

Université du Québec
INRS-Eau

**ESTIMATION DE LA VALEUR DE L'INFORMATION HYDROLOGIQUE
À L'AIDE DE PROBABILITÉS IMPRÉCISES**

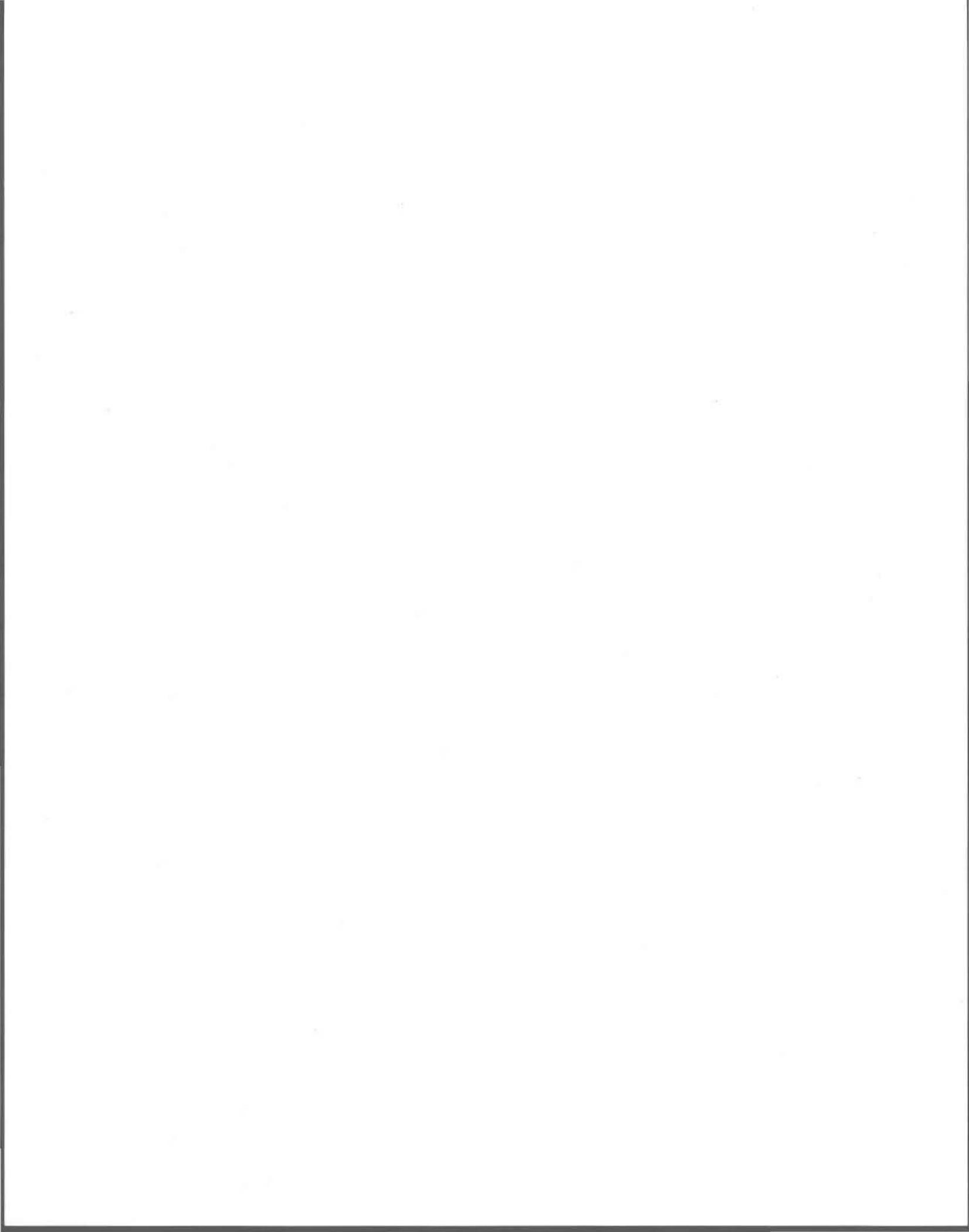
Par
Vincent Fortin
M.Sc. en sciences de l'eau

Thèse présentée pour l'obtention
du grade de Philosophiæ doctorat (Ph.D.)
en sciences de l'eau

Jury d'évaluation

Président du jury et examinateur externe	M. Bertrand Munier, directeur Groupe de recherche sur le risque, l'info. et la décision École Normale Supérieure de Cachan
Examineur externe	M. Jean-Marc Azaïs, professeur Laboratoire de Statistique et Probabilités Université Paul Sabatier
Examineur externe	M. Christian Genest, professeur Département de mathématiques et de statistique Université Laval
Examineur interne	M. Éric Parent, professeur École Nationale du Génie Rural, des Eaux et des Forêts
Expert invité	M. René Roy, chargé d'équipe Prévision des apports, Hydro-Québec
Examineur interne et codirecteur de recherche	M. Jacques Bernier, professeur invité INRS-Eau
Directeur de recherche	M. Bernard Bobée, professeur INRS-Eau

Thèse soutenue le 31 octobre 1997

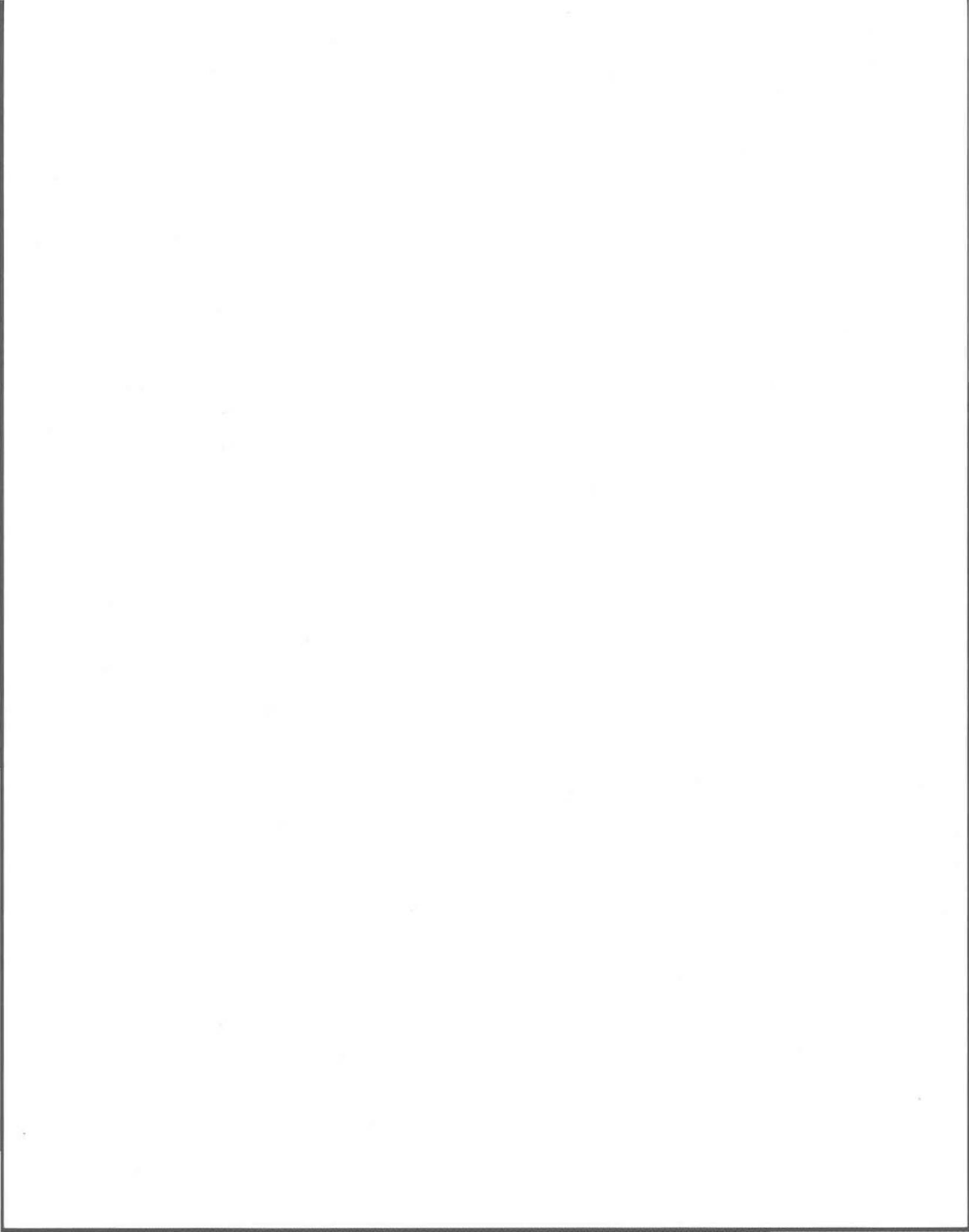


RESUME

La théorie bayésienne de la décision peut être utilisée pour effectuer une prévision de la valeur d'une information avant son observation, que l'on peut comparer à son coût d'acquisition. Cependant, cela demande de pouvoir représenter l'information a priori à l'aide d'une distribution de probabilité, ce qui est en pratique souvent très difficile. Pour faciliter la modélisation de l'information a priori, nous proposons l'emploi de la théorie des prévisions inférieures cohérentes, et nous définissons le concept de valeur de l'information dans le cadre de cette théorie. Ces idées sont illustrées par un exemple concret, qui consiste à évaluer la valeur d'une mesure du débit moyen annuel à un site où l'on envisage de construire un barrage et une centrale hydroélectrique. La théorie des prévisions inférieures cohérentes est ensuite comparée à l'approche bayésienne robuste, à la théorie de Dempster-Shafer et à la théorie des ensembles flous.

ABSTRACT

Bayesian decision theory can be used to evaluate the worth of additional data prior to its collection. To use this approach, it must be possible to model prior information using a probability distribution. This is often difficult, in particular when there is little prior information. An alternative is to use a coherent lower prevision to model prior information. The objective of this thesis is to define the concept of data worth when prior information is modeled by a coherent lower prevision. The definition is illustrated by an example, which consists in evaluating the economic value of mean annual flow measurements at a site where a dam and a hydroelectric power plant are planned. Finally, for the problem of evaluating the worth of data, the theory of coherent lower previsions is compared to Bayesian sensitivity analysis, Dempster-Shafer theory, and fuzzy sets theory.



REMERCIEMENTS

Plusieurs personnes m'ont soutenu pendant mes études de doctorat. Mentionnons d'abord mon directeur de thèse, Bernard Bobée, et mes codirecteurs, Peter Rasmussen et Jacques Bernier. J'ai aussi effectué deux stages pendant ces trois années d'étude: un premier à l'Université d'Arizona avec le Pr. Lucien Duckstein, et un second à l'École nationale du génie rural, des eaux et des forêts (France) sous la direction d'Eric Parent. Le premier m'a permis de cerner mon sujet et le second de rédiger ma thèse. Il me faut enfin souligner que les commentaires des examinateurs externes ont permis d'améliorer de façon significative la qualité de ce document.

Mes études ont été financées par une variété d'organismes. J'ai obtenu deux bourses d'études: l'une du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), l'autre du Fonds pour la formation de chercheur et l'aide à la recherche du Québec (FCAR). Mes stages à l'étranger ont été financés par le Gouvernement du Québec (Ministère des relations internationales, Ministère de l'éducation et Fonds FCAR). Des compléments de bourse m'ont aussi été octroyés par l'INRS et la Chaire industrielle en hydrologie statistique de l'INRS-Eau, qui est financée par le CRSNG, l'INRS et Hydro-Québec.

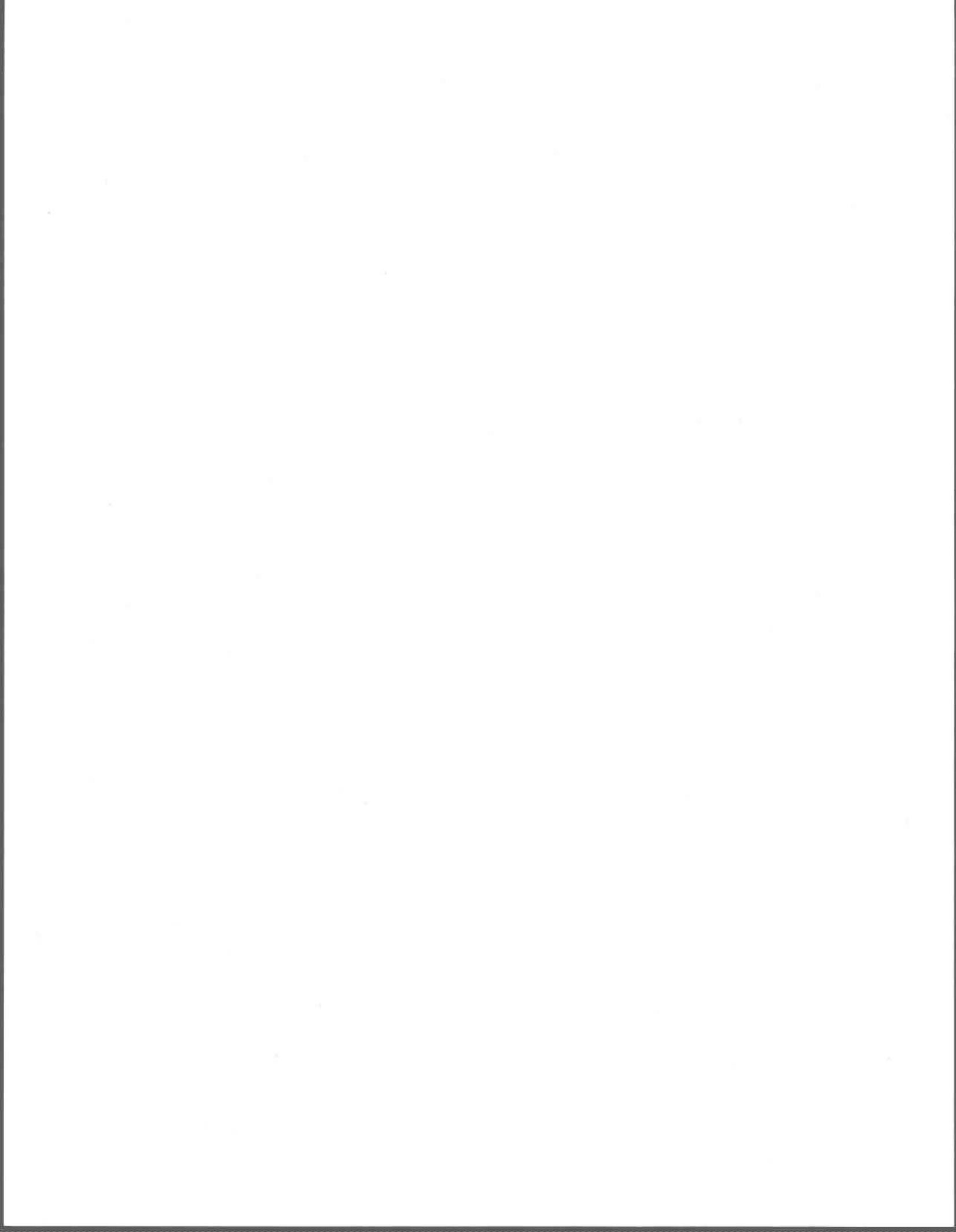


TABLE DES MATIERES

INTRODUCTION	1
1. ESTIMATION BAYÉSIENNE DE LA VALEUR DE L'INFORMATION	3
1.1 ESPACE DE PROBABILITÉS SUBJECTIVES	4
1.1.1 <i>Interprétation d'une probabilité subjective</i>	5
1.1.2 <i>Principe d'évitement des pertes sûres</i>	6
1.1.3 <i>Espérance mathématique et prévision</i>	8
1.1.4 <i>Construction d'un espace de probabilités subjectives</i>	9
1.1.5 <i>Distribution prévisionnelle</i>	10
1.1.6 <i>Distribution cumulée</i>	11
1.1.7 <i>Règle de Bayes</i>	11
1.1.8 <i>Indépendance et échangeabilité</i>	13
1.1.9 <i>Exemple d'utilisation de la règle de Bayes</i>	14
1.2 THÉORIE BAYÉSIENNE DE LA DÉCISION	17
1.2.1 <i>Utilité d'une action</i>	17
1.2.2 <i>Principe de maximisation de l'utilité espérée</i>	19
1.2.3 <i>Risque et regret d'une action</i>	20
1.2.4 <i>Aversion au risque</i>	21
1.2.5 <i>Exemple d'application de la théorie bayésienne de la décision</i>	21
1.2.5.1 <i>Analyse a priori de la décision</i>	22
1.2.5.2 <i>Analyse a posteriori de la décision</i>	22
1.3 ESTIMATION DE LA VALEUR D'UNE INFORMATION	23
1.3.1 <i>Valeur a posteriori d'une information</i>	23
1.3.2 <i>Valeur prévisionnelle d'une information</i>	23
1.3.3 <i>Borne maximale pour la valeur prévisionnelle de l'information</i>	25
1.3.4 <i>Valeur prévisionnelle d'un échantillon aléatoire</i>	25
1.3.5 <i>Exemple de prévision de la valeur de l'information</i>	26
1.3.5.1 <i>Valeur de l'information parfaite</i>	26
1.3.5.2 <i>Valeur de l'information en fonction de la taille d'échantillon</i>	28
1.4 ANALYSE BAYÉSIENNE DE RÉFÉRENCE	30
1.4.1 <i>Principe d'invariance de Jeffreys</i>	31
1.4.2 <i>Principe de maximisation de l'entropie</i>	31
1.4.3 <i>Principe de maximisation du regret espéré</i>	31

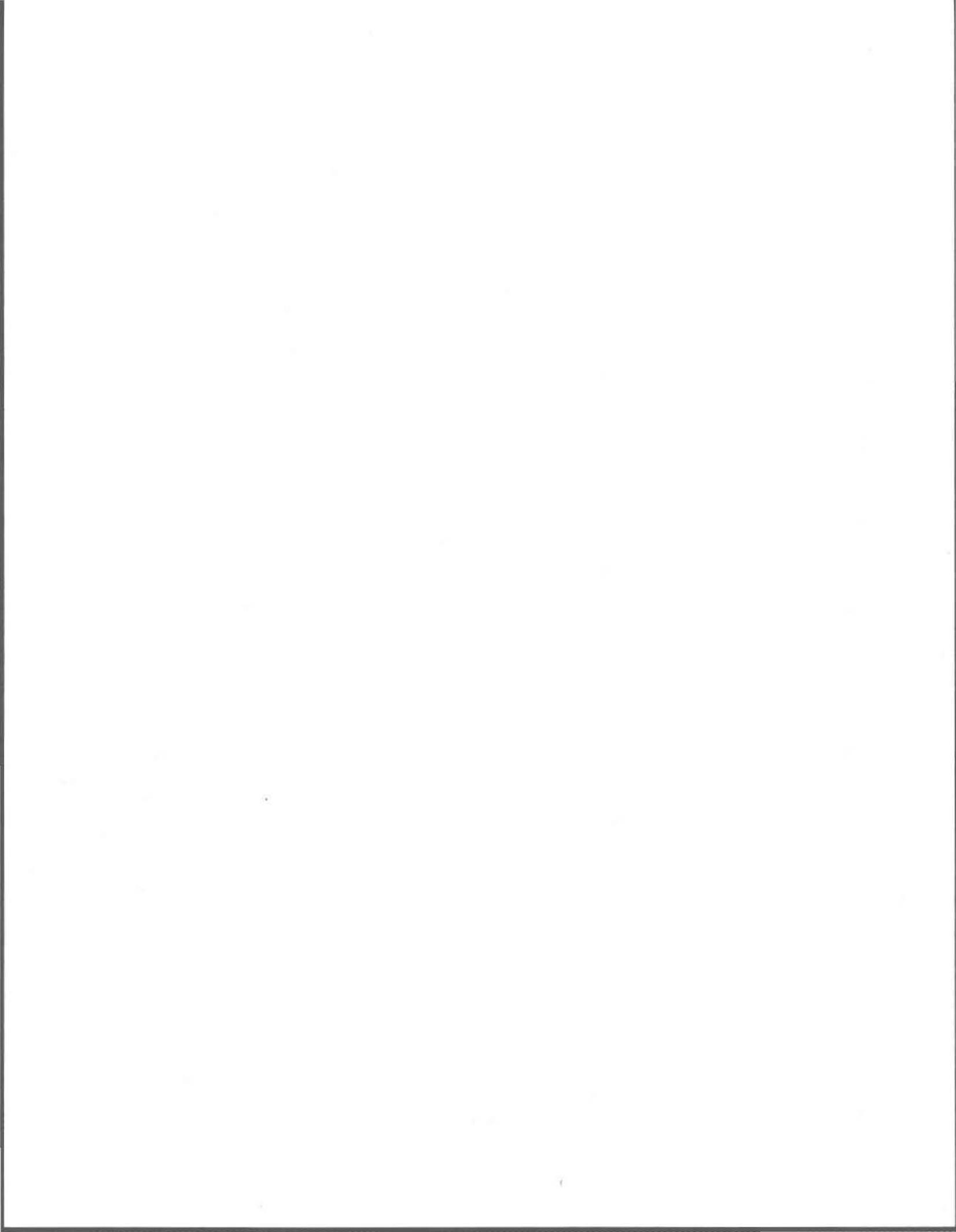
1.5 DÉVELOPPEMENTS DE LA MÉTHODOLOGIE BAYÉSIENNE POUR L'ESTIMATION DE LA VALEUR DE L'INFORMATION HYDROLOGIQUE	32
1.6 CRITIQUE DE L'APPROCHE BAYÉSIENNE.....	34
1.6.1 Critique du principe de précision	34
1.6.2 Critique du principe de maximisation de l'utilité espérée.....	36
1.7 APPROCHE BAYÉSIENNE ROBUSTE	37
1.7.1 Estimation robuste de la valeur de l'information	38
1.7.2 Mise en oeuvre de l'approche bayésienne robuste	38
1.7.3 Principe de précision idéale	39
2. PRÉVISION IMPRÉCISE DE LA VALEUR DE L'INFORMATION	41
2.1 THÉORIE DES PRÉVISIONS INFÉRIEURES COHÉRENTES.....	41
2.1.1 Axiomes de la théorie des prévisions inférieures cohérentes	42
2.1.2 Probabilités imprécises cohérentes.....	47
2.1.3 Principe d'extension des prévisions inférieures	47
2.1.4 Théorème de l'enveloppe inférieure.....	49
2.1.5 Règle de Bayes généralisée	50
2.1.6 Événements de probabilité inférieure nulle.....	50
2.1.7 Cas particulier des probabilités inférieures 2-monotones	51
2.1.8 Modélisation de l'ignorance et de la quasi-ignorance a priori.....	52
2.1.9 L'indécision: une conséquence de l'imprécision.....	52
2.1.10 Comparaison avec l'approche bayésienne robuste.....	53
2.1.11 Comparaison avec les modèles descriptifs à base de capacités de Choquet	54
2.1.12 Application pratique de la théorie des prévisions inférieures cohérentes	55
2.1.12.1 Exemple d'application du principe d'extension des prévisions inférieures	55
2.1.12.2 Exemple d'application de la règle de Bayes généralisée.....	59
2.2 PROPOSITION D'UNE APPROCHE PERMETTANT D'ESTIMER LA VALEUR DE L'INFORMATION DANS LE CADRE DE LA THÉORIE DES PRÉVISIONS INFÉRIEURES COHÉRENTES	64
2.2.1 Valeur imprécise de l'information	64
2.2.2 Valeur imprécise de l'information parfaite	66
2.2.3 Construction d'une règle de décision.....	66
2.3 ANALYSE DE SENSIBILITÉ DE LA VALEUR IMPRÉCISE DE L'INFORMATION.....	68
2.3.1 Lien avec la théorie des nombres flous	72
2.4 EXEMPLE DE PRÉVISION IMPRÉCISE DE LA VALEUR DE L'INFORMATION	76
2.4.1 Analyse du risque a priori	77
2.4.2 Analyse du risque a posteriori	78

2.4.3 Valeur a posteriori de l'information.....	82
2.4.4 Prévission floue de la valeur de l'information	82
3. VALEUR D'UNE INFORMATION HYDROLOGIQUE AU SITE DE CONSTRUCTION D'UN OUVRAGE HYDROÉLECTRIQUE.....	89
3.1 PROBLÉMATIQUE GÉNÉRALE.....	89
3.1.1 Aménagement hydroélectrique SM3	92
3.1.2 Mesures du débit disponibles	92
3.1.3 Transfert de l'information régionale	93
3.1.4 Caractéristiques hydrologiques et technico-économiques du projet Sainte-Marguerite-3	94
3.1.5 Sensibilité de la rentabilité et des coûts de construction au débit	95
3.2 PROBLÉMATIQUE ÉTUDIÉE.....	95
3.3 ESPACE DES ACTIONS	96
3.4 FONCTION DE COÛT	97
3.5 ANALYSE DU RISQUE A PRIORI	100
3.6 VALEUR DE L'INFORMATION PARFAITE.....	102
3.6.1 Modélisation de l'information a priori	102
3.6.2 Calcul de la valeur de l'information parfaite à l'aide du théorème de l'enveloppe inférieure	103
3.7 EVALUATION DU COÛT RELIÉ À L'OBSERVATION DU DÉBIT ANNUEL.....	107
3.8 ANALYSE DU RISQUE A POSTERIORI.....	108
3.9 VALEUR A POSTERIORI D'UNE OBSERVATION DU DÉBIT ANNUEL.....	110
3.10 PRÉVISION IMPRÉCISE DE LA VALEUR DE L'INFORMATION BASÉE SUR UNE HYPOTHÈSE DE NORMALITÉ DU DÉBIT ANNUEL.....	110
3.10.1 Modèle d'échantillonnage.....	111
3.10.2 Prévission de la valeur de l'information en fonction du module et de l'écart-type du débit annuel.....	113
3.10.3 Modélisation de l'information sur l'écart-type du débit annuel.....	114
3.10.4 Application du principe d'extension des prévisions inférieures.....	116
3.10.5 Résultats numériques	119
3.11 CONSTRUCTION D'UNE PRÉVISION SUPÉRIEURE DE LA VALEUR DE L'INFORMATION SANS L'HYPOTHÈSE DE NORMALITÉ DU DÉBIT ANNUEL	126
3.12 CONCLUSIONS GÉNÉRALES	128
4. DISCUSSION ET CONCLUSION	131
4.1 AVANTAGES DE LA THÉORIE DES PRÉVISIONS INFÉRIEURES COHÉRENTES SUR LA THÉORIE BAYÉSIENNE DES PROBABILITÉS	131
4.2 AUTRES MODÈLES D'INCERTITUDE PERMETTANT DE S'AFFRANCHIR DU PRINCIPE DE PRÉCISION	133

4.2.1	<i>Théorie de Dempster-Shafer</i>	133
4.2.1.1	Modélisation de l'incertitude	133
4.2.1.2	Éléments focaux, support et noyau	135
4.2.1.3	Combinaison de l'incertitude à l'aide de la règle de Dempster.....	135
4.2.1.4	Extension de la théorie de Dempster-Shafer aux domaines infinis	137
4.2.1.5	Inférence statistique	138
4.2.2	<i>Théorie des ensembles flous</i>	140
4.2.3	<i>Théorie des possibilités</i>	141
4.3	INFLUENCE DE LA REPRÉSENTATION DE L'INFORMATION A PRIORI	142
4.4	APPLICATION PRATIQUE DU PRINCIPE D'EXTENSION DES PRÉVISIONS INFÉRIEURES ET DE LA RÈGLE DE BAYES GÉNÉRALISÉE.....	145
4.5	BILAN ET PERSPECTIVES	147
4.5.1	<i>Aspects originaux du travail réalisé</i>	147
4.5.2	<i>Travaux ultérieurs à mener</i>	147
4.5.3	<i>Retombées du travail effectué</i>	148
	REFERENCES	151

LISTE DES TABLEAUX

TABLEAU 2.1: DÉCISION A PRIORI ET A POSTERIORI CORRESPONDANT À CHACUNE DES QUATRE ZONES IDENTIFIÉES SUR LA FIGURE 2.13	82
TABLEAU 3.1: CARACTÉRISTIQUES TECHNICO-ÉCONOMIQUES DES PROJETS ÉTUDIÉS POUR RÉALISER LES OBJECTIFS D'EXPORTATION D'ÉLECTRICITÉ (ADAPTÉ DE HYDRO-QUÉBEC, 1991)	90
TABLEAU 3.2: CARACTÉRISTIQUES DES VARIANTES ÉTUDIÉES DU PROJET SM3 (ADAPTÉ DE HYDRO-QUÉBEC, 1991).....	95
TABLEAU 3.3: PRIX DE REVIENT DE L'ÉLECTRICITÉ À SM3, EN FONCTION DU MODULE NATUREL À SM3, DU DÉBIT DÉTOURNÉ ET DU DÉBIT RÉSERVÉ (ADAPTÉ DE HYDRO-QUÉBEC, 1991 ET 1992).....	98



LISTE DES FIGURES

FIGURE 1.1: DISTRIBUTION A PRIORI, VRAISEMBLANCE ET DISTRIBUTION A POSTERIORI POUR LA PROBABILITÉ D'OBSERVER UN DÉBIT ENTRE 40 ET 50 M ³ /s	16
FIGURE 1.2: VALEUR RELATIVE DE L'INFORMATION CONTENUE DANS UN ÉCHANTILLON DE TAILLE N , CALCULÉE POUR $\alpha=4$, $\beta=12$ ET $q=1/6$	30
FIGURE 2.1: EXEMPLE DE DISTRIBUTION PRÉVISIONNELLE D'UNE VARIABLE ALÉATOIRE BINOMIALE BASÉE SUR UNE DISTRIBUTION A PRIORI BÊTA ($\alpha=1/4$, $\beta=3/4$, $N=4$)	56
FIGURE 2.2: EXEMPLE DE PRÉVISION IMPRÉCISE D'UNE VARIABLE ALÉATOIRE BINOMIALE	58
FIGURE 2.3: PROBABILITÉ IMPRÉCISE DE N'OBSERVER QUE DES ÉCHECS EN N ESSAIS INDÉPENDANTS	59
FIGURE 2.4: PROBABILITÉ IMPRÉCISE D'OBSERVER 4 ÉCHECS CONSÉCUTIFS, APRÈS AVOIR OBSERVÉ N ÉCHECS SUCCESSIFS, AVEC DEUX CONTRAINTES PRÉCISES	60
FIGURE 2.5: PROBABILITÉ IMPRÉCISE D'OBSERVER 4 ÉCHECS CONSÉCUTIFS, APRÈS AVOIR OBSERVÉ N ÉCHECS SUCCESSIFS, AVEC TROIS CONTRAINTES PRÉCISES	62
FIGURE 2.6: PROBABILITÉ SUPÉRIEURE D'OBSERVER 4 ÉCHECS CONSÉCUTIFS, APRÈS AVOIR OBSERVÉ N ÉCHECS EN N ESSAIS, FONDÉES SUR DEUX ET TROIS CONTRAINTES PRÉCISES, ET COMPARAISON AVEC UNE SOLUTION BAYÉSIENNE	63
FIGURE 2.7: FAMILLE DE PRÉVISIONS IMPRÉCISES SUR LE PARAMÈTRE D'UNE LOI BINOMIALE	70
FIGURE 2.8: RÉSULTAT D'UNE ANALYSE DE SENSIBILITÉ DE LA PROBABILITÉ D'OBSERVER QUATRE ÉCHECS CONSÉCUTIFS À L'IMPRÉCISION D'UNE PRÉVISION SUR LA PROBABILITÉ DE SUCCÈS	71
FIGURE 2.9: RÉSULTAT D'UNE ANALYSE DE SENSIBILITÉ DE LA PROBABILITÉ D'OBSERVER QUATRE ÉCHECS CONSÉCUTIFS AU DOMAINE DE DÉFINITION DE LA PROBABILITÉ DE SUCCÈS	72
FIGURE 2.10: LIENS ENTRE LES MODÈLES D'INCERTITUDE	75
FIGURE 2.11: VALEUR DU PARAMÈTRE κ DE LA RÈGLE DE DÉCISION À PARTIR DE LAQUELLE L'ACTION A_1 EST SÉLECTIONNÉE A PRIORI, EN FONCTION DU NIVEAU H DE LA COUPE DU RISQUE FLOU DE L'ACTION A_0	78
FIGURE 2.12: RISQUE FLOU A POSTERIORI DE L'ACTION A_0	80
FIGURE 2.13: VALEUR DU PARAMÈTRE κ DE LA RÈGLE DE DÉCISION À PARTIR DE LAQUELLE L'ACTION A_1 EST SÉLECTIONNÉE A POSTERIORI, EN FONCTION DU NIVEAU H DE LA COUPE DU RISQUE FLOU DE L'ACTION A_0	81
FIGURE 2.14: PRÉVISION FLOUE DE LA VALEUR DE L'INFORMATION POUR UNE RÈGLE DE DÉCISION MINIMIN ($\kappa=0$)	84
FIGURE 2.15: PRÉVISION FLOUE DE LA VALEUR DE L'INFORMATION LORSQUE $1/6 \leq \kappa \leq 1/3$	85
FIGURE 2.16: PRÉVISION IMPRÉCISE DE LA VALEUR DE L'INFORMATION EN FONCTION DE H , LORSQUE $\delta_0^{(\kappa)} = \alpha_0$ ET $\delta^{(\kappa)}(1) = \alpha_1$	86
FIGURE 2.17: COMPARAISON DE LA VALEUR DE L'INFORMATION ET DU COÛT RELIÉ À SON ACQUISITION	87
FIGURE 3.1: PLAN GÉNÉRAL DE L'AMÉNAGEMENT, PROJET SM3 (TIRÉ DE HYDRO-QUÉBEC, 1991)	91

