXPERTS		23
3.1	Formulation d'un avis d'expert	23
3.2	Cycle de vie et pronostic du risque de défaillance	23
3.3	Description des scénarios	25
3.4	Modélisation des scénarios	28
3.4.1	Application des scénarios sur le modèle binomial	28
3.4.2	Application des scénarios sur le modèle de Poisson	30
3.5	Construction des distributions a priori	31
3.5.1	Distributions a priori de départ	31
3.5.2	Distributions a priori selon le scénario pour le modèle binomial	33
3.5.3	Distribution a priori selon le scénario pour le modèle de Poisson	34
3.6	Prise en compte de la certitude d'un avis d'expert	36
CHAPITRE 4 RÉSULTATS ET ANALYSE DES SCÉNARIOS		38
4.1	Distributions a posteriori pour le modèle binomial	38
4.1.1	Distribution a posteriori selon le scénario 1	38
4.1.2	Distribution a posteriori selon le scénario 2	39
4.1.3	Distribution a posteriori selon le scénario 3	40
4.2	Distributions a posteriori pour le modèle de Poisson	40
4.2.1	Distribution a posteriori selon le scénario 1	40
4.2.2	Distribution a posteriori selon le scénario 2	41
4.2.3	Distribution a posteriori selon le scénario 3	42
4.3	Sélection et analyse du meilleur scénario selon le modèle de risque	43
4.3.1	Sélection du meilleur scénario pour le modèle binomial	43
4.3.2	Sélection du meilleur scénario pour le modèle de Poisson	44
4.4	Estimation et pronostic du risque de défaillance	46
4.5	Résultats et analyse du risque incluant la certitude de l'avis d'expert	49
4.5.1	Certitude - Type 1	49
4.5.2	Certitude - Type 2	50
4.5.3	Certitude - Type 3	50
CHAPITRE 5 CONCLUSION		52
RÉFÉRENCES		55

LISTE DES TABLEAUX

Tableau 2.1	BDD-1 Données des moteurs d'aéronefs	9
Tableau 2.2	BDD-2 Données des turbines à gaz	9
Tableau 3.1	Données de construction de la distribution a priori de départ : BDD-1	32
Tableau 3.2	Données de construction de la distribution a priori de départ : BDD-2	32
Tableau 3.3	Paramètres des distributions a priori pour le scénario 2 : BDD-1 . . .	33
Tableau 3.4	Paramètres des distributions a priori pour le scénario 2 : BDD-2 . . .	33
Tableau 3.5	Paramètres des distributions a priori pour le scénario 3 : BDD-1 . . .	34
Tableau 3.6	Paramètres des distributions a priori pour le scénario 3 : BDD-2 . . .	34
Tableau 3.7	Paramètres des distributions a priori pour le scénario 2 : BDD-1 . . .	35
Tableau 3.8	Paramètres des distributions a priori pour le scénario 2 : BDD-2 . . .	35
Tableau 3.9	Paramètres des distributions a priori pour le scénario 3 : BDD-1 . . .	35
Tableau 3.10	Paramètres des distributions a priori pour le scénario 3 : BDD-2 . . .	36
Tableau 4.1	Résultats du scénario 1 - modèle binomial : BDD-1	38
Tableau 4.2	Résultats du scénario 1- modèle binomial : BDD-2	39
Tableau 4.3	Résultats du scénario 2 - modèle binomial : BDD-1	39
Tableau 4.4	Résultats du scénario 2 - modèle binomial : BDD-2	39
Tableau 4.5	Résultats du scénario 3 - modèle binomial : BDD-1	40
Tableau 4.6	Résultats du scénario 3 - modèle binomial : BDD-2	40
Tableau 4.7	Résultats du scénario 1 - modèle de Poisson : BDD-1	41
Tableau 4.8	Résultats du scénario 1 - modèle de Poisson : BDD-2	41
Tableau 4.9	Résultats du scénario 2 - modèle de Poisson : BDD-1	41
Tableau 4.10	Résultats du scénario 2 - modèle de Poisson : BDD-2	42
Tableau 4.11	Résultats du scénario 3 - modèle de Poisson : BDD-1	42
Tableau 4.12	Résultats du scénario 3 - modèle de Poisson : BDD-2	42
Tableau 4.13	Comparaison des scénarios : BDD-1	43
Tableau 4.14	Comparaison des scénarios : BDD-2	44
Tableau 4.15	Calcul du DIC moyen : BDD-1	44
Tableau 4.16	Calcul du DIC moyen : BDD-2	44
Tableau 4.17	Comparaison des scénarios : BDD-1	45
Tableau 4.18	Comparaison des scénarios : BDD-2	45
Tableau 4.19	Calcul du DIC moyen : BDD-1	45
Tableau 4.20	Calcul du DIC moyen : BDD-2	45
Tableau 4.21	Risque de défaillance - Modèle binomial : Scénario 3 BDD-1	46

Tableau 4.22	Risque de défaillance - Modèle binomial : Scénario 3 BDD-2	47
Tableau 4.23	Risque de défaillance - Modèle de Poisson : Scénario 3 BDD-1	47
Tableau 4.24	Risque de défaillance - Modèle de Poisson : Scénario 2 BDD-2	48
Tableau 4.25	Résultats a posteriori : Certitude - Type 1 selon le scénario 3	49
Tableau 4.26	Résultats a posteriori : Certitude - Type 2 selon le scénario 3	50
Tableau 4.27	Résultats a posteriori : Certitude - Type 3 - Cas 1 selon le scénario 3	51
Tableau 4.28	Résultats a posteriori : Certitude - Type 3 - Cas 2 selon le scénario 3	51

LISTE DES FIGURES

Figure 2.1	Défaillance d'un système multi composants (Procaccia et al. (2011)) .	12
Figure 2.2	Distribution a priori bêta non informative	17
Figure 2.3	Distribution a priori Gamma non informative	17
Figure 3.1	Cycle de vie et avis d'experts	24
Figure 3.2	Diagramme des scénarios	25
Figure 3.3	Scénario 1. A priori constant	26
Figure 3.4	Scénario 2. A priori adaptatif	27
Figure 3.5	Scénario 3. A priori cumulatif	28
Figure 4.1	Risque de défaillance	48

LISTE DES SIGLES ET ABRÉVIATIONS

DIC	Deviance information criterion
MCMC	Markov Chain Monte Carlo
SEM	Stochastic Expectation Maximization
BRM	Bayesian Restoration Maximization

CHAPITRE 1 INTRODUCTION

1.1 Contexte général

Les défaillances des systèmes multi technologiques tels que les turboréacteurs, les locomotives de train à grande vitesse et les centrales nucléaires peuvent être à l'origine d'accidents mortels. Bien que les organismes règlementaires obligent les entreprises à suivre une démarche méthodologique assez rigoureuse, telle que l'analyse des modes de défaillances, de leurs effets et de leur criticité (AMDEC), pour estimer dès la conception la probabilité de défaillance de ces systèmes, il en reste que les entreprises doivent garantir que cette probabilité devrait rester en deçà d'une certaine valeur critique durant l'exploitation.

De ce fait, les entreprises doivent tenir à jour les calculs des probabilités de défaillance et ce tout en tenant compte des données enregistrées réellement sur leurs systèmes en cours d'exploitation. Ces données peuvent caractériser des durées de vie de certains composants ou sous-systèmes. Elles sont généralement en nombre très restreint. Cependant, beaucoup d'autres données proviennent de plusieurs sources dont les visites préventives règlementaires, les carnets de bord (heures de fonctionnement, incidents techniques, etc.), les diagnostics de défaillance après signalisation d'incidents et les rapports d'accidents. Ces données sont plus abondantes. Cependant, elles sont très hétérogènes et non uniformément formatées. Ces données représentent une connaissance spécifique de la situation examinée par les experts lors des interventions préventives ou après incidents. Elles sont généralement non exploitées à leurs justes valeurs.

Par ailleurs, les outils statistiques communément utilisés pour la modélisation du risque de défaillance tels que les méthodes d'estimation paramétrique ne sont pas adaptées pour les données de durées de vie en petits nombres. Il en résulte que les estimations du risque de défaillance comporteront une grande marge d'erreur et ne seront pas exploitables pour prévenir des accidents. Cependant, l'information consignée par les experts lors des interventions préventives ou après signalisation d'incidents est très importante et généralement difficilement exploitable par les techniques de modélisation statistique.

Le problème principal est d'actualiser le risque de défaillance en présence d'un nombre limité de données de défaillances et d'avis d'experts traduisant les connaissances apprises sur le comportement de ces systèmes via les interventions préventives et les diagnostics de défaillance après incidents. Il faut souligner que lors du diagnostic d'un incident, plusieurs causes peuvent être identifiées. Certaines causes sont liées directement à l'incident en cours et d'autres ne

le sont pas. Ces dernières sont appelées causes indirectes. Cependant, ces causes indirectes peuvent prendre de l'ampleur et causer ultérieurement des incidents sur d'autres composants ou sous-systèmes pouvant mener à une défaillance critique ou catastrophique.

L'objectif de recherche consiste à développer une méthodologie d'actualisation du risque de défaillance des systèmes multi technologiques basée sur un nombre limité d'incidents et une connaissance du comportement de ces systèmes détenue par des experts en maintenance. Ces connaissances sont très importantes dans l'actualisation du risque de défaillance, car elles traduisent le comportement réellement produit au moment de l'incident. Or, les experts peuvent formuler ou traduire de plusieurs manières leurs connaissances à propos d'un incident observé. La principale question de recherche consiste à déterminer la meilleure façon d'exploiter les avis d'experts dans l'actualisation du risque de défaillance des systèmes multi technologiques. À ce titre, plusieurs scénarios d'exploitation des avis d'experts peuvent être étudiés, comparés et analysés.

1.2 Problématique

L'objet de cette étude concerne les systèmes multi technologiques réparables nécessitant le maintien d'une fiabilité très élevée en cours d'exploitation. Le maintien d'un niveau élevé de fiabilité engendre beaucoup d'interventions préventives et par conséquent une faible quantité de données de défaillances. Cependant, beaucoup de données d'incidents ou de défauts mineurs sont généralement collectées lors des interventions préventives. Un incident mineur est défini comme un évènement anormal qui nécessite une investigation par des experts en maintenance en vue de déterminer ses causes directes et indirectes. Généralement, l'apparition d'un incident isolé peut ne pas être critique à première vue. Cependant, lorsque plusieurs incidents sont observés, cela pourrait conduire à une ou plusieurs défaillances critiques voire catastrophiques.

La problématique de recherche concerne l'exploitation des données d'incidents (défaillances et défauts de fonctionnement) afin de modéliser et d'actualiser le risque de défaillance d'une manière dynamique. L'investigation des experts en maintenance pour diagnostiquer les causes primaires d'incidents et les correctifs apportés constitue une connaissance acquise très importante. La principale question concerne l'exploitation de ces connaissances acquises au cours du temps. Il peut y avoir plusieurs possibilités d'exploitation de ces connaissances. Serait-il plus avantageux de cumuler les connaissances acquises au cours des périodes d'intervention antérieures pour prédire le comportement futur d'un système ou de se limiter uniquement aux plus récentes connaissances ? Ces différentes possibilités permettraient-elles d'établir de meilleurs pronostics pour éviter l'apparition des défaillances indésirables ?

Pour ce faire, les techniques d'estimation bayésienne sont particulièrement recommandées lorsque les données de défaillance sont limitées en nombre (Hamada et al. (2008)). Toutefois, ces techniques nécessitent la disponibilité de connaissances a priori sur les événements à prédire. Le recours à l'inférence bayésienne hiérarchique comme technique d'actualisation du risque de défaillance sur la base d'autres données pertinentes telles que les avis de diagnostic sur les incidents observés (défaillances et défauts de fonctionnement) en cours d'exploitation est tout à fait pertinent. Ces dernières peuvent contribuer à la modélisation des distributions a priori sur les paramètres d'intérêt du modèle de défaillance. Aussi, le paradigme bayésien fournit des estimateurs non biaisés des paramètres des distributions de défaillance contrairement aux autres techniques classiques telles que celles du maximum de vraisemblance, des moments et des moindres carrés. Cependant, la difficulté de l'inférence bayésienne réside dans la détermination de la distribution a priori du paramètre à partir d'un avis d'un expert (Procaccia et al. (2011)). Il s'agit particulièrement de tirer profit des connaissances acquises par les experts et de les utiliser comme information a priori sur le paramètre d'intérêt du modèle d'estimation du risque de défaillance. De plus, au-delà de cette élicitation, la question d'utiliser toutes les connaissances disponibles ou seulement une partie reste à investiguer.

En effet, la distribution a priori sur le paramètre d'intérêt peut ne pas être cohérente avec les données réellement observées, dites données de vraisemblance ou données de défaillance, ce qui entraînera des erreurs au niveau des estimations a posteriori sur le paramètre d'intérêt. Pour cette raison, dans ce travail de recherche, plusieurs scénarios de construction de distributions a priori sont proposés à partir de la combinaison des avis d'experts sur les incidents survenus en cours d'exploitation. Chaque scénario représente une façon particulière d'informer et de mettre à jour le risque de défaillance caractérisé par un modèle paramétrique de défaillance.

Par ailleurs, le paradigme de l'inférence bayésienne hiérarchique présente deux sources d'incertitudes : l'incertitude sur le modèle de défaillance et l'incertitude sur les distributions des paramètres d'intérêt. Le choix du meilleur paramètre dépendra de la valeur du p-value, plus ce critère est proche de 0,5 meilleur est l'ajustement de la distribution a posteriori par rapport à celle du paramètre d'intérêt. Le meilleur scénario sera celui qui donnera la plus petite valeur du DIC.

Dans cette étude, deux modèles d'estimation paramétrique du risque de défaillance sont considérés. Le premier est caractérisé par la loi binomiale ayant comme paramètre la probabilité d'occurrence d'une défaillance. Le second modèle est caractérisé par la loi de Poisson ayant comme paramètre le taux de défaillance. Pour chacun des modèles de calcul du risque de défaillance, trois scénarios de construction de l'a priori sur les paramètres d'intérêt sont pro-

posés. Chaque scénario renseigne sur le modèle de calcul du risque de défaillance de manière différente.

Le premier scénario considère un a priori constant d'une période à l'autre, l'information utilisée ne change pas au cours du temps. Le second scénario considère seulement l'information disponible au cours de la période d'intervention précédente pour établir l'a priori sur le paramètre d'intérêt au cours de la période subséquente. Le troisième scénario accumule l'information sur toutes les périodes antérieures pour établir l'a priori sur la période en cours. Nous considérons ces trois scénarios avec chacun des deux modèles d'estimation du risque de défaillance et ce tout au long de la durée d'exploitation d'un système. Cette durée est subdivisée en plusieurs périodes d'estimation du risque. D'une période à l'autre, chacun des trois scénarios de construction de l'a priori sur le paramètre d'intérêt est testé et évalué.

D'un point de vue pratique, les deux modèles de risque de défaillance permettent d'estimer périodiquement la probabilité de défaillance et le taux de défaillance en s'appuyant sur plusieurs scénarios de prise en compte de l'information a priori d'une période d'estimation à l'autre. Ces deux résultats permettent le suivi de l'évolution du risque au cours de la durée d'exploitation et d'ajuster les interventions de maintenance d'une période à l'autre afin de garder ce risque de défaillance en deçà d'un seuil critique acceptable.

1.3 Objectifs de recherche

Cette étude vise à modéliser, d'une manière dynamique, le risque de défaillance d'un système complexe au cours de sa durée d'exploitation afin d'établir le pronostic du risque de défaillance période par période, et ce en utilisant l'inférence bayésienne hiérarchique. Cette inférence fait intervenir, sous différents scénarios, les connaissances a priori sur les paramètres d'intérêt des modèles de risque. Deux exemples de systèmes sont considérés : les moteurs d'aéronefs et les turbines à gaz. Afin d'atteindre cet objectif général, les objectifs spécifiques suivants sont considérés :

- Proposer une méthodologie d'actualisation du risque de défaillance permettant de rendre compte de l'évolution, période par période, durant la durée d'exploitation d'un système, du risque de défaillance lié à une ou plusieurs causes directes ou indirectes ;
- Modéliser les différents scénarios de construction de la distribution a priori du paramètre du modèle choisi en exploitant les données sur les diagnostics d'incidents ;
- Évaluer la performance des scénarios proposés en fonction des critères statistiques tels que le DIC et le p-value Bayésien et retenir le meilleur scénario ;
- Estimer le risque de défaillance pour une flotte de moteurs d'aéronefs et de turbines

à gaz.