FIGURE 3.2: DÉBIT ANNUEL À SM2 ET À SM3 POUR LA PÉRIODE 1958-1982 (ADAPTÉ DE HYDRO-QUÉBEC, 1992)	94
FIGURE 3.3: RELATION ENTRE LE PRIX DE REVIENT DE L'ÉLECTRICITÉ ET LE MODULE TOTAL À SM3	99
FIGURE 3.4: VALEUR MAXIMALE D'UNE CONNAISSANCE PARFAITE DU MODULE À SM3	106
FIGURE 3.5: REPRÉSENTATION DU DÉBIT ANNUEL À SM2 SUR PAPIER NORMAL POUR LA PÉRIODE 1958-1993	112
FIGURE 3.6: NOMBRE FLOU UTILISÉ POUR EFFECTUER UNE ANALYSE DE SENSIBILITÉ DE LA VALEUR D'UNE OBSERVATION DU DÉBIT ANNUEL À SM3 AU DOMAINE DE DÉFINITION DU COEFFICIENT DE VARIATION DE CE DÉBIT ANNUEL	115
FIGURE 3.7: APPROXIMATION DE LA PRÉVISION IMPRÉCISE DE LA VALEUR DE L'INFORMATION, POUR $N_0=1$, $C=4.6$ ¢/KWH ET $P=0.4$, EN FONCTION DE LA LONGUEUR $K=2^{M-1}+1$ DE LA SUITE $(X_N)_M$	119
FIGURE 3.8: PRÉVISION FLOUE DE LA VALEUR DE L'INFORMATION HYDROLOGIQUE À SM3 POUR $N_0=1$ ET $C+P=5.6$ ($M=4$), PAR EXEMPLE POUR LE PROJET ALTERNATIF NBR AVEC $P=1$	120
FIGURE 3.9: PRÉVISION FLOUE DE LA VALEUR DE L'INFORMATION HYDROLOGIQUE À SM3 POUR $N_0=1$ ET $C+P=5.4$ ($M=4$), PAR EXEMPLE POUR LE PROJET ALTERNATIF GRANDE-BALEINE AVEC $P=1$	121
FIGURE 3.10: PRÉVISION FLOUE DE LA VALEUR DE L'INFORMATION HYDROLOGIQUE À SM3 POUR $N_0=1$ ET $C+P=5.0$ ($M=4$)	122
FIGURE 3.11: PRÉVISION SUPÉRIEURE DE LA VALEUR DE L'INFORMATION POUR $H=0$ ET DIFFÉRENTES VALEURS DE N_0 , EN FONCTION DE $C+P$ ($M=2$)	123
FIGURE 3.12: PRÉVISION SUPÉRIEURE DE LA VALEUR DE L'INFORMATION POUR $H=0.5$ ET DIFFÉRENTES VALEURS DE N_0 , EN FONCTION DE $C+P$ ($M=2$)	124
FIGURE 3.13: PRÉVISION SUPÉRIEURE DE LA VALEUR DE L'INFORMATION POUR $H=1$ ET DIFFÉRENTES VALEURS DE N_0 , EN FONCTION DE $C+P$ ($M=2$)	125
FIGURE 3.14: NOMBRES FLOUS PRÉSENTANT LES VALEURS DE $C+P$ POUR LESQUELLES LA PRÉVISION SUPÉRIEURE DE LA VALEUR DE L'INFORMATION HYDROLOGIQUE DÉPASSE 0.1 ¢/KWH	126
FIGURE 3.15: PRÉVISION SUPÉRIEURE DE LA VALEUR DE L'INFORMATION HYDROLOGIQUE À SM3 EN FONCTION DE N_0 ET DE $C+P$, SANS HYPOTHÈSE SUR LA DISTRIBUTION D'ÉCHANTILLONNAGE	128
FIGURE 4.1: PROBABILITÉ IMPRÉCISE D'OBSERVER DEUX SUCCÈS CONSÉCUTIFS EN FONCTION D'UNE ESTIMATION PONCTUELLE M DU PARAMÈTRE DE LA LOI BINOMIALE	144

INTRODUCTION

Une gestion efficace de la ressource en eau requiert une connaissance adéquate de l'état passé, présent et futur de cette ressource. Cependant, le coût d'acquisition d'une information hydrologique peut être élevé. Il existe une méthode bien établie, l'analyse bayésienne, pour comparer ce coût à l'utilité de l'information (Bernier, 1990). Cette approche permet de modéliser l'incertitude et de mesurer son effet sur un processus décisionnel, dans la mesure où l'incertitude sur un événement peut être adéquatement représentée par un seul nombre, sa probabilité. Nous défendons la thèse voulant que deux nombres soient parfois plus efficaces pour modéliser l'incertitude: une probabilité inférieure, et une probabilité supérieure, c'est-à-dire un intervalle que l'on nomme une probabilité imprécise.

Pour modéliser l'ignorance, nous verrons que l'utilisation de probabilités imprécises paraît essentielle. Cependant, de telles situations se produisent rarement. Plus fréquemment, on dispose d'une information fragmentaire, de connaissances subjectives, difficiles à quantifier. Il est alors laborieux de fixer des probabilités subjectives pour l'ensemble des événements, ce qui rend fastidieuse l'utilisation du modèle bayésien. Nous proposons dans ces circonstances l'emploi de la théorie des prévisions inférieures cohérentes (Walley, 1991), une généralisation de la théorie bayésienne des probabilités qui permet de modéliser plus aisément des informations imprécises.

L'objectif de la thèse consiste à développer une méthode d'estimation de la valeur de l'information hydrologique basée sur la théorie des prévisions inférieures cohérentes, et à montrer son intérêt à l'aide d'un exemple réaliste. Hydro-Québec a entrepris sur la rivière Sainte-Marguerite la construction d'un réservoir et d'une centrale hydroélectrique. Au moment de la conception des ouvrages, aucune information directe sur le débit au site n'était disponible. Par contre, une information pertinente mais incertaine était disponible en aval, près de l'exutoire. Nous montrerons comment la théorie des prévisions inférieures cohérentes peut permettre de modéliser l'information régionale, et de quantifier la valeur d'une information hydrologique au site.

Ce rapport se divise en quatre chapitres. Dans le premier, nous présentons la solution apportée par la théorie bayésienne au problème de l'estimation de la valeur de l'information,

avec une revue de la littérature reliée à l'application de cette méthode en hydrologie. Les hypothèses sous-jacentes sont discutées et certaines faiblesses sont relevées. Le chapitre suivant présente la théorie des prévisions inférieures cohérentes et propose une extension de cette théorie permettant de baser une estimation de la valeur de l'information sur une information a priori imprécise. Le troisième chapitre montre comment l'application de cette théorie peut être utile pour la planification d'un projet hydroélectrique. Le dernier chapitre est consacré à une discussion des résultats obtenus.

1. ESTIMATION BAYESIENNE DE LA VALEUR DE L'INFORMATION

Dans le cadre de la théorie bayésienne de la décision, il est possible de quantifier la valeur d'une information de façon prévisionnelle, c'est-à-dire avant son observation. Cette mesure de la valeur de l'information peut ensuite être comparée au coût d'échantillonnage, de façon à permettre de décider s'il vaut la peine d'échantillonner. Évidemment, la valeur d'une information dépend de l'usage qu'on en fait. D'un point de vue opérationnel, on peut considérer qu'une information n'a de valeur que dans la mesure où elle permet de prendre de meilleures décisions. Raiffa et Schlaifer (1961) proposent un modèle, basé sur la théorie de l'utilité (Savage, 1972), permettant d'évaluer sous cet angle la valeur d'une information.

Soit $S=\{a\}$ un ensemble d'actions parmi lesquelles un individu veut effectuer un choix, sachant que les conséquences de ces actions sont fonctions d'un paramètre inconnu $\theta \in \Theta$, que l'on nomme l'état de la nature. Soit aussi $x \in \Omega$ une information qui peut permettre de préciser l'idée que l'on a de la valeur de θ . Pour fixer les idées, on peut imaginer une rivière dont on veut exploiter le potentiel hydroélectrique. L'ensemble des actions pourrait correspondre à différents projets (un seul réservoir, un ensemble de réservoirs et de centrales en cascade, une centrale au fil de l'eau, ...). Imaginons aussi que la rentabilité de ces projets dépend du débit annuel moyen θ . On peut alors observer la valeur x du débit annuel une année donnée pour obtenir une idée de θ , puisque $E(x)=\theta$.

La valeur d'une information supplémentaire dépend de l'information dont dispose déjà un individu sur θ et sur le lien entre x et θ . On suppose que l'individu peut formaliser cette information par une mesure de probabilité P , définie sur une tribu \mathbf{A} (i.e. σ -algèbre) du produit cartésien $\Theta \times \Omega$, de telle façon que $P(A)$, pour $A \in \mathbf{A}$, exprime le degré de confiance subjectif de l'individu dans l'occurrence de l'événement A . On suppose aussi que l'individu accepte d'agir de façon à maximiser l'utilité espérée, c'est à dire qu'il existe une fonction d'utilité $U(a, \theta)$ telle que:

- avant d'observer x , l'individu préfère l'action \hat{a} qui maximise $E[U(a, \theta)]$, i.e. l'espérance de la fonction d'utilité par rapport à la mesure P ;
- après avoir observé x , l'individu préfère l'action $\hat{a}_{(x)}$ qui maximise $E[U(a, \theta)|x]$, i.e. l'espérance conditionnelle de la fonction d'utilité par rapport à la mesure P .

Ceci revient à dire que l'individu respecte les postulats de von Neumann et Morgenstern (1953) ou de Savage (1972), qui sont équivalents. Après avoir admis ces hypothèses et identifié la fonction d'utilité $U(a, \theta)$, on obtient une prévision $XVSI$ de l'utilité de l'information x en calculant l'augmentation espérée de l'utilité de la décision optimale apportée par x :

$$XVSI = E \left\{ E[U(\hat{a}_{(x)}, \theta) | x] - E[U(\hat{a}, \theta)] \right\} \quad (1.1)$$

Dans ce chapitre, nous voulons d'abord présenter en détail et illustrer par un exemple la solution de Raiffa et Schlaifer (1961). Ceci sera suivi d'une revue de littérature des développements théoriques reliés à cette approche et des applications les plus significatives en hydrologie. Nous discuterons ensuite des problèmes rencontrés en pratique pour fixer des probabilités subjectives, des solutions proposées dans le cadre de l'approche bayésienne, et de la validité du principe de maximisation de l'utilité espérée.

1.1 Espace de probabilités subjectives

La définition bayésienne de la valeur d'une information repose sur la modélisation, à l'aide d'un espace de probabilités, des connaissances dont un individu dispose sur l'état de la nature θ et de la précision de l'expérience qu'il entend mener pour mieux cerner la valeur de θ . Rappelons qu'un **espace de probabilités** est un triplet $(\mathfrak{S}, \mathbf{A}, P)$, où (cf. Bartle, 1995):

- \mathfrak{S} est un ensemble, que l'on nomme le **référéntiel**;
- \mathbf{A} est une **tribu** du référéntiel \mathfrak{S} , c'est-à-dire une famille de sous-ensembles de \mathfrak{S} qui contient au moins l'ensemble \mathfrak{S} et l'ensemble vide \emptyset , et qui est de plus fermée pour les opérations de complément et de réunion dénombrable:
 - (i) $\mathfrak{S}, \emptyset \in \mathbf{A}$,
 - (ii) $A \in \mathbf{A} \Rightarrow A^c \in \mathbf{A}$ où $A^c = \mathfrak{S} - A$,
 - (iii) $(A_n) \subseteq \mathbf{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathbf{A}$;
- P est une **mesure de probabilité** définie sur l'ensemble des éléments de \mathbf{A} , c'est à dire une fonction définie de \mathbf{A} vers $[0, 1]$, additive pour toute suite dénombrable d'ensembles disjoints de \mathbf{A} , qui associe la mesure 0 à l'ensemble vide \emptyset et la mesure 1 au référéntiel:
 - (i) $P(\emptyset) = 0, P(\mathfrak{S}) = 1$;

$$(ii) \quad \forall A \in \mathbf{A}, 0 \leq P(A) \leq 1;$$

$$(iii) \quad \forall (A_n)_{n \in \mathbf{N}} \subseteq \mathbf{A}, \text{ si } A_i \cap A_j = \emptyset \text{ lorsque } i \neq j, \text{ alors } P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

Exemple: Soit le référentiel $\mathfrak{S} = \{a, b, c\}$. On vérifie que $\mathbf{A} = \{\emptyset, \{a\}, \{b, c\}, \mathfrak{S}\}$ est une tribu de \mathfrak{S} , même si elle ne contient pas tous les sous-ensembles de \mathfrak{S} . Soit $P: \mathbf{A} \rightarrow [0, 1]$ une fonction définie ainsi: $P(\emptyset) = 0$, $P(\{a\}) = 1/2$, $P(\{b, c\}) = 1/2$, $P(\mathfrak{S}) = 1$. On vérifie que P est une mesure de probabilité sur \mathbf{A} , de telle sorte que $(\mathfrak{S}, \mathbf{A}, P)$ est un espace de probabilité. On remarque que cet espace de probabilité ne permet pas d'attribuer une probabilité aux ensembles $\{b\}$ et $\{c\}$.

On peut utiliser un espace de probabilité $(\mathfrak{S}, \mathbf{A}, P)$ pour modéliser l'information disponible sur une quantité φ dont la valeur appartient à un ensemble \mathfrak{S} . Dans ce contexte, on nomme les éléments de \mathbf{A} des **événements**, et on associe à chaque événement $A \in \mathbf{A}$ une probabilité $P(A)$ qui représente le degré de confiance d'un individu dans l'occurrence de l'événement A , c'est à dire dans la proposition logique $\varphi \in A$. Pour qu'un tel modèle soit utile en pratique, il faut proposer une interprétation de ce degré de confiance pour laquelle les trois propriétés d'une mesure de probabilité sont justifiées.

1.1.1 Interprétation d'une probabilité subjective

La valeur numérique d'une probabilité peut être fixée de différentes façons par un individu. Par exemple, on peut interpréter de façon opérationnelle la probabilité $P(A)$ d'un événement A en terme de prix d'achat ou de vente d'un pari. Supposons une première loterie pour laquelle un grand nombre N de billets $(a_n, n=1, 2, \dots, N)$ sont émis: un seul donne droit à un prix, et chaque billet a_n a la même chance d'être tiré. Supposons aussi l'existence d'une seconde loterie pour laquelle un seul billet b est émis, qui donne droit à un prix équivalent à celui de la première loterie seulement si l'événement A se produit. On peut alors mesurer $P(A)$ en demandant à un individu de déterminer le nombre maximal r de billets a_n de la première loterie qu'il serait prêt à échanger pour obtenir le billet b de la seconde loterie, ou encore en lui demandant de spécifier le nombre minimal r' de billets a_n de la première loterie qu'il demanderait pour vendre le billet b de la seconde loterie, en admettant qu'il le possède déjà. Évidemment, pour qu'un seul nombre soit suffisant pour représenter le degré de confiance d'un individu dans l'événement A , il faut nécessairement avoir $r=r'$. Il s'agit du **principe de précision**. Le réalisme

de cette hypothèse dépend du degré de connaissance qu'un individu a de l'expérience aléatoire. En admettant cette hypothèse, on définit alors $P(A) = r/N = r'/N$. Ainsi, $P(A)$ correspond à la fois au prix maximal d'achat et au prix minimal de vente d'un pari sur l'événement A , en proportion de la valeur de la récompense obtenue si A se produit. En somme, on mesure la probabilité d'un événement en déterminant des équivalences entre une première loterie dont le mécanisme est entièrement connu, et une seconde qui dépend de l'événement A . Il s'agit, comme pour toute mesure, d'effectuer un étalonnage.

On pourrait supposer que deux individus, disposant de la même base de connaissances, devraient logiquement arriver à proposer des probabilités semblables pour un même événement. Cependant, l'intuition et l'expérience ont un rôle à jouer dans la spécification d'une probabilité, et conséquemment deux individus peuvent fort bien proposer des probabilités différentes pour un même événement, apparemment sur la base de la même information. Puisqu'il n'est pas réaliste d'espérer que chacun d'eux pourra expliquer dans le menu détail comment sa perception du problème le mène à spécifier une probabilité particulière, il faut accepter qu'une probabilité n'a pas de valeur logique acceptable par tous, seulement une valeur personnelle et **subjective**.

Il importe de réaliser que l'interprétation proposée de la probabilité d'un événement n'est pas liée à sa fréquence d'occurrence. On peut par exemple associer une probabilité à la proposition "Le Nil est le plus long fleuve du monde". Même pour des événements dont on peut mesurer la fréquence, il n'est pas requis que la probabilité spécifiée par un individu corresponde à cette fréquence.

Cette interprétation opérationnelle de la valeur numérique d'une probabilité impose que la probabilité d'un événement soit comprise entre 0 et 1, et suggère que l'on fixe $P(\emptyset) = 0$ et $P(\mathcal{S}) = 1$, puisqu'il n'y a aucun intérêt à acheter un pari sur l'événement impossible \emptyset ou à vendre un pari sur l'événement certain \mathcal{S} . En considérant l'achat et la vente d'un ensemble de paris, on peut aussi proposer une justification de l'additivité des probabilités.

1.1.2 Principe d'évitement des pertes sûres

Dans la suite de ce rapport, on notera par $A(\varphi)$ la **fonction indicatrice** d'un événement A , fonction qui vaut 0 si $\varphi \notin A$ et qui vaut 1 si $\varphi \in A$. Avec l'interprétation que nous proposons

d'une probabilité, la fonction indicatrice $A(\varphi)$ de l'événement A représente le gain résultant d'un pari sur A , après observation de φ . On peut ainsi représenter par $G(\varphi, A) = A(\varphi) - P(A)$ le **bénéfice net** obtenu par l'achat pour un prix $P(A)$ d'un pari sur un événement A , après observation de φ . Remarquons que la vente pour un prix $P(A)$ du pari $A(\varphi)$ se soldera, après observation de φ , par un bénéfice net de $P(A) - A(\varphi)$, soit $-G(\varphi, A)$.

Même si différents individus peuvent spécifier différentes valeurs pour la probabilité d'un même événement, certaines contraintes raisonnables sur la valeur de ces probabilités peuvent être proposées. En particulier, on peut imaginer que le bénéfice net de l'achat ou de la vente d'un pari ne devrait pas être systématiquement négatif, ce qui revient à demander que les deux contraintes suivantes soient respectées: $\sup_{\varphi} G(\varphi, A) \geq 0$ et $\sup_{\varphi} -G(\varphi, A) \geq 0$. On peut

aussi faire une hypothèse plus forte, et imposer que pour toute séquence finie de paris $(A_n(\varphi), n=1, 2, \dots, k)$, le bénéfice net de toute combinaison d'achat ou de vente de ces paris ne puisse être systématiquement négatif:

$$\sup_{\varphi} \sum_{n=1}^k \pm G(\varphi, A_n) \geq 0, \quad \forall (A_n) \subseteq \mathbf{A} \quad (1.2)$$

Cette hypothèse, nommée **principe d'évitement des pertes sûres**, signifie en fait qu'il n'est pas rationnel pour un individu de s'engager à perdre systématiquement au jeu, quelque soit la valeur de la variable aléatoire φ . Il doit donc exister au moins une valeur de φ pour laquelle les engagements pris par un individu à travers la spécification des probabilités sur les événements $A_n(\varphi)$ lui permettent au moins d'obtenir après observation de φ des gains équivalents au montant total qu'il a investi avant d'observer φ . Le principe d'évitement des pertes sûres permet de justifier la règle d'addition des probabilités:

$$P(\bigcup_{n=1}^k A_n) = \sum_{n=1}^k P(A_n), \quad \forall (A_n, n=1, 2, \dots, k) \subseteq \mathbf{A} \quad (1.3)$$

Démonstration: Soit $(A_n, n=1, 2, \dots, k) \subseteq \mathbf{A}$ une suite finie d'événements disjoints, i.e. $A_i \cap A_j = \emptyset$ si $i \neq j$, et soit $A = \bigcup_{n=1}^k A_n$. Considérons la transaction constituée de la vente du pari A , suivie de l'achat successif de tous les paris $(A_n, n=1, 2, \dots, k)$. Le bénéfice net de cette transaction est donné par $-G(A, \varphi) + \sum_{n=1}^k G(A_n, \varphi)$, ou encore $P(A) - A(\varphi) + \sum_{n=1}^k [A_n(\varphi) - P(A_n)]$. Cette expression peut être simplifiée, puisque $A(\varphi) = \sum_{n=1}^k A_n(\varphi)$, étant donné que les

événements sont disjoints. On obtient donc finalement comme bénéfice net $P(A) - \sum_{n=1}^k P(A_n)$. Le principe d'évitement des pertes sûres implique que cette quantité doit être positive ou nulle, d'où l'on tire $P(A) \geq \sum_{n=1}^k P(A_n)$. On obtient l'inégalité inverse en appliquant la même démarche pour l'achat du pari A , suivi de la vente de tous les paris $(A_n, n=1, 2, \dots, k)$. ■

Le principe d'évitement des pertes sûres ne permet cependant pas de justifier l'additivité des probabilités pour des suites dénombrables d'événements. Ce passage à la limite est plutôt une commodité mathématique qui permet d'utiliser les résultats de la théorie de l'intégration et de la mesure pour la manipulation de probabilités subjectives.

1.1.3 Espérance mathématique et prévision

On nomme **espérance mathématique** l'intégrale d'une fonction (intégrable) à valeur réelle $Z: \mathcal{G} \rightarrow \mathbf{R}$, par rapport à une mesure de probabilité P , quantité que l'on note $E^P(Z)$:

$$E^P(Z) = \int Z dP \quad (1.4)$$

On omettra d'indexer l'espérance par rapport à la mesure de probabilité lorsque cela ne porte pas à confusion, en écrivant plus simplement $E(Z) = \int Z dP$. On observe que l'espérance mathématique de la fonction indicatrice $A(\varphi)$ donne la probabilité de l'événement correspondant: $E[A(\varphi)] = \int A(\varphi) dP = \int_A dP(\varphi) = P(A)$. Ceci suggère une interprétation utile d'une espérance mathématique, et une généralisation du concept de pari. Jusqu'à maintenant, nous avons supposé qu'un pari était représenté par une fonction indicatrice $A(\varphi)$, puisqu'il s'agissait de parier sur un événement pour gagner un prix de valeur unitaire. On peut interpréter en fait toute fonction Z comme un pari donnant droit à un prix de valeur $Z(\varphi)$ après observation de φ . Dans cette optique, $E(Z)$ s'interprète à la fois comme le prix maximal d'achat de ce pari, et comme le prix minimal de vente. On peut donc dire qu'il s'agit d'une **prévision** de la valeur du pari Z avant observation de φ .

En utilisant les propriétés de l'intégrale (cf. Bartle, 1995), on montre que $E(-Z) = -E(Z)$ et que $\sum_{n=1}^k E(Z_n) = E(\sum_{n=1}^k Z_n)$, ce qui constitue la règle d'addition des prévisions.

1.1.4 Construction d'un espace de probabilités subjectives

Rappelons d'abord la définition d'une distribution de probabilité lorsque le référentiel est un sous-ensemble (mesurable) \mathfrak{S} de \mathbf{R}^n . On peut montrer que toute fonction intégrable non négative $f: \mathfrak{S} \rightarrow \mathbf{R}$, $f \geq 0$, génère une mesure (cf. Bartle, 1995). Ceci signifie que la fonction $\mu(A) = \int \dots \int_A f(z_1, \dots, z_n) dz_1 \dots dz_n$ est une mesure. Si de plus l'intégrale de cette fonction sur \mathfrak{S} est unitaire, alors f génère une mesure de probabilité. On dira alors que f est une **distribution de probabilité**. Inversement, pour toute mesure de probabilité P absolument continue, il existe une distribution de probabilité f qui permet de la générer.

Pour modéliser à l'aide d'un espace de probabilités les connaissances que possède un individu sur l'état de la nature $\theta \in \Theta$ et l'information supplémentaire apportée par une observation $x \in \Omega$, il faut construire une mesure de probabilité P sur une tribu adéquate de l'espace-produit $\mathfrak{S} = \Theta \times \Omega$. Dans la grande majorité des situations, il est possible de générer P à partir d'une distribution de probabilité. Pour simplifier la notation, supposons que Θ et Ω sont des sous-ensembles (mesurables) de \mathbf{R}^n . Dans ce cas, pour toute mesure de probabilité P absolument continue, il existe une distribution de probabilité $f: \Theta \times \Omega \rightarrow \mathbf{R}$ telle que:

$$P(B \times C) = \int_B \int_C f(x, \theta) dx d\theta, \quad \forall (B \times C) \in \mathbf{A} \quad (1.5)$$

On nomme $f(x, \theta)$ la distribution de probabilité jointe. Il est plus facile pour un individu de spécifier $f(x, \theta)$ plutôt que P , mais cela reste une tâche souvent trop ardue. On préfère donc en pratique obtenir la distribution jointe à partir de fonctions plus simples.

À partir de la distribution de probabilité jointe, on peut obtenir des distributions de probabilité **marginales** $f(x)$ et $b(\theta)$ pour x et θ qui permettent de générer respectivement les probabilités des événements de la forme $\Theta \times C$ et $B \times \Omega$:

$$P(B) = P(B \times \Omega) = \int_B b(\theta) d\theta, \quad b(\theta) = \int_{\Omega} f(x, \theta) dx, \quad \forall (B \times \Omega) \in \mathbf{A} \quad (1.6)$$

$$P(C) = P(\Theta \times C) = \int_C f(x) dx, \quad f(x) = \int_{\Theta} f(x, \theta) d\theta, \quad \forall (\Theta \times C) \in \mathbf{A} \quad (1.7)$$

En somme, on arrive à construire à partir de $f(x, \theta)$ des espaces de probabilités pour les référentiels Θ et Ω , en considérant la restriction de \mathbf{A} et de P à ces référentiels. La distribution $b(\theta)$ résume les connaissances subjectives de l'individu sur l'état de la nature θ , et la distribution $f(x)$ synthétise les idées a priori de l'individu sur la valeur de x qu'il prévoit

observer. On ne peut malheureusement pas construire $f(x, \theta)$ à partir de $b(\theta)$ et $f(x)$, car on ne connaît alors pas le lien entre θ et x .

On peut montrer qu'en divisant la distribution jointe par une distribution marginale, on obtient une autre distribution de probabilité, que l'on qualifie de **conditionnelle**. De telles distributions peuvent être utilisées pour construire la distribution jointe. Par exemple, pour θ fixé, le rapport $f(x|\theta) = f(x, \theta)/b(\theta)$ constitue une distribution de probabilité sur Ω (dans la mesure où $b(\theta) > 0$). La famille de distributions $L_x: \theta \rightarrow f(x|\theta)$ peut être interprétée comme un modèle pour la distribution d'échantillonnage de x , θ représentant alors les paramètres de ce modèle. Ceci suggère une procédure pratique pour construire un espace de probabilités subjectives sur $\Theta \times \Omega$:

- résumer l'information disponible sur l'état de la nature θ à l'aide d'une distribution de probabilité $b(\theta)$;
- choisir une famille paramétrique de distributions $L_x(\theta)$, telle que $L_x(\theta) = f(x|\theta)$ correspond à la distribution d'échantillonnage de x pour chaque valeur de θ ,
- obtenir la distribution de probabilité jointe par le produit de $b(\theta)$ et $f(x|\theta)$:

$$f(x, \theta) = f(x|\theta) \cdot b(\theta);$$
- générer P par intégration de la distribution jointe $f(x, \theta)$.

Pour simplifier, nous nommerons **loi de probabilité** toute famille paramétrique de distributions de probabilité.

1.1.5 Distribution prévisionnelle

On remarque que la distribution marginale $f(x) = \int_{\Theta} f(x, \theta) d\theta$ peut s'écrire sous la forme $f(x) = \int_{\Theta} f(x|\theta) b(\theta) d\theta$, puisque $f(x, \theta) = f(x|\theta) \cdot b(\theta)$. Selon la définition de l'espérance mathématique, on peut aussi écrire $f(x) = E^{b(\theta)}[f(x|\theta)]$. Ainsi, $f(x)$ correspond à la prévision de la distribution d'échantillonnage. C'est pourquoi on qualifie cette distribution de **prévisionnelle**.