1.4 Hypothèses de travail

Afin d'atteindre les objectifs spécifiques, nous supposons que :

- Les données de défaillance et de défauts recueillies sur les systèmes étudiés sont représentatives de l'ensemble de la flotte appartenant à la même famille.
- Les systèmes étudiés sont des individus indépendants. Par conséquent, la défaillance d'un individu de la flotte ne cause pas la défaillance d'un autre.
- Les systèmes sont constitués de plusieurs composants ou sous-systèmes réparables. La réparation d'une défaillance est parfaite. Le système est aussi bon que neuf après chaque intervention.

1.5 Organisation du mémoire

Ce mémoire est structuré en quatre chapitres. Le Chapitre 2 présente une revue de littérature sur les modèles d'estimation du risque de défaillance ainsi que les méthodes statistiques de traitement des données de défaillance fortement censurées telles que les méthodes d'augmentation des données et les méthodes d'inférence bayésienne. Les modèles binomial et de Poisson sont particulièrement revus ainsi que les concepts d'inférence bayésienne hiérarchique et ses techniques d'échantillonnage statistique. Les critères de performance d'un modèle, le DIC et le p-value bayésien, sont présentés.

Le Chapitre 3 décrit la méthodologie de construction des distributions a priori sur le paramètre d'intérêt de chaque modèle de risque de défaillance considéré. Les différents scénarios de modélisation des avis d'experts sont présentés et justifiés. Pour chaque scénario, la distribution sur le paramètre d'intérêt est modélisée et mise en œuvre.

Le Chapitre 4 présente les résultats obtenus pour chaque scénario et ce pour les flottes de moteurs d'aéronefs et des turbines à gaz. Ce chapitre discute et analyse ces résultats en se basant sur les critères de performance, le DIC et le p-value Bayésien. Une comparaison basée sur le test de Friedman est utilisée pour comparer la performance de chaque scénario de façon globale dans le pronostic du risque d'une période à une autre.

Enfin, le Chapitre 5 présente une conclusion générale, les limites de la méthodologie proposée et les perspectives d'amélioration.

CHAPITRE 2 SYSTÈMES À L'ÉTUDE ET REVUE DE LITTÉRATURE

Ce chapitre présente les deux systèmes à l'étude ainsi que les données disponibles pour chaque système. Il présente également une revue de littérature des modèles paramétriques d'estimation de risque et leurs applications. Les méthodes classiques d'estimation des paramètres des modèles de risque et les techniques d'estimation bayésienne du risque sont discutées. Finalement, les méthodes de simulation de Monte-Carlo par chaînes de Markov (MCMC) et les critères d'ajustement et de sélection des modèles dans une approche bayésienne sont présentés.

2.1 Systèmes à l'étude : moteurs d'aéronefs et turbines à gaz

L'étude concerne deux types de systèmes : les moteurs à compression utilisés pour la propulsion des aéronefs et les turbines à gaz exploitées dans le transport de gaz dans les oléoducs. Ces systèmes sont très sollicités en industrie. Ils nécessitent une attention particulière des services de maintenance car ils doivent garder une fiabilité assez élevée par rapport à d'autres systèmes. Aussi, la défaillance d'un moteur d'aéronef peut occasionner des pertes en vies humaines. Celle d'une turbine à gaz engendre l'arrêt de l'export du gaz naturel aux clients et peut occasionner, des pertes financières qui peuvent atteindre des dizaines de milliers de dollars par jour.

Un moteur d'aéronef est constitué d'une turbosoufflante qui souffle le flux d'air à l'entrée d'un compresseur. Ce dernier aspire l'air et le comprime pour le ramener à une température et une pression optimales requises par la combustion. À la suite de l'injection du carburant, l'énergie calorifique produite par la combustion du carburant et de l'air comprimé est transformée par la turbine à gaz en énergie mécanique. La turbine entraîne à son tour le compresseur et la turbosoufflante.

Dans un chantier d'exploitation gazière, les turbines à gaz sont sollicitées en permanence. Elles entraînent des compresseurs centrifuges pour augmenter la pression du gaz à exporter et ainsi faciliter son déplacement dans les oléoducs. Une unité de compression du gaz se compose de trois compresseurs indépendants entraînés chacun par une turbine à gaz. Une turbine à gaz est constituée principalement d'un compresseur axial, d'une chambre de combustion, d'une turbine de puissance et de systèmes auxiliaires. La turbine à gaz convertit l'énergie thermique à la sortie de la chambre de combustion en énergie mécanique. L'air ambiant passe par le compresseur axial entraîné au début par le système de démarrage. Ensuite, en présence du

gaz naturel et d'une source d'ignition dans la chambre de combustion, l'air comprimé est transformé en gaz chaud à haute pression entraînant à grande vitesse la génératrice de gaz (GG) et la turbine de puissance (PT) grâce à l'air chaud dégagé par la GG.

Ces deux systèmes présentent une différence importante quant à la façon de les considérer dans l'étude : les moteurs d'aéronefs fonctionnent de manières complètement indépendantes alors que les turbines à gaz doivent fonctionner simultanément afin d'obtenir la pression suffisante pour transporter le gaz.

2.2 Traitement et organisation des données

Un effort très important a été déployé pour construire les deux bases de données nécessaires à l'étude. L'objectif est d'extraire les données sur les défauts, les défaillances et les durées d'exposition des systèmes sous étude à partir de plusieurs sources d'informations sur les moteurs d'aéronefs et de turbines à gaz. Les défaillances sont rares sur ces deux systèmes. D'autant plus que pour les moteurs d'aéronefs, les données de défaillances sont hautement protégées pour des raisons de sécurité. Donc, nous avons regroupé les données de défaillance de l'ensemble d'une flotte de moteurs d'un même constructeur, opérant dans des conditions d'utilisation presque similaires et soumis aux mêmes procédures d'intervention de maintenance.

Pour comptabiliser des défauts, nous n'avons pas uniquement pris en compte les défauts qui ont déclenchés des alarmes mais aussi ceux qui ont été décelés lors des inspections ainsi que les défauts potentiels. En effet, lors de l'inspection, si l'ingénieur de maintenance détecte une anomalie au niveau d'un composant qui peut éventuellement déclencher ultérieurement une alarme et causer l'arrêt du système, cette anomalie est considérée comme un défaut. Ainsi, deux types de défauts sont considérés :

1. Alarme et défaut confirmé suite à un diagnostic car une alarme peut être déclenchée à cause d'un problème avec un capteur et non pas le composant en question.
2. Défaut n'ayant pas causé une alarme, mais confirmé suite à une inspection (cause potentielle d'arrêt dans le futur).

L'ensemble des données sur les défaillances, les défauts, les réparations et les remplacements a été extrait à partir des rapports journaliers et mensuels de panne et d'interventions ainsi que des demandes de travaux curatifs (bons de travail). Souvent, les documents d'inspection et de diagnostic consultés sont subjectifs. Plusieurs champs lexicaux expliquent en détails le type de maintenance effectué et la manière avec laquelle le défaut a été décelé ainsi que

son mode défaillance. Les données de diagnostic correspondent aux causes potentielles de défaillance : ces données sont récoltées lors des inspections et des réparations correctives. L'évènement principal recherché à travers tous ces documents, et ce pour les deux systèmes étudiés est l'arrêt de fonctionnement.

Lors du traitement des données brutes, nous n'avons pas considéré les délais logistiques pour manque de pièces de rechange, d'absence de la main-d'œuvre spécialisée et d'erreurs de diagnostic qui ont retardé la réparation du défaut. Aussi, nous supposons que la réparation est instantanée suite à l'apparition du défaut. De ce fait, le défaut détecté suite à une alarme est compté uniquement lors de sa première apparition même s'il continue à apparaître dans les données et qu'il n'y a pas eu d'intervention pour le corriger.

Par ailleurs, les périodes de rodage et de démarrage de test n'ont pas été retenues pour la construction des bases de données. Les systèmes sont considérés dans les phases de vie utile et de vieillissement seulement. De ce fait, les données collectées correspondent au fonctionnement des systèmes après la stabilisation du processus afin d'avoir des analyses en régime permanent. Les rapports journaliers et mensuels de maintenance qu'engendrent toutes les actions de maintenance nous ont servis pour la construction des nos bases de données.

En résumé, les données disponibles pour les moteurs d'aéronef et les turbines à gaz concernent :

- Durée d'exposition T . Cette durée représente le temps de fonctionnement du système.
- Nombre de défaillances X . Cette variable donne le nombre de défauts ayant déclenché l'alarme et causé l'arrêt du système.
- Nombre de défauts Y . Cette variable donne le nombre de défauts détectés lors de l'inspection et corrigés par crainte qu'ils déclenchent une alarme ultérieurement.

Pour la flotte d'aéronefs, la durée d'exposition T représente le cumul des temps de fonctionnements de tous les moteurs en service durant la période concernée. Cette durée est approximative. Elle est calculée sur la base d'un nombre moyen d'heures de fonctionnement d'un moteur multiplié par le nombre de moteurs en service. Ce dernier est donné par le nombre de livraisons des avions entre 1997 et 2014 (General Aviation Manufacturer Association (GAMA) <http://www.gama.aero/publications>).

Les tableaux 2.1 et 2.2 présentent la structure des données BDD-1 pour les moteurs d'aéronefs et BDD-2 pour les turbines à gaz.

Tableau 2.1 BDD-1 Données des moteurs d'aéronefs

i Période	X Nb de déf.(IFSD)	Y Nb de défauts	T Durée d'exposition (h)	N Nb de moteurs
1	2	3	3820165	14800
2	5	2	4089164	14800
3	2	1	3658563	14800
4	2	5	4216473	14800
5	2	6	4502048	14800
6	7	10	4088984	14800
7	4	9	4348631	14800

Tableau 2.2 BDD-2 Données des turbines à gaz

i Période	X Nb de déf.(SD)	Y Nb de défauts	T Durée d'exposition (h)	N Nb de turbines
1	2	10	2804	5000
2	3	14	2959	5000
3	9	18	2829	5000
4	5	7	4763	5000
5	5	11	5896	5000
6	14	4	6940	5000
7	4	12	4929	5000
8	10	7	8804	5000
9	2	10	1667	5000
10	6	19	1000	5000

2.3 Revue des modèles de pronostic du risque de défaillance

Les modèles de pronostic de risque sont multiples et s'adressent à plusieurs domaines d'ingénierie. En effet, Gaver and O'Muircheartaigh (1987) traite le cas de 10 pompes similaires dans une centrale nucléaire. Pour chaque pompe i ($i \in \{1, \dots, 10\}$), une durée d'utilisation T_i et un nombre de défaillances X_i sont enregistrés. L'auteur propose un modèle de taux de défaillance caractérisé par un processus de Poisson homogène de paramètre λ_i . Chaque paramètre d'intérêt est considéré comme une variable aléatoire ayant une distribution a priori qui suit la loi Gamma à deux hyper-paramètres $\lambda_i \sim G(1.8, \beta)$. Le paramètre de forme est fixé à 1,8 alors que le paramètre d'échelle β suit une loi Gamma $\beta \sim G(\gamma, \zeta)$. Les valeurs

du couple d'hyper-paramètres (γ, ζ) sont données par un expert. La résolution du modèle du taux de défaillance a posteriori utilise l'approche bayésienne hiérarchique et l'échantillonnage MCMC de Gibbs.

Zaidi et al. (2012) propose la modélisation de la fiabilité des systèmes complexes à l'aide d'un réseau bayésien. Ces systèmes ne peuvent être représentés ni par des diagrammes de fiabilité ni des arbres de défaillance. Le modèle consiste à estimer la fiabilité en définissant la relation causale entre les composants du système. La structure du système peut être connue ou inconnue. Pour ce faire, un système (S) constitué de trois composants indépendants (A, B, C) a été étudié. Les données de défaillances du système et des trois composants ont été enregistrées. L'expert se prononce sur les probabilités conditionnelles qui caractérisent les relations causales dans le réseau bayésien selon les différents scénarios de défaillance des composants menant à la défaillance du système. La fiabilité du système est donnée par la somme de toutes les probabilités conditionnelles afférentes.

Le modèle proposé par Shafaghi (2008) étudie le risque de défaillance de cuves d'entreposage exploitées par une société de pétrole et de gaz. Les données de défaillances sur 187 cuves sous pression ont été collectées sur une période de 15 ans. Le modèle cherche à estimer le taux de défaillance d'une cuve selon plusieurs modes de défaillance. Ainsi, le nombre de défaillances est caractérisé par un processus de Poisson homogène $X \sim Poisson(\lambda t)$. Le paramètre λ suit a priori une distribution Gamma, $\lambda \sim G(\alpha, \beta)$. Afin de déterminer cette distribution a priori, l'analyste élicite plusieurs avis d'experts exprimés sur le taux de défaillance moyen d'une cuve sous pression et attribue un facteur de rang et ensuite un poids selon la crédibilité de l'avis. La valeur moyenne du taux de défaillance est la somme des taux de défaillance attribués par chaque avis multiplié par leurs poids. Cette valeur constitue la moyenne de la distribution a priori et la détermination de ses hyper-paramètres en découle. Ainsi, la distribution a posteriori du paramètre représentant le taux de défaillance actualisé est déduite à partir de l'actualisation des paramètres de la distribution a priori avec les données de défaillances et la moyenne de la distribution a priori. Cette forme d'a priori considère une double subjectivité provenant de la détermination de la moyenne du paramètre d'intérêt et le jugement que l'analyste porte par rapport à ces différents avis.

Wang et al. (2015) proposent un modèle basé sur le modèle de Kijima (1989) pour le pronostic des défaillances pour une flotte de turbines auxiliaires (APU). Le modèle tient compte de deux facteurs : l'effet de la réparation puisque l'APU peut faire l'objet d'une réparation mineure, majeure ou d'un remplacement et ce dépendamment du mode de défaillance et de l'âge virtuel de l'équipement après la réparation. Ce modèle suppose que les APUs ne peuvent pas être modélisées par des processus de renouvellement ou des processus de Poisson

non homogène, et ce parce qu’après réparation l’équipement revient à un état entre aussi bon que neuf et aussi mauvais que vieux. La turbine est décomposée en trois sous-systèmes indépendants. Par conséquent, la probabilité de défaillance d’une turbine est obtenue par la somme des probabilités de défaillance de chacun des sous-systèmes. Les données de durées de vie sont nécessaires au développement de ce modèle. Le modèle est basé sur l’estimation ponctuelle contrairement à l’inférence bayésienne qui fournit une distribution du paramètre. Les modèles précédents partagent l’usage des durées de défaillance et/ou du nombre de défaillances enregistrés durant une certaine durée d’exploitation. Le focus est fait sur la manière de construire la distribution a priori sur le paramètre d’intérêt à partir des informations de défaillance disponibles dans l’historique. Dans le cas qui nous concerne, plusieurs données de maintenance sont disponibles et la question que nous nous posons est la suivante. Comment pouvons-nous exploiter cette information de maintenance dans la formulation des avis d’experts sur le risque de défaillance d’un système.

2.4 Modélisation du risque de défaillance

Dans nos deux cas d’études, comme les évènements de défaillance sont rares, nous avons choisi les lois binomiale et de Poisson. La loi binomiale est une loi discrète qui sert à modéliser le nombre de succès ou d’échec parmi n essais indépendants. La loi de Poisson est une loi discrète qui sert à modéliser un comptage des évènements pendant une période de temps T . Chacune de ces deux lois est définie par un seul paramètre : la probabilité de défaillance p d’un système pour la loi binomiale et le taux de défaillance d’un système par unité de temps λ pour le loi de Poisson. Comme ces deux lois supposent que les évènements sont indépendants et qu’il n’y a aucun d’effet d’âge, par conséquent les paramètres p et λ sont constants (Siu and Kelly (1998)).

Le choix de la loi binomiale repose sur le fait que nous nous intéressons à la défaillance d’un certain nombre de systèmes d’une flotte parmi plusieurs. Ici, chaque système de la flotte représente une répétition. Ainsi, la fonction de masse de la loi binomiale est donnée par :

$$P(Y = y) = C_y^n \cdot p^y \cdot (1 - p)^{n-y}, \quad \forall y \in \{0, 1, 2, \dots, n\}. \quad (2.1)$$

Avec :

- p : probabilité d’échec (probabilité d’arrêt).
- y : valeur observée de la variable aléatoire Y représentant le nombre d’échec (arrêt).

La fonction de répartition de la loi binomiale, nommée aussi fonction de risque, est donnée par :

$$F_Y(y) = \sum_{k=0}^y C_k^n \cdot p^k \cdot (1-p)^{n-k}, \quad \forall y \geq 0 \quad (2.2)$$

Le choix de la loi de Poisson (processus de poisson homogène) repose sur l'hypothèse qu'il s'agit d'un système réparable et que la réparation est parfaite. Il s'agit d'une hypothèse plausible car la fiabilité des systèmes considérés dans cette étude est très élevée. Il y en a très peu de défaillances sur une longue période d'exploitation. Le processus de poisson est souvent utilisé pour modéliser la défaillance d'un système multi-composant réparable. Selon Procaccia et al. (2011), le taux de défaillance constant est justifié par le fait que la réparation d'un composant défaillant d'un système ne change pas significativement le taux de défaillance du système au complet comme l'illustre la figure 2.1.

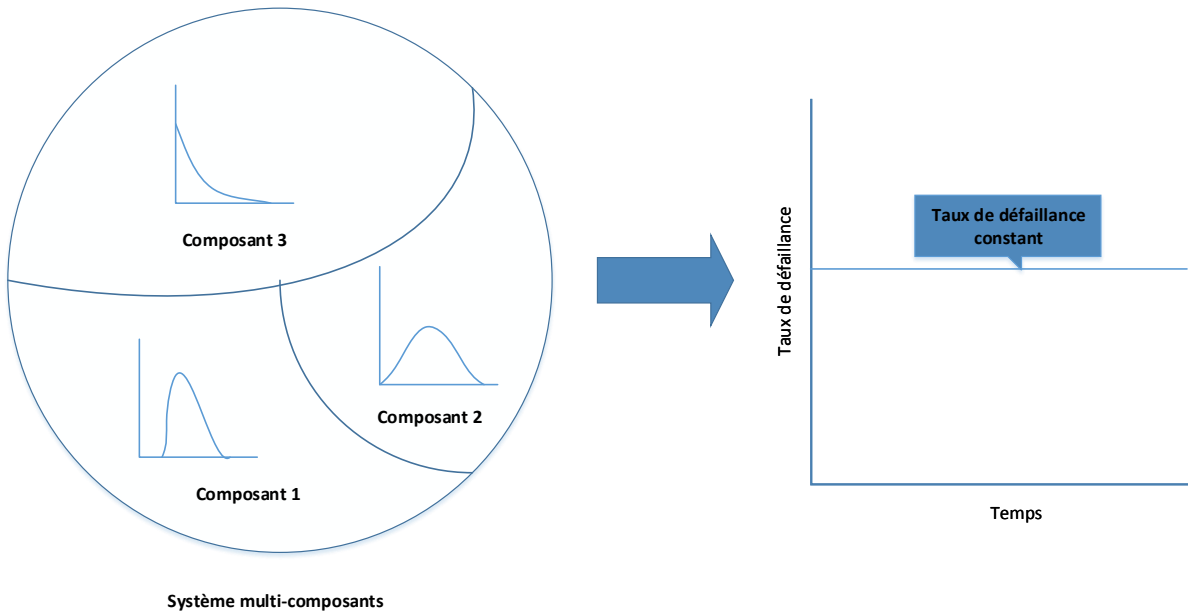


Figure 2.1 Défaillance d'un système multi composants (Procaccia et al. (2011))

Ainsi, la fonction de masse est donnée par :

$$P(Y = y) = \frac{(\lambda T)^y \cdot \exp(-\lambda T)}{y!}, \quad \forall y \in \mathbb{N}. \quad (2.3)$$

Avec :

- λ : taux de défaillance (taux d'arrêt).
- λT : paramètre de la loi de poisson.
- y : valeur observée de la variable aléatoire Y représentant le nombre des évènements pendant un intervalle de temps T .

La fonction de répartition est donnée par :

$$F_Y(y) = \sum_{k=0}^y \frac{(\lambda T)^k \cdot \exp(-\lambda T)}{k!}, \quad \forall y \geq 0. \quad (2.4)$$

2.5 Estimation du risque de défaillance de systèmes hautement fiables

Sans perte de généralité, les données de vie collectées durant l'exploitation d'un système peuvent être, soient des données de défaillance, soient des données censurées. Les données de défaillance sont obtenues lorsque le système perd une de ses fonctions principales. Le système n'est plus fonctionnel. Les données censurées sont des durées de vie incomplètes. Le système est observé à un moment donné, il était fonctionnel. Le système n'est plus observé par la suite. Une typologie des données de durée de vie est présentée dans (Bacha et al. (1998)). Pour la plupart des systèmes hautement fiables, il y a très peu de données de défaillance comparativement aux données censurées. Par rapport à nos deux cas d'études, les données de durée vie censurées peuvent concerner les composants ayant un défaut diagnostiqué lors de l'inspection et ensuite réparé.

Le traitement statistique des données de défaillance pour établir la distribution paramétrique des durées de vie d'un système est généralement fait par la technique de maximum de vraisemblance. Cette technique donne une estimation non biaisée des paramètres de la distribution lorsque le nombre de données de défaillance est relativement grand. Toutefois, lorsque ce nombre est très petit, cette méthode donne une estimation biaisée des paramètres. Dans ce cas, d'autres techniques peuvent être utilisées pour transformer les données censurées en données de défaillance.

Les méthodes Stochastic Expectation Maximization (SEM) et Bayesian Restoration Maximization (BRM) sont particulièrement proposées pour transformer les données censurées en des données de défaillance (Bacha et al. (1998)). Ces méthodes se basent sur le principe de ré-échantillonnage des données censurées afin de les transformer en des données de défaillance. Ainsi, l'estimation des paramètres de la distribution des durées de vie par la technique de maximum de vraisemblance donne des estimés non biaisés de ces derniers. Par conséquent, cela diminue l'erreur sur les paramètres estimés (Bacha and Celeux (1996)).

2.5.1 Stochastic Expectation Maximization (SEM)

La méthode SEM permet de remplacer les données censurées T_c par des données complètes T_d simulées à partir de la distribution conditionnelle $f(T_d|T_c, \alpha^i)$, $\{T_1, T_2, \dots, T_r, T_{r+1}, \dots, T_n\}$ où $\{T_1, \dots, T_r\}$ représente l'ensemble des données complètes et $\{T_{r+1}, \dots, T_n\}$ représente l'ensemble des données censurées (Bacha et al. (1998)).

- n : nombre de systèmes réparables.
- T_d : données complètes (temps de défaillance) et $g(T_d, \alpha)$ sa distribution.
- T_c : données censurées (temps de censure) et $h(T_c, \alpha)$ sa distribution

Selon Mihalache et al. (2004), l'algorithme SEM consiste à :

- Calculer la distribution conditionnelle $f(T_d|T_c, \alpha^i) = \frac{g(T_d, \alpha^i)}{h(T_c, \alpha^i)}$ sachant que la valeur initiale du paramètre α^0 doit être choisie d'une manière judicieuse.
- Simuler un échantillon $T_s = T_d^{(i+1)}$ à partir de la distribution conditionnelle $f(T_d|T_c, \alpha^i)$.
- Estimer les paramètres $\alpha^{(i+1)}$ de l'échantillon complété par la méthode de maximum de vraisemblance.

Les trois étapes de l'algorithme SEM sont répétées M fois et un ensemble $\{\alpha^1, \dots, \alpha^M\}$ est obtenu. À partir de cet ensemble, la quantité α est estimée de la manière suivante :

$$\hat{\alpha} = \frac{1}{L} \sum_{i=M-L+1}^M \alpha^i, \text{ avec } L \text{ représentant les dernières itérations des } M \text{ itérations.}$$

D'après Bacha et al. (1998), un minimum de 500 itérations pour M et de 300 pour L sont requis afin d'atteindre un régime stationnaire des paramètres et d'avoir des estimateurs fiables.

2.5.2 Bayesian Restoration Maximization (BRM)

Contrairement à la technique SEM, la technique BRM utilise la distribution a priori du paramètre dans son processus d'augmentation des données. En effet, à partir de la distribution a priori $\Pi(\alpha)$, le paramètre α est déterminé. Il sera introduit dans la distribution conditionnelle $\Pi(T|\alpha)$. À partir de cette dernière distribution, les données complètes T_d sont simulées (Bacha et al. (1998)). Les étapes de l'algorithme sont les suivantes :

- Choisir α à partir de la distribution a priori $\Pi(\alpha)$.
- Simuler les données complètes à partir de la distribution $\Pi(T|\alpha)$, ces données simulées viendront remplacer les temps de censure dans l'échantillon de données.
- Estimer le paramètre α , pour le nouvel échantillon de données, par le biais de la technique de maximum de vraisemblance.

La répétition des étapes de l'algorithme permet d'obtenir un échantillon du paramètre

$\alpha = (\alpha_1, \dots, \alpha_M)$ estimé pour chaque donnée complète simulée. La distribution a posteriori $\Pi(\alpha|T)$ est déterminée grâce à la méthode des noyaux et le principe de lissage maximum (Terrell (1990)).

Un estimé du paramètre à partir de sa distribution a posteriori est calculé grâce à l'échantillonnage préférentiel de la manière suivante selon Mihalache (2007) :

$$\hat{\alpha} = \frac{\sum_{i=1}^M w_i \alpha^i}{\sum_{i=1}^M w_i} \quad (2.5)$$

avec : $w_i = \frac{\Pi(\alpha^i)}{\Pi(\alpha^i|T)}$.

D'après Geweke (1989), l'échantillonnage préférentiel converge presque toujours.

En général, les techniques d'augmentation de données SEM et BRM donnent des estimations des paramètres sans biais (Bacha M (1994)). Toutefois, elles nécessitent la connaissance des temps de défaillance et de censure, d'autant plus que la technique SEM ne permet pas l'utilisation de la distribution a priori. Or, comme nous souhaitons tester différents manières de construire la distribution a priori et analyser son impact sur les résultats a posteriori et que nous disposons uniquement des durées d'exposition regroupant plusieurs systèmes, les précédentes techniques sont, par conséquent, inapplicables dans notre situation. L'inférence bayésienne représente une technique plus générale, nous la présentons dans la section suivante.

2.6 Technique d'inférence bayésienne

L'inférence bayésienne décrit l'incertitude d'un expert par rapport au paramètre d'intérêt exprimée dans la distribution a priori en amont de la collecte des données de vraisemblance. Cette incertitude varie en fonction des données, ce changement est notable au niveau de la distribution a posteriori (Dezfuli et al. (2009)).

La technique d'inférence bayésienne s'applique pour les grands et petits échantillons de données et permet, grâce au processus de mise à jour et d'actualisation périodique, de surveiller de près les systèmes (Ion et al. (2006)). Elle considère les paramètres à estimer comme des variables aléatoires et leur attribue une distribution représentant l'incertitude sur ce paramètre.

Contrairement aux techniques fréquentistes, l'avantage de l'inférence bayésienne est qu'elle ne donne pas d'estimation biaisée en cas d'une pénurie de données (Ion et al. (2006)). En

effet, même dans le cas d'une distribution a priori vague, la variance du paramètre inconnu sera plus petite que celle donnée par les méthodes classiques (moindres carrées, maximum de vraisemblance,...) (Billy et al.).

Grâce à l'inférence bayésienne, nous allons d'une part pallier au problème de manque de données et d'autre part intégrer l'expertise des spécialistes pour l'estimation du risque de défaillance. Il en découle la combinaison de la réalité, tirée des observations et des expériences, et des informations supplémentaires généralement négligées pour l'estimation du risque. En effet, l'avis d'un expert résume son expérience passée et les données de vraisemblance viennent la compléter.

La distribution a posteriori est déterminée grâce la formule de Bayes :

$$\pi(\theta|Y) = \frac{\pi(Y|\theta) \cdot \pi(\theta)}{\int \pi(Y|\theta) \cdot \pi(\theta) \cdot d\theta} \quad (2.6)$$

Avec :

- Y : données observées.
- θ : paramètre du modèle.
- $\pi(Y|\theta)$: fonction de vraisemblance.
- $\pi(\theta)$: distribution a priori.
- $\pi(\theta|Y)$: distribution a posteriori.

2.6.1 Fonction de vraisemblance

Le rôle de la fonction de vraisemblance est d'incorporer les données dans un modèle qui décrit le comportement aléatoire du système (Siu and Kelly (1998)). Le choix du modèle est lié à la nature du système (réparable, dégradable, etc.) et de la variable aléatoire que nous cherchons à modéliser.

La fonction de vraisemblance associée à la probabilité de défaillance p pour la distribution binomiale est :

$$L(p) = \prod_{i=1}^k C_{y_i}^n \cdot p^{y_i} \cdot (1-p)^{n-y_i} \quad (2.7)$$

La fonction de vraisemblance de λ pour un processus de Poisson homogène est :

$$L(\lambda) = \prod_{i=1}^k \frac{(\lambda T)^{y_i} \cdot \exp(-\lambda T)}{y_i!} \quad (2.8)$$

$y_i, i = 1, \dots, k$ sont les données observées.

2.6.2 Distribution a priori

La distribution a priori du paramètre modélise les connaissances qualitatives a priori ainsi que l'incertitude sur ce paramètre. Elle doit contenir des informations et des opinions avant la collecte des données de vraisemblance. Pour ce faire, un choix judicieux s'impose car il peut affecter les résultats et l'efficacité de la technique d'inférence bayésienne en dépend. L'effet de la distribution a priori sur la distribution a posteriori diminue avec l'augmentation des données de vraisemblance (Siu and Kelly (1998)).

A priori non informatif

Un a priori non informatif donne une information vague sur le paramètre d'intérêt. Ce qui résulte d'un manque d'information a priori. Ce manque d'information est le résultat d'une pénurie d'expertise dans le domaine ou l'exploitation d'un nouveau système dont nous n'avons pas encore enregistré des données historiques.

Dans le cas d'une distribution binomiale, l'a priori non informatif correspond à une distribution $\text{bêta}(1, 1)$ (Kleyner et al. (2015)) fig(2.2.). Alors que la distribution $\text{gamma}(0.5, 0.0001)$ (Zaidi et al. (2012)) fig(2.3.) constitue un a priori non informatif pour la loi de Poisson.

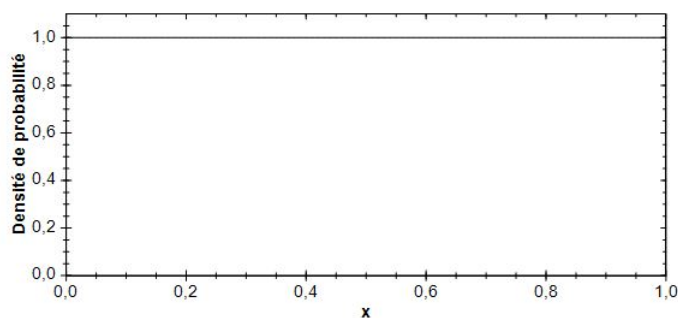


Figure 2.2 Distribution a priori bêta non informative

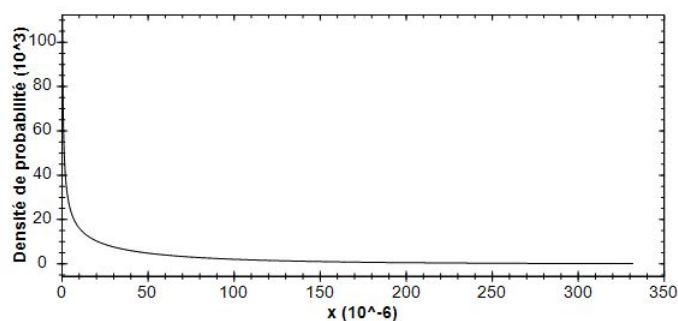


Figure 2.3 Distribution a priori Gamma non informative

Il est déconseillé d'utiliser une distribution a priori non informative lorsque le nombre de données de vraisemblance est faible. De ce fait, l'a priori non informatif peut être utilisé dans le cas d'un important échantillon de données de vraisemblance pour qu'il n'ait pas d'influence sur la distribution a posteriori du paramètre.

A priori informatif

La construction de la distribution a priori informative se base sur des données historiques d'un même système ou d'un système de technologie similaire, de l'expérience acquise par les spécialistes et par les intuitions. D'après Kelly and Smith (2011), la distribution a priori peut être développée à partir de trois (3) informations comme :

- La médiane ou la moyenne et une valeur maximale.
- La valeur maximale et minimale.
- La moyenne et l'écart-type.

2.6.3 Lois conjuguées

La distribution a priori conjuguée à la fonction de vraisemblance permet d'avoir une distribution a posteriori appartenant à la même famille que la distribution a priori. Les deux lois conjuguées à la distribution binomiale et de Poisson sont respectivement la loi bêta et la loi gamma qui sont des lois continues.

Loi bêta

Soit Y la variable aléatoire modélisée par la distribution binomiale de paramètres N et p et soit y sa réalisation. Nous supposons que le paramètre p suit une loi bêta de paramètres α et β et nous écrivons $p \sim beta(\alpha, \beta)$.

La fonction de densité de la loi bêta est :

$$f(p) = \frac{p^{\alpha-1} \cdot (1-p)^{\beta-1}}{beta(\alpha, \beta)} \cdot \mathbb{1}_{[0,1]}(p) \quad (2.9)$$

La moyenne d'une loi bêta est $\frac{\alpha}{\alpha + \beta}$ et sa variance est $\frac{\alpha\beta}{(\alpha + \beta)^2(1 + \alpha + \beta)}$

Les deux paramètres α et β de la distribution bêta présentent respectivement le nombre de succès (ou échec) et le nombre d'échecs (ou succès) sur les N répétitions (Kelly and Smith (2011)). α est le nombre de défaillances et $\alpha + \beta$ est le nombre de répétitions (nombre de systèmes de la flotte).