1.1.6 Distribution cumulée

Lorsqu'une distribution de probabilité $f(x)$ est définie sur la droite des réels \mathbf{R} , il est utile de définir la distribution cumulée correspondante $F(x)$, qui représente la probabilité de l'événement $(-\infty, x]$:

$$F(x) = P((-\infty, x]) = \int_a^x f(t) dt \quad (1.8)$$

1.1.7 Règle de Bayes

Dans la majorité des problèmes d'inférence statistique, on ne s'intéresse pas à la distribution jointe $f(x, \theta)$, mais on cherche plutôt à construire une distribution de probabilité pour θ en tenant compte d'une information d'échantillonnage x . C'est-à-dire que l'on cherche à modéliser ce que l'observation x nous apprend sur la valeur de θ . Une façon de résoudre ce problème consiste à étudier la distribution conditionnelle $b(\theta|x) = f(x, \theta)/f(x)$, qui constitue une distribution de probabilité sur Θ pour x fixé, dans la mesure où $f(x) > 0$. Puisque $f(x, \theta) = f(x|\theta) \cdot b(\theta)$, on montre que:

$$b(\theta|x) = \frac{f(x|\theta)b(\theta)}{\int_{\Theta} f(x|\theta)b(\theta)d\theta}, \text{ si } f(x) > 0 \quad (1.9)$$

Démonstration: On sait que $b(\theta|x) \cdot f(x) = f(x, \theta) = f(x|\theta) \cdot b(\theta)$, d'où l'on tire:

$$b(\theta|x) = \frac{f(x|\theta)b(\theta)}{f(x)}, \text{ si } f(x) > 0$$

On peut enfin remplacer $f(x)$ par sa définition: $f(x) = \int_{\Theta} f(x, \theta)d\theta = \int_{\Theta} f(x|\theta)b(\theta)d\theta$. ■

On nomme cette équation la règle de Bayes, car elle peut être utilisée comme règle de mise à jour d'une distribution de probabilité $b(\theta)$ sur l'état de la nature θ après observation d'une information x . Vu sous cet angle:

- on nomme $b(\theta)$ la **distribution a priori** de l'état de la nature, avant observation de x ;
- on nomme $b(\theta|x)$ la **distribution a posteriori** de l'état de la nature, après observation de x .

On remarque que l'information x n'influence la distribution a posteriori qu'à travers la fonction $f(x|\theta)$. Lorsque l'on utilise la règle de Bayes pour mettre à jour une distribution a priori après observation d'une information x , on doit calculer $b(\theta|x)$ pour les différentes valeurs de θ , mais pour une seule valeur de x . Ainsi, au niveau de la règle de Bayes, $f(x|\theta)$ est

utilisé comme une fonction de θ , pour x fixé, et non l'inverse. C'est pourquoi on écrit souvent le modèle $f(x|\theta)$ sous la forme d'une fonction de θ , $L_x(\theta)=f(x|\theta)$, que l'on nomme la vraisemblance de θ . On peut ainsi écrire la règle de Bayes sous la forme simplifiée suivante:

$$b(\theta|x) \propto b(\theta) \cdot L_x(\theta) \quad (1.10)$$

La distribution a posteriori dépend donc de deux fonctions, l'une représentant l'information a priori et l'autre représentant l'information d'échantillonnage. La règle de Bayes agit comme un processeur séquentiel d'information, permettant de combiner les deux sources d'information. Si, par la suite, de nouvelles informations y deviennent disponibles, la règle de Bayes pourra de nouveau être appliquée. La distribution a posteriori $b(\theta|x)$ obtenue lors de la première application de la règle de Bayes tiendra alors le rôle de la distribution a priori pour cette nouvelle application de la règle de Bayes, puisque cette distribution modélise l'ensemble de l'information disponible avant observation de y .

Pour $b(\theta)$ et $f(x|\theta)$ fixé, on nomme **espérance a priori** d'une fonction (intégrable) $Z:\Theta \rightarrow \mathbf{R}$ l'intégrale $E(Z)=\int_{\Theta} Z(\theta)b(\theta)d\theta$, et **espérance a posteriori** l'intégrale $E(Z|x)=\int_{\Theta} Z(\theta)b(\theta|x)d\theta$. Il est intéressant de remarquer que $E(Z|x)$ est une fonction (intégrable) de Ω vers \mathbf{R} . On peut montrer que $E^{f(x)}[E(Z|x)]=E^{b(\theta)}(Z)$, c'est-à-dire l'espérance a priori de Z .

Démonstration:

$$\begin{aligned} E^{f(x)}[E(Z|x)] &= \int_{\Omega} \int_{\Theta} Z(\theta)b(\theta|x)f(x)d\theta dx, \quad \text{par définition} \\ &= \int_{\Omega} \int_{\Theta} Z(\theta)b(\theta)f(x|\theta)d\theta dx, \quad \text{règle de Bayes} \\ &= \int_{\Theta} \int_{\Omega} Z(\theta)b(\theta)f(x|\theta)dx d\theta, \quad \text{Th. de Fubini} \\ &= \int_{\Theta} Z(\theta)b(\theta) \left\{ \int_{\Omega} f(x|\theta)dx \right\} d\theta \\ &= \int_{\Theta} Z(\theta)b(\theta)d\theta, \quad \text{car } \int_{\Omega} f(x|\theta)dx = 1 \text{ par définition} \\ E^{f(x)}[E(Z|x)] &= E^{b(\theta)}(Z) \quad \blacksquare \end{aligned}$$

Walley (1991) nomme cette propriété le **principe de conglomérabilité**.

1.1.8 Indépendance et échangeabilité

La modélisation de l'information a priori par une distribution $b(\theta)$ est souvent difficile, entre autres parce que les paramètres du vecteur θ n'ont pas toujours un sens physique, et ne peuvent généralement pas être observés directement. On peut donc remettre en question la pertinence d'une analyse basée sur la combinaison d'un modèle paramétrique $f(x|\theta)$ et d'une distribution a priori $b(\theta)$ comme moyen de prendre en compte une information a priori subjective, lorsque la variable à laquelle on s'intéresse véritablement est x .

Une alternative à cette approche consiste à construire directement la distribution $f(x)$ à partir des informations a priori. Cette façon de procéder se complique cependant lorsque l'on s'intéresse à une suite $\mathbf{x}=(x_n, n=1,2,\dots,k)$ d'observations. Il faudrait alors construire explicitement la distribution jointe de cette suite $f(\mathbf{x})=f(x_1,x_2,\dots,x_k)$, tâche dont la difficulté croît rapidement avec k . Une façon de simplifier l'analyse consiste à supposer que les observations sont indépendantes, c'est-à-dire que la distribution jointe $f(x_1,x_2,\dots,x_k)$ est donnée par le produit des distributions marginales $f(x_n)$:

$$f(\mathbf{x})=f(x_1,x_2,\dots,x_k)=\prod_{n=1}^k f(x_n) \quad (1.11)$$

On dira dans ce cas que $\mathbf{x}=(x_n, n=1,2,\dots,k)$ constitue un échantillon aléatoire. Cette hypothèse peut être acceptable pour représenter l'incertitude sur des observations de phénomènes distincts séparés dans le temps et dans l'espace, mais est généralement plus difficile à justifier lorsque la suite (x_n) représente un ensemble de variables ayant des liens causaux plus importants. Par exemple, lorsque (x_n) est une série chronologique représentant l'évolution de la valeur d'une quantité physique au cours du temps, l'hypothèse d'indépendance est souvent trop forte.

Si les observations ont toutes le même domaine de définition, on peut dans certains cas au moins faire l'hypothèse qu'elles sont échangeables, c'est-à-dire que la valeur de $f(x_1,x_2,\dots,x_k)$ ne dépend pas de l'ordre dans lequel sont présentées les informations. Dans ce cas, peut associer la même valeur de f à toute permutation de la suite (x_n) . Ceci peut simplifier la construction de la distribution jointe, mais peut aussi permettre de justifier l'emploi d'un modèle paramétrique et d'une distribution a priori. En effet, dans les situations où l'hypothèse d'échangeabilité est acceptable quel que soit le nombre k d'observations formant la suite,

Loève (1978) montre que, sous certaines conditions de régularité, il existe un modèle paramétrique $f(x_1, x_2, \dots, x_k | \theta)$ et une distribution a priori $b(\theta)$ telle que:

$$f(x_1, x_2, \dots, x_k) = \int \prod_{n=1}^k f(x_n | \theta) \cdot b(\theta) d\theta \quad (1.12)$$

En conséquence, un individu qui construit directement une distribution jointe $f(x_1, x_2, \dots, x_k)$ pour des observations qu'il considère échangeables quelle que soit la longueur de la série agit donc comme s'il fixait un modèle paramétrique $f(x_n | \theta)$ pour chaque x_n , comme s'il supposait l'indépendance entre les observations pour θ fixé, et comme s'il avait en tête une distribution a priori sur ces paramètres.

1.1.9 Exemple d'utilisation de la règle de Bayes

La mécanique de l'inférence bayésienne peut être illustrée par un exemple emprunté à Caselton et Luo (1992). Nous utiliserons quelques propriétés de la loi bêta, résumées par Bernardo et Smith (1994). La loi bêta est constituée de la famille de distributions suivantes:

$$b(\theta; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}, \alpha, \beta > 0 \quad (1.13)$$

où α et β sont les paramètres de la loi et $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ correspond à la fonction gamma.

Rappelons que $\Gamma(n+1) = n!$ pour n entier, et que $z \cdot \Gamma(z) = \Gamma(z+1)$ (Abramowitz et Stegun, 1970).

La moyenne $E(\theta)$ et la variance $V(\theta)$ de la loi bêta sont respectivement données par les expressions suivantes:

$$E(\theta) = \frac{\alpha}{\alpha + \beta} \quad (1.14)$$

$$V(\theta) = \frac{\alpha \cdot \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)} \quad (1.15)$$

Imaginons maintenant qu'une localité située en bordure d'un cours d'eau soit protégée contre les crues par une digue permettant de contenir un débit de $40 \text{ m}^3/\text{s}$. Dans le cadre d'une étude sur la pertinence de relever le niveau de cette digue de façon à contenir un débit de $50 \text{ m}^3/\text{s}$, la municipalité demande à un hydrologue d'établir la probabilité qu'une année quelconque le débit soit supérieur à $40 \text{ m}^3/\text{s}$ sans dépasser $50 \text{ m}^3/\text{s}$ (car la nouvelle digue serait plus utile que l'ancienne uniquement dans ces conditions). Dénotons par θ cette probabilité. En se basant sur un échantillon \mathbf{x} de n observations du débit maximum annuel et sur son

expérience, l'hydrologue pourrait, par introspection, fournir directement une valeur précise pour cette probabilité subjective, mais l'utilisation de la règle de Bayes peut lui permettre:

- d'explicitier la méthode qu'il utilise pour tenir compte de l'information d'échantillonnage \mathbf{x} ;
- de représenter par une distribution $b(\theta|\mathbf{x})$ l'incertitude qui subsiste sur la valeur de θ .

Supposons que l'hydrologue considère que les observations du débit maximum annuel sont indépendantes et identiquement distribuées. Dans ces conditions, on montre que la probabilité que le débit maximum annuel se situe exactement à r reprises entre 40 et 50 m³/s au cours d'une période de n années suit une loi binomiale de paramètre θ .

$$f(r, n|\theta) = \binom{n}{r} \theta^r (1 - \theta)^{n-r}, \text{ où } \binom{n}{r} = \frac{n!}{r!(n-r)!} \quad (1.16)$$

En accord avec la terminologie usuelle pour le modèle binomial, on peut qualifier de succès l'observation d'un débit maximum annuel entre 40 et 50 m³/s. On dispose donc d'un modèle pour la distribution d'échantillonnage du nombre r de succès, qui peut aussi être mis sous la forme d'une fonction de vraisemblance $L_{r,n}(\theta) = f(r, n|\theta)$. Pour appliquer la règle de Bayes, il reste à modéliser l'information a priori sur l'état de la nature θ par une distribution de probabilité $b(\theta)$. Pour faciliter les calculs, on peut choisir $b(\theta)$ dans la famille de distributions correspondant à la loi bêta (θ, α, β) . Cette famille de distributions a l'avantage d'être conjuguée par rapport à la loi binomiale, ce qui signifie que lorsque l'on combine à l'aide de la règle de Bayes une distribution a priori de la famille bêta avec une loi binomiale, on obtient une distribution a posteriori bêta de paramètres modifiés. Dans ce cas, si les paramètres de la distribution a priori sont α et β , et que l'on observe r succès sur une période de n années, alors les paramètres de la distribution a posteriori sont $\alpha+r$ et $\beta+n-r$ (Bernardo et Smith, 1994).

L'hydrologue peut ainsi, en fixant les paramètres α et β , construire rapidement $b(\theta)$ et obtenir facilement $b(\theta|\mathbf{x})$. Pour fixer ces paramètres, il est utile d'étudier l'influence qu'ils ont sur l'espérance a priori et a posteriori de θ . On montre que:

$$E(\theta) = \frac{\alpha}{\alpha + \beta}, \quad E(\theta|\mathbf{x}) = \frac{r + \alpha}{n + \alpha + \beta} \quad (1.17)$$

Le rapport de α à la somme de α et β fournit donc une estimation de la probabilité θ . On observe aussi que, pour le calcul de $E(\theta|\mathbf{x})$, le poids du paramètre α est analogue à celui de r , alors que le poids de $(\alpha+\beta)$ est analogue à celui de n . Ainsi, la somme de α et β devrait

refléter le poids de l'information a priori dont dispose l'hydrologue, mesuré dans une unité homogène à une taille d'échantillon.

Imaginons que l'hydrologue considère, après avoir observé le site, que la probabilité de succès est de l'ordre de 10%, et que ce jugement devrait avoir un poids équivalent à dix années d'observation du débit maximum annuel. Dans ce cas, il pourrait fixer $\alpha=1$ et $\beta=9$. Supposons aussi qu'un échantillon de $n=6$ années soit disponible, pendant lesquelles on a observé $r=3$ succès. On peut alors combiner l'information a priori de l'hydrologue, modélisée par une distribution $b(\theta, \alpha=1, \beta=9)$, avec la vraisemblance $L_{3,6}(\theta)$ pour obtenir une distribution a posteriori $b(\theta|\mathbf{x}; \alpha_{(\mathbf{x})}=4, \beta_{(\mathbf{x})}=12)$.

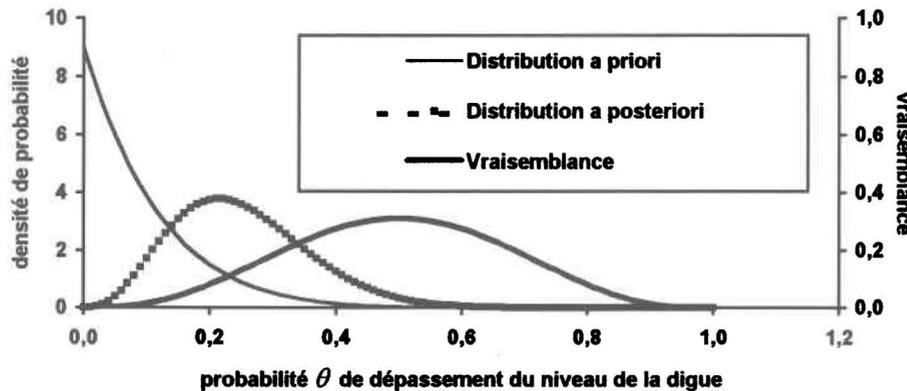


Figure 1.1: Distribution a priori, vraisemblance et distribution a posteriori pour la probabilité d'observer un débit entre 40 et 50 m³/s

La Figure 1.1 permet de visualiser la forme des distributions a priori et a posteriori, ainsi que celle de la vraisemblance. La distribution a priori accorde plus de poids aux faibles valeurs de θ , mais la vraisemblance favorise les valeurs de θ proches de $\frac{1}{2}$. La distribution a posteriori, qui combine les deux sources d'information, pointe vers une valeur de θ plus près de $\frac{1}{4}$.

On peut calculer la variance des distributions a priori et a posteriori à partir des propriétés de la loi bêta énoncées au début de la section. On obtient une variance de $V(\theta)=9/1100$ pour la distribution a priori, et une variance de $V(\theta|\mathbf{x})=9/816$ pour la distribution a posteriori. Il est intéressant de noter que la variance de la distribution a posteriori est plus élevée que celle de la distribution a priori. Ainsi, l'opinion a posteriori de l'hydrologue est en

quelque sorte plus vague que son opinion a priori. Ceci se produit parce que la fréquence observée $r/n=1/2$ est assez différente de la prévision a priori $E(\theta)=1/10$. On peut vérifier que si aucun débordement n'avait été observé durant les $n=6$ années d'observations, la variance a posteriori aurait été environ deux fois plus faible que la variance a priori.

1.2 Théorie bayésienne de la décision

Après avoir introduit le concept d'utilité d'une action, nous montrons comment un modèle d'incertitude à base de probabilités subjectives peut être utilisé pour prévoir et comparer les conséquences de différentes actions. Notre présentation est inspirée de celle de Lindley (1994).

1.2.1 Utilité d'une action

Imaginons qu'un individu soit tenu de choisir entre un ensemble $S=(a_n)$ d'actions possibles. La théorie bayésienne de la décision propose une méthode pour classer ces actions. Cette technique consiste d'abord à évaluer les conséquences de ces actions. Chaque action peut évidemment avoir plusieurs effets différents. Soit $c(a)$ un vecteur représentant chacun des effets de l'action a . Prenons l'exemple des rejets d'un polluant organique d'une usine dans un cours d'eau. Les conséquences, en fonction de la quantité rejetée, pourraient être mesurés par exemple:

- par le coût du traitement nécessaire pour limiter la quantité rejetée à ce niveau;
- par un paramètre de la qualité de l'eau en aval de l'usine, par exemple le taux d'oxygène;
- par le niveau de protestation de la population contre l'usine.

Pour mieux comprendre le concept d'utilité d'une action, imaginons dans un premier temps que l'état de la nature est connu avec précision. Ainsi, supposons que pour chaque action a on peut établir exactement la valeur de ce vecteur de conséquences $c(a)$. Avec de telles conséquences, il devient difficile de comparer deux actions car cela demande de combiner plusieurs critères pour obtenir une seule valeur mesurant l'utilité de chaque action. On peut par contre facilement comparer deux actions très particulières: celle dont les conséquences sont les moins défavorables, et celle dont les conséquences sont les plus désastreuses. Pour les critères énumérés précédemment, on peut identifier les conséquences les moins défavorables et les plus désastreuses:

Critère	Valeur la moins défavorable (action b_1)	Valeur la plus désastreuse (action b_0)
coût du traitement	0\$	100% de la capacité d'investissement
taux d'oxygène en aval	le même qu'en amont	0%
niveau de protestation	0%	100%

Évidemment, les deux situations sont hypothétiques, mais on peut quand même imaginer des actions qui pourraient mener à de telles conséquences. Soit b_0 une action ayant les conséquences les plus défavorables et b_1 une action ayant les conséquences les moins défavorables. N'importe quel gestionnaire saurait quelle action choisir entre b_0 et b_1 , mais typiquement de telles actions ne feront pas partie de l'ensemble S des choix possibles. Pour comparer les différentes actions d'un ensemble S , on peut construire une fonction U à valeur réelle telle que $U(a) > U(b)$ si et seulement si a est préféré à b . Étant donné les conséquences des actions b_0 et b_1 , on pourrait fixer par convention $U(b_0) = 0$ et $U(b_1) = 1$. On devrait alors logiquement avoir, pour toute action $a \in S$, $U(b_0) \leq U(a) \leq U(b_1)$.

De la même façon que nous avons imaginé qu'il existait des actions b_0 et b_1 menant à des conséquences extrêmes, imaginons maintenant toute une famille d'actions intermédiaires b_q , $q \in [0, 1]$, dont les conséquences dépendent d'un tirage au sort, c'est-à-dire que l'action b_q conduit avec probabilité q aux mêmes conséquences que l'action b_1 , et avec probabilité $1 - q$ aux mêmes conséquences que l'action b_0 . Encore une fois, un décideur n'aurait pas de difficulté à choisir entre les actions de la famille b_q . En effet, il est naturel de supposer qu'il préférerait les actions menant aux conséquences $c(b_1)$ avec une probabilité plus élevée. Malheureusement, les actions de l'ensemble S ne font habituellement pas partie de la famille b_q .

La définition de l'utilité $U(a)$ d'une action $a \in S$ repose sur l'hypothèse que pour tout $a \in S$ il existe une action imaginaire $b_{q(a)}$ que le décideur considère équivalente, c'est à dire qu'il est indifférent entre choisir a ou $b_{q(a)}$. Si c'est le cas, on peut en effet définir l'utilité d'une action a par la probabilité $q(a)$ de l'action imaginaire $b_{q(a)}$ qui lui est associée: $U(a) = q(a)$.

Cette façon de procéder pour classer les actions permet de transformer un problème où l'on doit choisir entre différentes actions $a \in S$ ayant des conséquences difficiles à comparer mais connues précisément en un problème où l'on doit choisir entre des actions b_q ayant des

conséquences faciles à comparer mais aléatoires. Chaque action $a \in \mathcal{S}$ étant caractérisée par la probabilité $q(a)$ d'occurrence de l'ensemble des conséquences les moins défavorables $c(b_1)$, il paraît naturel de maximiser la probabilité d'occurrence q des conséquences les moins défavorables, c'est-à-dire de choisir une action \hat{a} d'utilité maximale:

$$U(\hat{a}) = \sup_{a \in \mathcal{S}} q(a) = \sup_{a \in \mathcal{S}} U(a) \quad (1.18)$$

En pratique, il n'est pas nécessaire de construire une fonction d'utilité comprise entre 0 et 1, puisque l'action maximisant toute transformation linéaire de $q(a)$ est la même. Cependant, pour que l'interprétation de l'utilité donnée ici soit possible, il faut au moins que la fonction d'utilité soit bornée. Notons que Robert (1992) propose une définition plus générale et plus rigoureuse de l'utilité, qui ne demande pas de supposer que cette fonction soit bornée.

1.2.2 Principe de maximisation de l'utilité espérée

Il est souvent difficile d'établir les conséquences exactes $c(a)$ d'une action a . En effet, ces conséquences dépendent d'un certain nombre de variables non contrôlées qui ont une influence sur les conséquences. On peut représenter ces variables par un vecteur $\theta \in \Theta$, que nous avons nommé précédemment l'état de la nature. Pour θ connu, admettons qu'il soit possible de déterminer les conséquences $c(a, \theta)$ d'une action. En utilisant le procédé expliqué à la section précédente, il est alors aussi possible d'établir l'utilité $U(a, \theta)$ de chaque action pour θ connu: $U(a, \theta) = q(a|\theta)$. Ceci permet de choisir une action d'utilité maximale pour θ connu. Cependant, en pratique, il faut choisir une action sans connaître précisément θ .

Si l'incertitude sur θ est modélisée par une distribution $b(\theta)$, on peut obtenir $q(a)$, donc $U(a)$, par intégration: $U(a) = q(a) = \int q(a|\theta) b(\theta) d\theta$. Lorsque les conséquences des actions dépendent d'un état de la nature inconnu θ , l'action optimale \hat{a} est donc celle qui maximise l'espérance de l'utilité:

$$U(\hat{a}) = \sup_{a \in \mathcal{S}} E[U(a, \theta)] \quad (1.19)$$

Il importe de remarquer que nous sommes arrivés à ce résultat sans imaginer une répétition du processus de décision: ce n'est pas l'utilité moyenne qui est maximisée, mais bien l'utilité ponctuelle de la décision.

1.2.3 Risque et regret d'une action

Dans plusieurs problèmes, l'utilité des actions n'est jamais positive. Pour éviter de travailler avec des quantités négatives, il est alors plus intéressant de définir une mesure de désutilité $g(a, \theta) = -U(a, \theta)$, que l'on nomme aussi fonction de coût. On nomme la prévision $R(a)$ de la désutilité d'une action a le risque de cette action:

$$R(a) = E^{b(\theta)}[g(a, \theta)] \quad (1.20)$$

Dans ce cas, l'action optimale \hat{a} est celle qui minimise le risque:

$$R(\hat{a}) = \inf_{a \in S} R(a) \quad (1.21)$$

On qualifie de risque de Bayes la désutilité de l'action optimale. On peut aussi s'intéresser à la différence entre le coût d'une action et le coût minimal en supposant θ connu. On nomme cette quantité le regret de l'action. Pour éviter toute confusion inutile, nous utiliserons la notation anglo-saxonne qui consiste à représenter le regret par la fonction $OL(a, \theta)$, qui signifie *Opportunity Loss*:

$$OL(a, \theta) = g(a, \theta) - \inf_{a \in S} g(a, \theta) \quad (1.22)$$

En pratique, on ne connaît pas exactement la valeur de θ , et il n'est possible que d'effectuer une prévision du regret, que nous dénoterons $XOL(a)$, pour *eXpected Opportunity Loss*:

$$XOL(a) = R(a) - E^{b(\theta)} \left[\inf_{a \in S} g(a, \theta) \right] \quad (1.23)$$

Ces calculs de risque et de prévision du regret forment ce que l'on nomme l'analyse a priori, puisqu'ils se basent sur la distribution a priori $b(\theta)$. Si l'on observe une information x , on peut alors mettre à jour cette distribution par application de la règle de Bayes pour obtenir une distribution a posteriori $b(\theta|x)$. Ceci permet de déterminer le risque a posteriori $R(a|x)$, l'action optimale a posteriori $\hat{a}_{(x)}$, le risque de Bayes a posteriori $R(\hat{a}_{(x)}|x)$ et le regret espéré a posteriori $XOL(a|x)$:

$$R(a|x) = E^{b(\theta|x)}[g(a, \theta)] \quad (1.24)$$

$$R(\hat{a}_{(x)}|x) = \inf_{a \in A} R(a|x) \quad (1.25)$$

$$XOL(a|x) = R(a|x) - E^{b(\theta|x)} \left[\inf_{a \in \mathcal{S}} g(a, \theta) \right] \quad (1.26)$$

1.2.4 Aversion au risque

Même lorsqu'on peut trouver un nombre unique pour mesurer les conséquences d'une action, il ne s'agit généralement pas d'une fonction d'utilité ou de désutilité. L'exemple le plus frappant est la valeur monétaire d'un bien. Lorsqu'on prend la peine d'établir la relation entre l'utilité d'une action et sa valeur monétaire, on s'aperçoit généralement que cette relation est concave, c'est-à-dire marginalement décroissante (Berger, 1985). Cette attitude se nomme l'aversion au risque: une augmentation de la valeur monétaire d'une action se traduit par une augmentation moins importante de l'utilité. En particulier, ceci signifie que les gains monétaires importants ont une utilité qui tend à plafonner, alors que les pertes importantes ont une désutilité qui croît très rapidement. Ainsi, un individu ayant une aversion au risque préfère un gain certain à un pari de même espérance. Ceci explique pourquoi il est souvent raisonnable d'acheter une assurance pour éviter des pertes importantes. En effet, bien qu'il s'agisse d'une action ayant normalement une espérance de gain négative, la désutilité associée aux pertes importantes peut dépasser le coût de l'assurance.

1.2.5 Exemple d'application de la théorie bayésienne de la décision

Reprenons l'exemple, introduit à la section 1.1.9, d'une localité protégée par une digue contre les crues de débit inférieur ou égal à $40\text{m}^3/\text{s}$. On envisage de relever cette digue de façon à pouvoir contenir un débit de $50\text{m}^3/\text{s}$. Si l'on évalue que le coût annuel de construction et d'entretien de la digue correspond à une proportion q des dommages engendrés par une inondation, serait-il rentable de relever le niveau de la digue?

On doit choisir entre une action α_1 consistant à relever le niveau de la digue pour une fraction q du coût annuel d'une inondation, et une action α_0 consistant à accepter le risque supplémentaire d'inondation. Si on néglige l'aversion au risque et si on accepte les hypothèses d'indépendance et de distribution identique des débits maximums annuels, l'utilité de cette action α_0 correspond à la probabilité θ d'observer une crue printanière entre 40 et $50\text{m}^3/\text{s}$, et l'on peut résumer toute l'information d'échantillonnage par une séquence $\mathbf{x}=(x_n)$ de variables de Bernoulli avec $x_n=1$ si le débit maximum annuel appartient à l'intervalle $(40,50]$, et $x_n=0$

autrement. La décision de moindre risque dépend de l'information disponible sur θ , qui constitue alors le paramètre d'une loi binomiale $f(r, n | \theta)$ régissant la probabilité d'observer r succès, i.e. r débits maximums annuels entre 40 et 50m³/s, sur une période de n années.

Supposons que l'information a priori puisse être encore une fois représentée par une distribution bêta, mais de paramètres $\alpha=1$ et $\beta=9$, que $r=3$ débordements de la digue se soient produits au cours des $n=6$ dernières années, mais que le débit n'a jamais dépassé 50m³/s.

1.2.5.1 Analyse a priori de la décision

En tenant compte uniquement de l'information a priori on veut choisir, entre a_0 et a_1 , l'action de risque minimal. Les fonctions de coût de ces deux actions sont respectivement $g(a_0, \theta) = \theta$ et $g(a_1, \theta) = q$, exprimées en proportion des dommages liés à une inondation annuelle du territoire. Le risque $R(a_0)$ de l'action a_0 est donc identique à la prévision $E^{b(\theta)}(\theta)$, alors que le risque $R(a_1)$ de l'action a_1 est précisément égal à q . Puisque $E^{b(\theta)}(\theta) = \alpha / (\alpha + \beta) = 1/10$, le risque de l'action a_0 est plus faible que celui de l'action a_1 si et seulement si $q > 1/10$. Par exemple, pour $q = 1/6$, le coût d'une nouvelle digue serait supérieur à la prévision du coût relié aux inondations que cette digue permettrait d'éviter, de sorte qu'il ne serait pas utile de la construire.

1.2.5.2 Analyse a posteriori de la décision

Comparons maintenant les actions a_0 et a_1 en tenant compte de l'information d'échantillonnage \mathbf{x} . Comme nous l'avons expliqué à la section 1.1.9, la distribution a posteriori obtenue dans ce cas est une loi bêta de paramètres $\alpha_{(\mathbf{x})} = 4$ et $\beta_{(\mathbf{x})} = 12$. Le risque a posteriori de l'action a_0 , qui correspond à la prévision a posteriori de θ , peut donc être calculé ainsi: $R(a_0 | \mathbf{x}) = E^{b(\theta | \mathbf{x})}(\theta) = \alpha_{(\mathbf{x})} / [\alpha_{(\mathbf{x})} + \beta_{(\mathbf{x})}] = 1/4$. L'action optimale a posteriori consiste alors à construire la digue si et seulement si $R(a_0 | \mathbf{x}) > R(a_1 | \mathbf{x})$, c'est-à-dire si et seulement si $q < 1/4$. Dans ce problème, la décision optimale est influencée par l'information d'échantillonnage. Si l'on veut valoriser les efforts requis pour effectuer une campagne d'échantillonnage, il est intéressant de quantifier cette influence.

1.3 Estimation de la valeur d'une information

La façon naturelle d'estimer la valeur d'une information, dans le cadre de la théorie bayésienne de la décision, consiste à en mesurer l'utilité. En effet, l'acquisition d'une information x permet de mettre à jour la distribution $b(\theta)$ modélisant l'incertitude sur l'état de la nature, et peut amener à modifier la décision optimale. Cette façon de procéder a été proposée par Raiffa et Schlaifer (1961), et fut introduite en hydrologie par Davis et Dvoranchik (1971) et par Davis et al. (1972).

1.3.1 Valeur a posteriori d'une information

A posteriori, la définition de la valeur d'une information x est simple: il s'agit de la réduction du risque apportée par une information nouvelle:

$$VI(x) = R(\hat{a}|x) - R(\hat{a}_{(x)}|x) = E^{b(\theta|x)}[g(\hat{a}, \theta) - g(\hat{a}_{(x)}, \theta)] \quad (1.27)$$

Il est important de noter que la valeur a posteriori d'une information x est définie par la différence entre le risque **a posteriori** de l'action optimale **a priori** et le risque de Bayes a posteriori. Il ne s'agit pas de la différence entre les risques de Bayes, que l'on peut noter $\Delta R(x) = R(\hat{a}) - R(\hat{a}_{(x)}|x)$. Alors que $\Delta R(x)$ peut être positif ou négatif, selon que l'information obtenue x corrobore l'idée a priori modélisée par $b(\theta)$ ou s'en éloigne largement, il est clair de par sa définition que $VI(x)$ ne peut être que positif ou nul. En effet, le risque de Bayes a posteriori correspondant au risque a posteriori minimal, on ne peut obtenir une quantité négative en le soustrayant du risque a posteriori de n'importe quelle action. Remarquons que la valeur a posteriori d'une information ainsi que la réduction du risque sont mesurés en unités d'utilité. On peut en fait interpréter $VI(x)$ comme l'utilité de l'action α^* consistant à observer x , évaluée a posteriori: $VI(x) = U(\alpha^*, x)$.

1.3.2 Valeur prévisionnelle d'une information

Il est en général nécessaire de mesurer la valeur de l'information de façon prévisionnelle, c'est-à-dire avant observation de x . Puisque $VI(x)$ correspond à une utilité conditionnelle, on peut définir la valeur prévisionnelle d'une information par l'espérance de $VI(x)$:

$$XVSI = E^{f(x)}[VI(x)] \quad (1.28)$$

Nous notons cette quantité $XVSI$ en suivant à nouveau la notation anglo-saxonne, c'est-à-dire en utilisant l'abréviation pour *eXpected Value of Sample Information*. On peut aussi définir $XVSI$ comme la prévision de la réduction du risque, ou encore la prévision de la réduction du regret espéré:

$$XVSI = E^{f(x)}[\Delta R(x)] = R(\hat{a}) - E^{f(x)}[R(\hat{a}_{(x)}|x)] \quad (1.29)$$

$$XVSI = XOL(\hat{a}) - E^{f(x)}[XOL(\hat{a}_{(x)}|x)] \quad (1.30)$$

Ces deux autres définitions sont équivalentes à la première.

Démonstration: Puisque $R(\hat{a})$ et $R(\hat{a}|x)$ constituent respectivement l'espérance a priori et a posteriori de la fonction $g(\hat{a}, \theta)$, on sait que $E^{f(x)}[R(\hat{a}|x)] = R(\hat{a})$ par application du principe de conglomérabilité. Ceci suffit pour démontrer l'équivalence entre les définitions (1.28) et (1.29). En effet:

$$E[VI(x)] - E[\Delta R(x)] = E[R(\hat{a}|x) - R(\hat{a}_{(x)}|x)] - E[R(\hat{a}) - R(\hat{a}_{(x)}|x)] = E[R(\hat{a}|x)] - R(\hat{a}) = 0.$$

En remplaçant dans (1.30) le regret espéré par sa définition [eq. (1.23)], on obtient:

$$\begin{aligned} XOL(\hat{a}) - E^{f(x)}[XOL(\hat{a}_{(x)}|x)] &= \\ R(\hat{a}) - E^{b(\theta)}\left[\inf_{a \in S} g(a, \theta)\right] - E^{f(x)}[R(\hat{a}_{(x)}|x)] &+ E^{f(x)}\left[E^{b(\theta|x)}\left[\inf_{a \in S} g(a, \theta)\right]\right] \\ = E^{f(x)}[\Delta R(x)] - E^{b(\theta)}\left[\inf_{a \in S} g(a, \theta)\right] &+ E^{f(x)}\left[E^{b(\theta|x)}\left[\inf_{a \in S} g(a, \theta)\right]\right] \end{aligned}$$

Pour montrer que la différence des regrets est équivalente à l'espérance de la différence des risques, il suffit donc de montrer que $E^{f(x)}\left[E^{b(\theta|x)}\left[\inf_{a \in S} g(a, \theta)\right]\right] = E^{b(\theta)}\left[\inf_{a \in S} g(a, \theta)\right]$. C'est aussi une conséquence du principe de conglomérabilité. Les trois définitions sont donc équivalentes. ■

Il sera parfois utile d'indexer $XVSI$ par la distribution a priori et la fonction de coût utilisées pour déterminer la valeur de l'information. Ainsi, $XVSI_{b,g}$ fera référence à la prévision de la valeur de l'information basée sur la distribution a priori $b(\theta)$ et la fonction de coût $g(a, \theta)$.

Davis et al. (1979) ont étendu la définition de la valeur de l'information aux situations où il n'est pas possible de choisir l'action minimisant le risque de Bayes a posteriori. Supposons qu'une règle $\delta(x)$ soit employée pour choisir l'action a posteriori, au lieu de

l'action $\hat{a}_{(x)}$ minimisant le risque de Bayes. Dans ce cas, la valeur de l'information peut être mesurée par:

$$XVSI(\delta) = R(\hat{a}) - E^{f(x)}[R(\delta(x)|x)] \quad (1.31)$$

Cette quantité peut être négative, ce qui implique qu'une information supplémentaire peut être nuisible lorsque l'on prend une décision sous-optimale, c'est-à-dire qui ne minimise pas le risque de Bayes.

1.3.3 Borne maximale pour la valeur prévisionnelle de l'information

En pratique, même si les trois définitions sont équivalentes, la plus facile à appliquer est généralement $XVSI = R(\hat{a}) - E[R(\hat{a}_{(x)}|x)]$. Cependant, le calcul du regret espéré de l'action optimale a priori, $XOL(\hat{a})$, est intéressant puisqu'il s'agit d'une borne maximale pour la valeur prévisionnelle de l'information:

$$0 \leq XVSI \leq XOL(\hat{a}) \quad (1.32)$$

Etant donné cette relation, on nomme souvent $XOL(\hat{a})$ la **valeur de l'information parfaite**.

Démonstration: Par construction, $OL(\hat{a}_{(x)}, \theta) = g(\hat{a}_{(x)}, \theta) - \inf_{a \in S} g(a, \theta) \geq 0$. Ainsi, l'espérance

a posteriori du regret de l'action optimale a posteriori est aussi positif ou nul:

$$XOL(\hat{a}_{(x)}|x) = E^{b(\theta|x)}[OL(\hat{a}_{(x)}, \theta)] \geq 0$$

Il s'ensuit que la prévision de ce regret espéré ne peut être négative:

$$E^{f(x)}[XOL(\hat{a}_{(x)}|x)] \geq 0$$

Or, cette quantité est retranchée à $XOL(\hat{a})$ pour obtenir $XVSI$ [eq. (1.30)]. Ainsi, $XOL(\hat{a})$ ne peut être inférieure à $XVSI$. ■

1.3.4 Valeur prévisionnelle d'un échantillon aléatoire

L'information dont on cherche à prévoir la valeur peut être un échantillon aléatoire, c'est-à-dire une suite $\mathbf{x}=(x_n)$ d'observations échangeables. Dans ce cas, il peut être intéressant de chercher à déterminer comment $XVSI$ varie avec la taille de l'échantillon. Si n dénote la taille de l'échantillon \mathbf{x} , cela revient à étudier une fonction $XVSI(n)$. De par la définition de la valeur de

l'information, on comprend que $XVSI(n)$ est une fonction non décroissante: une information supplémentaire ne peut avoir une utilité négative. Par contre, cette fonction n'est pas nécessairement marginalement décroissante. En divisant $XVSI(n)$ par la valeur de l'information parfaite, on obtient une fonction $RXVSI(n)=XVSI(n)/XOL(\hat{a})$ qui permet d'évaluer la **valeur relative de l'information** contenue dans un échantillon de taille n .

1.3.5 Exemple de prévision de la valeur de l'information

Pour le problème discuté précédemment (sections 1.1.9 et 1.2.5) de dimensionnement d'une digue, on peut estimer la valeur d'une ou plusieurs observations supplémentaires du débit maximum annuel, en tenant compte de l'information a priori et de l'information d'échantillonnage. Ainsi, pour les fins de calcul de la valeur d'une information supplémentaire, la distribution a posteriori obtenue précédemment sera utilisée comme distribution a priori, puisqu'elle synthétise toute l'information disponible. Rappelons qu'il s'agit d'une distribution bêta $b(\theta; \alpha, \beta)$ de paramètres $\alpha=4$ et $\beta=12$. Rappelons aussi qu'il faut choisir entre deux actions a_0 et a_1 dont les fonctions de coût sont respectivement $g(a_0, \theta)=\theta$ et $g(a_1, \theta)=q$. Le risque a priori de ces actions est donc donné par $R(a_0)=\theta$ et $R(a_1)=q$.

Si l'on calcule la valeur de l'information, c'est pour décider entre (1) agir immédiatement, sur la base de l'information a priori, et (2) retarder la construction de la digue de façon à obtenir plus d'information pour prendre une meilleure décision.

Pour rendre l'exemple plus général, nous ferons les calculs pour tout α et β tels que $E(\theta)=\alpha/(\alpha+\beta)\geq q$. Le cas $E(\theta)<q$ est moins intéressant puisque l'action optimale a priori serait a_0 , action qui consiste à ne pas construire la digue. Dans ce cas, l'estimation de la valeur de l'information permettrait de choisir entre (1) ne pas construire la digue et (2) retarder la construction de la digue pour obtenir plus d'information. En pratique, cela revient au même.

1.3.5.1 Valeur de l'information parfaite

Avant d'établir la relation entre la taille d'échantillon n et la valeur de l'information $XVSI(n)$, on peut calculer le regret $XOL(\hat{a})=XOL(a_1)$, qui permet de borner la valeur de l'information:

$$XOL(\hat{a}) = q \cdot F(q; \alpha, \beta) - \frac{\alpha}{\alpha + \beta} \cdot F(q; \alpha + 1, \beta) \quad (1.33)$$

où $F(q; \alpha, \beta)$ correspond à la distribution cumulée de la loi bêta de paramètres α et β . En utilisant une table de la loi bêta incomplète (Abramowitz et Stegun, 1970), on obtient par exemple $XOL(\hat{a}) \cong 0.01$ pour $q=1/6$, $\alpha=4$ et $\beta=12$.

Démonstration: On peut développer $XOL(\hat{a})$ ainsi:

$$\begin{aligned} XOL(\hat{a}) &= R(\hat{a}) - E^{b(\theta)} \left[\inf_{a \in S} g(a, \theta) \right] = q - E^{b(\theta)} [\min(\theta, q)] \\ &= q - \int_0^q \min(\theta, q) b(\theta) d\theta - \int_q^1 \min(\theta, q) b(\theta) d\theta \\ &= q - \int_0^q \theta \cdot b(\theta) d\theta - q \int_q^1 b(\theta) d\theta \\ &= q - \int_0^q \theta \cdot b(\theta; \alpha, \beta) d\theta - q [1 - F(q; \alpha, \beta)] \end{aligned}$$

On peut aussi montrer que $\theta \cdot b(\theta; \alpha, \beta) = \frac{\alpha}{\alpha + \beta} \cdot b(\theta; \alpha + 1, \beta)$:

$$\begin{aligned} \theta \cdot b(\theta; \alpha, \beta) &= \theta \cdot \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \cdot \theta^{\alpha-1} (1 - \theta)^{\beta-1} \\ &= \frac{\alpha}{\alpha + \beta} \cdot \frac{\Gamma(\alpha + \beta + 1)}{\Gamma(\alpha + 1) \Gamma(\beta)} \cdot \theta^{\alpha} (1 - \theta)^{\beta-1}, \quad \text{car } \Gamma(z + 1) = z \cdot \Gamma(z) \\ &= \frac{\alpha}{\alpha + \beta} \cdot b(\theta; \alpha + 1, \beta) \end{aligned}$$

Si on remplace $\theta \cdot b(\theta; \alpha, \beta)$ par $\frac{\alpha}{\alpha + \beta} \cdot b(\theta; \alpha + 1, \beta)$ dans la définition de $XOL(\hat{a})$, on obtient:

$$\begin{aligned} XOL(\hat{a}) &= q - \frac{\alpha}{\alpha + \beta} \int_0^q b(\theta; \alpha + 1, \beta) d\theta - q + q \cdot F(q; \alpha, \beta) \\ &= -\frac{\alpha}{\alpha + \beta} F(q; \alpha + 1, \beta) + q \cdot F(q; \alpha, \beta) \quad \blacksquare \end{aligned}$$

Rappelons que l'unité de mesure de la valeur de l'information parfaite est la même que celle de la fonction de coût, soit une proportion des dommages encourus lors d'une inondation. Si l'on considère que, pour obtenir une année de mesure supplémentaire du débit, on s'expose avec probabilité θ à une inondation qui serait contenue par une digue plus élevée mais que par contre on économise $R(\hat{a})$, on peut donc mesurer le coût de l'attente par $XCS = E(\theta) - R(\hat{a})$. (*XCS* est l'abréviation de *eXpected Cost of Sampling*.) On peut enfin comparer $XOL(\hat{a})$ à XCS pour déterminer s'il vaut la peine d'obtenir des mesures supplémentaires du débit avant de rehausser le niveau de la digue. Par exemple, pour $q=1/6$, $\alpha=4$ et $\beta=12$, la décision est claire car $XOL(\hat{a}) \cong 0.01$ et $XCS=1/12$: l'information disponible est suffisante, et l'on devrait entreprendre immédiatement les travaux pour rehausser la digue.

1.3.5.2 Valeur de l'information en fonction de la taille d'échantillon

En appliquant la définition (1.29), il est possible d'établir la relation entre n et la valeur de l'information, $XVSI$. Cette quantité dépend de la distribution prévisionnelle $f(r, n) = \int_{\Theta} f(r, n|\theta) \cdot b(\theta, \alpha, \beta) d\theta$. Lorsque la distribution a priori est une loi bêta, la distribution prévisionnelle $f(r, n)$ du nombre de succès r en n tirages suit une loi bêta-binomiale (Bernardo et Smith, 1994):

$$f(r, n; \alpha, \beta) = \binom{n}{r} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + n)} \cdot \frac{\Gamma(\alpha + r) \cdot \Gamma(\beta + n - r)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \quad (1.34)$$

Connaissant la distribution prévisionnelle, on montre que $XVSI(n)$ est donnée par l'équation suivante:

$$XVSI(n) = \begin{cases} 0 & \text{si } n < \frac{\alpha \cdot (1 - q) - \beta \cdot q}{q} \\ \sum_{r=0}^{\lfloor q(\alpha + \beta + n) - \alpha \rfloor} \left[q - \frac{\alpha + r}{\alpha + \beta + n} \right] f(r, n; \alpha, \beta) & \text{si } n \geq \frac{\alpha \cdot (1 - q) - \beta \cdot q}{q} \end{cases} \quad (1.35)$$

où $\lfloor z \rfloor$ dénote la partie entière de z .

Démonstration: En appliquant la définition (1.29), on obtient:

$$XVSI(n) = R(\hat{a}) - E^{f(r, n; \alpha, \beta)} \left[\inf_{a \in S} R(a|r, n) \right]$$

où $R(a|r, n)$ correspond au risque a posteriori de l'action a , mesuré après observation de r succès en n essais. Par hypothèse, $R(\hat{a}) = R(a_1) = R(a_1|r, n) = q$. De plus, $R(a_0|r, n) = \frac{\alpha + r}{\alpha + \beta + n}$. On peut donc écrire:

$$XVSI(n) = q - \sum_{r=0}^n \min\left(\frac{\alpha + r}{\alpha + \beta + n}, q\right) f(r, n; \alpha, \beta)$$

On montre ensuite que:

$$\min\left(\frac{\alpha + r}{\alpha + \beta + n}, q\right) = \begin{cases} \frac{\alpha + r}{\alpha + \beta + n} & \text{si } r < q(\alpha + \beta + n) - \alpha \\ q & \text{si } r \geq q(\alpha + \beta + n) - \alpha \end{cases}$$

En combinant cette dernière équation avec la définition de $XVSI(n)$, on obtient:

Si $q(\alpha + \beta + n) - \alpha < 0$:

$$XVSI(n) = q - \sum_{r=0}^n q \cdot f(r, n; \alpha, \beta) = q - q = 0$$

Si $q(\alpha + \beta + n) - \alpha \geq 0$:

$$XVSI(n) = q - \sum_{r=0}^{\lfloor q(\alpha + \beta + n) - \alpha \rfloor} \frac{\alpha + r}{\alpha + \beta + n} f(r, n; \alpha, \beta) - \sum_{r=\lfloor q(\alpha + \beta + n) - \alpha \rfloor + 1}^n q \cdot f(r, n; \alpha, \beta)$$

On peut finalement simplifier cette dernière équation:

$$\begin{aligned} XVSI(n) &= q \cdot \left\{ 1 - \sum_{r=\lfloor q(\alpha + \beta + n) - \alpha \rfloor + 1}^n f(r, n; \alpha, \beta) \right\} - \sum_{r=0}^{\lfloor q(\alpha + \beta + n) - \alpha \rfloor} \frac{\alpha + r}{\alpha + \beta + n} f(r, n; \alpha, \beta) \\ &= q \cdot \left[\sum_{r=0}^{\lfloor q(\alpha + \beta + n) - \alpha \rfloor} f(r, n; \alpha, \beta) - \sum_{r=0}^{\lfloor q(\alpha + \beta + n) - \alpha \rfloor} \frac{\alpha + r}{\alpha + \beta + n} f(r, n; \alpha, \beta) \right] \\ &= \sum_{r=0}^{\lfloor q(\alpha + \beta + n) - \alpha \rfloor} \left[q - \frac{\alpha + r}{\alpha + \beta + n} \right] f(r, n; \alpha, \beta) \quad \blacksquare \end{aligned}$$

Par exemple, pour $q=1/6$, $\alpha=4$ et $\beta=12$, on obtient:

$$XVSI(n) = \begin{cases} 0 & \text{si } n < 8 \\ \sum_{r=0}^{\lfloor \frac{n-8}{6} \rfloor} \left[\frac{1}{6} - \frac{4+r}{16+n} \right] f(r, n; \alpha = 4, \beta = 12) & \text{si } n \geq 8 \end{cases} \quad (1.36)$$

La Figure 1.2 montre la forme de la fonction $RXVSI(n) = XVSI(n) / XOL(\hat{a})$, tracée pour $q=1/6$, $\alpha=4$, $\beta=12$ et n variant de 1 à 30.

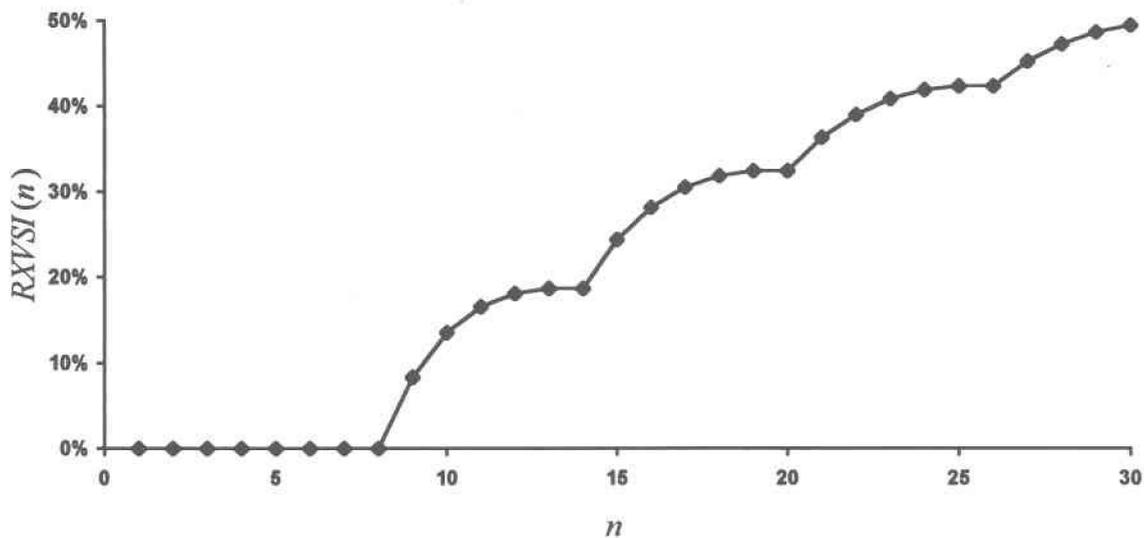


Figure 1.2: Valeur relative de l'information contenue dans un échantillon de taille n , calculée pour $\alpha=4$, $\beta=12$ et $q=1/6$

On observe que plus de trente années d'observation du débit maximum annuel sont nécessaires pour récupérer 50% de la valeur de l'information parfaite. Ce graphique illustre aussi le fait que la valeur de l'information n'est pas marginalement décroissante. La valeur de l'information est d'ailleurs nulle pour $n < 9$.

1.4 Analyse bayésienne de référence

La possibilité de modéliser une information a priori à l'aide d'une distribution de probabilité $b(\theta)$ constitue à la fois une force et une faiblesse de l'approche bayésienne, parce qu'il peut être difficile de fixer cette distribution et que l'analyse peut être sensible à ce choix. Il arrive souvent que l'on soit indifférent face aux différentes valeurs que peut prendre θ , et que l'on cherche une distribution qui ne favorise pas de région particulière de l'espace des paramètres. Dans ces situations, il n'est pas utile de dépenser trop d'énergies à choisir une distribution, et une méthode ad hoc peut suffire, dans la mesure où elle est adaptée au problème. Nous en présenterons trois: le principe d'invariance de Jeffreys (1961), le principe de maximisation de l'entropie (Jaynes, 1968) et le principe de maximisation du regret espéré (Bernardo, 1979).

1.4.1 Principe d'invariance de Jeffreys

Le principe d'invariance de Jeffreys consiste à choisir, pour un référentiel fixé $\Theta = \{\theta\}$, une distribution $b(\theta)$ de façon que les inférences soient peu influencées par une transformation de variable $\varphi(\theta)$. Jeffreys (1961) montre que cet objectif peut être atteint en choisissant $b(\theta)$ proportionnel à la racine carrée de la mesure d'information de Fisher $J(\theta)$:

$$b(\theta) \propto J^{1/2}(\theta), \quad J(\theta) = -E^{f(x|\theta)} \left[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right] \quad (1.37)$$

Cette solution peut se généraliser à une distribution a priori possédant plusieurs paramètres. La distribution a priori ainsi obtenue est relativement peu sensible à des transformations de variable.

1.4.2 Principe de maximisation de l'entropie

Pour choisir une distribution a priori lorsque l'information a priori semble insuffisante pour suggérer une distribution précise, Jaynes (1968) propose de choisir une distribution a priori qui maximise une mesure d'entropie, sous les contraintes imposées par l'information a priori. Dans le cas d'une variable réelle θ , l'entropie d'une distribution $b(\theta)$ est définie de la façon suivante:

$$H(b) = - \int b(\theta) \log \frac{b(\theta)}{h(\theta)} d\theta \quad (1.38)$$

où $h(\theta)$ est une «mesure d'invariance», censée modéliser une indifférence totale face à θ . Outre le problème du choix de $h(\theta)$, il faut mentionner que l'entropie n'est pas nécessairement bornée, et que la fonction $b(\theta)$ maximisant l'entropie n'est pas nécessairement unique. En général, cette approche fonctionne mieux lorsque certains moments de la distribution a priori sont fixés que lorsque des quantiles le sont (Berger, 1985).

1.4.3 Principe de maximisation du regret espéré

Bernardo (1979) suggère de limiter $b(\theta)$ à une famille paramétrique et de choisir ensuite $b(\theta)$ dans cette famille de façon à maximiser $XOL(\hat{a})$. En effet, la valeur d'une information d'échantillonnage devrait être d'autant plus grande que l'information a priori est vague. Cependant, selon le modèle et la fonction de coût choisie, il peut exister plusieurs distributions a priori $b(\theta)$ maximisant $XOL(\hat{a})$. Ces auteurs proposent aussi une fonction de coût standard

qui peut être utilisée avec cette approche pour les problèmes d'inférence. Considérons comme espace des actions l'ensemble S de toutes les distributions de probabilités strictement positives $S = \{q(\cdot) : q(\theta) > 0, \int q(\theta) d\theta = 1\}$ et la fonction de coût $g(q(\cdot), \theta) = \log q(\theta)$. On peut interpréter $q(\cdot)$ comme un estimateur global de θ . Pour cette fonction de coût particulière, Bernardo et Smith (1994) ont déterminé $b(\theta)$ en maximisant $XOL(\hat{\alpha})$ pour certains modèles $f(x|\theta)$ usuels, et présentent leurs résultats dans un format facilitant leur compréhension et leur utilisation.

1.5 Développements de la méthodologie bayésienne pour l'estimation de la valeur de l'information hydrologique

Suite aux travaux de Davis et Dvoranchik (1971) et de Davis et al. (1972), qui ont introduit la méthodologie bayésienne pour estimer la valeur de l'information en hydrologie, plusieurs chercheurs ont adopté cette approche décisionnelle. Il faut mentionner que l'approche bayésienne s'opposait alors à une solution fréquentiste du problème. Cette dernière solution, d'abord proposée par Dawdy et al. (1970), consistait à étudier la réduction de la variance et de l'écart quadratique moyen d'un estimateur avec l'augmentation de la taille d'échantillon. Sur la base de simulations intensives, comme celles effectuées par Slack et al. (1975) et Wallis et al. (1976), on montrait alors généralement le peu d'influence d'une information supplémentaire sur la précision d'un estimateur. Ces études, fondées sur des hypothèses de lois parentes difficilement justifiables, ont provoqué un certain scepticisme chez les hydrologues face à l'intérêt des modèles statistiques en hydrologie opérationnelle (Klemes, 1977). La faiblesse de l'approche fréquentiste, qui ne considère que des situations générales et hypothétiques, et qui ne permet pas de mesurer la valeur de l'information en termes économiques, est finalement critiquée par Dawdy (1979), qui reconnaît la supériorité théorique et opérationnelle de l'approche bayésienne, malgré les difficultés techniques liées à sa mise en oeuvre. L'organisation météorologique mondiale (WMO, 1990), dans un rapport portant sur l'analyse coûts-bénéfices des données hydrologiques, signale que l'approche bayésienne constitue la solution la plus adéquate pour effectuer cette analyse en présence d'incertitude.

Un aspect nouveau a été abordé par Szidarovsky et al. (1976), qui ont montré la sensibilité de la valeur de l'information hydrologique aux incertitudes économiques.

L'intégration des incertitudes économiques dans l'analyse bayésienne de la valeur de l'information est discutée par Duckstein et al. (1978).

Davis et al. (1979) se sont attardés au problème de l'estimation de la valeur de l'information lorsqu'il est impossible de prendre une décision optimale, c'est-à-dire une décision de risque minimal. Ceci se produit lorsque le décideur bayésien doit respecter des contraintes hors de son contrôle, comme des règlements gouvernementaux, mais également par manque de rigueur dans le développement de la règle de décision et dans la construction de la fonction d'utilité. Ces auteurs montrent que, dans ce cas, la valeur de l'information peut être négative. Pour mener à bien une estimation de la valeur de l'information et s'assurer d'une utilisation optimale de l'information, Davis et Nnaji (1982) discutent des points cruciaux auxquels il faut s'attarder, et présentent une application de la méthodologie à la conception d'un système d'alerte destiné à prévenir les automobilistes du fort débit d'un cours d'eau que traverse la route.

Winkler et al. (1983) ont étudié la valeur des prévisions météorologiques et hydrologiques pour les pratiques agricoles, car celles-ci ont une influence sur le choix des cultures. En effet, selon la quantité d'eau disponible pour l'irrigation, différentes céréales peuvent être cultivées. Le choix de la culture à privilégier doit cependant se faire à l'avance sur la base d'une prévision de la quantité d'eau qui sera réellement disponible. Plus le printemps avance, meilleure est cette prévision mais moins grande est la liberté d'action. Il devient donc important de prendre la meilleure décision au temps le plus opportun.

Krzysztofowicz (1986) s'est attaqué à ce problème à l'aide d'un processus markovien d'arrêt optimal qui permet d'optimiser en même temps le moment de la décision et le choix de la culture en fonction de la quantité d'eau prévue pour l'irrigation. Krzysztofowicz et Watada (1986) ont appliqué cette même méthodologie au problème plus général d'une décision qui dépend du volume prévu de la crue printanière, et qui doit être prise le plus tôt possible (par exemple la planification des cultures, l'évacuation de la crue à un barrage, ou l'alerte d'une inondation imminente). Sachant que le service national de météorologie des Etats-Unis fournit cinq ou six de ces prévisions avant et pendant la période de la crue et que ces prévisions s'améliorent avec le temps, ces auteurs montrent qu'il est possible de déterminer, par

programmation dynamique, le moment optimal pour prendre la décision, et la valeur de cette décision optimale.

Plus récemment, Bernier (1990) a montré comment le concept de valeur de l'information pouvait être utilisé pour comparer l'utilité des informations locales et spatiales, de façon à proposer des méthodes efficaces de transfert d'information dans les analyses hydrologiques régionales. Enfin, Parent et al. (1995, 1996) ont montré le potentiel d'une analyse bayésienne de la valeur de l'information pour le contrôle de la qualité.

Parallèlement à ces développements méthodologiques, plusieurs applications ont été réalisées. Citons dans le domaine de l'hydrologie de surface Jacobi (1975), Lane et al. (1979), et Attanasi et Karlinger (1979, 1982). Plusieurs applications intéressantes en hydrogéologie doivent aussi être mentionnées: Gates et Kisiel (1974), Reichards et Evans (1989), Freeze et al. (1992), James et Freeze (1993) ainsi que James et Gorelick (1994).

1.6 Critique de l'approche bayésienne

L'approche bayésienne est basée sur deux hypothèses importantes:

1. qu'il est possible de modéliser l'incertitude sur l'état de la nature à l'aide d'un espace de probabilités subjectives;
2. qu'il est préférable de choisir l'action de risque minimal, c'est-à-dire celle qui maximise l'utilité espérée.

Nous discuterons séparément de la validité des ces deux hypothèses.

1.6.1 Critique du principe de précision

Rappelons d'abord que le principe de précision revient à supposer que le prix d'achat maximal et le prix de vente minimal spécifiés par un individu pour un pari quelconque devraient toujours se confondre. Cette hypothèse revient à supposer que l'incertitude sur un événement et l'utilité d'une décision peuvent être résumées par un seul nombre. En effet, rappelons que l'utilité d'une décision est définie par une transformation linéaire de la probabilité d'occurrence d'un événement particulier.

Le principe de précision est particulièrement critiqué pour la modélisation de l'ignorance (Shafer, 1976; Walley, 1991), c'est-à-dire pour la modélisation de propositions sur lesquelles un individu ne possède à peu près aucune information. Prenons un exemple en

hydrogéologie: un citoyen, ayant acheté une maison de campagne isolée, veut y creuser un puits artésien et se demande s'il trouvera assez d'eau pour s'y alimenter. Il considère donc l'ensemble des états de la nature $\Theta = \{\theta_0, \theta_1\}$, où θ_0 signifie qu'il n'y a pas assez d'eau, et θ_1 signifie qu'il y a assez d'eau. Etant donné qu'il n'a que peu de connaissances en hydrogéologie et peu d'expérience de la vie à la campagne, il ne sait pas comment fixer les probabilités $P(\{\theta_0\})$ et $P(\{\theta_1\})$. Parce que leur somme doit être unitaire et qu'il n'a aucune raison de favoriser l'une des deux alternatives, certains auteurs affirment que l'individu doit suivre le principe d'indifférence (Laplace, 1812) et fixer $P(\{\theta_0\}) = P(\{\theta_1\}) = 1/2$. En effet, accorder une plus forte probabilité à l'un des deux événements semblerait vouloir dire qu'il a une raison de croire plus en cet événement, alors que ce n'est pas le cas. Le principe d'indifférence milite donc, lorsque l'information a priori est si vague, en faveur d'une distribution a priori uniforme.