L'actualisation de la distribution a priori en fonction des données de vraisemblance est faite

de la manière suivante (Kerman et al. (2011)) :

$$\begin{cases} \alpha_{post} = \alpha_{priori} + y \\ \beta_{post} = \beta_{priori} + N - y \end{cases} \quad (2.10)$$

y étant la réalisation de la variable aléatoire Y .

Loi Gamma

Soit Y la variable aléatoire modélisée par la distribution de Poisson de paramètre λ . Soit y sa réalisation et t la durée pendant laquelle Y s'est réalisée. Nous supposons que le paramètre λ suit une loi gamma de paramètres α et β et nous notons $\lambda \sim \text{gamma}(\alpha, \beta)$.

La fonction de densité d'une loi gamma est donnée par :

$$f(\lambda) = \frac{\beta}{\Gamma(\alpha)} (\beta\lambda)^{\alpha-1} \exp(-\beta\lambda) \mathbb{1}_{[0,+\infty]}(\lambda) \quad (2.11)$$

Sa moyenne est $\frac{\alpha}{\beta}$ et sa variance est $\frac{\alpha}{\beta^2}$

Les deux paramètres α et β de la distribution gamma présentent respectivement le nombre d'évènements attendus et la durée pendant laquelle ces évènements se sont réalisés. L'actualisation de la distribution a priori en fonction des données de vraisemblance est faite de la manière suivante (Shafaghi (2008)) :

$$\begin{cases} \alpha_{post} = \alpha_{priori} + y \\ \beta_{post} = \beta_{priori} + t \end{cases} \quad (2.12)$$

2.6.4 Techniques de simulation par chaines de Markov

Ces techniques permettent de générer un échantillon aléatoire à partir de la distribution à posteriori dont la nature est inconnue et ce pour contourner les difficultés de calcul d'intégrales dans le cas des distributions non conjuguées. Une chaîne de Markov est une série de nombres où chaque nombre dépend du point précédent de la série. L'échantillon généré ainsi que le temps de convergence dépendent des points de départ de la chaîne. Ainsi, nous devons choisir un point de départ qui soit dans le support de la distribution a priori. Une façon de minimiser l'influence de ces points est d'éliminer la première partie de l'échantillon, appelé période d'échauffement, et qui correspond à la durée de stabilisation de la chaîne.

Le logiciel libre Winbugs offre la possibilité de générer les points initiaux de la chaîne de Markov à partir de la distribution a priori du paramètre. Dezfuli et al. (2009) proposent 1000

itérations pour l'échauffement dans le cas d'une distribution ayant un seul paramètre.

La convergence de la chaîne de Markov dépend du nombre d'itérations : plus le nombre d'itérations est grand plus nous sommes sûrs d'avoir une bonne estimation du paramètre. D'après Spiegelhalter et al. (2003), la chaîne converge quand l'erreur de la chaîne Monte Carlo est inférieure à 5% de l'écart type de l'échantillon du paramètre.

Parmi les algorithmes de simulation, nous présentons les plus utilisés, l'algorithme de Gibbs et l'algorithme de Metropolis-Hastings.

Algorithme de Gibbs

Pour cet algorithme itératif, nous devons déterminer les distributions conditionnelles de chaque paramètre en fonction du restant des paramètres et des données.

1. Initialiser $\beta^0 = (\beta_1^0, \beta_2^0)$ et $t = 1$.
2. Générer β_1^t à partir de la distribution conditionnelle $p(\beta_1|\beta_2^{t-1})$.
3. Générer β_2^t à partir de la distribution conditionnelle $p(\beta_2|\beta_1^t)$.
4. Répéter les étapes 2 à 3 pour obtenir l'échantillon du vecteur des paramètres.

Algorithme de Metropolis-Hastings

La différence entre l'algorithme de Gibbs et de Metropolis-Hastings réside dans le fait que l'algorithme de Gibbs accepte toutes les valeurs de l'échantillon du paramètre alors que l'algorithme de Metropolis-Hastings peut rejeter une valeur et considérer celle de l'itération d'avant en faisant une sélection suivant un taux d'acceptation. La distribution conditionnelle dans le cas de Gibbs est remplacée par une distribution instrumentale pour la méthode de Metropolis-Hastings. La difficulté de la méthode de Metropolis-Hastings réside dans le choix de la distribution instrumentale. Dans notre travail, le choix de cette distribution ne pose pas de problème car le logiciel Winbugs s'en occupe.

Les étapes de l'algorithme Metropolis-Hastings sont les suivantes :

1. Initialiser β_0 et $t = 1$.
2. Générer un point β à partir d'une distribution instrumentale.
3. Calculer le ratio des distributions a posteriori avec le nouveau et l'ancien point de la séquence :

$$r(\beta, \beta_{t-1}) = \frac{a \text{ posteriori}(\beta)}{a \text{ posteriori}(\beta_{t-1})} \quad (2.13)$$

4. Calculer la probabilité d'acceptation :

$$\alpha(\beta, \beta_{t-1}) = \min(r(\beta, \beta_{t-1}), 1) \quad (2.14)$$

5. Générer un point y à partir de la distribution uniforme dans l'intervalle $[0, 1]$.
6. Si $y < \alpha(\beta, \beta_{t-1})$ alors $\beta_t = \beta$ sinon $\beta_t = \beta_{t-1}$.

2.6.5 Critères de sélection des modèles

La cohérence entre les données de vraisemblance et la distribution a priori du paramètre est capitale afin d'avoir un modèle représentant fidèlement les données observées. En conséquence, nous recourons à des indicateurs de mesure de la qualité d'ajustement du modèle tels que le critère du p-value bayésien et le critère de sélection de modèles (le critère d'information de déviance DIC). Ce sont les deux critères statistiques les plus utilisés.

DIC (Deviance information criterion)

Le DIC est un critère d'information de la déviance permettant de comparer les modèles. Le modèle avec le plus petit DIC est considéré comme le meilleur modèle. Ce critère est préconisé pour les modèles hiérarchiques (Hamada et al. (2008)), il prend en considération la moyenne a posteriori et la complexité du modèle, il se calcule de la manière suivante :

$$DIC = \bar{D} + p_D \quad (2.15)$$

\bar{D} représente la moyenne de la déviance a posteriori avec $D = -2\log(\text{vraisemblance})$, il nous indique l'ajustement du modèle aux données.

p_D représente le nombre des paramètres estimés et par conséquent modélise la complexité du modèle. Il se calcule à partir de la moyenne de la déviance a posteriori à laquelle on retranche la déviance de la moyenne à posteriori.

P-value bayésien

Le p-value bayésien nous informe sur la capacité du modèle de reproduire les données observées. Il vérifie la concordance entre les données prédites, grâce à la distribution a posteriori, et les données observées. Il est basé sur la statistique de chi-carré :

$$\chi^2 = \sum \frac{(x_i - \text{moyenne}(x_i))^2}{\text{variance}(x_i)} \quad (2.16)$$

La statistique appliquée sur les valeurs observées donne :

$$X_{obs}^2 = \sum \frac{(x_i^{obs} - \text{moyenne}(x_i^{obs}))^2}{\text{variance}(x_i^{obs})} \quad (2.17)$$

La statistique appliquée sur les valeurs prédites donne :

$$X_{pred}^2 = \sum \frac{\left(x_i^{pred} - \text{moyenne}(x_i^{pred})\right)^2}{\text{variance}(x_i^{pred})} \quad (2.18)$$

Par la suite, nous comparons les deux statistiques relatives aux valeurs observées et prédites.

La probabilité d'avoir $X_{obs}^2 \leq X_{pred}^2$ correspond au p-value dont la valeur doit être proche de 0,5 ce qui correspond à la réplication des données observées par le modèle (Dezfuli et al. (2009)) .

En résumé, les critères statistiques du p-value et du DIC doivent être, respectivement, le plus proche de 0,5 et le plus petit possible afin de mesurer la performance d'un a priori par rapport à la vraisemblance. Toutefois, il ne faut pas confondre l'usage de ces deux critères de performance : le p-value bayésien permet de vérifier seul la qualité d'ajustement de l'a posteriori par rapport à la vraisemblance alors que le DIC nécessite au minimum deux distributions a priori afin de pouvoir choisir le plus petit DIC des deux. De ce fait, le p-value bayésien est un critère de la qualité d'ajustement entre la distribution a posteriori et la vraisemblance. Cependant, le DIC est un critère pour la sélection de modèles.

CHAPITRE 3 SCÉNARIOS ET MODÉLISATION DES AVIS D'EXPERTS

Ce chapitre décrit la méthodologie de construction des distributions a priori sur le paramètre d'intérêt de chaque modèle de risque de défaillance considéré : le modèle binomial et de Poisson. Les différents scénarios de modélisation des avis d'experts sont présentés et justifiés. Pour chaque scénario, la distribution sur le paramètre d'intérêt est modélisée et mise en œuvre.

3.1 Formulation d'un avis d'expert

Un expert peut formuler son avis sur un comportement probable d'un système sous forme d'un a priori en se basant sur plusieurs sources d'information. Ses sources se résument généralement aux documents du constructeur, aux programmes d'exploitation et à l'historique de maintenance. Toutefois, les interventions de maintenance peuvent avoir un effet négatif sur le comportement d'un système lorsqu'elles ne sont pas réalisées de manière parfaite (maintenance efficace).

Pour les deux cas d'études, nous considérons que le régime permanent est atteint. Les programmes d'exploitation gardent une certaine stabilité au cours du temps. Les interventions de maintenance sont efficaces : les causes à l'origine des défaillances et défauts détectés suite à un diagnostic sont identifiées et corrigées. Ces interventions n'engendrent aucun effet sur le comportement du système. Elles permettent de le remettre dans un état opérationnel selon les spécifications du constructeur et selon les normes de sécurité en vigueur. Il en résulte que l'expérience acquise en termes de connaissances du comportement d'un système est corrélée au nombre d'interventions.

De ce fait, l'usage de l'information disponible sur le nombre d'interventions au cours du cycle de vie d'un système peut servir comme moyen pour formuler un avis sur le comportement attendu d'un système. Bien entendu, un avis est une façon de percevoir une réalité, en l'occurrence du nombre de défaillances attendu, avant que cette réalité ne se concrétise.

3.2 Cycle de vie et pronostic du risque de défaillance

Le pronostic du risque est un processus qui concerne toute la durée de vie d'un système, et ce dès la mise en service jusqu'au moment de son retrait définitif à la fin de sa vie utile. Cette durée est généralement décomposée en plusieurs périodes de temps ayant la même longueur pour faciliter la gestion du cycle de vie d'un système. Cette gestion cyclique cherche à évaluer

l'état du système, au cours d'une période donnée, tenant compte de son état actuel établi après inspection et de son historique de maintenance dans le but de prévoir les interventions au cours de la ou des périodes subséquentes. Ces interventions doivent assurer la pérennité du système sur une longue période de temps. Ce processus s'arrête lorsque le système devient non sécuritaire à l'usage, les durées de réparation sont relativement longues ou que les coûts de maintenance dépassent le coût de remplacement du système.

Ainsi, le processus de pronostic du risque doit s'arrimer avec la gestion cyclique du système. L'évaluation de risque de défaillance période par période devient nécessaire. La figure 3.1 présente le cycle de vie d'un système décomposé en plusieurs périodes, les données a priori disponibles pour chaque période de temps et la notation des fonctions de pronostic de risque avant et après une période donnée du cycle de vie.

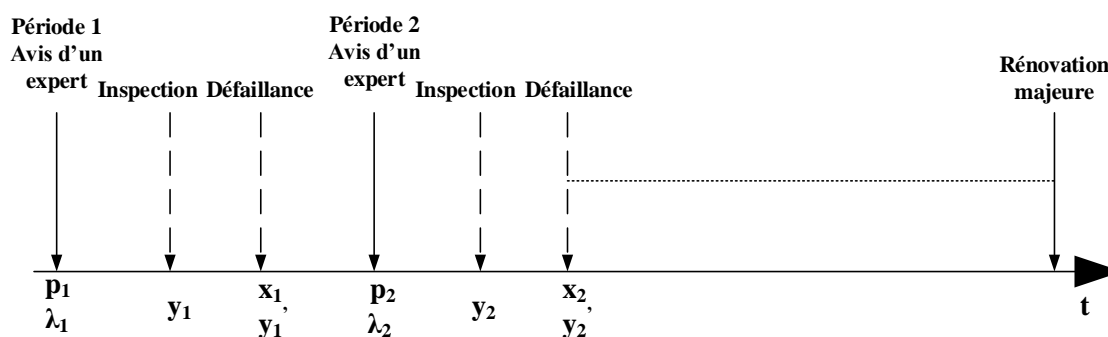


Figure 3.1 Cycle de vie et avis d'experts

Pour une flotte de systèmes identiques, le pronostic du risque devient un élément très important dans la prise de décision en matière d'investissements en maintenance ou de correction d'une dérive négative par rapport à la performance attendue d'une flotte. Par conséquent, les différentes façons de considérer les données disponibles sur les périodes antérieures du cycle de vie dans le pronostic du risque de défaillance pour la période subséquente permettent de construire les différents scénarios relatant ainsi les avis a priori sur les paramètres d'intérêt des modèles de risque de défaillance.

La figure 3.2 présente les trois scénarios considérés. Ils sont décrits dans la section suivante.

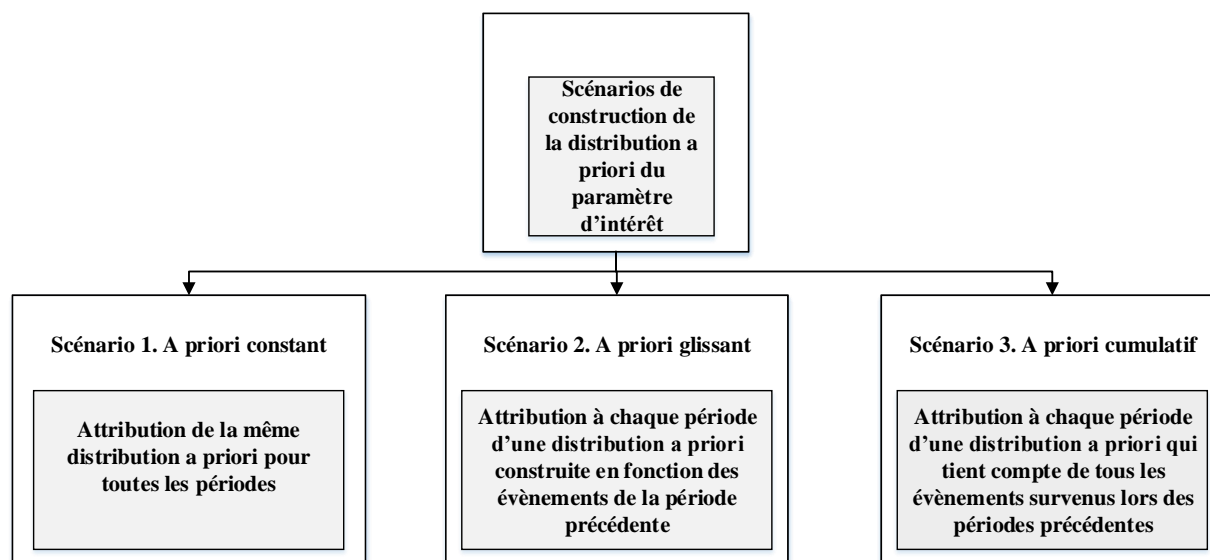


Figure 3.2 Diagramme des scénarios

3.3 Description des scénarios

Compte tenu du processus de gestion du cycle de vie par période, deux variables aléatoires sont réalisées à chaque période (i) : le nombre de défaillances ($X_i=x_i$) et le nombre de défauts ($Y_i=y_i$). L'idée d'un scénario est de pouvoir exploiter cette information disponible pour prédire le nombre de défaillances de la période subséquente ($i+1$). Ainsi trois scénarios sont distingués.

Scénario 1. A priori constant d'une période à l'autre

Dans ce scénario, l'expérience acquise sur toutes les périodes antérieures à une période donnée sert à construire un avis unique a priori sur le nombre de défaillances pour toutes les périodes subséquentes.

Ce scénario exprime un avis a priori sur le paramètre d'intérêt pour une période donnée qui ne change pas d'une période à l'autre. L'expert se sert de son expérience acquise à travers toutes ses interventions précédentes de maintenance pour formuler son avis sur la distribution a priori sur le paramètre d'intérêt d'une période à l'autre tout au long du cycle de vie. Dans ce scénario, l'expert considère que toutes les interventions de maintenance antérieures lui

permettent de se faire une idée sur le nombre de défaillances attendu durant la période subséquente.