Les choses se compliquent lorsque cet individu réalise que trouver de l'eau ne réglera pas tout: en effet, à moins que le puits artésien soit jaillissant, il lui faudra installer une pompe. En approfondissant le problème, il arrive donc à un nouvel espace de possibilités: $\Theta = \{\theta_0, \theta_{1a}, \theta_{1b}\}$, où θ_{1a} signifie qu'il y a de l'eau mais qu'il faut une pompe, alors que θ_{1b} signifie que le puits est jaillissant, et qu'il n'y a pas besoin de pompe. Comment doit-il maintenant distribuer les probabilités entre les trois événements élémentaires? S'il respecte le principe d'indifférence présenté plus haut, il posera $P(\{\theta_0\}) = 1/3$ et sera en contradiction avec lui-même. Sinon, il devra distribuer différemment les probabilités et exprimer ainsi une idée a priori en faveur de l'absence d'eau. Dans tous les cas, il est difficile de justifier les nombres fixés. Ce que cet exemple montre, c'est qu'il est impossible de concilier, en utilisant des probabilités, le principe d'indifférence avec un principe d'invariance de représentation, c'est-à-dire cette idée que la probabilité d'un événement ne devrait pas être influencée par une modification du référentiel.

De façon générale, les situations où l'on se trouve dans une ignorance totale sont plutôt rares. En pratique, on dispose au moins d'un peu d'expérience et d'intuition. Dans l'exemple précédent, on saurait probablement en fonction de la région du globe où l'on se trouve s'il est possible de trouver une source jaillissante. Il n'en reste pas moins que fixer une valeur précise pour $P(\{\theta_{1b}\})$ pose problème. Il est donc difficile en pratique de justifier une distribution a

priori $b(\theta)$ dans les situations rares de quasi-ignorance et dans les situations plus fréquentes où l'information est difficile à préciser.

D'un point de vue bayésien, une solution naturelle consiste à modéliser l'incertitude sur la distribution a priori par une distribution de probabilité. Lindley (1968) suggère que l'emploi d'une approche hiérarchique de ce type peut suffire à résoudre les difficultés reliées au principe de précision. Il faut tout de même réussir à établir une distribution a priori de second niveau $b_2(\alpha)$ sur les paramètres de la distribution a priori $b(\theta|\alpha)$. Pour choisir les paramètres de b_2 , on fera cependant face au même type de problème. Berger (1985) souligne cependant que les résultats d'une analyse bayésienne sont peu sensibles aux valeurs des paramètres de second niveau.

1.6.2 Critique du principe de maximisation de l'utilité espérée

Le principe de maximisation de l'utilité espérée est critiqué par l'école de l'économiste Maurice Allais (voir par exemple Allais, 1953 et Shafer, 1986). Ces chercheurs, sur la base de résultats expérimentaux (par exemple Slovic et Tversky, 1974), montrent qu'une grande proportion d'individus peuvent violer sciemment ce principe sans se trouver irrationnels. Ainsi, dans certaines situations, il n'existe pas de couple (mesure de probabilité, fonction d'utilité) qui permette de représenter des comportements pourtant bien réfléchis. À la lumière de ces résultats, on peut difficilement soutenir que la théorie bayésienne de la décision a un important pouvoir descriptif.

Les théoriciens de l'école de Savage (1972) accordent peu d'importance à ces résultats expérimentaux, car ils soutiennent que la théorie bayésienne de la décision a plutôt une valeur normative: les seules décisions «rationnelles» seraient celles qui peuvent être obtenues en maximisant l'espérance d'une fonction d'utilité. Dans cette perspective, les écarts observés entre le modèle bayésien et les résultats expérimentaux sont causés par des erreurs de jugement de la part du décideur, qu'il n'est pas intéressant de modéliser. Il est vrai que la théorie bayésienne de la décision constitue une façon rationnelle de prendre une décision, en ce sens que les conclusions auxquelles on arrive en appliquant cette théorie ne sont que des déductions logiques obtenues à partir des hypothèses faites au départ. Mais chacun doit s'assurer qu'il trouve personnellement ces hypothèses compatibles avec sa vision du problème.

Au prochain chapitre, nous proposerons une définition du concept de valeur de l'information qui ne repose pas sur le principe de précision. Pour appliquer notre définition, il nous faudra à un certain moment décrire le comportement du décideur. On sait d'ores et déjà que la théorie bayésienne de la décision sera à cet effet insuffisante.

Une panoplie de modèles ne reposant pas sur la maximisation de l'utilité espérée ont été proposés pour décrire le comportement d'un décideur (Munier, 1989; Munier et Parent, 1995). Par exemple, Quiggin (1982) suggère de modéliser l'attitude du décideur face à l'incertitude en effectuant une transformation de l'axe des probabilités cumulées, de la même façon que l'on effectue une transformation de l'axe des conséquences pour obtenir une mesure d'utilité. Il s'agit du modèle à dépendance de rang, aussi nommé modèle dichotomique. Une autre approche, suggérée par Schmeidler (1989), consiste à supposer qu'un décideur agit comme s'il modélisait son incertitude à l'aide d'une capacité de Choquet (53-54), c'est à dire une fonction d'ensemble non décroissante, et comme s'il choisissait la décision maximisant l'intégrale de Choquet d'une fonction d'utilité. Wakker (1990) montre que ce modèle englobe le modèle dichotomique. Cependant, si l'on fait seulement l'hypothèse qu'un décideur préfère toujours avoir plus de chances de gain que moins, alors ces modèles sont équivalents (on nomme cette hypothèse peu controversée le respect de la **dominance stochastique du premier ordre**).

1.7 Approche bayésienne robuste

Si l'on accepte le principe de précision, il faut au minimum vérifier la sensibilité de l'analyse à cette hypothèse fondamentale. L'idée de l'approche bayésienne robuste, défendue par Berger (1985) est d'identifier une classe $M=\{b(\theta)\}$ de distributions a priori et une classe $C=\{g(a,\theta)\}$ de fonctions de coût parmi lesquelles est censée se trouver la «vraie» distribution subjective $b_0(\theta)$ et la «vraie» fonction de coût $g_0(a,\theta)$. Les inférences sont alors effectuées pour toutes les combinaisons de distribution a priori et de fonction de coût précises issues du produit cartésien $M \times C$. On obtient alors des bornes inférieure et supérieure pour la probabilité d'un événement et le risque d'une action:

$$\underline{P}(A) = \inf_{b \in M} \int_A b(\theta) d\theta, \quad \overline{P}(A) = \sup_{b \in M} \int_A b(\theta) d\theta \quad (1.39)$$

$$\underline{R}(a) = \inf_{\substack{b \in M \\ g \in C}} \int_A g(a, \theta) b(\theta) d\theta, \quad \bar{R}(a) = \sup_{\substack{b \in M \\ g \in C}} \int_A g(a, \theta) b(\theta) d\theta \quad (1.40)$$

L'approche bayésienne robuste peut entraîner une indécision. En effet, il peut s'avérer impossible de déterminer l'action de risque minimal. Soit $S=\{a\}$ l'espace des actions. Pour chaque distribution a priori b et chaque fonction de coût g , l'action $\hat{a}_{b,g}$ minimisant le risque de Bayes peut être déterminée (il peut y en avoir plus d'une, mais elles ont alors toutes le même risque, et on peut en choisir une de façon arbitraire). D'un point de vue d'analyse de sensibilité, on sait seulement que la «vraie» action optimale \hat{a} fait partie de l'ensemble $\hat{S}=\{\hat{a}_{b,g}\}$.

1.7.1 Estimation robuste de la valeur de l'information

Une estimation robuste de la valeur de l'information est obtenue en considérant l'ensemble des valeurs de $XVSI$ correspondant aux distributions a priori et aux fonctions de coût pour lesquelles on effectue une analyse de sensibilité. On obtient alors une borne inférieure et une borne supérieure pour $XVSI$:

$$\inf XVSI = \inf_{\substack{b \in M \\ g \in C}} XVSI_{b,g} \quad (1.41)$$

$$\sup XVSI = \sup_{\substack{b \in M \\ g \in C}} XVSI_{b,g} \quad (1.42)$$

1.7.2 Mise en oeuvre de l'approche bayésienne robuste

Plusieurs méthodes ont été proposées pour construire la classe $\{b(\theta)\}$ des distributions a priori:

- **approche paramétrique:** choisir une famille paramétrique $L(\alpha) \rightarrow b(\theta|\alpha)$ pour la distribution a priori et faire varier les paramètres α (Berger, 1985);
- **construction d'un voisinage par contamination:** identifier une distribution a priori possible $b_0(\theta)$ et considérer la classe $M(b_0, \varepsilon, L) = \{b : b(\theta) = \varepsilon b_0(\theta) + (1-\varepsilon) \cdot \pi, \pi \in L\}$, c'est-à-dire la contamination de b_0 à un niveau ε par des distributions π appartenant à une classe L , par exemple l'ensemble de toutes les distributions (Berger, 1985);

- **contraintes sur les quantiles:** choisir une famille de lois M , par exemple la famille de toutes les lois continues et unimodales, et considérer toutes les distributions $b(\theta) \in M$ qui respectent un ensemble de contraintes sur leurs quantiles, de la forme $\{F(\theta_n) = p_n, n=1, 2, \dots, k\}$ (O'Hagan et Berger, 1988);
- **contraintes sur les moments:** spécifier les premiers moments de la distribution, et considérer toutes les distributions a priori qui respectent ces contraintes (Robert, 1992);
- **contrainte sur la dérivée:** choisir toutes les lois continues par morceau $M = \{b(\theta)\}$ qui varient, en dehors des points de discontinuité, relativement peu rapidement; on pourra par exemple choisir b tel que $|d[\log b(\theta)]/d\theta| \leq c$, où c est une constante qui limite la concentration de l'information a priori (Walley, 1993);

D'autres approches pourraient aussi être considérées. L'approche paramétrique est souvent la plus facile à appliquer, mais elle peut ne pas être satisfaisante si la famille paramétrique choisie est trop restreinte. La construction d'un voisinage par contamination peut s'avérer difficile à mettre en oeuvre lorsque la classe L est large. Notons que l'imposition de contraintes sur la dérivée ne permet d'obtenir une distribution a posteriori intégrable que si la décroissance asymptotique de la vraisemblance est au moins exponentielle. Par contre, cette contrainte sur la dérivée du logarithme de $b(\theta)$ fait que la classe M est équivalente à toutes fins pratiques à une classe de fonctions exponentielles par morceau, ce qui simplifie l'application de la méthode.

1.7.3 Principe de précision idéale

L'approche bayésienne robuste repose en quelque sorte sur le principe de précision, puisque son application revient à supposer qu'il existe une distribution a priori et une fonction de coût précises permettant de modéliser correctement le problème, mais que ces dernières ne peuvent être identifiées, par manque de temps ou de ressources. Walley (1991) nomme cette hypothèse le **principe de précision idéale**. Ceci mène à un résultat paradoxal. En effet, c'est dans les situations d'ignorance ou de quasi-ignorance a priori qu'il est le plus difficile de représenter l'incertitude par une distribution de probabilités. Il semblerait donc que moins on a d'information a priori, plus il faut de temps et de ressources pour l'analyser correctement à l'aide de la théorie bayésienne.

Etant donné les difficultés théoriques et pratiques reliées à la modélisation d'une information a priori subjective à l'aide de probabilités précises, même lorsque l'on prend la peine d'analyser la sensibilité des résultats à cette hypothèse, nous nous proposons d'explorer un autre modèle d'incertitude, la théorie des prévisions inférieures cohérentes. Ce modèle plus général permet non seulement de s'affranchir du principe de précision, mais facilite aussi la prise en compte d'une information subjective a priori.

2. PREVISION IMPRECISE DE LA VALEUR DE L'INFORMATION

La théorie bayésienne de la décision suggère une mesure concrète de l'utilité d'une information, à condition de pouvoir représenter l'incertitude par une mesure de probabilités précises et d'accepter de prendre la décision qui maximise une fonction d'utilité. Il est possible, dans le cadre de la théorie des prévisions inférieures cohérentes (Walley, 1991), de définir l'utilité d'une information sans utiliser le principe de précision. Cette théorie s'apparente à l'approche bayésienne robuste. Cependant, l'interprétation des résultats obtenus peut s'effectuer sans supposer l'existence d'une «vraie» distribution a priori. Nous verrons que cette nuance influence la définition de la valeur de l'information au sens de la théorie des prévisions inférieures cohérentes.

Dans un premier temps, nous présenterons les fondements de la théorie des prévisions inférieures cohérentes. Nous proposerons ensuite une définition de la valeur de l'information adaptée à cette théorie, qui sera illustrée par un exemple simple.

2.1 Théorie des prévisions inférieures cohérentes

Au coeur de la théorie des prévisions inférieures cohérentes se trouve le concept de prévision, plus fondamental que celui de probabilité. On suppose que pour un pari $X(\theta)$ donné, un individu est capable de spécifier un prix d'achat et un prix de vente, mesurés en unités d'utilité. Une prévision sur un pari $X(\theta)$ est ainsi constituée de deux nombres: un prix maximal d'achat, $\underline{E}(X)$, et un prix minimal de vente $\bar{E}(X)$. Lorsque ces deux nombres sont confondus, la prévision est précise. Autrement, l'intervalle $[\underline{E}(X), \bar{E}(X)]$ constitue la **prévision imprécise** du pari $X(\theta)$. Lorsque ce pari correspond à la fonction indicatrice d'un événement A , on obtient la **probabilité imprécise** de cet événement. On nomme $\underline{E}(X)$ la **prévision inférieure** de $X(\theta)$ et $\bar{E}(X)$ la **prévision supérieure** de ce même pari.

Il peut être plus facile d'émettre des prévisions que de fixer des probabilités, en particulier des prévisions imprécises. En effet, même un individu ayant peu de connaissances formelles en statistique peut effectuer dans un domaine qu'il connaît bien une prévision sur la valeur d'une quantité observable.

En pratique, on retrouve même dans les revues scientifiques des estimations présentés sous la forme d'intervalles, sans que soit mentionnée l'interprétation que l'on doit donner à ces intervalles. Pour donner quelques exemples, nous avons parcouru un numéro récent de la *Revue des Sciences de l'Eau* (vol. 9, no 4, 1996). Sans faire une recherche exhaustive, nous avons extrait de chaque article de cette revue un exemple d'estimation sous forme d'intervalle:

- temps d'une analyse de la quantité de carbone organique dissous biodégradable (CODB) à l'aide d'un réacteur à piston: 10 à 40 min. (Reuter et al., 1996);
- altitude de la couche d'inversion thermique au dessus de l'île de la Réunion: entre 2000 et 2500m, parfois aussi bas que 1500m (Barcelo et Coudray, 1996);
- gradient de vitesse optimal pour un flottateur à air dissous: 3000 à 4000s⁻¹, certainement pas plus de 5000s⁻¹ (Gallinari et al., 1996);
- concentration de la flore totale dans un réseau de distribution d'eau: 10⁴ à 10⁵ cellules/ml (Nakache et al., 1996);
- teneur en tritium dans les précipitations: 10 à 30 unités-tritium, parfois plus de 40 (Olive et al., 1996);
- contribution des acides aminés totaux à la composition du carbone organique dissous biodégradable d'un perméat de nanofiltration: 50 à 60% (Agbekodo et al., 1996).

Dans certains cas, il est clair que l'intervalle indiqué n'est pas le domaine de définition de la variable, puisque des observations à l'extérieur de l'intervalle sont mentionnées. Une interprétation en terme d'intervalle de confiance ou d'intervalle de crédibilité est aussi problématique puisqu'elle nécessite l'attribution d'une probabilité à l'intervalle. L'interprétation en terme de prévision imprécise est peut-être plus naturelle: il n'est pas exclu que la variable aléatoire soit à l'extérieur de l'intervalle, mais on serait prêt à parier qu'elle s'y trouve.

2.1.1 Axiomes de la théorie des prévisions inférieures cohérentes

La théorie des prévisions inférieures cohérentes est basée sur six axiomes qui tentent de caractériser un individu cohérent, c'est-à-dire un individu qui ne se contredit jamais. Nous présentons brièvement ces axiomes, qui sont justifiés de façon plus détaillée par Walley (1991).

Représentons par $G(\theta, X) = X(\theta) - \underline{E}(X)$ le **bénéfice net de l'achat d'un pari** et par $\bar{G}(\theta, X) = \bar{E}(X) - X(\theta)$ le **bénéfice net de la vente** de ce même pari. On voit que $\bar{G}(\theta, -X) = X(\theta) - [-\bar{E}(-X)]$. Autrement dit, le bénéfice de la vente d'un pari $-X(\theta)$ est le même que celui de l'achat du pari $X(\theta)$ au prix $-\bar{E}(-X)$: la vente d'un pari peut être interprétée comme l'achat d'un pari différent. Un individu peut donc spécifier le prix d'achat d'un pari $X(\theta)$ par $\underline{E}(X)$ ou par $-\bar{E}(-X)$. Ce résultat suggère un premier axiome, afin d'assurer un minimum de cohérence:

Axiome 1 - lien entre les prévisions inférieures et supérieures:

$$\underline{E}(X) = -\bar{E}(-X) \quad (2.1)$$

Ainsi, la connaissance des prévisions inférieures est suffisante pour permettre d'établir la prévision imprécise de n'importe quel pari à partir des prévisions inférieures. Dans la suite de cette section, nous nous concentrerons sur les prévisions inférieures, sachant que les prévisions supérieures peuvent être obtenues à l'aide de l'équation (2.1).

Le second axiome de la théorie des prévisions inférieures cohérentes consiste à affirmer qu'un individu devrait être prêt à payer pour un pari $X(\theta)$ au moins ce qu'il est certain d'en retirer, c'est-à-dire $\inf X(\theta)$:

Axiome 2 - principe d'évitement des gains certains:

$$\underline{E}(X) \geq \inf_{\theta \in \Theta} X(\theta) \quad (2.2)$$

Il s'agit d'une condition plus forte que le principe d'évitement des pertes sûres, qu'on pourrait nommer le principe d'évitement des gains certains. En effet, si un individu paye un montant $\underline{E}(X) < \inf X(\theta)$ pour un pari $X(\theta)$, son bénéfice net $X(\theta) - \underline{E}(X)$ sera nécessairement positif. On peut entre autres justifier cet axiome en affirmant qu'il est à peu près inutile de spécifier un prix d'achat $\underline{E}(X) < \inf X(\theta)$, puisqu'aucun individu rationnel n'acceptera de vendre le pari $X(\theta)$ s'il est certain de perdre dans la transaction. D'une certaine façon, ce second axiome est équivalent à l'application du principe d'évitement des pertes sûres au monde extérieur: choisir $\underline{E}(X) \geq \inf X(\theta)$ permet de s'assurer que l'individu duquel on achète le pari $X(\theta)$ ne subira pas de pertes sûres. On peut aussi exprimer ce second axiome sous la forme

$$\inf_{\theta \in \Theta} G(\theta, X) \leq 0.$$

Un troisième axiome est relié à l'hypothèse que les paris et les montants pariés sont mesurés en unités d'utilité, ce qui implique que le prix d'achat d'un pari devrait être proportionnel à la valeur de ce pari, puisque l'aversion au risque est déjà prise en compte. On peut exprimer cela par l'équation suivante:

Axiome 3 - homogénéité positive:

$$\underline{E}(\lambda \cdot X) = \lambda \cdot \underline{E}(X), \lambda > 0 \quad (2.3)$$

Le dernier axiome suggère une certaine cohérence interne de la part d'un individu rationnel. Considérons deux paris $X(\theta)$ et $Y(\theta)$ pour lesquels un individu accepte de payer respectivement les montants $\underline{E}(X)$ et $\underline{E}(Y)$. Supposons aussi que cet individu soit disposé à payer au maximum un montant $\underline{E}(X+Y) < \underline{E}(X) + \underline{E}(Y)$ pour le pari $(X+Y)(\theta) = X(\theta) + Y(\theta)$. S'il achète simultanément les deux paris $X(\theta)$ et $Y(\theta)$, il se retrouve dans la même situation que s'il avait acheté directement le pari $(X+Y)(\theta)$ pour un prix $\underline{E}(X) + \underline{E}(Y) > \underline{E}(X+Y)$. On peut donc manipuler les prévisions effectuées par cet individu sur les paris $X(\theta)$ et $Y(\theta)$ pour le forcer à se contredire et à payer un montant supérieur à ce qu'il avait envisagé pour le pari $(X+Y)(\theta)$. Les prévisions inférieures d'un individu qui évite ce type d'incohérence respecteront l'axiome suivant:

Axiome 4 - super-additivité:

$$\underline{E}(X+Y) \geq \underline{E}(X) + \underline{E}(Y) \quad (2.4)$$

En utilisant la propriété d'associativité des fonctions à valeur réelle, on montre par induction que pour toute suite finie de k paris $(X_n, n=1, 2, \dots, k)$ il existe $\varepsilon \geq 0$ tel que:

$$\underline{E}\left(\sum_{n=1}^k X_n\right) = \sum_{n=1}^k \underline{E}(X_n) + \varepsilon \quad (2.5)$$

On peut aussi obtenir une relation semblable pour le bénéfice net d'une suite de paris:

$$\underline{G}\left(\sum_{n=1}^k X_n\right) + \varepsilon = \sum_{n=1}^k \underline{G}(X_n), \quad \varepsilon \geq 0 \quad (2.6)$$

Lorsque des prévisions imprécises spécifiées sur un espace linéaire de paris $X = \{X\}$ respectent ces quatre axiomes, on dira qu'elles sont cohérentes sur X , et que la fonction $\underline{E}: X \rightarrow \mathbb{R}$ constitue une prévision inférieure cohérente sur X . Rappelons que X est un espace linéaire si $X(\theta), Y(\theta) \in X$ et $\lambda \in \mathbb{R} \Rightarrow (X+Y)(\theta) \in X$ et $(\lambda X)(\theta) \in X$. Lorsque X n'est pas linéaire, on introduit un axiome supplémentaire, qui termine de caractériser la cohérence:

Axiome 5 - cohérence:

Pour tout $k \geq 0$, pour toute suite de paris $(X_n, n=0,1,2,\dots,k) \subseteq \mathbf{X}$, et pour toute suite de nombres non négatifs $(\lambda_n, n=0,1,2,\dots,k)$:

$$\sup_{\theta \in \Theta} \left[\sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n) - \lambda_0 \underline{G}(\theta, X_0) \right] \geq 0 \quad (2.7)$$

Lorsque $k=0$, on retrouve l'axiome 2. Lorsque $\lambda_0=0$ et $\lambda_1=\lambda_2=\dots=\lambda_k=1$, on obtient:

$$\sup_{\theta \in \Theta} \sum_{n=1}^k \underline{G}(\theta, X_n) \geq 0, \quad \text{pour tout } k \geq 0, (X_n) \subseteq \mathbf{X} \quad (2.8)$$

On reconnaît le principe d'évitement des pertes sûres. Ainsi, la cohérence implique l'évitement des pertes sûres et des gains certains. L'axiome permet aussi d'éviter qu'un individu se contredise en spécifiant des prix d'achats de paris. En effet, imaginons qu'un individu fixe des prix maximums d'achats $\underline{E}(X_n)$ pour une suite $(X_n, n=1,2,\dots,k)$ de paris et viole l'axiome de cohérence. Dans ce cas, on peut arriver à le contraindre à payer un prix strictement supérieur à $\underline{E}(X_0)$ pour le pari X_0 , en prenant avantage des prix d'achats spécifiés pour les autres paris.

Démonstration: Supposons que l'axiome de cohérence soit violé pour une suite (X_n) de paris, ce que l'on peut représenter par l'équation suivante:

$$(\forall \theta \in \Theta) \quad \sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n) < \lambda_0 \cdot \underline{G}(\theta, X_0)$$

Dans ce cas, on peut toujours choisir $\delta > 0$ suffisamment petit de façon que l'équation suivante soit vérifiée:

$$(\forall \theta \in \Theta) \quad \sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n) \leq \lambda_0 \cdot [\underline{G}(\theta, X_0) - \delta]$$

En remplaçant le bénéfice net par sa définition, on obtient:

$$(\forall \theta \in \Theta) \quad \sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n) \leq \lambda_0 \cdot X_0(\theta) - \lambda_0 \cdot (\underline{E}(X_0) + \delta)$$

Une autre façon de dire qu'un individu est prêt à payer $\underline{E}(X)$ pour un pari $X(\theta)$ consiste à dire qu'il est prêt à accepter le pari $\underline{G}(\theta, X) = X(\theta) - \underline{E}(X)$ si on le lui offre gratuitement. Par définition, nous dirons que tout pari qu'un individu est prêt à accepter gratuitement est désirable. Ainsi, le bénéfice net $\underline{G}(\theta, X)$ de tout pari X est désirable. Nous ferons l'hypothèse que si un individu trouve un pari Y désirable, alors il trouvera aussi le pari $Y + \varepsilon$ désirable si $\varepsilon \geq 0$.

Ceci paraît naturel puisque le pari $Y+\varepsilon$ ne peut pas lui rapporter moins que le pari Y . S'il considère avantageux d'accepter Y , alors il devrait trouver au moins aussi avantageux d'accepter $Y+\varepsilon$. Pour les mêmes raisons, nous ferons l'hypothèse que si Y est désirable et si $Y(\theta) \leq Z(\theta)$ pour tout $\theta \in \Theta$, alors le pari Z est désirable, puisqu'il rapportera plus que Y quel que soit θ .

Puisque le bénéfice net de tout pari est désirable, en particulier $\underline{G}(\theta, \sum_{n=1}^k \lambda_n X_n)$ est désirable. Selon l'axiome 4:

$$\underline{G}(\theta, \sum_{n=1}^k \lambda_n X_n) + \varepsilon = \sum_{n=1}^k \underline{G}(\theta, \lambda_n X_n) \text{ où } \varepsilon \geq 0$$

Ainsi, $\sum_{n=1}^k \underline{G}(\theta, \lambda_n X_n) = \sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n)$ est désirable, de même que $\lambda_0 \cdot X_0(\theta) - \lambda_0 \cdot (\underline{E}(X_0) + \delta)$, qui domine $\sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n)$. Ceci implique, par définition, que l'individu serait prêt à payer un prix $\lambda_0 \cdot (\underline{E}(X_0) + \delta)$ pour le pari $\lambda_0 \cdot X_0$. Selon l'axiome 3, il serait aussi prêt à payer $(\underline{E}(X_0) + \delta)$ pour le pari X_0 , c'est-à-dire un montant strictement supérieur à $\underline{E}(X_0)$, le montant maximal qu'il avait fixé pour ce pari. ■

Axiome 6 - conglomérabilité

Pour présenter l'axiome de conglomérabilité, il faut d'abord définir le concept de prévision inférieure conditionnelle. Soit $\underline{E}(X|H)$ le montant maximal qu'un individu est prêt à payer pour un pari X , sachant que ce pari sera annulé si l'événement H ne se produit pas. On nomme $X|H$ un pari conditionnel, dont le bénéfice net est donné par $\underline{G}(\theta, X|H) = H(\theta) \cdot [X(\theta) - \underline{E}(X|H)]$, et $\underline{E}(X|H)$ sa **prévision inférieure conditionnelle**.

L'axiome de conglomérabilité spécifie que si \mathbf{H} constitue une partition de Θ et que le pari conditionnel $X|H$ est désirable pour tout $H \in \mathbf{H}$, c'est-à-dire qu'un individu est prêt à accepter le pari $X|H$ si on lui offre gratuitement, alors il devrait toujours considérer le pari X désirable. D'un point de vue décisionnel, cela revient à supposer que si un individu considère une action avantageuse quelle que soit l'hypothèse qu'il puisse faire sur l'état de la nature, alors il devrait trouver cette action avantageuse même s'il ne connaît pas l'état de la nature.

2.1.2 Probabilités imprécises cohérentes

Soit $A(\theta)$ un pari correspondant à la fonction indicatrice d'un événement A , i.e. $A(\theta)=1$ si $\theta \in A$, et $A(\theta)=0$ sinon. Pour distinguer ce type particulier de pari, on représente sa prévision inférieure par $\underline{P}(A)$, que l'on nomme aussi probabilité inférieure de l'événement A . De la même façon, on représente sa prévision supérieure par $\overline{P}(A)$, que l'on nomme la probabilité supérieure de l'événement A . Dans le cadre de la théorie des prévisions inférieures cohérentes, l'intervalle $[\underline{P}(A), \overline{P}(A)]$ peut être interprété directement, sans supposer l'existence d'une probabilité exacte $P(A)$.

Lorsque l'information sur une variable aléatoire est modélisée par une mesure de probabilité P , la prévision de tout pari (mesurable) $X(\theta)$ est entièrement déterminée par P : il s'agit de l'intégrale $E^P(X) = \int X dP$. La situation est différente pour les prévisions imprécises; on peut définir deux prévisions inférieures cohérentes différentes \underline{E}_1 et \underline{E}_2 qui génèrent cependant des probabilités imprécises semblables (Walley, 1991). Dans le cadre de la théorie des prévisions inférieures cohérentes, le concept de prévision imprécise ne peut donc être dérivé de celui de probabilité imprécise.

2.1.3 Principe d'extension des prévisions inférieures

Il est souvent intéressant de déterminer quelles sont les implications d'une prévision inférieure \underline{E} cohérente sur \mathbf{X} pour un pari $Z(\theta)$ ne faisant pas partie de \mathbf{X} . Ceci demande d'effectuer une extension du domaine de \underline{E} à l'ensemble de paris $\mathbf{Z} = \mathbf{X} \cup \{Z\}$. Si l'on veut que la prévision inférieure \underline{E} reste cohérente sur \mathbf{Z} , $\underline{E}(Z)$ doit respecter certaines contraintes. On peut montrer que la plus petite valeur de $\underline{E}(Z)$ qui permet à \underline{E} de rester cohérente sur \mathbf{Z} est obtenue en résolvant l'équation suivante, que l'on nomme le principe d'extension des prévisions inférieures (Walley, 1991):

$$\underline{E}(Z) = \sup \left\{ \inf_{\theta \in \Theta} \left[Z(\theta) - \sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n) \right], \text{ pour tout } k \geq 0, (X_n) \subseteq \mathbf{X}, \lambda_n \geq 0 \right\} \quad (2.9)$$

(La somme de $n=1$ à k est considérée valoir zéro pour $k=0$.)

Démonstration: Pour être cohérente sur $\mathbf{X} \cup \{Z\}$, la prévision $\underline{E}(Z)$ doit respecter l'axiome 5:

$$\sup_{\theta \in \Theta} \left[\sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n) - \lambda_0 [Z(\theta) - \underline{E}(Z)] \right] \geq 0, \quad \text{pour tout } k \geq 0, (X_n) \subseteq \mathbf{X}, \lambda_n \geq 0$$

Puisque \underline{E} est cohérent sur \mathbf{X} , cette inégalité est toujours vérifiée pour $\lambda_0=0$. On peut donc supposer que $\lambda_0 > 0$. Puisque l'on ne modifie pas la valeur logique de l'inégalité en divisant tous les termes par λ_0 , on peut sans perte de généralité supposer que $\lambda_0=1$. En isolant $\underline{E}(Z)$ on obtient alors:

$$\underline{E}(Z) \geq \inf_{\theta \in \Theta} \left[\sum_{n=1}^k Z(\theta) - \lambda_n \underline{G}(\theta, X_n) \right], \quad \text{pour tout } k \geq 0, (X_n) \subseteq \mathbf{X}, \lambda_n \geq 0$$

De façon équivalente, on peut écrire:

$$\underline{E}(Z) \geq \sup \left\{ \inf_{\theta \in \Theta} \left[Z(\theta) - \sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n) \right] : k \geq 0, (X_n) \subseteq \mathbf{X}, \lambda_n \geq 0 \right\}$$

En choisissant $\underline{E}(Z)$ égal au terme de droite, on fixe sa valeur de la façon la plus conservatrice possible tout en préservant la cohérence. ■

La valeur de $\underline{E}(Z)$ ainsi obtenue résume les implications de la prévision inférieure \underline{E} cohérente sur \mathbf{X} pour le pari Z . Par l'application de ce principe d'extension, il est en théorie possible d'étendre ainsi une prévision inférieure cohérente à tout ensemble de paris.