Ainsi, l'avis d'expert traduit un comportement du risque de défaillance qui ne subit pas l'effet de l'augmentation ou la diminution des interventions de maintenance. De ce fait, l'hypothèse prise par l'expert consiste à dire qu'une certaine stabilité comportementale des processus d'intervention en maintenance est atteinte. Ces interventions ne viennent pas influencer la distribution a priori sur le paramètre d'intérêt pour les périodes subséquentes (figure 3.3)

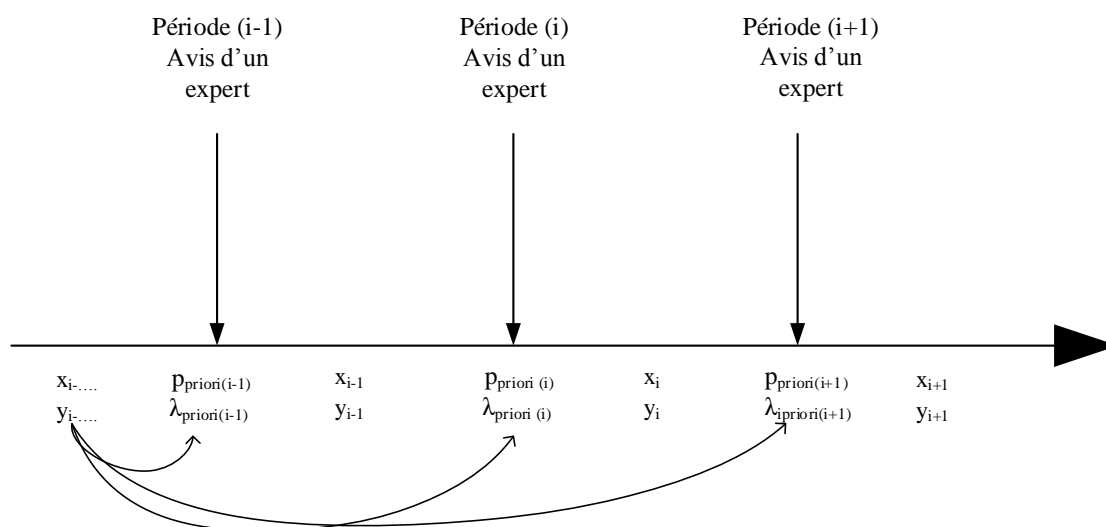


Figure 3.3 Scénario 1. A priori constant

Scénario 2. A priori adaptatif d'une période à l'autre

Dans ce scénario, juste l'expérience acquise à la période précédente sert à construire l'avis a priori sur le nombre de défaillances de la période en cours.

Ce scénario traduit la situation d'un expert qui se sert uniquement de son expérience acquise à travers ses interventions au cours de la période antérieure pour construire son avis pour la période en cours. Ainsi, l'expert réajuste son avis d'une période à l'autre. Dans ce scénario, l'expert considère que seules les interventions de maintenance réalisées à la période antérieure influenceront le pronostic du nombre de défaillances au cours de la période en cours.

L'hypothèse de l'expert est basée sur le principe markovien : son avis sur la distribution a priori du paramètre d'intérêt ne dépend que du nombre d'interventions de maintenance réalisées au cours de la période précédente. Les interventions de maintenance précédant la

période antérieure n'ont aucune influence sur le comportement du nombre de défaillances de la période en cours. L'expert peut avoir une perception pessimiste au fait que le nombre de défaillances de la période en cours suivra la même tendance que la période antérieure.

Ainsi, l'avis d'expert est réévalué d'une période à l'autre sur la base de son expérience acquise au cours de la période précédente. L'augmentation ou la diminution des nombres de défaillances et de défauts enregistrés viennent influencer d'une période à l'autre la construction de la distribution a priori sur le paramètre d'intérêt (figure 3.4).

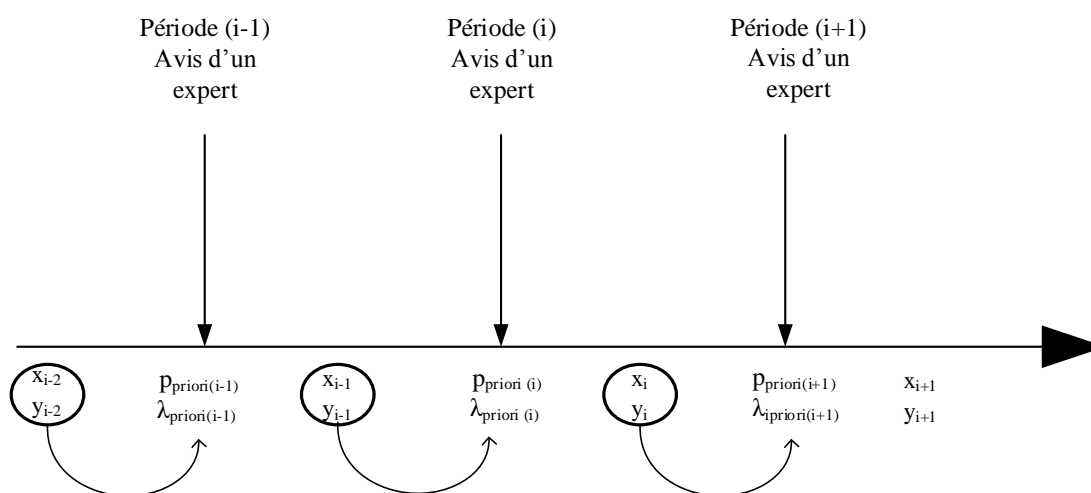


Figure 3.4 Scénario 2. A priori adaptatif

Scénario 3. A priori cumulatif d'une période à l'autre

Dans ce scénario, le cumul de l'information au cours des périodes précédentes sert à construire l'avis a priori sur le nombre de défaillances de la période en cours.

Ce scénario traduit la situation d'un expert qui se sert de son expérience cumulative acquise, période après période, pour formuler son avis sur le paramètre d'intérêt pour la période en cours. L'hypothèse prise par l'expert consiste à dire que le pronostic du nombre de défaillances pour une période donnée n'est pas seulement le résultat des interventions de maintenance durant la période précédente mais également de celles remontant à toutes les autres périodes.

Ainsi, l'expert attribue à chaque nouvelle période à l'étude un a priori qui se base sur le nombre moyen d'interventions de maintenance enregistrées pendant toutes les périodes qui précèdent (figure 3.5).

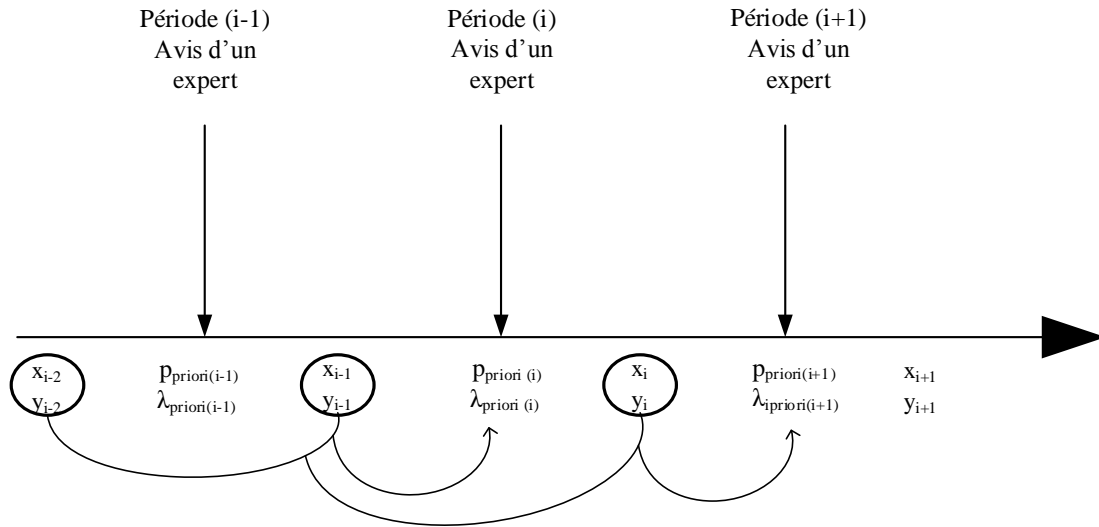


Figure 3.5 Scénario 3. A priori cumulatif

3.4 Modélisation des scénarios

3.4.1 Application des scénarios sur le modèle binomial

Pour appliquer les scénarios sur le modèle de risque binomial, les variables suivantes sont définies :

- i : périodes constantes de plusieurs mois $i = 1, 2, \dots$
- N : nombre de systèmes en marche de la flotte pendant la période i .
- T : durée totale de fonctionnement en heures de tous les systèmes de la flotte.
- t_i : valeur observée de T à la période i .
- Y : nombre de défauts (diagnostic, inspection).
- y_i : valeur observée de Y à la période i .
- X : nombre de systèmes défaillants parmi N .
- x_i : valeur observée de X à la période i .
- p : probabilité de défaillance du système.
- $X \sim Bin(N, p)$: distribution binomiale de paramètres N et p .
- $p \sim Beta(\alpha, \beta)$: distribution a priori de p de forme bêta avec les paramètres α et β .

Nous supposons que $X \sim Bin(N, p)$ et que $p \sim Beta(\alpha, \beta)$.

Scénario 1 - modèle binomial

La même distribution a priori est attribuée au paramètre p au début de chaque période. Cet a priori est actualisé avec les données de défaillances de chaque nouvelle période.

Nous nous situons à la période i :

$$\left\{ \begin{array}{l} p_{\text{apriori}}(i) \sim \text{Beta}(\alpha, \beta) \\ p_{\text{apost}}(i) \sim \text{Beta}(\alpha + x_i, \beta + N - x_i) \\ p_{\text{apriori}}(i + 1) \sim \text{Beta}(\alpha, \beta) \\ p_{\text{apost}}(i + 1) \sim \text{Beta}(\alpha + x_{i+1}, \beta + N - x_{i+1}) \end{array} \right.$$

Scénario 2 - modèle binomial

À chaque début de période, les paramètres de la distribution a priori du paramètre p sont déterminés à partir de la moyenne des défauts et de défaillances de la période précédente et actualisés avec les données de défaillance de la période en cours.

Nous nous situons à la période i :

$$\left\{ \begin{array}{l} p_{\text{apriori}}(i) \sim \text{Beta}(\alpha, \beta) \\ p_{\text{apost}}(i) \sim \text{Beta}(\alpha + x_i, \beta + N - x_i) \\ p_{\text{apriori}}(i + 1) \sim \text{Beta}((y_i + x_i)/2, N - (y_i + x_i)/2) \\ p_{\text{apost}}(i + 1) \sim \text{Beta}(((y_i + x_i)/2) + x_{i+1}, 2N - ((y_i + x_i)/2) - x_{i+1}) \\ p_{\text{apriori}}(i + 2) \sim \text{Beta}((y_{i+1} + x_{i+1})/2, N - ((y_{i+1} + x_{i+1})/2)) \\ p_{\text{apost}}(i + 2) \sim \text{Beta}(((y_{i+1} + x_{i+1})/2) + x_{i+2}, 2N - ((y_{i+1} + x_{i+1})/2) - x_{i+2}) \end{array} \right.$$

Scénario 3 - modèle binomial

À chaque début de période, les paramètres de la distribution a priori du paramètre p sont déterminés à partir de la moyenne des nombres de défauts et de défaillance des périodes précédentes et actualisés avec les données de défaillance de la période en cours.

Nous nous situons à la période i :

$$\left\{ \begin{array}{l} p_{apriori}(i) \sim Beta(\alpha, \beta) \\ p_{apost}(i) \sim Beta(\alpha + x_i, \beta + N - x_i) \\ p_{apriori}(i + 1) \sim Beta((y_i + x_i)/2, N - (y_i + x_i)/2) \\ p_{apost}(i + 1) \sim Beta(((y_i + x_i)/2) + x_{i+1}, 2N - ((y_i + x_i)/2) - x_{i+1}) \\ p_{apriori}(i + 2) \sim Beta((y_i + x_i + y_{i+1} + x_{i+1})/4, N - (y_i + x_i + y_{i+1} + x_{i+1})/4) \\ p_{apost}(i + 2) \sim Beta((y_i + x_i + y_{i+1} + x_{i+1})/4 + x_{i+2}, 2N - ((y_i + x_i + y_{i+1} + x_{i+1})/4) - x_{i+2}) \end{array} \right.$$

3.4.2 Application des scénarios sur le modèle de Poisson

Pour appliquer les scénarios sur le modèle de Poisson, les variables suivantes sont définies :

- λ_i : nombre moyen de défauts par heure.
- T : durée totale de fonctionnement en heures de tous les systèmes de la flotte.
- t_i : valeur observée de T à la période i .
- X : nombre de défaillances pour le nombre de systèmes en marche de la flotte.
- Y : nombre de défauts (diagnostic, inspection).
- Z : nombre de défauts (transformation de Y sur un intervalle prescrit).
- x_i : valeur observée de X à la période i .
- y_i : valeur observée de Y à la période i .
- z_i : valeur observée de Z à la période i .

Nous supposons que $X \sim poisson(\lambda t)$ et que $\lambda \sim Gamma(\alpha, \beta)$.

Scénario 1 - modèle de Poisson

La même distribution a priori est donnée au paramètre λ au début de chaque période, cet a priori est actualisé avec les données de défaillance de chaque nouvelle période.

Nous nous situons à la période i :

$$\left\{ \begin{array}{l} \lambda_{apriori}(i) \sim Gamma(\alpha, \beta) \\ \lambda_{apost}(i) \sim Gamma(\alpha + x_i, \beta + t_i) \\ \lambda_{apriori}(i + 1) \sim Gamma(\alpha, \beta) \\ \lambda_{apost}(i + 1) \sim Gamma(\alpha + x_{i+1}, \beta + t_{i+1}) \end{array} \right.$$

Scénario 2 - modèle de Poisson

Au début de chaque période, les paramètres de la distribution a priori du paramètre λ sont déterminés à partir de la moyenne des données de diagnostic et de défaillances de la période précédente et actualisée avec les données de défaillance de la période en cours.

Nous nous situons à la période i :

$$\left\{ \begin{array}{l} \lambda_{apriori}(i) \sim Gamma(\alpha, \beta) \\ \lambda_{apost}(i) \sim Gamma(\alpha + x_i, \beta + t_i) \\ \lambda_{apriori}(i + 1) \sim Gamma((y_i + x_i)/2, t_i) \\ \lambda_{apost}(i + 1) \sim Gamma((y_i + x_i)/2 + x_{i+1}, t_i + t_{i+1}) \\ \lambda_{apriori}(i + 2) \sim Gamma((y_{i+1} + x_{i+1})/2, t_{i+1}) \\ \lambda_{apost}(i + 2) \sim Gamma((y_{i+1} + x_{i+1})/2 + x_{i+2}, t_{i+1} + t_{i+2}) \end{array} \right.$$

Scénario 3 - modèle de Poisson

À chaque début de période, les paramètres de la distribution a priori du paramètre λ sont déterminés à partir de la moyenne des données de diagnostic et de défaillances des périodes précédentes et actualisés avec les données de défaillance de la période en cours.

Nous nous situons à la période i :

$$\left\{ \begin{array}{l} \lambda_{apriori}(i) \sim Gamma(\alpha, \beta) \\ \lambda_{apost}(i) \sim Gamma(\alpha + x_i, \beta + t_i) \\ \lambda_{apriori}(i + 1) \sim Gamma((y_i + x_i)/2, t_i) \\ \lambda_{apost}(i + 1) \sim Gamma((y_i + x_i)/2 + x_{i+1}, t_i + t_{i+1}) \\ \lambda_{apriori}(i + 2) \sim Gamma((y_i + x_i + y_{i+1} + x_{i+1})/4, t_{i+1}) \\ \lambda_{apost}(i + 2) \sim Gamma(((y_i + x_i + y_{i+1} + x_{i+1})/4) + x_{i+2}, t_{i+1} + t_{i+2}) \end{array} \right.$$

3.5 Construction des distributions a priori

3.5.1 Distributions a priori de départ

Afin de pouvoir comparer la contribution de chaque scénario de construction de l'a priori dans l'estimation du risque de défaillance a posteriori, nous avons choisi d'avoir les mêmes distributions a priori de départ pour les trois scénarios. Cela permettra de mieux apprécier l'apport des données récentes dans la précision du risque a postérieur.

Pour ce faire, nous avons utilisé les données de défaillances et de défauts de fonctionnement des trois (3) premières périodes de l'historique pour déterminer les paramètres de la distribution a priori de la période initiale (tableaux 3.1 et 3.2). Cette distribution de départ est la même pour la période 4 de chacun des trois (3) scénarios.