Il y a cependant des difficultés importantes liées à l'application pratique du principe d'extension. Si \mathbf{X} comporte k paris et si θ est un vecteur de m paramètres, la résolution de cette équation nécessite l'optimisation d'une fonction comportant $k+m$ variables. Lorsque k est infini, il peut être plus pratique d'utiliser le théorème de l'enveloppe inférieure, détaillé ci-après.

Il est important de comprendre que les prévisions obtenues par application du principe d'extension ne sont pas les seules prévisions qui permettent de conserver la cohérence avec les prévisions initiales. Ce que l'on obtient en appliquant cette règle, c'est la prévision minimale que peut fixer un individu pour préserver la cohérence. Il faut voir le principe d'extension comme un mécanisme permettant de compléter un espace \mathbf{X} de façon automatique, et qui donne une estimation conservatrice de la prévision inférieure de Z que pourrait fixer un individu.

On peut montrer que la prévision inférieure cohérente obtenue par l'application du principe d'extension du pari $\underline{E}(X+\alpha)$, où α est une constante, est donnée par $\underline{E}(X)+\alpha$.

Démonstration: la démonstration découle des propriétés des opérateurs sup et inf:

$$\begin{aligned} \underline{E}(X + \alpha) &= \sup \left\{ \inf_{\theta \in \Theta} \left[X(\theta) + \alpha - \sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n) \right] : n \geq 0, (X_n) \subseteq \mathbf{X}, \lambda_n \geq 0 \right\} \\ &= \sup \left\{ \inf_{\theta \in \Theta} \left[X(\theta) - \sum_{n=1}^k \lambda_n \underline{G}(\theta, X_n) \right] : n \geq 0, (X_n) \subseteq \mathbf{X}, \lambda_n \geq 0 \right\} + \alpha \end{aligned}$$

$$\underline{E}(X + \alpha) = \underline{E}(X) + \alpha \quad \blacksquare$$

2.1.4 Théorème de l'enveloppe inférieure

Les six axiomes de la théorie des prévisions inférieures cohérentes permettent de démontrer un théorème souvent utile en pratique pour construire des prévisions imprécises cohérentes, mais qui permet aussi de mieux comprendre la portée des axiomes. Ce théorème a été démontré dans le cas d'un référentiel fini par Smith (1961) et dans le cas d'un espace mesurable par Williams (1976).

Soit un ensemble \mathbf{X} des paris constitué de l'ensemble de toutes les fonctions mesurables par rapport à une tribu (ou σ -algèbre) \mathbf{A} du référentiel Θ . Une prévision inférieure \underline{E} est cohérente sur \mathbf{X} si et seulement si il existe une classe $\mathbf{M}=\{P\}$ de mesures de probabilité définies sur la tribu \mathbf{A} telle que:

$$\underline{E}(X) = \inf_{P \in \mathbf{M}} E^P(X) \text{ pour tout } X \in \mathbf{X} \quad (2.10)$$

Nous nommerons ce résultat le théorème de l'enveloppe inférieure, puisqu'il signifie qu'un ensemble de prévisions imprécises bornent toujours des prévisions précises obtenues à partir d'un ensemble de mesures de probabilité. Si \mathbf{X} ne constitue pas l'ensemble des fonctions mesurables par rapport à une tribu, on peut en théorie appliquer le principe d'extension des prévisions inférieures pour étendre \mathbf{X} à un domaine plus vaste pour lequel le théorème s'applique.

Le théorème de l'enveloppe inférieure signifie que la prévision inférieure d'un pari au sens de Walley (1991) peut toujours être obtenue à partir d'une classe de mesures de probabilité précises. Walley (1991) souligne que Good (1983) utilise ce résultat pour justifier l'approche bayésienne robuste, puisque la spécification partielle d'une mesure de probabilité à travers un ensemble de prévisions imprécises cohérentes revient à effectuer une analyse de sensibilité dans une classe de mesures de probabilité. Cet argument semble reposer sur

l'hypothèse implicite qu'il existe bel et bien une mesure de probabilité précise inconnue. En accord avec Walley (1991), nous croyons qu'il est nuisible de faire cette hypothèse, puisque dans les situations d'ignorance et de quasi-ignorance aucune mesure de probabilité ne peut être justifiée. Il est préférable d'interpréter directement les prévisions inférieures et supérieures comme des engagements à acheter ou à vendre des paris.

On peut utiliser le théorème de l'enveloppe inférieure pour construire une prévision inférieure cohérente \underline{E} en spécifiant directement une classe \mathbf{M} de mesures de probabilité. Cette approche n'est cependant pas toujours à privilégier car il est difficile en général de comprendre toutes les implications d'une classe de mesures de probabilité sur les valeurs des prévisions inférieures.

2.1.5 Règle de Bayes généralisée

Pour conserver la cohérence entre les prévisions inférieures non conditionnelles et conditionnelles, Walley (1991) montre que la prévision de ce bénéfice net doit être nulle, condition qui constitue la règle de Bayes généralisée:

$$\underline{E}[G(\theta, X|H)] = \underline{E}\{H \cdot [X - \underline{E}(X|H)]\} = 0 \quad (2.11)$$

La démonstration de la règle de Bayes généralisée est assez technique. On pourra consulter Walley (1991) pour comprendre comment cette règle découle des axiomes. Notons simplement que la démonstration repose principalement sur l'axiome de conglomérabilité.

Pour appliquer la règle de Bayes généralisée, il faut trouver une valeur $\mu = \underline{E}(X|H)$ telle que $\underline{E}[H \cdot (X - \mu)] = 0$. Cette valeur est unique dans la mesure où la probabilité inférieure $\underline{P}(H)$ de l'hypothèse H est strictement positive.

Démonstration: Soit $\varepsilon > 0$. Selon l'axiome de super-additivité:

$$\underline{E}\{H \cdot [X - (\mu - \varepsilon)]\} \geq \underline{E}[H \cdot (X - \mu)] + \varepsilon \cdot \underline{P}(H) > \underline{E}[H \cdot (X - \mu)]$$

Il s'ensuit que $\underline{E}[H \cdot (X - \mu)]$ est une fonction strictement décroissante en μ lorsque $\underline{P}(H) > 0$. Dans ce cas, si la valeur de μ telle que $\underline{E}[H \cdot (X - \mu)] = 0$ existe, elle est unique. ■

2.1.6 Événements de probabilité inférieure nulle

Dans le cas où $\underline{P}(H) = 0$, la règle de Bayes généralisée ne donne pas une solution unique, et on peut seulement affirmer que $\underline{E}(X|H) \geq \inf X(\theta)$. Cette solution est inacceptable, car dans les

problèmes d'inférence statistique on doit souvent obtenir des prévisions conditionnelles à des événements de probabilité inférieure nulle. En effet, si θ correspond aux paramètres du modèle $f(x|\theta)$ d'une variable aléatoire continue x , alors typiquement $P(\{x\}|\theta)=0$ pour tout x . On peut résoudre ce problème en remplaçant la fonction indicatrice de l'événement $H=\{x\}$ par la vraisemblance $L_x(\theta)=f(x|\theta)$, et écrire la règle de Bayes généralisée sous la forme suivante:

$$\underline{E}[G(\theta, X|x)] = \underline{E}\{L_x(\theta) \cdot [X(\theta) - \underline{E}(X|H)]\} = 0 \quad (2.12)$$

Cette équation a une solution unique dans la mesure où $f(x|\theta) > 0$. Comme dans le cas de la théorie bayésienne des probabilités, cette façon de procéder n'est justifiable que dans la mesure où l'on considère que la vraisemblance constitue un lissage d'une distribution discrète.

Le principe d'extension peut aussi être généralisé pour permettre de calculer une prévision non conditionnelle $\underline{E}(Z)$ à partir d'un ensemble de prévisions conditionnelles à des événements de probabilité nulle, mais dont la vraisemblance est strictement positive:

$$\underline{E}(Z) = \sup \left\{ \inf_{\theta \in \Theta} \left[Z(\theta) - \sum_{n=1}^k \lambda_n G(\theta, X_n | L_{x_n}) \right] : k \geq 0, (X_n) \subseteq \mathbf{X}, L_{x_n} > 0, \lambda_n \geq 0 \right\} \quad (2.13)$$

2.1.7 Cas particulier des probabilités inférieures 2-monotones

L'application de la règle de Bayes généralisée est plus facile lorsque les probabilités inférieures, induites par une prévision inférieure cohérente \underline{E} , sont 2-monotones, i.e. si:

$$\underline{P}(A \cap B) + \underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B) \quad \forall A, B \subseteq \Theta \quad (2.14)$$

Dans cette situation particulière, il est utile de définir la fonction de répartition inférieure $\underline{F}_{X|H}$ et la fonction de répartition supérieure $\bar{F}_{X|H}$ d'un pari $X(\theta)$ par:

$$\underline{F}_X(\theta) = \underline{P}(\{\omega: X(\omega) \leq \theta\}|H) \quad (2.15)$$

$$\bar{F}_X(\theta) = \bar{P}(\{\omega: X(\omega) \leq \theta\}|H) \quad (2.16)$$

En effet, la prévision conditionnelle imprécise $[\underline{E}(X|H), \bar{E}(X|H)]$ peut alors être obtenue par intégration de ces fonctions (Walley, 1991):

$$\underline{E}(X|H) = \int \theta \cdot d\bar{F}_{X|H}(\theta) \quad (2.17)$$

$$\bar{E}(X|H) = \int \theta \cdot d\underline{F}_{X|H}(\theta) \quad (2.18)$$

Le calcul des probabilités imprécises conditionnelles est aussi facilité (Walley, 1996b):

$$\underline{P}(A|H) = \frac{\underline{P}(A \cap H)}{\underline{P}(A \cap H) + \overline{P}(A^c \cap H)} \quad (2.19)$$

$$\overline{P}(A|H) = \frac{\overline{P}(A \cap H)}{\overline{P}(A \cap H) + \underline{P}(A^c \cap H)} \quad (2.20)$$

Notons que certains auteurs qualifient de convexes, plutôt que de 2-monotones, les probabilités imprécises qui respectent (2.14).

2.1.8 Modélisation de l'ignorance et de la quasi-ignorance a priori

La théorie des prévisions inférieures cohérentes permet de modéliser de façon satisfaisante l'ignorance a priori. Il suffit de choisir $\underline{E}(X) = \inf X(\theta)$ pour tout pari $X(\theta)$. Ceci permet d'obtenir une probabilité inférieure nulle $\underline{P}(A) = 0$ pour tout événement autre que le référentiel Θ . Cette représentation de l'ignorance a priori respecte à la fois le principe d'indifférence et le principe d'invariance de représentation, puisque $\underline{P}(A)$ ne dépend ni de A , ni du choix de Θ .

En pratique, ce modèle est peu utile car les situations d'ignorance complète sont rares, et les inférences obtenues à partir d'une ignorance totale sont très vagues. En effet, la classe \mathbf{M} de mesures de probabilité correspondant à ces prévisions inférieures est constituée de l'ensemble de toutes les mesures de probabilité.

Il est cependant possible de choisir les prévisions inférieures de façon que $\underline{E}(\theta) = \inf \theta$ et que $\underline{P}(A) = 0$ pour tout événement, sans que \mathbf{M} soit constitué d'un ensemble si étendu de distributions. De telles prévisions inférieures ne modélisent plus une ignorance complète, mais respectent au moins pour les probabilités les principes d'indifférence et d'invariance de représentation. Walley (1996a) présente un tel modèle de quasi-ignorance pour la distribution multinomiale.

2.1.9 L'indécision: une conséquence de l'imprécision

Les conclusions d'une analyse de la décision basée sur la théorie des prévisions inférieures cohérentes ne sont pas aussi définitives que celles issues d'une analyse bayésienne, car il n'est pas possible d'établir une relation d'ordre total entre les actions. On peut seulement comparer deux actions a et a' si le risque inférieur de l'une dépasse le risque supérieur de l'autre, c'est-à-dire si les intervalles $[\underline{R}(a), \overline{R}(a)]$ et $[\underline{R}(a'), \overline{R}(a')]$ ne se chevauchent pas. Nous dirons qu'une

action a est préférable à une action a' ssi $\underline{R}(a') > \bar{R}(a)$, et qu'une action a' est sous-optimale s'il existe une action a qui lui soit préférée. Soit S' l'ensemble des actions sous-optimales. Notons que S' peut être vide. Une analyse de la décision basée sur une évaluation imprécise du risque des actions permet uniquement d'éliminer les actions de l'ensemble S' , et d'identifier ainsi une classe $\hat{S} = S - S'$ d'actions préférables.

2.1.10 Comparaison avec l'approche bayésienne robuste

Le principal avantage de la théorie des prévisions inférieures cohérentes sur l'approche bayésienne robuste est qu'elle permet de modéliser l'ignorance et la quasi-ignorance a priori de façon satisfaisante, puisque les probabilités et prévisions imprécises peuvent être interprétées directement, sans faire l'hypothèse peu réaliste dans ces situations qu'il existe une «vraie» distribution a priori. Cette théorie donne donc une justification théorique à l'emploi pour ces situations de techniques bayésiennes robustes.

L'approche bayésienne robuste, applicable d'un point de vue théorique lorsqu'il est raisonnable d'accepter au moins l'existence d'une distribution a priori précise, peut être assimilée à une analyse de l'incertitude «en profondeur d'abord»: on tente d'identifier une distribution précise $b_0(\theta)$ et on construit un voisinage autour d'elle. Au contraire, la théorie des prévisions inférieures cohérentes correspond plus à une analyse de l'incertitude «en largeur d'abord»: une classe de distributions \mathbf{M} est délimitée à partir d'un ensemble de contraintes, sans jamais avoir à préciser autant notre pensée. Selon le type de problème, une recherche «en profondeur d'abord» ou «en largeur d'abord» peut être plus appropriée.

Lorsqu'il existe une imprécision à la fois sur la distribution a priori et sur la fonction de coût, une différence plus importante apparaît entre la théorie des prévisions inférieures et l'approche bayésienne robuste. En effet, lorsqu'il faut prendre une décision, l'approche «en largeur d'abord» de la théorie des prévisions inférieures ne permet que d'éliminer les actions a dont le risque inférieur $\underline{R}(a) = \underline{E}[g(a, \theta)]$ dépasse le risque supérieur $\bar{R}(a') = \bar{E}[g(a', \theta)]$ d'une autre action a' . Il se peut fort bien qu'une action a ne soit optimale pour aucune combinaison de distribution a priori et de fonction de coût précises, mais qu'il n'existe aucune action a' telle que $\underline{R}(a) \geq \bar{R}(a')$. Dans ce cas, du point de vue de la théorie des prévisions inférieures cohérentes, rien ne permet de discréditer l'action a . Avec une interprétation d'analyse de

sensibilité, on conclurait pourtant que l'action a ne peut pas être la «véritable» action optimale. L'approche bayésienne robuste est donc plus restrictive.

2.1.11 Comparaison avec les modèles descriptifs à base de capacités de Choquet

Schmeidler (1989) a proposé un modèle descriptif du comportement d'un décideur basé sur des capacités de Choquet. Rappelons qu'une capacité de Choquet est une fonction d'ensemble croissante, i.e. une fonction C associant un nombre entre 0 et 1 à tous les éléments d'une tribu \mathbf{A} de telle façon que $C(A) \geq C(B)$ si $A \supseteq B$ pour tout $A, B \in \mathbf{A}$. Une mesure de probabilité est donc une capacité de Choquet, mais l'inverse est faux puisque les capacités de Choquet ne sont pas nécessairement additives. Pour la construction d'un modèle descriptif, la souplesse accrue des capacités de Choquet est utile. Choquet (53-54) explique comment l'on peut définir l'intégrale de Choquet d'une fonction mesurable f par rapport à une capacité de Choquet C . Schmeidler (1989) suggère que l'on peut approcher le comportement d'un décideur en supposant qu'il agit de façon à maximiser l'intégrale d'une fonction d'utilité $U: (S, \Theta) \rightarrow \mathbf{R}$ par rapport à une capacité de Choquet $C: \mathbf{A} \rightarrow [0, 1]$, définie sur une tribu \mathbf{A} du référentiel Θ . Autrement dit, il existe une fonction $V: S \rightarrow \mathbf{R}$ telle que $V(a) = \int U(a, \theta) dC(\theta)$, C est une capacité de Choquet, et $V(\hat{a}) = \inf_{a \in S} V(a)$.

Ce modèle se rapproche de la théorie des prévisions inférieures, car on peut montrer qu'une probabilité inférieure cohérente constitue toujours une capacité de Choquet, et qu'une prévision inférieure $\underline{E}(X)$ correspond à l'intégrale de Choquet $\int X d\underline{P}$, lorsque \underline{P} est 2-monotone (Chateauneuf et Jaffray, 1989). La différence principale entre le modèle de Schmeidler (1989) et celui de Walley (1991) est que le premier permet, une fois que la capacité C et la fonction d'utilité U sont connues, de construire un ordre total sur les actions. Avec la théorie des prévisions inférieures cohérentes, on peut seulement établir un ordre partiel. Dans certaines situations, l'indécision résultant de l'application de la théorie des prévisions inférieures cohérentes est acceptable, parce qu'elle permet d'identifier des situations où l'information disponible est insuffisante pour comparer l'utilité des décisions. Dans d'autres situations, en

particulier pour l'estimation de la valeur de l'information, nous verrons qu'il faut choisir, et conséquemment qu'il faut que le modèle permette d'en arriver à une décision non équivoque. Même si l'on perd alors la cohérence (au sens de Walley), il est alors utile de considérer un modèle descriptif, tel que celui de Schmeidler.

2.1.12 Application pratique de la théorie des prévisions inférieures cohérentes

On peut envisager principalement deux façons d'appliquer la théorie des prévisions inférieures cohérentes:

- définir une prévision inférieure cohérente par une famille de mesures de probabilités \mathbf{M} , et appliquer le théorème de l'enveloppe inférieure pour effectuer les inférences nécessaires;
- effectuer un ensemble de prévisions imprécises, s'assurer de leur cohérence, et effectuer les inférences nécessaires à l'aide du principe d'extension des prévisions inférieures et de la règle de Bayes généralisée.

La première approche est très liée, au moins du point de vue des méthodes de calculs, aux développements récents de l'approche bayésienne robuste. Les applications de la théorie des prévisions inférieures cohérentes, à ce jour peu nombreuses (Walley et Campello de Souza, 1990; Walley, 1993; Walley, 1996a; Walley et al. 1996), ont favorisé cette première approche. Nous nous intéresserons aussi à la seconde.

2.1.12.1 Exemple d'application du principe d'extension des prévisions inférieures

Il est utile de montrer par un exemple comment la théorie peut être appliquée de cette façon, et ce qu'elle peut apporter. Soit $(x_j, j=1,2,\dots,n)$ une suite de variables de Bernoulli indépendantes et r le nombre de succès observés. Conditionnelle à la connaissance de la probabilité θ d'observer un succès, la distribution de r est binomiale $f(r,n|\theta) = \binom{n}{r} \theta^r (1-\theta)^{n-r}$. La représentation de l'incertitude totale sur r dépend cependant aussi de l'information disponible sur θ . Supposons que l'analyste puisse résumer son information a priori par deux prévisions précises:

- $E(\theta)=1/4$;
- $E(\theta|x_1=0)=E(\theta|r=0,n=1)=1/8$.

Que sait-on sous ces hypothèses de la probabilité d'observer r succès en n essais? Une analyse bayésienne demanderait de choisir une distribution a priori $b(\theta)$ pour le paramètre θ . Une solution simple numériquement consiste à choisir une loi bêta respectant les deux informations a priori disponibles, car la loi bêta est conjuguée par rapport à la binomiale, de sorte que la distribution a posteriori est une bêta de paramètres modifiés lorsque la distribution a priori est bêta. À partir de l'équation (1.17), on peut vérifier que la seule loi de la famille bêta qui respecte les deux prévisions précises $E(\theta)=1/4$ et $E(\theta|r=0,n=1)=1/8$ a pour paramètres $\alpha=1/4$ et $\beta=3/4$. La distribution prévisionnelle $f(r,n,\alpha,\beta)$ est alors bêta-binomiale, de sorte que la probabilité d'observer r succès en n essais peut être calculée précisément de façon analytique. Par exemple, la Figure 2.1 présente $f(r,n,\alpha,\beta)$ pour $n=4$.

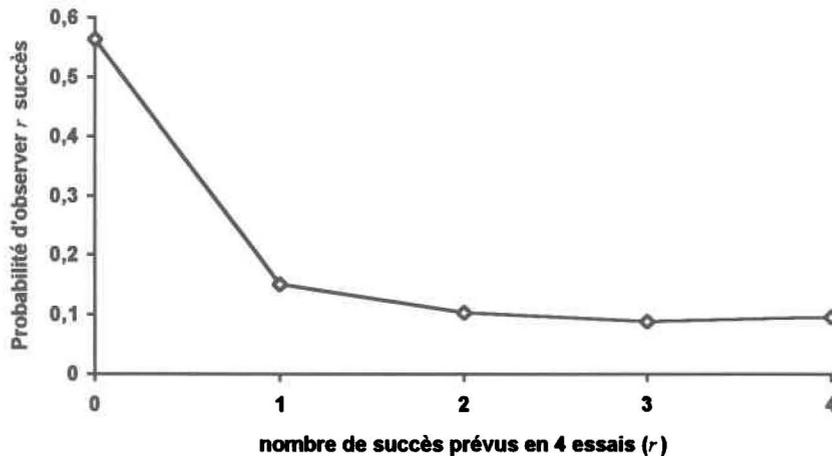


Figure 2.1: Exemple de distribution prévisionnelle d'une variable aléatoire binomiale basée sur une distribution a priori bêta ($\alpha=1/4$, $\beta=3/4$, $n=4$)

On peut se demander quelle est l'influence de l'hypothèse de la loi a priori bêta sur la distribution prévisionnelle. En effet, nous avons choisi de façon arbitraire une loi de la famille bêta respectant les deux contraintes $E(\theta)=1/4$ et $E(\theta|x_1=0)=1/8$ uniquement pour simplifier les calculs. Avec la théorie des prévisions inférieures cohérentes, on peut déterminer les probabilités imprécises cohérentes avec ces contraintes sans faire d'hypothèse supplémentaire sur la distribution a priori, sans même supposer qu'il existe une distribution a priori permettant de modéliser correctement l'information disponible.

L'information a priori peut être représentée par quatre prévisions inférieures. En effet, la contrainte $E(\theta)=1/4$ signifie que la prévision inférieure et la prévision supérieure de θ doivent toutes deux être fixées à $1/4$: $\underline{E}(\theta)=\bar{E}(\theta)=1/4$. La contrainte sur la prévision supérieure peut être transformée, par application de l'axiome 1, en une prévision inférieure: $\underline{E}(-\theta)=-1/4$. De la même façon, la seconde contrainte implique que $\underline{E}(\theta|x_1=0)=1/8$ et que $\underline{E}(-\theta|x_1=0)=-1/8$.

On peut alors appliquer le principe d'extension des prévisions inférieures [eq. (2.9)] pour obtenir $\underline{f}(r,n)=\underline{E}(L_{r,n})$ et $\bar{f}(r,n)=\bar{E}(L_{r,n})=-\underline{E}(-L_{r,n})$, où $L_{r,n}$ est la vraisemblance de θ pour r succès observés en n essais: $L_{r,n}(\theta)=f(r,n|\theta)$. Ceci demande d'abord de calculer le bénéfice net, ou pari marginal, relié à chaque prévision:

$$\underline{G}(\theta)=\theta-1/4$$

$$\underline{G}(-\theta)=-\theta+1/4$$

$$\underline{G}(\theta|x_1=0)=L_{0,1}(\theta)\cdot(\theta-1/8)=(1-\theta)\cdot(\theta-1/8)$$

$$\underline{G}(-\theta|x_1=0)=L_{0,1}(\theta)\cdot(-\theta+1/8)=(1-\theta)\cdot(-\theta+1/8)$$

L'équation à résoudre pour obtenir $\underline{f}(r,n)$ est donc la suivante:

$$\begin{aligned} \underline{f}(r,n) &= \sup_{\lambda_n \geq 0} \left\{ \inf_{\theta \in \Theta} \left[L_{r,n}(\theta) - \lambda_1 \underline{G}(\theta) - \lambda_2 \underline{G}(-\theta) - \lambda_3 \underline{G}(\theta|x_1=0) - \lambda_4 \underline{G}(-\theta|x_1=0) \right] \right\} \\ &= \sup_{\lambda_n \geq 0} \left\{ \inf_{\theta \in \Theta} \left[\binom{n}{r} \theta^r (1-\theta)^{n-r} + (\lambda_2 - \lambda_1) \left(\theta - \frac{1}{4} \right) + (\lambda_4 - \lambda_3) (1-\theta) \left(\theta - \frac{1}{8} \right) \right] \right\} \end{aligned} \quad (2.21)$$

Il est possible de simplifier davantage cette expression en remplaçant $(\lambda_2 - \lambda_1)$ par k_1 et $(\lambda_4 - \lambda_3)$ par k_2 , les paramètres k_1 et k_2 pouvant être négatifs ou positifs:

$$\underline{f}(r,n) = \sup_{k_1, k_2} \left\{ \inf_{\theta \in \Theta} \left[\binom{n}{r} \theta^r (1-\theta)^{n-r} + k_1 \left(\theta - \frac{1}{4} \right) + k_2 (1-\theta) \left(\theta - \frac{1}{8} \right) \right] \right\} \quad (2.22)$$

On peut obtenir de la même façon $\bar{f}(r,n)$:

$$\bar{f}(r,n) = \inf_{k_1, k_2} \left\{ \sup_{\theta \in \Theta} \left[\binom{n}{r} \theta^r (1-\theta)^{n-r} + k_1 \left(\theta - \frac{1}{4} \right) + k_2 (1-\theta) \left(\theta - \frac{1}{8} \right) \right] \right\} \quad (2.23)$$

La Figure 2.2 présente les probabilités imprécises obtenues. On observe que les valeurs correspondant à l'hypothèse d'une distribution bêta a priori sont comprises entre les

probabilités inférieures et supérieures, en accord avec le théorème de l'enveloppe inférieure. L'imprécision est assez importante, de l'ordre de 10%.

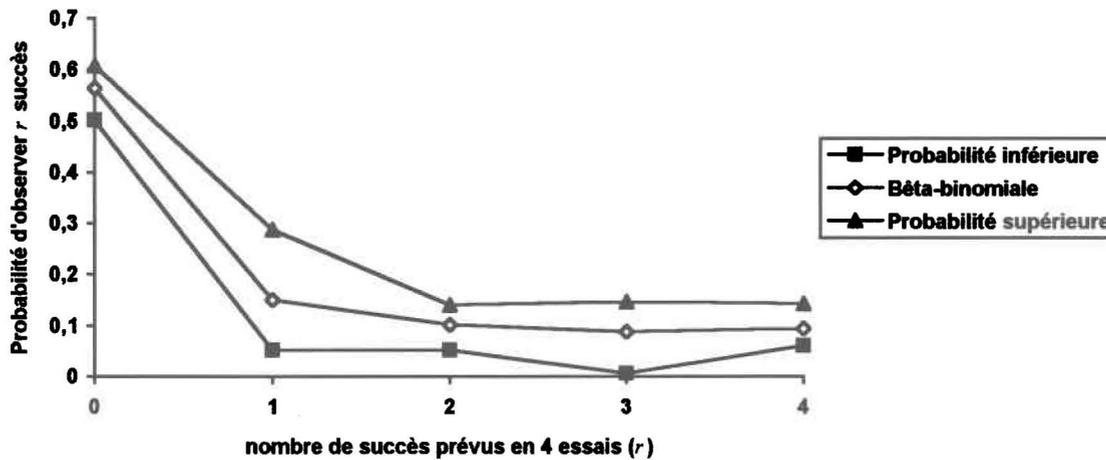


Figure 2.2: Exemple de prévision imprécise d'une variable aléatoire binomiale

Il est intéressant d'observer comment évoluent ces prévisions imprécises pour r fixé, en fonction de n . La Figure 2.3 montre les probabilités inférieures et supérieures pour l'événement consistant à n'observer aucun succès en n essais, ainsi que l'estimation obtenue en faisant l'hypothèse d'une loi bêta respectant les contraintes. Pour de faibles valeur de n , l'imprécision est faible, et l'approximation bayésienne conduisant à la loi bêta-binomiale peut être acceptable. Par contre, lorsque n devient grand, l'imprécision est beaucoup plus importante, et la différence entre la probabilité supérieure et la probabilité inférieure tend vers 60%.

Ce phénomène est analogue à un problème d'extrapolation. L'analyse bayésienne est basée sur un modèle d'incertitude simple, à deux paramètres, qui est en quelque sorte «calé» sur deux informations a priori. A partir de ce modèle, on effectue des prévisions sur le comportement statistique d'échantillons de plus en plus long. Lorsque la taille d'échantillon est inférieure ou égale au nombre de prévisions effectuées a priori ($n \leq 2$), il n'y a aucune imprécision sur la probabilité de n'observer que des échecs; le modèle bayésien, fondé sur l'hypothèse d'une distribution a priori particulière, est suffisant. Par contre, pour $n > 2$, le modèle bayésien est trop simple pour représenter toute l'incertitude. Alors que la borne inférieure décroît de façon exponentielle, la probabilité précise issue du modèle bêta-binomial décroît en $n^{-1/4}$. La distribution a priori bêta impose donc une vitesse de décroissance précise

pour $f(r=0, n)$, alors que plusieurs types de décroissance sont compatibles avec les contraintes imposées par l'information a priori.

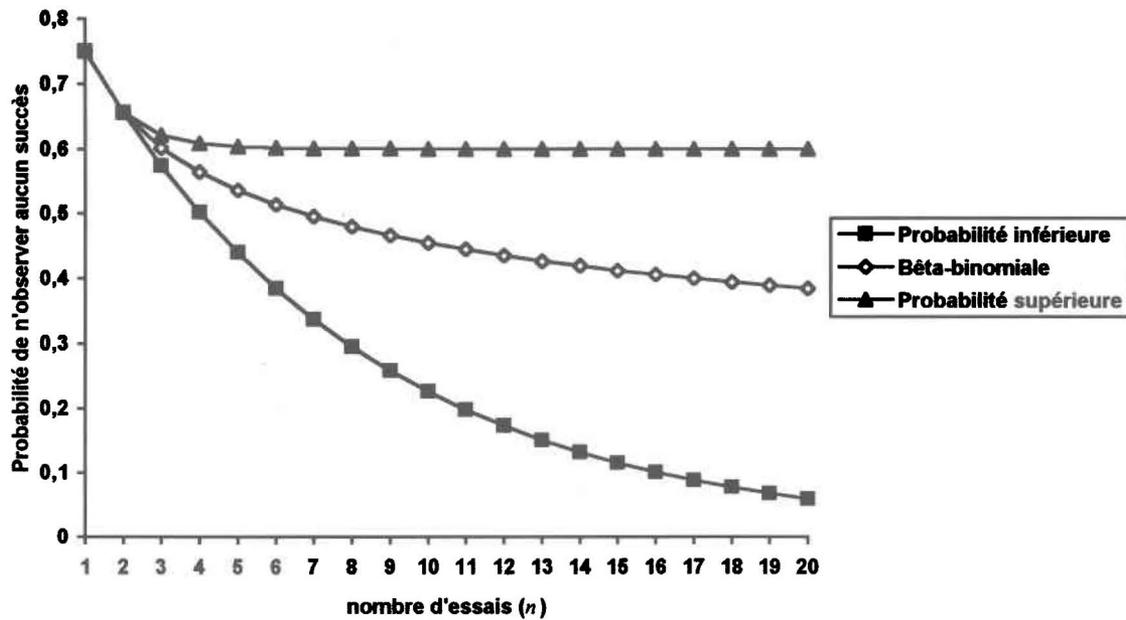


Figure 2.3: Probabilité imprécise de n'observer que des échecs en n essais indépendants

2.1.12.2 Exemple d'application de la règle de Bayes généralisée

Au lieu de prévoir la probabilité de n'observer que des échecs en fonction de n à partir d'une information a priori limitée, on peut aussi observer comment évolue cette probabilité pour n fixé lorsque l'information a priori s'accumule. On peut effectuer une telle prévision conditionnelle en appliquant la règle de Bayes généralisée.

Supposons que l'on dispose, en plus de l'information a priori modélisée à la section précédente, d'un échantillon aléatoire constitué de k succès en N essais. Que peut-on dire de la probabilité d'observer r succès au cours des n prochains tirages? Calculons par exemple la probabilité (imprécise) d'observer quatre échecs successifs ($r=0, n=4$), après observation de N échecs en N essais ($k=0$), pour N variant de 0 à 10. Il s'agit donc de calculer les prévisions imprécises suivantes:

$$\underline{f}(r=0, n=4|k=0, N) = \underline{E}[L_{0,4}|L_{0,N}] \quad (2.24)$$

$$\bar{f}(r=0, n=4|k=0, N) = -\underline{E}[-L_{0,4}|L_{0,N}] \quad (2.25)$$

Selon la règle de Bayes généralisée [eq. (2.11)], $E[L_{0,4}|L_{0,N}]$ correspond à la valeur μ telle que $E\{L_{0,N} \cdot [L_{0,4} - \mu]\} = 0$. Sachant que $0 \leq \mu \leq 1$, nous avons résolu cette équation en appliquant la méthode numérique Regula-Falsi sur cet intervalle, et nous avons procédé de façon semblable pour la prévision supérieure.

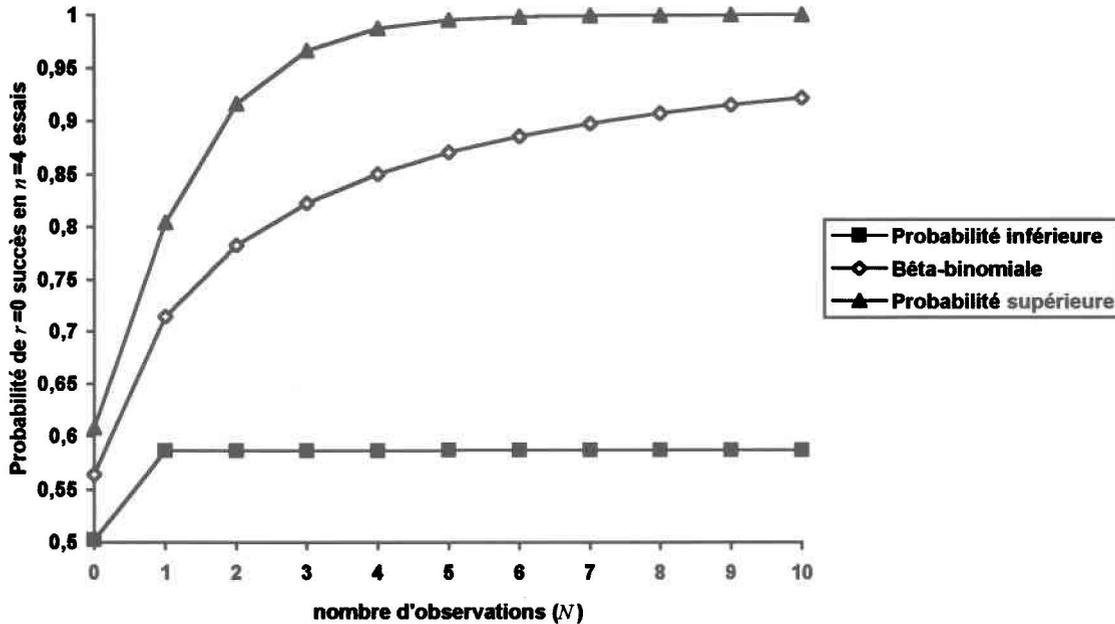


Figure 2.4: Probabilité imprécise d'observer 4 échecs consécutifs, après avoir observé N échecs successifs, avec deux contraintes précises

La Figure 2.4 présente les résultats obtenus en solutionnant selon cette procédure la règle de Bayes généralisée. On a aussi tracé les valeurs obtenues à partir de la distribution bêta-binomiale. On voit d'abord que les probabilités précises issues de l'approche bayésienne respectent les bornes inférieures et supérieures, en accord avec le théorème de l'enveloppe inférieure. Cependant, chose surprenante, la borne inférieure ne converge pas vers la fréquence observée. Comment interpréter ce résultat?

Du point de vue de la théorie des prévisions inférieures cohérentes, il n'y a aucune difficulté. En effet, nous savons que ces résultats ont été obtenus par application du principe d'extension, et que ce dernier ne donne réellement qu'une estimation conservatrice de la

probabilité imprécise. On sait seulement qu'un individu cohérent avec lui-même devrait fixer des probabilités au moins aussi précises que les résultats obtenus de façon automatique.

Si l'on utilise une interprétation bayésienne robuste, ces résultats signifient que les contraintes ne permettent pas d'éliminer des distributions a priori qui sont «inacceptables» parce que trop concentrées en certains points de l'espace. Il faudrait ajouter des contraintes supplémentaires.

Nous proposons une troisième façon de voir les choses. L'imprécision des probabilités dans l'exemple précédent provient du fait que les contraintes modélisant l'information a priori ne sont pas suffisantes pour spécifier une distribution a priori. Les résultats indiquent que les contraintes sont cohérentes avec des modèles qui convergent plus vite vers la fréquence observée que la distribution bêta-binomiale, et aussi avec des modèles qui ne convergent pas vers cette fréquence. Selon nous, un individu qui choisit de faire converger $f(r = 0, n = 4 | k = 0, N)$ très lentement vers 1, malgré l'observation d'une longue séquence d'observations identiquement nulles, doit posséder une information a priori ayant un poids substantiel par rapport à l'information d'échantillonnage. On peut alors interpréter le fait que ce même individu n'a spécifié que deux prévisions précises a priori en supposant qu'il dispose d'une information a priori importante qu'il n'a pas su ou qu'il n'a pas voulu modéliser.

Un individu peut utiliser les résultats précédents pour réaliser qu'il n'a pas codé toute l'information a priori dont il dispose. Peut-être peut-il porter un jugement explicite sur la convergence des probabilités imprécises vers la fréquence observée? Par exemple, il pourrait ajouter la contrainte $E(\theta | k=0, N) = 1/[4(N+1)]$. L'ajout de cette troisième contrainte est suffisante pour permettre aux probabilités imprécises de converger vers la fréquence observée, comme le montre la Figure 2.5.

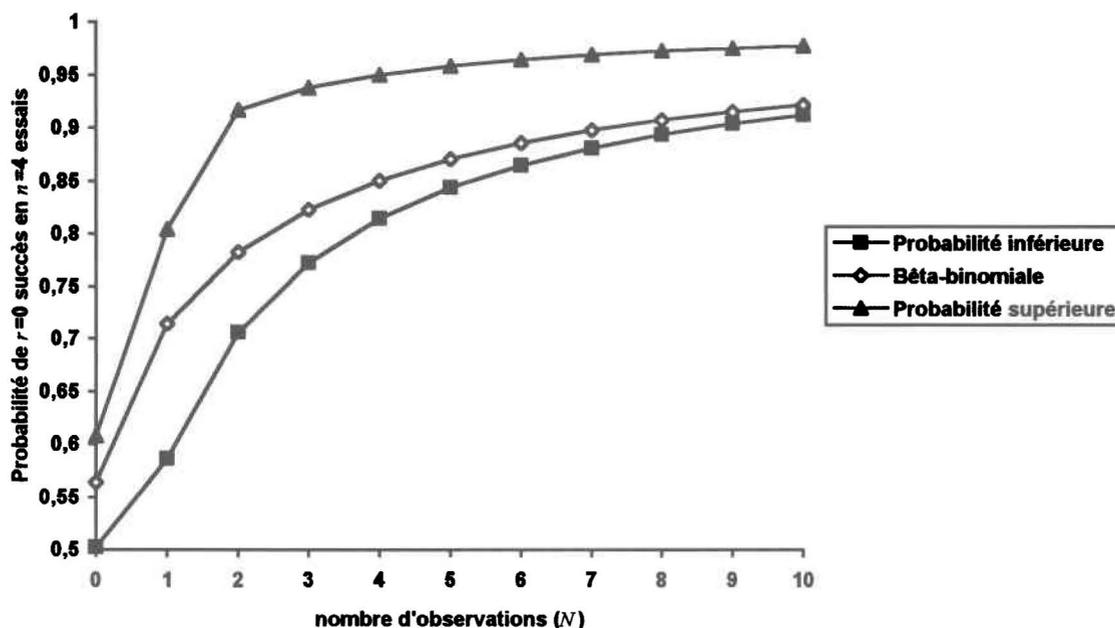


Figure 2.5: Probabilité imprécise d'observer 4 échecs consécutifs, après avoir observé N échecs successifs, avec trois contraintes précises

Après introspection, un individu peut cependant aussi réaliser qu'il est incapable de proposer d'autres contraintes reflétant une réelle information a priori. Dans ce cas, les deux contraintes initiales constituant l'ensemble de l'information a priori. En comparaison, l'information d'échantillonnage devient alors plus importante, et il semble naturel de se tourner plus rapidement vers la fréquence observée pour estimer la probabilité. Dans les situations où les contraintes déjà explicitées mobilisent toute l'information a priori, nous suggérons donc que la courbe qui converge le plus rapidement vers la fréquence observée, tout en restant cohérente avec les probabilités inférieures et supérieures calculées, constitue un modèle acceptable du peu d'information a priori disponible. Ce modèle est celui qui accorde le poids le plus faible à l'information a priori, mesuré en «unités de probabilité». Dans le cas de l'exemple que nous avons présenté, la borne supérieure de la probabilité constituerait le modèle le moins informatif compatible avec les contraintes, puisqu'il n'est pas possible de concevoir une distribution de probabilité précise qui permette de respecter les contraintes et de converger plus vite vers la fréquence observée.

En pratique, il est plutôt rare d'être certain d'avoir modélisé toute l'information a priori, de sorte que l'approche que nous venons de proposer pour retenir un modèle précis minimisant le poids de l'information a priori sera rarement applicable. Par contre, cette approche peut être utilisée pour mesurer l'effet de l'ajout d'une contrainte. En combinant les la Figure 2.4 et la Figure 2.5, on peut par exemple comparer les probabilités supérieures $\bar{f}(r = 0, n = 4 | k = 0, N)$ correspondant aux situations où l'on fixe deux, puis trois contraintes précises, avec le modèle bayésien bêta-binomial (Figure 2.6).

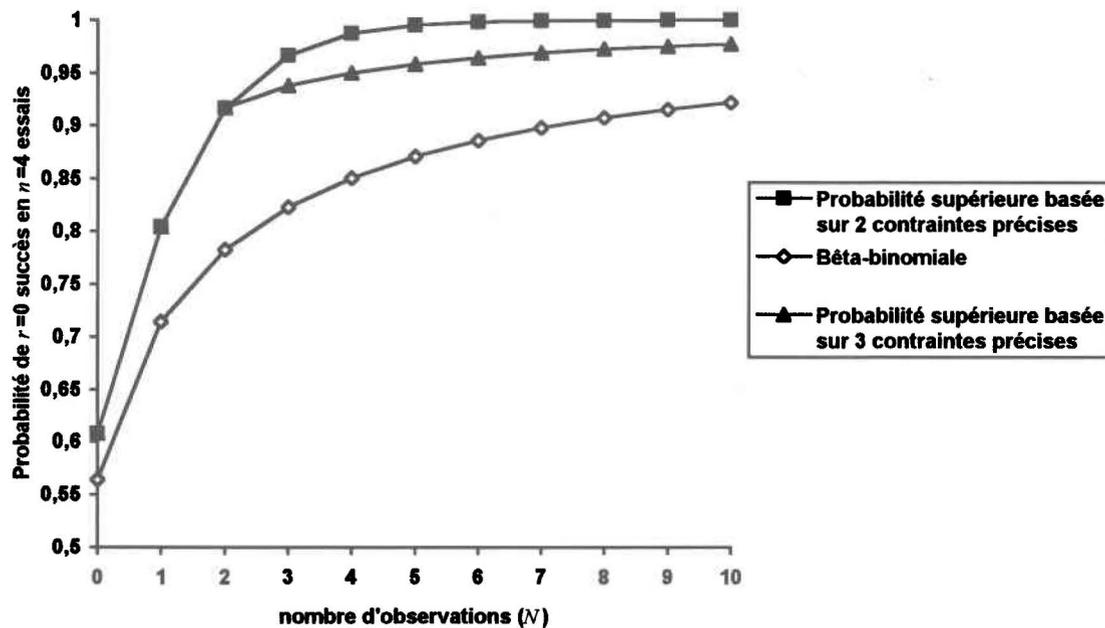


Figure 2.6: Probabilité supérieure d'observer 4 échecs consécutifs, après avoir observé N échecs en N essais, fondées sur deux et trois contraintes précises, et comparaison avec une solution bayésienne

On voit qu'en fonction du niveau d'information a priori, la probabilité converge plus ou moins rapidement vers la fréquence observée. On peut ainsi mesurer le poids d'une information a priori supplémentaire, représentée par une contrainte supplémentaire ou par le choix d'une distribution a priori, en terme de réduction de la vitesse de convergence vers la fréquence observée.

2.2 Proposition d'une approche permettant d'estimer la valeur de l'information dans le cadre de la théorie des prévisions inférieures cohérentes

La définition du concept de valeur de l'information est plus problématique dans le cadre de la théorie des prévisions inférieures cohérentes, principalement à cause de l'indécision que peut entraîner l'imprécision. Si l'on adoptait un point de vue d'analyse de sensibilité, on pourrait utiliser la même définition de la valeur de l'information que celle proposée pour une approche bayésienne robuste, par application du théorème de l'enveloppe inférieure, c'est-à-dire déterminer la valeur de l'information au sens bayésien pour chaque distribution a priori compatible avec les contraintes, et construire un ensemble $\{XVSI\}$ de valeurs possibles. Cette approche est problématique pour deux raisons. D'abord, d'un point de vue pratique, cela impose l'utilisation du théorème de l'enveloppe inférieure comme méthode d'inférence, alors que l'application du principe d'extension est souvent plus facile à mettre en oeuvre. Mais la critique la plus forte de cette solution est qu'elle fait perdre au concept de valeur de l'information son caractère opérationnel.

En effet, cette définition est applicable selon nous uniquement dans la situation particulière où l'on veut évaluer, à mi-parcours du processus de modélisation de l'information a priori, la valeur de l'information. On suppose que, par la suite, on sera en mesure d'identifier une distribution a priori précise, et que la valeur de l'information sera alors donnée par l'un des éléments de l'ensemble $\{XVSI\}$. En pratique, il faut une définition plus générale applicable dans la situation où l'information a priori ne permet pas d'identifier une seule action optimale.

2.2.1 Valeur imprécise de l'information

Si un individu réalise que son information a priori ne lui permet que d'identifier un ensemble d'actions préférables \hat{S} , il pourra soit ne rien faire, soit choisir une action parmi cet ensemble d'une façon ou d'une autre. Dans le reste de ce travail, nous ferons l'hypothèse qu'**il faut choisir**. Malgré l'indécision engendrée par l'imprécision, l'individu doit prendre une décision. Nous représenterons la stratégie employée par un individu pour effectuer ce choix par une fonction δ qui dépend de l'information disponible, et aussi de facteurs plus arbitraires, que nous ne détaillerons pas explicitement. Nous écrirons donc δ_0 pour représenter la règle de

décision basée sur l'information a priori, et $\delta(x)$ pour représenter la règle de décision a posteriori. Ceci permet de définir la prévision imprécise de la valeur de l'information à partir de la différence du coût $g(\delta_0, \theta)$ de l'action sélectionnée a priori et du coût $g(\delta(x), \theta)$ de l'action sélectionnée a posteriori:

$$\underline{VSI}(x, \delta) = \underline{E}[g(\delta_0, \theta) - g(\delta(x), \theta) | L_x] \quad (2.26)$$

$$\overline{VSI}(x, \delta) = \overline{E}[g(\delta_0, \theta) - g(\delta(x), \theta) | L_x] \quad (2.27)$$

Pour θ connu, la prévision de cette valeur imprécise de l'information s'obtient en prenant l'espérance par rapport à la distribution d'échantillonnage:

$$\underline{XVSI}(\theta, \delta) = E^{f(x|\theta)}[\underline{VSI}(x, \delta)] \quad (2.28)$$

$$\overline{XVSI}(\theta, \delta) = E^{f(x|\theta)}[\overline{VSI}(x, \delta)] \quad (2.29)$$

La prévision inférieure de $\underline{XVSI}(\theta, \delta)$ et la prévision supérieure de $\overline{XVSI}(\theta, \delta)$ permettent d'obtenir la valeur prévisionnelle imprécise de l'information sans connaître θ :

$$\underline{XVSI}(\delta) = \underline{E}[\underline{XVSI}(\theta, \delta)] \quad (2.30)$$

$$\overline{XVSI}(\delta) = \overline{E}[\overline{XVSI}(\theta, \delta)] \quad (2.31)$$

Walley (1991) montre que cette approche est numériquement équivalente à une prévision imprécise jointe de $g(\delta_0, \theta) - g(\delta(x), \theta)$, c'est-à-dire en travaillant directement dans l'espace-produit des référentiels de θ et x . La solution consistant à déterminer d'abord la valeur a posteriori de l'information est plus simple à appliquer sur le plan numérique.

La définition que nous proposons de la valeur de l'information dans le contexte de la théorie des prévisions inférieures cohérentes dépend d'une règle de décision arbitraire permettant au décideur de régler les situations d'indécision. Il pourrait s'agir d'une règle randomisée, ou obéissant à des principes autres. Il sera évidemment intéressant d'étudier la sensibilité de la valeur de l'information à la règle de décision.

Selon la règle de décision choisie, la valeur de l'information pourra être positive ou négative. Le choix d'une règle de décision compatible avec une distribution a priori permet d'éviter cette situation, mais un tel comportement indiquerait plutôt qu'une information a priori importante n'a pas été modélisée. En effet, si l'on choisit d'agir de façon conforme à une

analyse bayésienne de la décision, c'est peut-être parce que l'on a assez d'information pour fixer une distribution a priori.

Notre définition de la valeur de l'information est à rapprocher de celle de Davis et al. (1979), qui ont généralisé le concept bayésien usuel aux situations où l'on prend une décision non optimale. Dans ces circonstances, la valeur de l'information peut aussi être négative, et dépend de la règle de décision utilisée.

2.2.2 Valeur imprécise de l'information parfaite

On peut proposer une définition de la valeur de l'information parfaite dans le cadre de la théorie des prévisions inférieures cohérentes. Rappelons que le regret OL d'une règle de décision δ correspond à la différence entre le coût de la décision a priori δ_0 et le coût minimal:

$OL(\delta, \theta) = g(\delta_0, \theta) - \inf_{a \in S} g(a, \theta)$. Notons $[\underline{XOL}(a), \overline{XOL}(a)]$ la prévision imprécise de ce regret:

$$\underline{XOL}(\delta) = \underline{E} \left[g(\delta_0, \theta) - \inf_{a \in S} g(a, \theta) \right] \quad (2.32)$$

$$\overline{XOL}(\delta) = \overline{E} \left[g(\delta_0, \theta) - \inf_{a \in S} g(a, \theta) \right] \quad (2.33)$$

Ces deux quantités constituent respectivement des bornes supérieures pour $\underline{XVSI}(\delta)$ et $\overline{XVSI}(\delta)$, et peuvent ainsi être interprétées comme une prévision imprécise de la valeur d'une information parfaite.

2.2.3 Construction d'une règle de décision

On peut proposer une famille de règles de décision non randomisées susceptibles de représenter une variété de comportements. Considérons d'abord deux cas extrêmes. Un décideur pourrait envisager une stratégie \underline{E} -minimax, consistant à choisir, parmi les actions de l'ensemble \hat{S} , celle qui minimise le risque maximal: $\bar{R}(\delta) = \inf_{a \in \hat{S}} \bar{R}(a)$. On remarque qu'en

l'absence d'information a priori, cette règle de décision est similaire à l'approche minimax en analyse classique de la décision (Lehmann, 1983), puisque dans ce cas $\hat{S} = S$. A l'opposé, lorsque l'information est suffisante pour préciser une distribution a priori, on retombe sur l'approche bayésienne décisionnelle, puisque \hat{S} ne contient alors que des actions de risque

identique. La stratégie \underline{E} -minimax est donc intéressante puisqu'elle permet de généraliser une approche décisionnelle classique et la théorie bayésienne de la décision (Walley, 1991). Ce n'est cependant pas la seule règle de décision possible.

On peut aussi envisager une stratégie \underline{E} -minimin, consistant à choisir, parmi les actions de l'ensemble \hat{S} , celle qui minimise le risque minimal: $\underline{R}(\delta) = \inf_{a \in \hat{S}} \underline{R}(a)$. Entre ces deux extrêmes, on peut aussi envisager une règle $\delta^{(\kappa)}$ consistant à comparer les actions sur la base d'un point situé entre $\underline{R}(a)$ et $\bar{R}(a)$, à une fraction κ de la distance entre ces deux bornes:

$$\underline{R}(\delta^{(\kappa)}) + \kappa \cdot [\bar{R}(\delta^{(\kappa)}) - \underline{R}(\delta^{(\kappa)})] = \inf_{a \in \hat{S}} \{ \underline{R}(a) + \kappa \cdot [\bar{R}(a) - \underline{R}(a)] \} \quad (2.34)$$

On observe que la règle \underline{E} -minimax est obtenue pour $\kappa=1$ et la règle \underline{E} -minimin pour $\kappa=0$. Si l'on sait que la règle de décision fera partie de cette famille, on peut proposer une estimation conservatrice de la valeur de l'information en prenant le maximum et le minimum de la valeur imprécise de l'information par rapport à κ .

Ce que nous proposons, à travers l'équation (2.34), c'est réellement un modèle descriptif du comportement d'un décideur. Pour justifier ce modèle, il faudrait idéalement le valider de façon expérimentale. Cependant, les résultats obtenus récemment par Jaffray et Philippe (1997) militent dans une certaine mesure en faveur de ce modèle. Ceux-ci montrent que si les conditions suivantes sont respectées:

1. le comportement d'un individu est conforme au modèle de Schmeidler (cf. section 2.1.11), c'est-à-dire qu'il maximise une fonction $V: S \rightarrow \mathbf{R}$ qui correspond à l'intégrale de Choquet d'une fonction d'utilité par rapport à une capacité de Choquet C ;
2. V est une fonction d'utilité au sens de von Neumann et Morgenstern (1953);
3. les probabilités inférieures \underline{P} sont 2-monotones (i.e. \underline{P} est une capacité convexe);
4. l'ensemble S des actions est assez riche;
5. $\underline{R}(a) \leq V(a) \leq \bar{R}(a)$ pour tout $a \in S$;

alors:

1. il existe $0 \leq \kappa \leq 1$ tel que $C = \underline{P} + \kappa \cdot [\bar{P} - \underline{P}]$;
2. $V(a) = \underline{R}(a) + \kappa \cdot [\bar{R}(a) - \underline{R}(a)]$.

Ceci signifie que le modèle du comportement d'un décideur suggéré par (2.34) correspond à un cas particulier du modèle de Schmeidler, que Jaffray (1989) nomme la «généralisation du critère de l'utilité espérée aux choix dans l'incertain régulier». Ceci ne signifie pas que toutes les règles de décisions appartenant à la famille (2.34) sont raisonnables. En particulier, certaines de ces règles peuvent ne pas respecter le principe de la dominance stochastique du premier ordre au sens de Sarin et Wakker (1992). Il serait intéressant de pouvoir limiter l'ensemble de règles de décisions considérées à celles qui respectent ce principe de rationalité assez fondamental, car cela permettrait de réduire l'imprécision sur la prévision de la valeur de l'information.

2.3 Analyse de sensibilité de la valeur imprécise de l'information

S'il faut effectuer une analyse de sensibilité sur la règle de décision utilisée, il faut aussi considérer la précision des jugements a priori modélisés par les contraintes fixées. Dans l'exemple que nous avons présenté précédemment, les contraintes fixées étaient précises: la prévision inférieure et la prévision supérieure coïncidaient. En général, ce ne sera pas le cas. L'information a priori ne justifiera toujours pas des jugements aussi précis.

Par exemple, considérons un individu qui cherche à établir la valeur d'un paramètre hydrogéologique, tel que la porosité totale ω d'un milieu. En consultant la littérature et en observant le terrain, il peut obtenir un intervalle pour cette valeur, par exemple 25 à 35% pour un sable fin. Cet intervalle peut être assimilé à une prévision imprécise $\underline{E}(\omega)=25\%$ et $\bar{E}(\omega)=35\%$, ce qui signifie que l'individu serait prêt à acheter un pari $X(\omega)=\omega$ pour un montant 0.25, et prêt à vendre ce pari pour un montant de 0.35. Notons qu'il serait aussi possible de modéliser cette information en affirmant qu'il est probable que la valeur de ω soit entre 25 et 35%, ce qu'on pourrait modéliser par exemple par $\underline{P}(0.25 \leq \omega \leq 0.35)=0.5$, ce qui constitue aussi une prévision imprécise.

Une critique courante, par les théoriciens bayésiens, des approches fondées sur les probabilités ou les prévisions imprécises est que l'on doit fixer deux nombres précis plutôt qu'un seul pour représenter l'incertitude sur un événement ou pour fixer les prix d'achat et de vente d'un pari. Nous croyons que ce surcroît de complexité est utile pour faciliter la représentation d'informations a priori subjectives. Cependant, lorsque l'information a priori ne

suggère pas de valeurs précises pour les bornes inférieures et supérieures d'une prévision, il importe de vérifier la sensibilité de l'analyse aux valeurs choisies pour ces bornes.

Soit $X=\{X\}$ un ensemble de paris pour lesquels on accepte d'établir des prévisions imprécises de la forme $[\underline{E}(X), \overline{E}(X)]$. Soit aussi $[\underline{E}(Z), \overline{E}(Z)]$ la prévision imprécise d'un pari $Z(\theta)$ obtenue par application du principe d'extension des prévisions inférieures. Une façon d'étudier la sensibilité des bornes de l'intervalle $[\underline{E}(Z), \overline{E}(Z)]$ aux prévisions imprécises effectuées sur les paris appartenant à X consiste à remplacer chaque prévision imprécise sur $X \in X$ par une famille d'intervalles, par exemple en construisant une fonction $\tilde{E}_h(X)$ associant à chaque valeur d'un paramètre h un intervalle $[\underline{E}_h(X), \overline{E}_h(X)]$. Pour h fixé, on pourra alors calculer, à l'aide de la théorie des prévisions inférieures, une prévision inférieure cohérente $\underline{E}_h(Z)$ et une prévision supérieure cohérente $\overline{E}_h(Z)$ pour $Z(\theta)$. On obtiendra donc une fonction $\tilde{E}_h(Z)$ permettant d'étudier la sensibilité de la prévision imprécise de $Z(\theta)$ à l'imprécision des prévisions effectuées sur les paris de l'ensemble X . Par convention, nous limiterons h à l'intervalle $[0,1]$, et choisirons $\tilde{E}_h(X)$ de façon que $h_1 \leq h_2 \Leftrightarrow \tilde{E}_{h_1}(X) \supseteq \tilde{E}_{h_2}(X)$. Ceci permettra une représentation graphique intéressante.

Supposons par exemple que l'on veuille prévoir la probabilité d'obtenir quatre échecs consécutifs lors de la répétition d'expériences de Bernoulli indépendantes ayant toutes la même probabilité de succès θ . Pour θ connu, la probabilité d'observer r succès en n essais est donnée par la loi binomiale, de sorte que la probabilité de quatre échecs consécutifs est $f(r=0, n=4 | \theta) = (1-\theta)^4$. En fonction de l'information disponible a priori sur θ , on peut obtenir une prévision imprécise de la probabilité d'observer quatre échecs consécutifs en appliquant la théorie des prévisions inférieures cohérentes. Supposons encore que l'on accepte uniquement d'effectuer une prévision imprécise sur θ de la forme $[\underline{E}(\theta), \overline{E}(\theta)]$ autour de $1/4$, sans trop savoir quelle précision accorder à cette prévision. On pourrait utiliser la fonction $\tilde{E}_h(\theta)$ suivante pour effectuer une analyse de sensibilité:

$$\tilde{E}_h(\theta) = [\underline{E}_h(\theta), \overline{E}_h(\theta)] = \left[\frac{h}{4}, \frac{4-3h}{4} \right] \quad (2.35)$$

En faisant varier h entre 0 et 1, on peut ainsi passer d'une prévision non informative $\tilde{E}_0(\theta)=[0,1]$ à une prévision précise $\tilde{E}_1(\theta)=1/4$. Nous avons porté $\underline{E}_h(\theta)$ et $\bar{E}_h(\theta)$ en fonction de h sur le même graphique [Figure 2.7]. Il n'est pas utile d'identifier les deux courbes, puisque l'on sait que $\underline{E}_h(\theta) \leq \bar{E}_h(\theta)$.

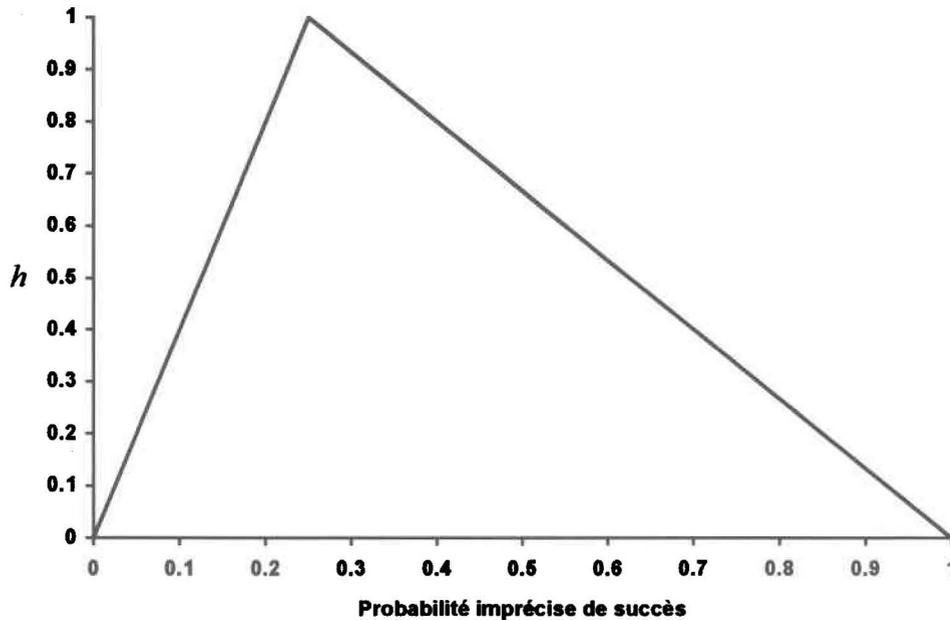


Figure 2.7: Famille de prévisions imprécises sur le paramètre d'une loi binomiale

On peut alors calculer, par application du principe d'extension des prévisions inférieures, la prévision imprécise du pari $Z(\theta)=(1-\theta)^4$, qui correspond à la vraisemblance de quatre échecs consécutifs. On obtient pour chaque valeur de h une probabilité imprécise différente, qui permet de construire une fonction $\tilde{P}_h(r=0, n=4) = \tilde{E}_h(Z)$, représentée graphiquement par la Figure 2.8.