Tableau 3.1 Données de construction de la distribution a priori de départ : BDD-1

i	X	Y	T	N
Période	Nb de déf.(IFSD)	Nb de défauts	Durée d'exposition (h)	Nb de moteurs
1	2	3	3820165	14800
2	5	2	4089164	14800
3	2	1	3658563	14800

Les distributions a priori de départ des paramètres des deux modèles binomial et de Poisson pour la base de données des moteurs d'aéronefs sont :

$$p_{\text{départ}} \sim \text{b\^eta}(3, 14797)$$

$$\lambda_{\text{départ}} \sim \text{gamma}(3, 3855964)$$

Tableau 3.2 Données de construction de la distribution a priori de départ : BDD-2

i	X	Y	T	N
Période	Nb de déf.(IFSD)	Nb de défauts	Durée d'exposition (h)	Nb de moteurs
1	2	10	2804	5000
2	3	14	2959	5000
3	9	18	2829	5000

Les distributions a priori de départ des paramètres des deux modèles binomial et de Poisson pour la base de données des turbines à gaz sont :

$$p_{\text{départ}} \sim \text{beta}(10, 4990)$$

$$\lambda_{\text{départ}} \sim \text{gamma}(10, 2864)$$

3.5.2 Distributions a priori selon le scénario pour le modèle binomial

Scénario 1 - modèle binomial

Ce scénario 1 considère une distribution a priori inchangée d'une période à l'autre. De ce fait, pour l'échantillon BDD-1, la distribution a priori sur la probabilité de défaillance p suit une loi bêta de paramètres : 3 et 14797. Ces paramètres restent les mêmes durant toutes les périodes de 4 à 7.

Pour l'échantillon BDD-2, les paramètres de la loi bêta sont : 10 et 4990. Ces paramètres sont maintenus pour toutes les périodes de 4 à 10.

Scénario 2 - modèle binomial

Ce scénario considère un a priori à mémoire courte, ne tenant compte que des nombres de défaillances et de défauts de fonctionnement récents au cours de la dernière période d'observation. Ainsi, nous avons autant de distributions a priori sur le paramètre d'intérêt que de périodes d'estimation. Les tableaux 3.3 et 3.4 donnent les différentes distributions période par période pour les échantillons BDD-1 et BDD-2.

Tableau 3.3 Paramètres des distributions a priori pour le scénario 2 : BDD-1

Période i	α	β
4	2	14798
5	4	14796
6	4	14796
7	9	14791

Tableau 3.4 Paramètres des distributions a priori pour le scénario 2 : BDD-2

Période i	α	β
4	10	4990
5	6	4994
6	8	4992
7	3	4997
8	8	4992
9	19	4981
10	6	4994

Scénario 3 - modèle binomial

Ce scénario considère un a priori ayant une mémoire cumulative qui tient compte de toutes les défaillances et défauts de fonctionnement enregistrés durant toutes les périodes d'observation antérieures. Ainsi, nous avons à chaque fois une distribution a priori sur la probabilité de défaillance qui cumule l'information de toutes les périodes antérieures. Les paramètres des distributions a priori pour les échantillons BDD-1 et BDD-2 à chaque période d'estimation sont récapitulés respectivement dans les tableaux 3.5 et 3.6.

Tableau 3.5 Paramètres des distributions a priori pour le scénario 3 : BDD-1

Période i	α	β
4	3	14797
5	3	14797
6	3	14797
7	4	14796

Tableau 3.6 Paramètres des distributions a priori pour le scénario 3 : BDD-2

Période i	α	β
4	10	4990
5	9	4991
6	9	4991
7	8	4992
8	8	4992
9	9	4991
10	9	4991

3.5.3 Distribution a priori selon le scénario pour le modèle de Poisson

Scénario 1 - modèle de Poisson

Pour le modèle de Poisson, la distribution du taux de défaillance a priori est donnée par la loi gamma dont les paramètres sont : 3 et 3855964 pour l'échantillon BDD-1, et 10 et 2864 pour l'échantillon BDD-2. Ces paramètres restent les mêmes durant toutes les périodes, respectivement, de 4 à 7 pour BDD-1 et de 4 à 10 pour BDD-2.

Scénario 2- modèle de Poisson

Les tableaux 3.7 et 3.8 donnent les différentes distributions de la loi gamma, d'une période à l'autre, pour les échantillons BDD-1 et BDD-2.

Tableau 3.7 Paramètres des distributions a priori pour le scénario 2 : BDD-1

Période i	α	β
4	3	3855964
5	4	4216473
6	4	4502048
7	9	4088984

Tableau 3.8 Paramètres des distributions a priori pour le scénario 2 : BDD-2

Période i	α	β
4	10	2864
5	6	4763
6	8	5896
7	3	6940
8	8	4929
9	19	8804
10	6	1667

Scénario 3- modèle de Poisson

Les tableaux 3.9 et 3.10 donnent les différentes distributions de la loi gamma considérant le cumul de la connaissance accumulée sur toutes les périodes antérieures à la période actuelle pour les échantillons BDD-1 et BDD-2.

Tableau 3.9 Paramètres des distributions a priori pour le scénario 3 : BDD-1

Période i	α	β
4	3	3855964
5	3	3946091,25
6	3	4057282,6
7	4	4062566,17

Tableau 3.10 Paramètres des distributions a priori pour le scénario 3 : BDD-2

Période i	α	β
4	10	2864
5	9	3338,75
6	9	3850,2
7	8	4365,17
8	8	4445,71
9	9	4990,5
10	9	4621,22

3.6 Prise en compte de la certitude d'un avis d'expert

Dans cette section, nous considérons la certitude d'un avis d'expert lors de la formulation de son a priori sur l'un des paramètres d'intérêts à l'étude : la probabilité de défaillance p de la loi binomiale et le taux de défaillance de la loi de Poisson λ . Pour ce faire, nous avons choisi la loi bêta comme distribution a priori pour p et pour λ . Ce choix est motivé par la flexibilité de cette loi, son support, normalement entre 0 et 1 peut être modifié pour prendre n'importe quels autres valeurs entre 0 et 1.

Pour illustrer la certitude sur un avis a priori d'un expert, le modèle binomial est utilisé avec la loi bêta comme distribution a priori sur le paramètre p . Pour le modèle de Poisson, il suffit juste de substituer p par λ dans la démarche de calcul.

Afin de prendre en compte la certitude d'un avis d'expert sur le paramètre p , le support de loi bêta définissant la distribution a priori sur le paramètre d'intérêt ne sera plus entre 0 et 1, mais entre deux valeurs définies par l'expert. Ces deux valeurs encadrent de part et d'autre la formulation de son avis a priori sur le paramètre d'intérêt. Nous désignons par p_1 et p_2 les valeurs minimale et maximale du paramètre d'intérêt p , définissant le nouveau support de la distribution bêta.

Soit c la moyenne des défauts et des défaillances pendant toutes les périodes passées et L_1 et L_2 les deux bornes de la probabilité $F_X(c)$ définissant la certitude de l'expert par rapport au nombre de défaillances X à la prochaine période, F étant la fonction de répartition de la loi binomiale. Ainsi, nous allons construire une distribution a priori de la probabilité p de la façon suivante :

$$L_1 \leq P(X \leq c) \leq L_2 \quad (3.1)$$

Le support de la distribution a priori de p , noté $[p_1, p_2]$, est déterminé en résolvant l'inéquation (3.1).

Notons par Y la variable aléatoire qui modélise la probabilité de défaillance p . Nous supposons que $Y \sim \text{Beta}(\alpha, \beta)$. Sa distribution s'écrit comme suit :

$$f_Y(y) = \frac{y^{\alpha-1} \cdot (1-y)^{\beta-1}}{\text{beta}(\alpha, \beta)} \cdot \mathbb{1}_{[0,1]}(y) \quad (3.2)$$

Ainsi, le problème revient à déterminer une distribution pour la nouvelle variable aléatoire Z ayant pour support $[p_1, p_2]$ à partir de la variable initiale Y . La variable Z est obtenue à partir de Y en faisant la transformation linéaire suivante :

$$Z = p_1 + (p_2 - p_1) \cdot Y = H(Y) \quad (3.3)$$

Comme la fonction H est monotone sur l'intervalle $[0, 1]$, la fonction de densité de Z est déduite directement à partir de celle de Y . Elle est donnée comme suit :

$$f_Z(z) = \frac{\left(\frac{z-p_1}{p_2-p_1}\right)^{\alpha-1} \cdot \left(1 - \left(\frac{z-p_1}{p_2-p_1}\right)\right)^{\beta-1}}{(p_2 - p_1) \cdot \text{beta}(\alpha, \beta)} \mathbb{1}_{[p_1, p_2]}(z) \quad (3.4)$$

Ainsi, la moyenne et la variance de Z sont données par le système d'équations suivant :

$$\begin{cases} E[Z] = p_1 + (p_2 - p_1) \cdot E[Y] \\ V[Z] = (p_2 - p_1)^2 \cdot V[Y] \end{cases}$$

Les valeurs de la moyenne $E[Z]$ et de l'écart-type $V[Z]$ de la variable Z sont données par l'algorithme Metropolis-Hastings. De ce fait, les valeurs sont parfaitement déterminées et peuvent être utilisées dans la formulation de l'a priori défini par la loi bêta.

CHAPITRE 4 RÉSULTATS ET ANALYSE DES SCÉNARIOS

Ce chapitre présente les résultats de différents scénarios. Une comparaison de ces derniers est effectuée à l'aide du test de Friedman. Par la suite, une interprétation des résultats est effectuée. Enfin, le calcul du risque de défaillance et une analyse de la certitude des experts sont réalisés.

4.1 Distributions a posteriori pour le modèle binomial

4.1.1 Distribution a posteriori selon le scénario 1

La distribution a posteriori est complètement déterminée par sa moyenne et son écart-type. Les tableaux 4.1 et 4.2 présentent, respectivement, les résultats a posteriori de la probabilité de défaillance selon le modèle binomial, pour BDD-1 et BDD-2. Rappelons que le critère statistique du p-value permet à lui seul de mesurer la qualité d'un a priori par rapport à la vraisemblance lorsqu'il est proche de 0.5. Quant au critère du DIC, il ne peut pas être utilisé que pour comparer deux ou plusieurs modèles. Dans ce cas, le DIC du modèle le plus faible donne le plus proche a posteriori de la vraisemblance. Dans les résultats du tableau 4.1 (scénario 1 – modèle binomial – BDD-1), le p-value est proche de 0,5 pour toutes les périodes laissant conclure que le modèle binomial ajuste bien les données et que la distribution a priori est bonne. Par contre, selon les p-values de la BDD-2 (tableau 4.2), aucun ne sort de l'intervalle $[0.05, 0.95]$ et par conséquent le modèle de prédiction est bon pour l'ensemble des périodes.

Tableau 4.1 Résultats du scénario 1 - modèle binomial : BDD-1

Période i	Moyenne	Écart type	p-value	DIC
4	1,689E-04	7,625E-05	0,7181	3,55
5	1,689E-04	7,625E-05	0,7181	3,55
6	3,373E-04	1,068eE-04	0,4265	5,954
7	2,355E-04	8,884E-05	0,6317	4,483

Tableau 4.2 Résultats du scénario 1- modèle binomial : BDD-2

Période i	Moyenne	Écart type	p-value	DIC
4	1,5E-03	3,868E-04	0,4851	5,043
5	1,5E-03	3,868E-04	0,4851	5,043
6	1,1E-03	3,313E-04	0,1017	7,791
7	1,4E-03	3,735E-04	0,3895	5,298
8	2E-03	4,436E-04	0,6572	5,155
9	1,198E-03	3,47E-04	0,1856	6,463
10	1,603E-03	3,985E-04	0,5581	5,001

4.1.2 Distribution a posteriori selon le scénario 2

Les tableaux 4.3 et 4.4 donnent la moyenne et l'écart-type de la distribution a posteriori (scénario 2) de la probabilité de défaillance pour chaque période, respectivement pour BDD-1 et BDD-2. Les résultats du tableaux 4.3 montrent que les a priori sont informatifs et avec une bonne qualité d'ajustement de l'a posteriori par rapport à la vraisemblance. Dans les résultats du tableau 4.4 (scénario 2 – modèle binomial – BDD-2), le p-value bayésien de la période 9 est faible, soit de 0,01 : l'une des hypothèses du modèle binomiale est violée ou bien la distribution a priori n'est pas bonne.

Tableau 4.3 Résultats du scénario 2 - modèle binomial : BDD-1

Période i	Moyenne	Écart type	p-value	DIC
4	1,689E-04	7,625E-05	0,7181	3,55
5	2,025E-04	8,291E-05	0,6661	3,698
6	3,715E-04	1,116E-04	0,515	5,491
7	4,395E-04	1,218E-04	0,4654	5,074

Tableau 4.4 Résultats du scénario 2 - modèle binomial : BDD-2

Période i	Moyenne	Écart type	p-value	DIC
4	1,5E-03	3,868E-04	0,4851	5,043
5	1,1E-03	3,313E-04	0,6784	4,495
6	8,996E-04	2,972E-04	0,1794	6,230
7	6,974E-04	2,615E-04	0,6284	4,510
8	1,798E-03	4,215E-04	0,6022	5,381
9	2,099E-03	4,571E-04	0,01747	13,128
10	1,198E-03	3,47E-04	0,6674	4,767

4.1.3 Distribution a posteriori selon le scénario 3

Les tableaux 4.5 et 4.6 donnent la moyenne et l'écart-type de la distribution a posteriori de la probabilité de défaillance pour chaque période, respectivement pour BDD-1 et BDD-2 selon le scénario 3. Les résultats des p-values démontrent, pour BDD-1 et BDD-2, une bonne répliation des données observées par le modèle. .

Tableau 4.5 Résultats du scénario 3 - modèle binomial : BDD-1

Période i	Moyenne	Écart type	p-value	DIC
4	1,689E-04	7,625E-05	0,7181	3,55
5	1,689E-04	7,625E-05	0,7181	3,55
6	3,373E-04	1,068E-04	0,4265	5,954
7	2,7E-04	9,568E-05	0,6757	4,382

Tableau 4.6 Résultats du scénario 3 - modèle binomial : BDD-2

Période i	Moyenne	Écart type	p-value	DIC
4	1,5E-03	3,868E-04	0,4851	5,043
5	1,4E-03	3,735E-04	0,5483	4,811
6	1,001E-03	3,166E-04	0,1347	7,015
7	1,198E-03	3,47E-04	0,5344	4,698
8	1,798E-03	4,215E-04	0,6022	5,381
9	1,1E-03	3,313E-04	0,239	5,959
10	1,5E-03	3,868E-04	0,617	4,749

4.2 Distributions a posteriori pour le modèle de Poisson

4.2.1 Distribution a posteriori selon le scénario 1

Les tableaux 4.7 et 4.8 présentent les résultats de la moyenne et de l'écart-type du taux de défaillance a posteriori (scénario 1) de la probabilité de défaillance pour chaque période, respectivement pour BDD-1 et BDD-2. Les résultats du tableaux 4.7 montrent des valeurs de p-value bayésien variant entre 0,45 et 0,70. Cela représente des résultats acceptables. Toutefois, le p-value de la période 6 du tableau 4.8 est inférieur à 0,05, engendrant une sous prédiction du nombre de défaillance par le modèle.

Tableau 4.7 Résultats du scénario 1 - modèle de Poisson : BDD-1

Période i	Moyenne $\lambda_i T_i$	Écart type $\lambda_i T_i$	Moyenne λ_i	Écart type λ_i	p-value	DIC
4	2,603	1,163	6,174E-07	2,759E-07	0,7039	3,622
5	2,682	1,195	5,958E-07	2,654E-07	0,6969	3,652
6	5,147	1,631	12,59E-07	3,989E-07	0,4502	5,935
7	3,706	1,397	8,523E-07	3,213E-07	0,6438	4,531

Tableau 4.8 Résultats du scénario 1 - modèle de Poisson : BDD-2

Période i	Moyenne $\lambda_i T_i$	Écart type $\lambda_i T_i$	Moyenne λ_i	Écart type λ_i	p-value	DIC
4	9,394	2,431	1,972E-03	5,103E-04	0,2725	6,551
5	10,13	2,616	1,718E-03	4,436E-04	0,2138	7,480
6	7,797	2,351	1,124E-03	3,388E-04	0,0375	11,942
7	8,851	2,375	1,796E-03	4,818E-04	0,2034	7,114
8	15,08	3,399	1,713E-03	3,86E-04	0,3213	7,258
9	4,416	1,274	2,649E-03	7,645E-04	0,3844	4,579
10	4,147	1,04	4,147E-03	10,4E-04	0,4371	5,182

4.2.2 Distribution a posteriori selon le scénario 2

Les tableaux 4.9 et 4.10 donnent la moyenne et l'écart-type de la distribution a posteriori du taux de défaillance pour chaque période, respectivement pour BDD-1 et BDD-2 selon le scénario 2. Les p-values bayésiens varient entre 0,15 et 0,7. Les modèles de prédiction sont bons pour l'ensemble des périodes.

Tableau 4.9 Résultats du scénario 2 - modèle de Poisson : BDD-1

Période i	Moyenne $\lambda_i T_i$	Écart type $\lambda_i T_i$	Moyenne λ_i	Écart type λ_i	p-value	DIC
4	2,603	1,163	6,174E-07	2,759E-07	0,7039	3,622
5	3,097	1,261	6,878E-07	2,802E-07	0,6473	3,722
6	5,236	1,585	12,8E-07	3,876E-07	0,4703	5,747
7	6,701	1,871	15,41E-07	4,302E-07	0,4348	5,083

Tableau 4.10 Résultats du scénario 2 - modèle de Poisson : BDD-2

Période i	Moyenne $\lambda_i T_i$	Écart type $\lambda_i T_i$	Moyenne λ_i	Écart type λ_i	p-value	DIC
4	9,394	2,431	1,972E-03	5,103E-04	0,2725	6,551
5	6,089	1,848	1,033E-03	3,135E-04	0,6399	4,557
6	4,864	1,627	0,7009E-03	2,345E-04	0,1548	6,868
7	2,903	1,093	0,5889E-03	2,217E-04	0,5245	4,827
8	11,55	2,727	1,312E-03	3,098E-04	0,591	5,534
9	3,339	0,7277	2,003E-03	4,365E-04	0,5998	3,457
10	4,501	1,3	4,501E-03	13E-04	0,5095	5,142

4.2.3 Distribution a posteriori selon le scénario 3

Les tableaux 4.11 et 4.12 donnent la moyenne et l'écart-type de la distribution a posteriori du taux de défaillance pour chaque période, respectivement pour BDD-1 et BDD-2 selon le scénario 3. Les p-values bayésiennes sont proches de 0,5 pour BDD-1. Toutefois, la période 6 a un p-value très faible de 0,07 : l'avis de l'expert ne concorde pas avec la vraisemblance, l'a priori est très mauvais pour la période 6.

Tableau 4.11 Résultats du scénario 3 - modèle de Poisson : BDD-1

Période i	Moyenne $\lambda_i T_i$	Écart type $\lambda_i T_i$	Moyenne λ_i	Écart type λ_i	p-value	DIC
4	2,603	1,163	6,174E-07	2,759E-07	0,7039	3,622
5	2,656	1,185	5,899E-07	2,631E-07	0,6985	3,633
6	5,01	1,594	12,25E-07	3,898E-07	0,4288	5,967
7	4,128	1,457	9,493E-07	3,351E-07	0,6798	4,235

Tableau 4.12 Résultats du scénario 3 - modèle de Poisson : BDD-2

Période i	Moyenne $\lambda_i T_i$	Écart type $\lambda_i T_i$	Moyenne λ_i	Écart type λ_i	p-value	DIC
4	9,394	2,431	1,972E-03	5,103E-04	0,2725	6,551
5	6,503	1,765	2,245E-03	6,094E-04	0,618	4,554
6	6,422	2,042	0,9254E-03	2,943E-04	0,071	9,393
7	6,363	1,856	1,291E-03	3,765E-04	0,484	4,981
8	11,98	2,838	1,361E-03	3,224E-04	0,5614	5,609
9	2,75	0,8342	1,65E-03	5,004E-04	0,7544	3,198
10	2,668	0,6917	2,668E-03	6,917E-04	0,0961	7,599

4.3 Sélection et analyse du meilleur scénario selon le modèle de risque

Le test de Friedman est utilisé pour comparer les performances des distributions a priori sur toutes les périodes de tests disponibles par rapport au DIC. Ce test est largement utilisé pour vérifier statistiquement si les scénarios sont considérés comme différents ou non par rapport aux valeurs du DIC évalués sur toutes les périodes disponibles.

Le test de Friedman considère les deux hypothèses suivantes :

H_0 : les scénarios sont identiques H_1 : les scénarios sont différents.

La statistique du test consiste à calculer :

$$X_r^2 = \frac{12}{nk(k+1)} \sum R^2 - 3n(k+1) \quad (4.1)$$

Avec :

- α : seuil de signification du test.
- K : nombre de scénarios à comparer.
- n : nombre de traitement.
- R : somme des rangs du block.

4.3.1 Sélection du meilleur scénario pour le modèle binomial

Les tableaux 4.13 et 4.14 présentent les données du test de Friedman, respectivement pour les échantillons BDD-1 et BDD-2. Ces données permettent de sélectionner le meilleur scénario pour le modèle binomial.

Tableau 4.13 Comparaison des scénarios : BDD-1

Scénario i	Période 4	Période 5	Période 6	Période 7
1	3,55	3,55	5,954	4,483
2	3,55	3,698	5,491	5,074
3	3,55	3,55	5,954	4,382

La statistique du test donne : $X_{r\text{calculé}}^2=8,78 > X_r^2=7,815$ pour $\alpha=0,05$. L'hypothèse nulle est donc rejetée et par conséquent les scénarios ne sont pas identiques (p-value=0,032).

Le scénario 3 est le scénario qui a le plus petit DIC, le sélectionnant ainsi comme le meilleur scénario.

Tableau 4.14 Comparaison des scénarios : BDD-2

Scénario i	Période 4	Période 5	Période 6	Période 7	Période 8	Période 9	Période 10
1	5,043	5,043	7,791	5,298	5,155	6,463	5,001
2	5,043	4,495	6,230	4,510	5,381	13,128	4,767
3	5,043	4,811	7,015	4,698	5,381	5,959	4,749

La statistique du test est : $X_{r_{calculé}}^2=14,48 > X_r^2=12,592$ pour $\alpha=0,05$. L'hypothèse nulle est rejetée et par conséquent les scénarios sont différents (p-value=0,024).

En calculant le DIC moyen par scénario, nous obtenons les valeurs ci-après pour BDD-1 et BDD-2 (tableaux 4.15 et 4.16).

Tableau 4.15 Calcul du DIC moyen : BDD-1

Scénario i	DIC moyen
1	4,384
2	4,453
3	4,359

Tableau 4.16 Calcul du DIC moyen : BDD-2

Scénario i	DIC moyen
1	5,684
2	6,222
3	5,379

D'après les valeurs des DIC moyens pour les deux bases de données BDD-1 et BDD-2, nous pouvons conclure que le scénario 3 est le meilleur, suivi du scénario 2 et finalement le scénario 1 en termes du plus petit DIC moyen. Il est tout à fait plausible que le scénario 3 soit le meilleur. L'information cumulative semble l'emporter par rapport à une information glissante d'une période à l'autre. Le scénario 1 ne prévoit aucune nouvelle information pour la construction d'un a priori. Il est classé en dernier lieu. Le modèle de prédiction selon le scénario 3 pour les deux bases de données est le plus représentatif des données.

4.3.2 Sélection du meilleur scénario pour le modèle de Poisson

Les tableaux 4.17 et 4.18 présentent les données du test de Friedman, respectivement pour les échantillons BDD-1 et BDD-2. Ces données permettent de sélectionner le meilleur scénario

pour le modèle de Poisson.

Tableau 4.17 Comparaison des scénarios : BDD-1

Scénario i	Période 4	Période 5	Période 6	Période 7
1	3,622	3,652	5,935	4,531
2	3,622	3,722	5,747	5,083
3	3,622	3,633	5,967	4,235

La statistique du test est : X_r^2 calculé=9 > X_r^2 =7,815 pour $\alpha=0,05$. L'hypothèse nulle est rejetée et par conséquent les scénarios ne sont pas identiques (p-value=0,029).

Tableau 4.18 Comparaison des scénarios : BDD-2

Scénario i	Période 4	Période 5	Période 6	Période 7	Période 8	Période 9	Période 10
1	6,551	7,480	11,942	7,114	7,258	4,579	5,182
2	6,551	4,557	6,868	4,827	5,534	3,457	5,142
3	6,551	4,554	9,393	4,981	5,609	3,198	7,599

La statistique du test est : X_r^2 calculé=12,71 > X_r^2 =12,592 pour $\alpha=0,05$. L'hypothèse nulle est rejetée et par conséquent les scénarios sont différents. (p-value=0,047)

En calculant le DIC moyen pour chaque scénario, nous obtenons les valeurs présentées dans les tableaux 4.19 et 4.20.

Tableau 4.19 Calcul du DIC moyen : BDD-1

Scénario i	DIC moyen
1	4,435
2	4,543
3	4,364

Tableau 4.20 Calcul du DIC moyen : BDD-2

Scénario i	DIC moyen
1	7,158
2	5,276
3	5,983

D'après les valeurs des DIC moyens, nous pouvons conclure que le scénario 3 est le meilleur en terme du plus petit DIC moyen pour la première base de données (BDD-1) alors que

le scénario 2 l'emporte pour la deuxième base de données (BDD-2). Ceci est tout a fait explicable par le fait qu'il y a une importante variation de 57% entre les durées d'exposition pour des systèmes dans BDD-2 alors qu'il s'agit uniquement de 4% pour les systèmes de BDD-1.

4.4 Estimation et pronostic du risque de défaillance

Le pronostic du risque de défaillance a été réalisé selon le meilleur scénario sélectionné pour les échantillons de données BDD-1 et BDD-2. Le risque de défaillance est estimé pour une période donnée : le risque d'avoir 0, 1, 2, etc. défaillances. Cette fonction correspond exactement à la fonction de répartition en fiabilité. Elle utilise le cumul du risque. Cependant, le pronostic du risque de maintenance calcule le risque d'une période à l'autre. Les résultats obtenus sont présentés dans les tableaux 4.21, 4.22, 4.23 et 4.24.

Tableau 4.21 Risque de défaillance - Modèle binomial : Scénario 3 BDD-1

Période i	$F_X(0)$	$F_X(1)$	$F_X(2)$	$F_X(3)$	$F_X(4)$	$F_X(5)$
4	0,082091	0,28733	0,54387	0,75765	0,89123	0,95801
5	0,082091	0,28733	0,54387	0,75765	0,89123	0,95801
6	0,067861	0,40674	0,12528	0,26610	0,44186	0,61736
7	0,018379	0,091842	0,23865	0,43423	0,62962	0,78578
Période i	$F_X(6)$	$F_X(7)$	$F_X(8)$	$F_X(9)$	$F_X(10)$	
4	0,98583	0,99576	0,99886	0,99972	0,99994	
5	0,98583	0,99576	0,99886	0,99972	0,99994	
6	0,76337	0,86749	0,93246	0,96848	0,98646	
7	0,88977	0,94913	0,97877	0,99193	0,99719	

Tableau 4.22 Risque de défaillance - Modèle binomial : Scénario 3 BDD-2

Période i	$F_X(0)$	$F_X(1)$	$F_X(2)$	$F_X(3)$	$F_X(4)$	$F_X(5)$
4	2,2461E-10	5,2184E-09	6,0729E-08	4,7207E-07	2,7580E-06	1,2920E-05
5	9,8881E-10	2,1506E-08	2,3434E-07	1,7062E-06	9,3396E-06	4,1008E-05
6	3,6542E-07	5,7845E-06	4,5963E-05	2,4455E-04	9,8063E-04	3,1632E-03
7	1,9732E-08	3,7001E-07	3,4788E-06	2,1872E-05	1,0348E-04	3,9314E-04
8	2,7092E-12	7,4933E-11	1,0376E-09	9,5903E-09	6,6580E-08	3,7034E-07
9	8,4294E-08	1,4581E-06	1,2652E-05	7,3459E-05	3,2117E-04	1,1284E-03
10	2,2461E-10	5,2184E-09	6,0729E-08	4,7207E-07	2,7580E-06	1,292E-05
Période i	$F_X(6)$	$F_X(7)$	$F_X(8)$	$F_X(9)$	$F_X(10)$	
4	5,0563E-05	1,7007E-04	5,0206E-04	1,3218E-03	3,1431E-03	
5	1,5049E-04	4,7487E-04	1,3158E-03	3,2535E-03	7,2715E-03	
6	8,5559E-03	1,9976E-02	4,1135E-02	7,5980E-02	1,2762E-01	
7	1,2498E-03	3,4215E-03	8,2381E-03	1,7733E-02	3,4578E-02	
8	1,7195E-06	6,8556E-06	2,3962E-05	7,4606E-05	2,0953E-04	
9	3,3202E-03	8,4215E-03	1,8809E-02	3,7609E-02	6,8231E-02	
10	5,0563E-05	1,7007E-04	5,0206E-04	1,3218E-03	3,1431E-03	

Tableau 4.23 Risque de défaillance - Modèle de Poisson : Scénario 3 BDD-1

Période i	$F_X(0)$	$F_X(1)$	$F_X(2)$	$F_X(3)$	$F_X(4)$	$F_X(5)$
4	0,074051	0,26681	0,51768	0,73535	0,877	0,95074
5	0,070229	0,25676	0,50446	0,72377	0,86939	0,94674
6	0,0066709	0,040092	0,12381	0,26362	0,43874	0,61421
7	0,016115	0,082638	0,21994	0,40887	0,60385	0,76482
Période i	$F_X(6)$	$F_X(7)$	$F_X(8)$	$F_X(9)$	$F_X(10)$	
4	0,98273	0,99463	0,99850	0,99962	0,99991	
5	0,98098	0,99397	0,99828	0,99956	0,99990	
6	0,76072	0,86558	0,93125	0,96781	0,98612	
7	0,87557	0,94088	0,97458	0,99003	0,99641	

Tableau 4.24 Risque de défaillance - Modèle de Poisson : Scénario 2 BDD-2

Période i	$F_X(0)$	$F_X(1)$	$F_X(2)$	$F_X(3)$	$F_X(4)$	$F_X(5)$
4	8,3222E-05	8,6501E-04	4,5371E-03	1,6035E-02	4,3040E-02	9,3775E-02
5	2,2677E-03	1,6076E-02	5,8114E-02	1,4344E-01	2,7332E-01	4,3149E-01
6	7,7195E-03	4,5267E-02	1,3658E-01	2,8464E-01	4,6467E-01	6,3981E-01
7	5,4858E-02	2,1411E-01	4,4527E-01	6,6895E-01	8,3129E-01	9,2554E-01
8	9,636E-06	1,2093E-04	7,6367E-04	3,2382E-03	1,0383E-02	2,6889E-02
9	3,5472E-02	1,5391E-01	3,5165E-01	5,7174E-01	7,5545E-01	8,7814E-01
10	1,1098E-02	6,105E-02	1,7347E-01	3,4213E-01	5,3191E-01	7,0276E-01

Période i	$F_X(6)$	$F_X(7)$	$F_X(8)$	$F_X(9)$	$F_X(10)$
4	1,7321E-01	2,7981E-01	4,0499E-01	5,3564E-01	6,5838E-01
5	5,9201E-01	7,3164E-01	8,3791E-01	9,0981E-01	9,5359E-01
6	7,8179E-01	8,8044E-01	9,4042E-01	9,7284E-01	9,8861E-01
7	9,7115E-01	9,9006E-01	9,9692E-01	9,9914E-01	9,9978E-01
8	5,8662E-02	1,1109E-01	1,8678E-01	2,8391E-01	3,961E-01
9	9,4641E-01	9,7898E-01	9,9257E-01	9,9762E-01	9,993E-01
10	8,3092E-01	9,1333E-01	9,597E-01	9,8288E-01	9,9332E-01

La figure 4.1 présente un exemple de risque de défaillance pour les périodes 4, 5, 6 et 7, BDD-1 - modèle binomial

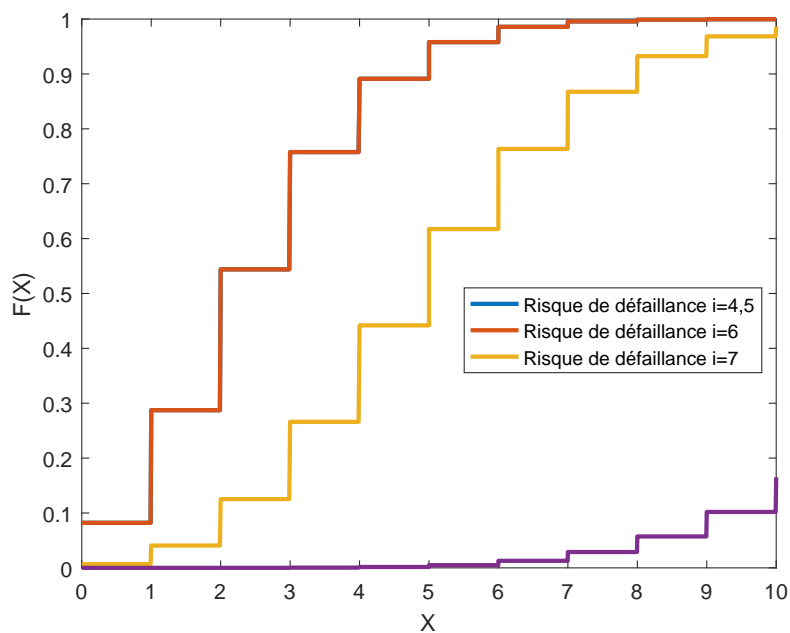


Figure 4.1 Risque de défaillance

4.5 Résultats et analyse du risque incluant la certitude de l'avis d'expert

Afin d'illustrer les résultats a posteriori, nous avons choisi de tester les trois scénarios selon le modèle binomial et le modèle de Poisson avec la même certitude de l'avis d'expert à travers toutes les périodes d'estimation. Ceci veut dire que l'expert ne change pas, d'une période à l'autre, les valeurs L_1 et L_2 donnant ses limites de confiance dans la formulation a priori du paramètre d'intérêt. Ainsi, pour tous les scénarios, les valeurs minimale et maximale du paramètre d'intérêt a priori sont déduites à partir de la fonction de répartition du modèle de risque selon le type de certitude : élevée, faible ou large.

En guise d'illustration, nous déterminerons les résultats a posteriori des paramètres d'intérêt, la probabilité de défaillance et le taux de défaillance pour seulement l'échantillon de données de la flotte de moteurs d'aéronefs BDD-1.

4.5.1 Certitude - Type 1

Dans ce type 1, l'expert pense avec une certitude "élevée" que le nombre de défaillances a priori pour la période en cours est inférieur ou égal à la moyenne des défauts et des défaillances ,noté a , de toutes les périodes précédentes.

$$0,8 \leq P(X \leq a) \leq 0,9$$

L'exécution de 31000 itérations d'une chaîne MCMC avec le scénario 3 sur l'échantillon de données BDD-2 donne les résultats suivants (tableau 4.25) :

Tableau 4.25 Résultats a posteriori : Certitude - Type 1 selon le scénario 3

Période i	Moyenne p	Écart type p	p-value	DIC
4	1,516E-03	3,776E-05	0,4056	4,479
5	1,35E-03	3,547E-05	0,5717	3,984
6	1,358E-03	3,522E-05	0,009266	10,353
7	1,184E-03	3,326E-05	0,5056	3,976
8	1,19E-03	3,273E-05	0,1143	6,453
9	1,347E-03	3,569E-05	0,07656	7,226
10	1,351E-03	3,522E-05	0,8508	3,751

4.5.2 Certitude - Type 2

Dans le type 2, l'expert pense avec une probabilité "faible" que le nombre de défaillances a priori pour la période en cours est inférieur ou égal à la moyenne des défauts et des défaillances ,noté a , de toutes les périodes précédentes.

$$0,2 \leq P(X \leq a) \leq 0,3$$

L'exécution de 31000 itérations d'une chaîne MCMC avec le scénario 3 sur l'échantillon de données BDD-2 donne les résultats suivants (tableau 4.26) :

Tableau 4.26 Résultats a posteriori : Certitude - Type 2 selon le scénario 3

Période i	Moyenne p	Écart type p	p-value	DIC
4	2,606E-03	3,906E-05	0,02773	9,968
5	2,386E-03	3,77E-05	0,0489	8,666
6	2,391E-03	3,749E-05	0,6096	4,826
7	2,163E-03	3,583E-05	0,04133	8,954
8	2,167E-03	3,596E-05	0,8636	4,225
9	2,384E-03	3,745E-05	0,004267	15,314
10	2,386E-03	3,789E-05	0,0945	7,279

4.5.3 Certitude - Type 3

Dans ce type 3, l'expert pense avec une certitude "large" que le nombre de défaillances a priori pour la période en cours est inférieur ou égal à la moyenne des défauts et des défaillances ,noté a , de toutes les périodes précédentes.

$$0,3 \leq P(X \leq a) \leq 0,9$$

Comme ce type de certitude donne le même support de la loi bêta, nous avons pensé utile de considérer deux cas de figures ayant chacun un mode différent. C'est comme si, nous avons deux experts donnant la même certitude à propos de l'a priori sur p , donc le même support de la loi bêta, mais avec deux modes différents. Ainsi, nous pouvons déplacer la certitude du côté de p_1 ou de p_2 dépendamment de l'expert. Certes ils donnent le même support mais pas la même moyenne de p a posteriori.

Comme les modes sont différents, le choix des paramètres de la loi bêta se fait en fonction de l'un des deux cas de figures suivants.

Cas 1. L'expert accorde plus de poids à la limite inférieure, par conséquent, $\alpha < \beta$.

L'exécution de 31000 itérations d'une chaîne MCMC avec le scénario 3 sur l'échantillon de données BDD-2 donne les résultats suivants (tableau 4.27) :

Tableau 4.27 Résultats a posteriori : Certitude - Type 3 - Cas 1 selon le scénario 3

Période i	Moyenne p	Écart type p	p-value	DIC
4	1,521E-03	4,885E-05	0,3994	4,511
5	1,356E-03	4,648E-05	0,5649	4,005
6	1,372E-03	5,193E-05	0,009933	10,223
7	1,191E-03	4,371E-05	0,501	4,001
8	1,202E-03	4,717E-05	0,123	6,384
9	1,352E-03	4,493E-05	0,0761	7,279
10	1,358E-03	4,691E-05	0,841	3,766

Cas 2. L'expert accorde plus de poids à la limite supérieure. En conséquence, $\alpha > \beta$.

L'exécution de 31000 itérations d'une chaîne MCMC avec le scénario 3 sur l'échantillon de données BDD-2 donne les résultats suivants (tableau 4.28) :

Tableau 4.28 Résultats a posteriori : Certitude - Type 3 - Cas 2 selon le scénario 3

Période i	Moyenne p	Écart type p	p-value	DIC
4	2,364E-03	5,284E-05	0,05186	8,528
5	2,156E-03	5,007E-05	0,08838	7,374
6	2,165E-03	4,673E-05	0,3751	5,358
7	1,945E-03	4,714E-05	0,075	7,633
8	1,952E-03	4,522E-05	0,9815	4,171
9	2,152E-03	5,165E-05	0,007966	13,431
10	2,157E-03	4,989E-05	0,1617	6,211

En intégrant la certitude de l'expert dans le scénario 3, nous remarquons que l'écart type de la distribution a posteriori du paramètre est passé de l'ordre de (10^{-4}) à (10^{-5}) et par conséquent la certitude a augmenté en définissant le support de la distribution a priori.

CHAPITRE 5 CONCLUSION

La problématique de recherche concernait l'exploitation de données de défaillances et de défauts issues de deux bases de données industrielles, les moteurs d'aéronefs et les turbines à gaz, dans la construction des avis a priori sur le paramètre d'intérêt du modèle de risque de défaillance. Pour chaque base de données, deux modèles d'estimation paramétrique du risque de défaillance sont considérés : la loi binomiale ayant comme paramètre la probabilité d'occurrence d'une défaillance et la loi de Poisson ayant comme paramètre le taux de défaillance. La loi de Poisson considère la durée d'exposition ou d'exploitation d'un système alors que la loi binomiale tient compte uniquement du nombre de systèmes défaillants parmi un ensemble en exploitation.

Pour chaque modèle de risque, nous avons développé, modélisé et évalué trois différents scénarios d'exploitation des données de défaillances et de défauts pour construire la distribution a priori sur le paramètre d'intérêt. Chaque scénario renseigne le modèle de calcul de risque de défaillance de manière différente. Le premier scénario considère un a priori constant d'une période à l'autre, l'information utilisée ne change pas au cours du temps. Le second scénario considère seulement l'information disponible au cours de la période d'intervention précédente pour établir l'a priori sur le paramètre d'intérêt au cours de la période subséquente. Le troisième scénario accumule l'information sur toutes les périodes antérieures pour établir l'a priori sur la période en cours.

D'un point de vue pratique, les deux modèles de risque de défaillance permettent d'estimer périodiquement la probabilité de défaillance et le taux de défaillance en s'appuyant sur plusieurs scénarios de prise en compte de l'information a priori d'une période d'estimation à l'autre. Ces deux résultats permettent le suivi de l'évolution du risque au cours de la durée d'exploitation et d'ajuster les interventions de maintenance d'une période à l'autre afin de garder ce risque de défaillance en deçà d'un seuil critique acceptable. Pour ce faire, nous avons exploité les techniques d'inférence bayésienne hiérarchique qui permettent la combinaison des connaissances a priori sur les défaillances et défauts antérieurs et les données de vraisemblance. Ainsi, les estimateurs a posteriori du paramètre d'intérêt sont non biaisés contrairement aux autres méthodes classiques telles que les méthodes du maximum de vraisemblance, des moments et des moindres carrés.

La qualité d'ajustement des distributions a posteriori a été évaluée par le critère statistique du p-value bayésien, et ce pour les deux bases de données des moteurs d'aéronefs et des turbines à gaz. La plupart des p-values bayésiennes sont proches de 0.5. Il y a très peu de

valeurs proches de 0 ou de 1. Cela veut dire qu'il s'agit d'un bon modèle de prédiction et appuie l'hypothèse que l'information disponible renseigne de manière objective la distribution a priori sur le paramètre d'intérêt. Dans le cas contraire où le p-value bayésien est proche de 0 ou 1, cela explique probablement que l'information a priori était non informative, donnant un a posteriori très loin de la vraisemblance ou que l'une des hypothèses du modèle de vraisemblance est violée. Cela est arrivé en particulier avec la base de données des turbines à gaz pour les trois scénarios à la période 6 où le p-value bayésien est en-dessous de 0.15.

Une étude comparative de ces différents scénarios, basée sur le critère statistique DIC, est réalisée afin d'évaluer la contribution de chacun des a priori dans la précision de pronostic du risque de défaillance. Usuellement, le DIC permet de sélectionner le modèle le plus approprié dans une démarche de modélisation bayésienne. Les résultats obtenus pour les données des moteurs d'aéronefs et des turbines à gaz montrent que le scénario 3, celui qui cumule l'information sur toutes les périodes antérieures, est le plus performant selon le DIC, et ce pour les deux modèles de risque de défaillance considérés. Toutefois, le scénario 2, celui ayant un a priori glissant d'une période à l'autre, est le meilleur pour la base de données des turbines à gaz, et ce pour le modèle de Poisson seulement. Cela démontre que le choix du scénario dépend de la base de données sur laquelle le scénario est appliqué et qu'il n'y a pas de scénario "idéal" à adopter quelque soit la nature des données de vraisemblance.

Par ailleurs, nous avons exploité les résultats de distributions a posteriori pour estimer le risque de défaillance à une période donnée. Ce risque calcule la probabilité cumulative d'avoir un certain nombre de systèmes défaillants parmi un ensemble en exploitation ou le taux de défaillance d'un système au cours du temps parmi d'autres similaires en exploitation. Cependant, le pronostic se fait d'une période à une autre. Ces résultats sont très importants lors du suivi de l'évolution du risque au cours du temps. À tout moment, il est possible d'établir un pronostic de défaillance, à partir d'une période courante vers une période subséquente, en fixant la probabilité ou le taux de défaillance attendu à la période subséquente, conditionnellement à la probabilité ou le taux de défaillance de la période en cours.

Enfin, une étude de la certitude de l'avis d'expert est prise en compte dans la formulation de son a priori. Trois types de certitudes ont été étudiés : élevée, faible ou large. En fait, ces trois types de certitudes permettent d'effectuer une analyse de sensibilité sur les modèles de risque a posteriori. Ce sont des résultats qui valident les plages de variabilité du paramètre d'intérêt a posteriori par rapport au nouveau support de la loi a priori qui est directement relié à la certitude de l'expert.

Cette recherche pourrait ouvrir sur d'autres problématiques de développement telles que l'exploitation des distributions a posteriori pour l'optimisation d'une stratégie de maintenance

qui chercherait à minimiser le nombre de défaillances d'une période à l'autre. La stratégie de maintenance ne serait plus basée sur l'âge, mais plutôt sur le comptage des nombres de défaillance et de défauts comme variables aléatoires à optimiser.

Également, il serait possible de considérer d'autres sources de connaissances que les nombres de défaillance et de défauts pour construire un avis a priori sur le paramètre d'intérêt. Ces connaissances pourraient être issues par exemple des études de modèles de dégradation. Aussi, la prise en compte de plusieurs experts qui donneraient leurs avis simultanément pourrait étendre le champ de la modélisation. Le problème se compliquerait davantage car la construction d'un seul avis d'experts à partir de plusieurs nous amèneraient à d'autres techniques de mélanges ou d'agrégation des modèles.

RÉFÉRENCES

- M. Bacha et G. Celeux, “Bayesian estimation of a weibull distribution in a highly censored and small sample setting”, Thèse de doctorat, INRIA, 1996.
- M. Bacha, G. Celeux, E. Idée, A. Lannoy, et D. Vasseur, *Estimation de modèles de durées de vie fortement censurées*. Éd. Eyrolles, 1998.
- I. E. L. A. V. D. Bacha M, Celeux G, “Bayesian estimation of weibull distribution parameters”, 1994.
- F. Billy, E. Remy, N. Bousquet, et G. Celeux, “Modelisation de la duree de vie par une loi de weibull : Comparaison de trois approches bayesiennes lifetime modelling with a weibull law : Comparison of three bayesian methods”.
- H. Dezfuli, D. Kelly, C. Smith, K. Vedros, et W. Galyean, “Bayesian inference for nasa probabilistic risk and reliability analysis”, 2009.
- D. P. Gaver et I. G. O’Muircheartaigh, “Robust empirical bayes analyses of event rates”, *Technometrics*, vol. 29, no. 1, pp. 1–15, 1987.
- J. Geweke, “Bayesian inference in econometric models using monte carlo integration”, *Econometrica : Journal of the Econometric Society*, pp. 1317–1339, 1989.
- M. S. Hamada, A. Wilson, C. S. Reese, et H. Martz, *Bayesian reliability*. Springer Science & Business Media, 2008.
- R. A. Ion, H. Kalishoek, D. M. Karydas, et I. Niturescu, “Bayesian model for early reliability prediction”, dans *RAMS’06. Annual Reliability and Maintainability Symposium, 2006*. IEEE, 2006, pp. 260–265.
- D. Kelly et C. Smith, *Bayesian inference for probabilistic risk assessment : a practitioner’s guidebook*. Springer Science & Business Media, 2011.
- J. Kerman *et al.*, “Neutral noninformative and informative conjugate beta and gamma prior distributions”, *Electronic Journal of Statistics*, vol. 5, pp. 1450–1470, 2011.
- M. Kijima, “Some results for repairable systems with general repair”, *Journal of Applied probability*, vol. 26, no. 01, pp. 89–102, 1989.

- A. Kleyner, D. Elmore, et B. Boukai, “A bayesian approach to determine test sample size requirements for reliability demonstration retesting after product design change”, *Quality Engineering*, vol. 27, no. 3, pp. 289–295, 2015.
- A. Mihalache, F. Guerin, M. Barreau, A. Todoskoff, et B. Dumon, “Reliability assessment of mechatronic systems : operating field data analysis”, dans *Industrial Technology, 2004. IEEE ICIT'04. 2004 IEEE International Conference on*, vol. 2. IEEE, 2004, pp. 1028–1033.
- A. G. Mihalache, “Modélisation et évaluation de la fiabilité des systèmes mécatroniques : application sur système embarqué”, Thèse de doctorat, Université d'Angers, 2007.
- H. Procaccia, E. Ferton, et M. Procaccia, *Fiabilité et maintenance des matériels industriels réparables et non réparables*. Lavoisier, 2011.
- A. Shafaghi, “Equipment failure rate updating—bayesian estimation”, *Journal of Hazardous Materials*, vol. 159, no. 1, pp. 87–91, 2008.
- N. O. Siu et D. L. Kelly, “Bayesian parameter estimation in probabilistic risk assessment”, *Reliability Engineering & System Safety*, vol. 62, no. 1, pp. 89–116, 1998.
- D. Spiegelhalter, A. Thomas, N. Best, et D. Lunn, “Winbugs user manual”, 2003.
- G. R. Terrell, “The maximal smoothing principle in density estimation”, *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 470–477, 1990.
- P. Wang, S. Wang, et L. Wang, “Gas turbine apu reliability modeling and failure forecasting”, dans *Reliability and Maintainability Symposium (RAMS), 2015 Annual*. IEEE, 2015, pp. 1–5.
- A. Zaidi, B. Ould Bouamama, et M. Tagina, “Bayesian reliability models of weibull systems : State of the art”, *International Journal of Applied Mathematics and Computer Science*, vol. 22, no. 3, pp. 585–600, 2012.