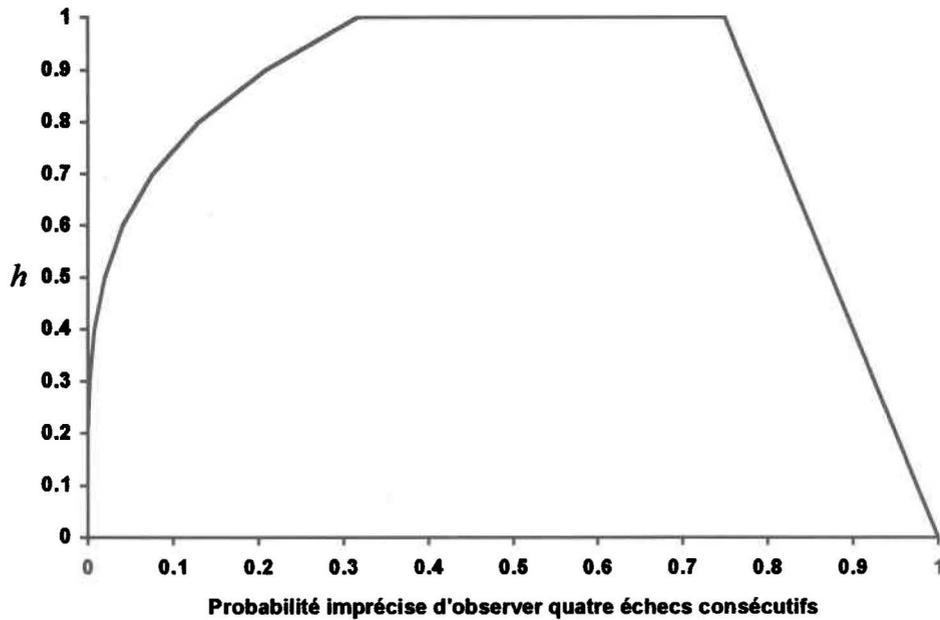


Figure 2.8: Résultat d'une analyse de sensibilité de la probabilité d'observer quatre échecs consécutifs à l'imprécision d'une prévision sur la probabilité de succès

On remarque d'abord que quelque soit le niveau de précision de la prévision sur θ , l'information a priori est insuffisante pour préciser exactement la probabilité d'observer quatre échecs consécutifs, puisque $\tilde{P}_1(r=0, n=4)=[0.32, 0.75]$. On voit aussi que la borne inférieure de $\tilde{P}_h(r=0, n=4)$ est plus sensible à une augmentation de l'imprécision que la borne supérieure. En fait, la borne inférieure de $\underline{P}_h(r=0, n=4)$ reste presque nulle tant que $h < 1/2$. Au contraire, la borne supérieure de $\bar{P}_h(r=0, n=4)$ est linéaire en h , comme $\bar{E}_h(\theta)$.

Au lieu d'analyser la sensibilité de $Z(\theta)$ à l'imprécision de la prévision sur θ , on pourrait aussi étudier sa sensibilité au domaine de définition de θ . Lorsque θ correspond au paramètre d'une loi $f(x|\theta)$, limiter le domaine de θ revient à faire une prévision imprécise sur x . Par exemple si θ est un paramètre de localisation, i.e. $\theta = E^{f(x|\theta)}(x)$, alors $[\underline{E}(x), \bar{E}(x)] = [\inf \theta, \sup \theta]$. Imaginons que la Figure 2.7 représente plutôt une famille de domaines de définition pour θ , modélisée par la fonction suivante:

$$\tilde{\Theta}_h = \left[\frac{h}{4}, \frac{4-3h}{4} \right] \quad (2.36)$$

Sachant uniquement que $\theta \in \tilde{\Theta}_h$, on peut appliquer le principe d'extension des prévisions inférieures pour obtenir une prévision imprécise pour le pari $Z(\theta) = (1-\theta)^4$:

$$\tilde{E}_h(Z) = \left[\inf_{\theta \in \tilde{\Theta}_h} (1-\theta)^4, \sup_{\theta \in \tilde{\Theta}_h} (1-\theta)^4 \right] = \left[\frac{81 \cdot h^4}{256}, \frac{(4-h)^4}{256} \right] \quad (2.37)$$

Le résultat de cette analyse de sensibilité, représenté par la Figure 2.9, est différent de celui obtenu en fixant seulement une prévision imprécise pour θ . La probabilité inférieure est la même, mais la probabilité supérieure est plus faible, de sorte que la prévision de Z est plus précise.

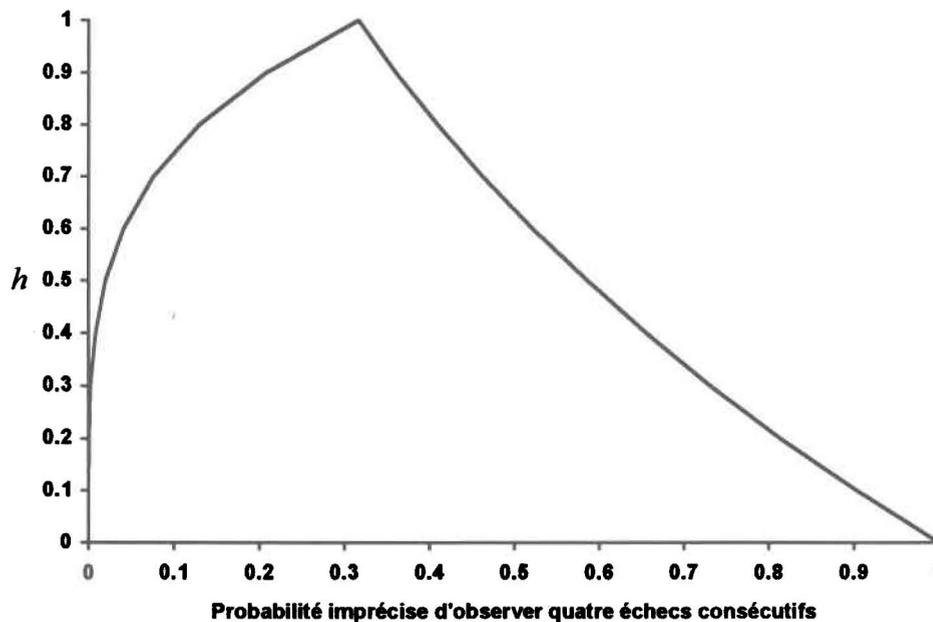


Figure 2.9: Résultat d'une analyse de sensibilité de la probabilité d'observer quatre échecs consécutifs au domaine de définition de la probabilité de succès

2.3.1 Lien avec la théorie des nombres flous

La théorie des prévisions inférieures, couplée à une analyse de sensibilité, peut être interprétée comme une généralisation de la théorie des nombres flous. Rappelons qu'un nombre flou \tilde{C} est une famille décroissante d'intervalles $\tilde{C} : [0,1] \rightarrow [a,b]$ qui associe à chaque nombre h de l'intervalle $[0,1]$ un intervalle fermé $\tilde{C}_h = [a_h, b_h]$ tel que $\tilde{C}_{h_1} \subseteq \tilde{C}_{h_2}$ si $h_1 \geq h_2$ (Kaufmann et

Gupta, 1991). \tilde{C}_h est souvent nommé la coupe de niveau h du nombre flou \tilde{C} . Dans notre démarche d'analyse de sensibilité, les fonctions $\tilde{E}_h(X)$, $X \in X$ peuvent être assimilées à des coupes de niveau h de nombres flous $\tilde{E}(X)$. Par contre, $\tilde{E}_h(Z)$ ne correspond pas nécessairement aux coupes de niveau h d'un nombre flou, puisqu'il n'y a aucune raison pour que la famille d'intervalles $[\underline{E}_h(Z), \overline{E}_h(Z)]$ soit décroissante en h . On peut considérer qu'il s'agit d'un nombre flou généralisé.

Les nombres flous sont habituellement combinés à l'aide du principe d'extension des nombres flous, qui revient à appliquer les règles du calcul d'erreur (Moore, 1979) à chaque coupe de niveau h . Ceci signifie que si \tilde{C} est un nombre flou, alors $f(\tilde{C})$ est le nombre flou correspondant à la famille d'intervalles $f(\tilde{C}_h)$ définie par l'équation suivante:

$$f(\tilde{C}_h) = \left[\inf_{x \in \tilde{C}_h} f(x), \sup_{x \in \tilde{C}_h} f(x) \right] \quad (2.38)$$

On remarque que le principe d'extension des nombres flous correspond à un cas particulier du principe d'extension des prévisions inférieures. En effet, si on modélise le domaine de définition d'un paramètre θ par un nombre flou $\tilde{\Theta}_h$, alors le nombre flou $\tilde{E}(Z)$ correspondant à la prévision d'un pari $Z(\theta)$ est obtenu de façon analogue par application du principe d'extension des prévisions inférieures:

$$\tilde{E}_h(Z) = \left[\inf_{\theta \in \tilde{\Theta}_h} Z(\theta), \sup_{\theta \in \tilde{\Theta}_h} Z(\theta) \right] \quad (2.39)$$

Le principe d'extension des prévisions inférieures est cependant plus puissant, puisqu'il permet de modéliser des informations a priori autres que le domaine de définition des paramètres.

Dans la suite de ce chapitre, nous utiliserons les concepts de prévisions floues et de probabilités floues pour faire référence à des fonctions $\tilde{E}_h(X)$ et $\tilde{P}_h(A)$ permettant de formaliser un type d'analyse de sensibilité sur les prévisions et les probabilités imprécises. Nous laissons au lecteur le soin de se souvenir de la définition que nous donnons de ces concepts et de l'interprétation que nous proposons. Certains trouveront que nous ne conservons du flou que le nom, que la nécessité de généraliser le concept de nombre flou, de même que l'emploi

du principe d'extension des prévisions inférieures au lieu de l'arithmétique floue nous éloignent trop de la théorie des nombres flous. En fait, nous croyons que le concept clé de la théorie des nombres flous est justement cette idée qu'une collection d'intervalles emboîtés permet de représenter une forme d'incertitude différente d'une distribution de probabilité. Nous présentons ici une utilisation concrète de nombres flous et proposons une interprétation précise. Cette interprétation nous amène à choisir une règle particulière pour transformer ces nombres flous, qui généralise l'arithmétique floue de Kaufmann et Gupta (1991). Ceci ne nous semble pas gênant, puisque les utilisateurs de nombres flous n'hésitent pas à proposer une variété de règles de combinaisons des nombres flous en fonction de leurs besoins (voir par exemple Bardossy et al. 1993, ou Bardossy et Duckstein, 1995).

En pratique, on peut penser utiliser de façon systématique des nombres flous triangulaires ou trapézoïdaux pour modéliser l'incertitude sur les prévisions a priori, et étudier la sensibilité de la valeur imprécise de l'information en vérifiant si les branches croissantes et décroissantes du nombre flou correspondant sont convexes ou concaves. Insistons encore sur le fait que l'incertitude est modélisée par des prévisions imprécises sur des paris, et que celles-ci sont représentées par des intervalles. Les nombres flous ne sont utilisés que pour effectuer une analyse de sensibilité sur ces intervalles, et ne donnent pas en soi d'information supplémentaire. Ce n'est que la façon dont l'imprécision de la valeur de l'information varie avec l'imprécision des prévisions a priori qui est intéressante. De plus, comme pour toute analyse de sensibilité, on ne peut considérer toutes les valeurs possibles des paramètres d'entrée. Il faut se limiter à un ensemble de prévisions imprécises a priori que l'on juge assez vaste pour couvrir celles qui seraient finalement choisies après une réflexion plus approfondie.

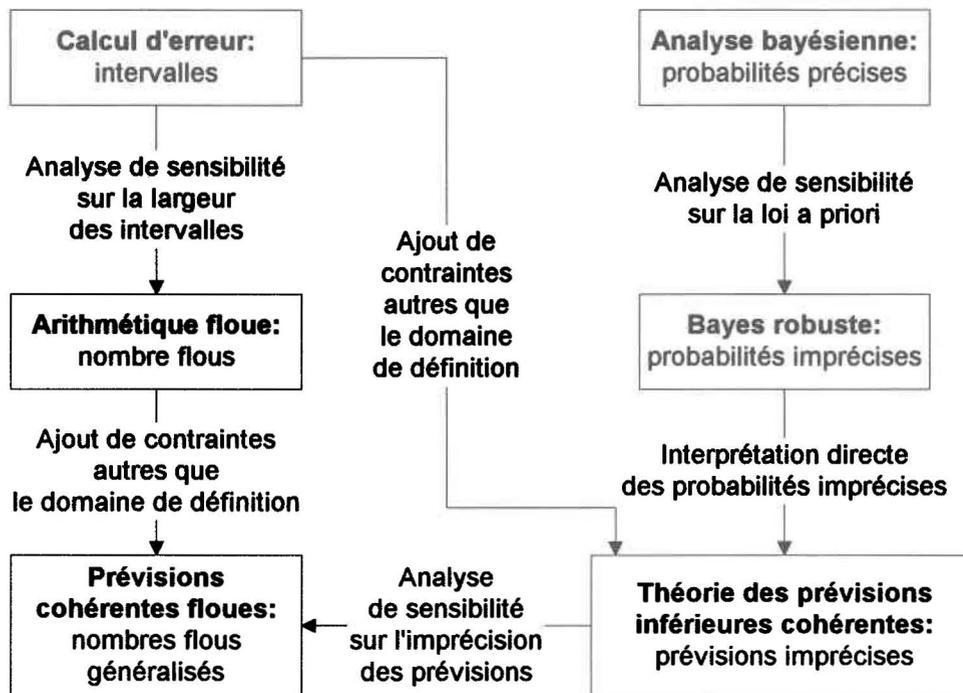


Figure 2.10: Liens entre les modèles d'incertitude

Les liens que nous avons établi entre l'analyse bayésienne, l'approche bayésienne robuste, le calcul d'erreur, la théorie des nombres flous et la théorie des prévisions inférieures cohérentes peuvent être résumés par la Figure 2.10. Une étude de sensibilité effectuée sur les probabilités a priori utilisées pour une analyse bayésienne conduit à une approche bayésienne robuste. Le passage à la théorie des prévisions inférieures cohérentes demande d'abandonner l'idée qu'il existe une mesure de probabilité, précise mais inconnue, permettant de modéliser l'information a priori. Ceci mène à considérer comme fondamental le concept de prévision imprécise. Une analyse de sensibilité sur l'imprécision des prévisions peut être effectuée en construisant des prévisions cohérentes floues. On peut aussi être amené à utiliser des prévisions cohérentes floues en généralisant un calcul d'erreur fondé sur l'arithmétique des intervalles, utile lorsque la seule information a priori correspond à une connaissance du domaine de définition des paramètres. D'abord, une analyse de sensibilité sur ce domaine de définition peut être effectuée en construisant des nombres flous et en appliquant le principe d'extension des nombres flous. Si l'on possède d'autres informations a priori, il faut alors se tourner vers la théorie des prévisions inférieures cohérentes.

2.4 Exemple de prévision imprécise de la valeur de l'information

L'exemple introduit au chapitre précédent peut servir à illustrer la définition proposée pour la valeur de l'information dans le contexte de la théorie des prévisions inférieures, et la technique d'analyse de sensibilité à l'aide de nombres flous. On peut rappeler cet exemple brièvement:

Imaginons qu'une localité située en bordure d'un cours d'eau soit protégée contre les crues printanières par une digue permettant de contenir un débit de $40\text{m}^3/\text{s}$. On envisage de relever cette digue de façon à pouvoir contenir un débit de $50\text{m}^3/\text{s}$. On évalue que le coût annuel de construction et d'entretien de la digue correspond au sixième des dommages engendrés par une inondation. Faut-il remonter immédiatement le niveau de la digue, ou serait-il plus sage d'attendre une année avant de prendre cette décision?

On doit choisir entre une action α_1 consistant à relever le niveau de la digue pour $1/6$ du coût annuel d'une inondation, et une action α_0 consistant à accepter le risque supplémentaire d'inondation. Si on néglige l'aversion au risque et si on accepte les hypothèses d'indépendance et de distribution identique des débits maximums annuels pendant la durée du projet, l'utilité de cette action α_0 correspond à la probabilité θ d'observer une crue printanière entre 40 et $50\text{m}^3/\text{s}$, et l'on peut résumer toute l'information d'échantillonnage par une séquence (x_n) de variables de Bernoulli avec $x_n=1$ si le débit maximum annuel appartient à l'intervalle $(40,50]$, et $x_n=0$ autrement. La décision optimale dépend de l'information disponible sur θ , qui constitue alors le paramètre d'une loi binomiale $f(r,n|\theta)$ régissant la probabilité d'observer r succès, i.e. r débits maximums annuels entre 40 et $50\text{m}^3/\text{s}$, sur une période de n années.

Imaginons que l'hydrologue chargé du projet ne possède pas d'information précise au site sur la fréquence des crues entre 40 et $50\text{m}^3/\text{s}$ pour cette rivière, mais soupçonne que la probabilité d'observer une telle crue soit assez élevée, de l'ordre de 25%. Nous supposons que l'hydrologue modélise son information a priori de la façon suivante:

- il accepte de fixer une prévision imprécise $[\underline{E}(\theta), \bar{E}(\theta)]$ pour θ qui contient la valeur 0.25;
- il diviserait par deux les bornes de cet intervalle s'il observait un débit maximum annuel à l'extérieur de l'intervalle $(40,50]$: $\underline{E}(\theta|r=0,n=1)=\underline{E}(\theta)/2$, $\bar{E}(\theta|r=0,n=1)=\bar{E}(\theta)/2$.

La première contrainte indique vers quelle valeur de θ pointe son information a priori, alors que la seconde spécifie le poids que doit avoir l'information a priori dans l'analyse. Dans

le cas présent, ce poids est assez faible, puisque l'hydrologue modifierait substantiellement son idée a priori après une seule observation supplémentaire du débit maximum annuel.

Ne connaissant pas exactement les valeurs de $\underline{E}(\theta)$ et $\bar{E}(\theta)$, nous étudierons la sensibilité de l'analyse aux bornes de cette prévision imprécise en considérant la prévision floue $\tilde{E}(\theta)$ représentée par la Figure 2.7.

Pour prévoir l'utilité d'une observation du débit maximum annuel, il faut d'abord étudier le risque a priori, par la suite le risque a posteriori, et enfin effectuer une prévision de leur différence.

2.4.1 Analyse du risque a priori

Il s'agit de choisir entre les actions a_0 et a_1 , en intégrant l'incertitude sur θ . Les fonctions de coût de ces deux actions sont respectivement $g(a_0, \theta) = \theta$ et $g(a_1, \theta) = 1/6$, exprimées en proportion des dommages liés à une inondation annuelle du territoire. Le risque $\tilde{R}(a_0)$ de l'action a_0 est donc identique à la prévision floue $\tilde{E}(\theta)$, alors que le risque $R(a_1)$ de l'action a_1 est précisément $1/6$. Dans la mesure où $h > 2/3$, on peut déterminer l'action de risque minimal, puisqu'on a alors $\underline{R}_h(a_0) = h/4 > 1/6 = R(a_1)$. L'action a_1 est toujours préférable dans ce cas.

Lorsque l'imprécision est plus importante ($h \leq 2/3$), les risques des deux actions ne sont plus aussi aisément comparables, et il n'est pas possible de sélectionner une action optimale a priori sur la base de l'information disponible. Il faut utiliser une autre règle de décision. Nous supposons que l'hydrologue accepte l'une des règles de décision paramétrées par l'équation (2.34). Dans ce cas, l'action $\delta_0^{(\kappa)}$ sélectionnée a priori dépend de la valeur de κ : l'action a_0 est sélectionnée pour de faibles valeurs de κ , alors que l'action a_1 est choisie pour de plus grandes valeurs de κ . On remarque qu'avec une règle de décision minimax ($\kappa=1$), l'action a_1 est toujours choisie, puisque $\bar{R}_h(a_0) > 1/6$ quel que soit h . La Figure 2.11 montre à partir de quelle valeur de κ l'action a_1 devient préférable, en fonction du niveau h de la coupe. On observe que dès que κ dépasse $1/6$, l'action sélectionnée est a_1 , quel que soit h .

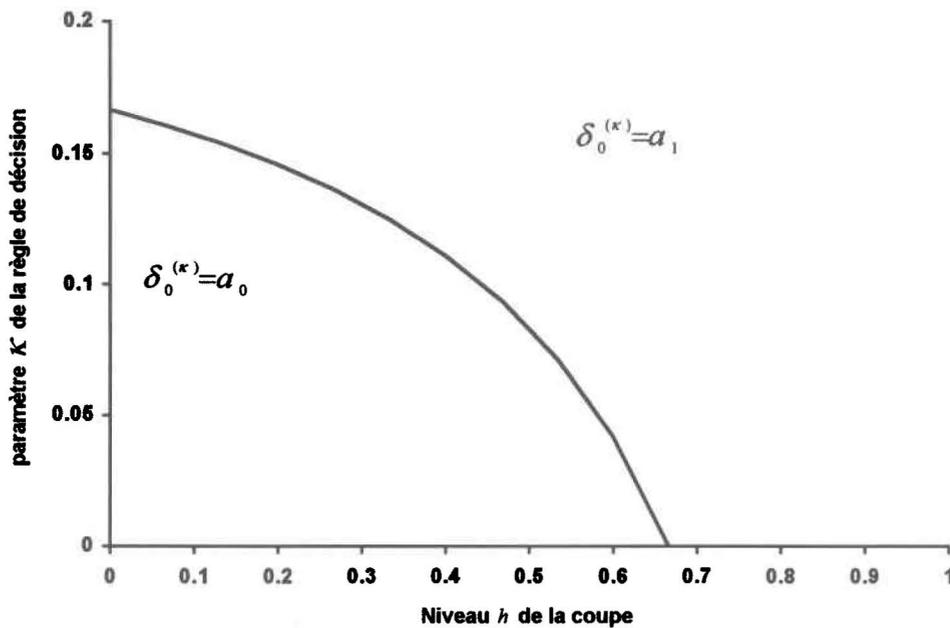


Figure 2.11: Valeur du paramètre κ de la règle de décision à partir de laquelle l'action a_1 est sélectionnée a priori, en fonction du niveau h de la coupe du risque flou de l'action a_0

2.4.2 Analyse du risque a posteriori

Pour déterminer quelle action sera sélectionnée après une observation du débit maximum annuel ($n=1$), il faut déterminer le risque a posteriori de chaque action pour chacune des deux observations possibles: $r=0$ ou $r=1$. Le risque a posteriori de l'action a_0 correspond à la prévision floue a posteriori de θ , donc $\tilde{R}(a_0|r, n) = \tilde{E}(\theta|r, n)$, alors que le risque a posteriori de l'action a_1 est toujours précisément $1/6$, puisqu'il ne dépend pas de θ . Pour $r=0$ et $n=1$, il est facile de déterminer le risque a posteriori de l'action a_0 , puisqu'il est fixé par hypothèse: $\tilde{R}(a_0|r=0, n=1) = \tilde{E}(\theta|r=0, n=1) = \tilde{E}(\theta)/2$.

La situation n'est pas aussi simple pour $r=1$; on doit appliquer la règle de Bayes généralisée [eq. (2.11)]. Pour chaque niveau h , $\underline{R}_h(a_0|r=1, n=1)$ est donné par la valeur μ telle que $\underline{E}[\theta(\theta-\mu)]=0$, puisque la vraisemblance est dans ce cas simplement θ . Pour résoudre cette équation, on doit calculer des prévisions inférieures cohérentes pour des paris de la forme $X(\theta)=\theta^2 - \theta \cdot \mu$. Ceci peut se faire en appliquant le principe d'extension. Pour un niveau h fixé, l'information a priori peut être représentée par quatre prévisions inférieures (on peut en

effet transformer les prévisions supérieures en prévisions inférieures par application du premier axiome):

1. $\underline{E}(\theta) = h/4$;
2. $\underline{E}(-\theta) = -\bar{E}(\theta) = 3h/4 - 1$;
3. $\underline{E}(\theta|r=0, n=1) = \underline{E}(\theta)/2 = h/8$;
4. $\underline{E}(-\theta|r=0, n=1) = -\bar{E}(\theta)/2 = 3h/8 - 1/2$.

On peut alors appliquer le principe d'extension des prévisions inférieures cohérentes [eq. (2.9)]:

$$\begin{aligned} \underline{E}[\theta(\theta - \mu)] = \sup_{\lambda_j > 0} \inf_{\theta} \{ & \theta(\theta - \mu) - \lambda_1(\theta - \underline{E}(\theta)) - \lambda_2(-\theta - \underline{E}(-\theta)) \\ & - \lambda_3 f(r=0, n=1|\theta)(\theta - \underline{E}(\theta|r=0, n=1)) \\ & - \lambda_4 f(r=0, n=1|\theta)(-\theta - \underline{E}(-\theta|r=0, n=1)) \} \end{aligned} \quad (2.40)$$

$$\begin{aligned} \underline{E}[\theta(\theta - \mu)] = \sup_{\lambda_j > 0} \inf_{\theta} \left\{ & \theta(\theta - \mu) - \lambda_1 \left(\theta - \frac{h}{4} \right) + \lambda_2 \left(\theta + \frac{3h}{4} - 1 \right) \right. \\ & \left. - \lambda_3 (1 - \theta) \left(\theta - \frac{h}{8} \right) + \lambda_4 (1 - \theta) \left(\theta + \frac{3h}{8} - \frac{1}{2} \right) \right\} \end{aligned} \quad (2.41)$$

On détermine $\bar{R}_h(a_0|r=1, n=1)$ de façon similaire, en appliquant le premier axiome [eq. (2.1)]. La Figure 2.12 présente les risques flous a posteriori ainsi obtenus. Notons que $\tilde{R}(a_0|r=1, n=1)$ a été obtenu en résolvant numériquement la règle de Bayes généralisée pour $h=0(0.1)1$. On remarque que l'imprécision sur $R(a_0|r=1, n=1)$ est beaucoup plus importante que l'imprécision sur $R(a_0|r=0, n=1)$. On pouvait s'attendre à un tel résultat, puisque $\tilde{R}(a_0|r=0, n=1)$ est fixé par l'hydrologue, alors que $\tilde{R}(a_0|r=1, n=1)$ est obtenu par application du principe d'extension, qui ne donne qu'une estimation conservatrice de la prévision imprécise que fixerait réellement un individu cohérent.

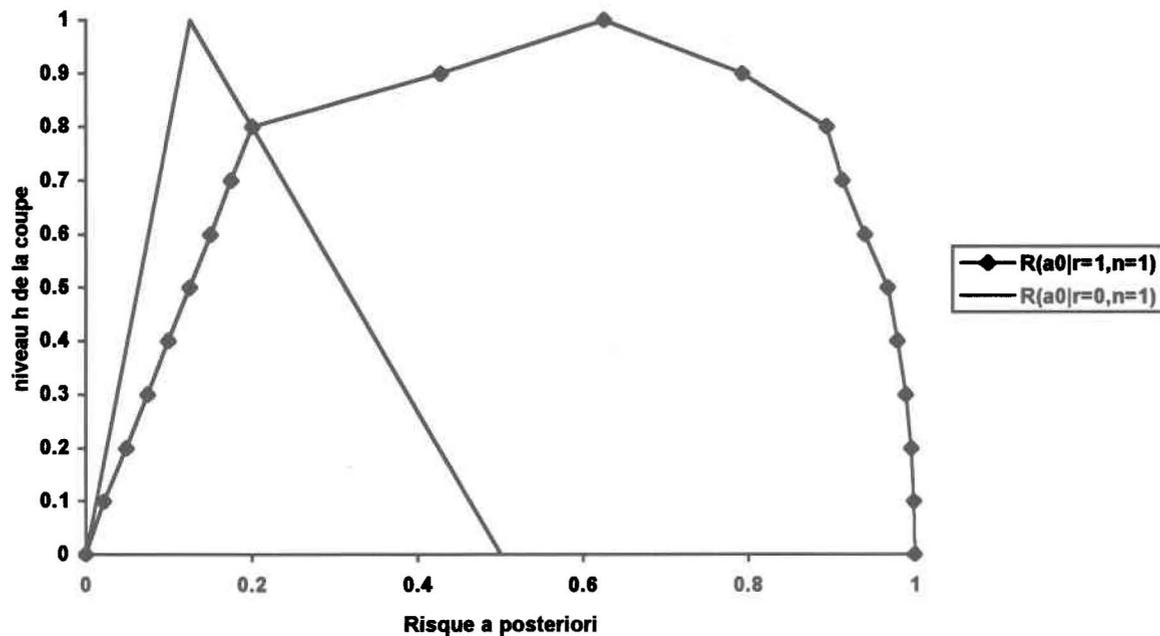


Figure 2.12: Risque flou a posteriori de l'action a_0

Comme dans le cas de l'analyse a priori, on ne peut déterminer l'action minimisant le risque que lorsque l'imprécision est relativement faible. Pour une règle de décision paramétrée selon l'équation (2.34), la Figure 2.13 montre à partir de quelle valeur de κ l'action a_1 est préférée à l'action a_0 . Dès que l'imprécision sur la prévision de θ est telle que $h \leq 8/9$, il n'est plus possible de déterminer une action optimale après observation de $r=0$ sur la base de l'information a priori. Il faut alors connaître la valeur de κ pour déterminer la décision $\delta^{(\kappa)}(r)$ à prendre. Nous avons inclus sur ce graphique la courbe obtenue a priori, et déjà décrite par la Figure 2.11.

On peut observer que la courbe calculée a priori se situe entre les deux courbes obtenues a posteriori. Ceci signifie que certaines combinaisons de décision a priori et de décision a posteriori sont incompatibles. Quelles que soient les valeurs de κ et h , on ne trouvera pas $\delta^{(\kappa)}(0)=a_1$ si $\delta_0^{(\kappa)}=a_0$, ni $\delta^{(\kappa)}(1)=a_0$ si $\delta_0^{(\kappa)}=a_1$. Cela semble logique: si l'on préfère a priori ne pas relever le niveau de la digue ($\delta_0^{(\kappa)}=a_0$), alors l'observation d'une crue printanière pour laquelle cette amélioration est inutile ($r=0$) devrait consolider la décision a priori ($\delta^{(\kappa)}(0)=a_0$). De la même façon, si l'hydrologue préfère a priori construire ($\delta_0^{(\kappa)}=a_1$),

alors l'observation d'une crue printanière qui aurait été contenue par une digue plus haute ($r=1$) ne devrait pas le faire changer d'idée ($\delta^{(\kappa)}(1)=a_1$).

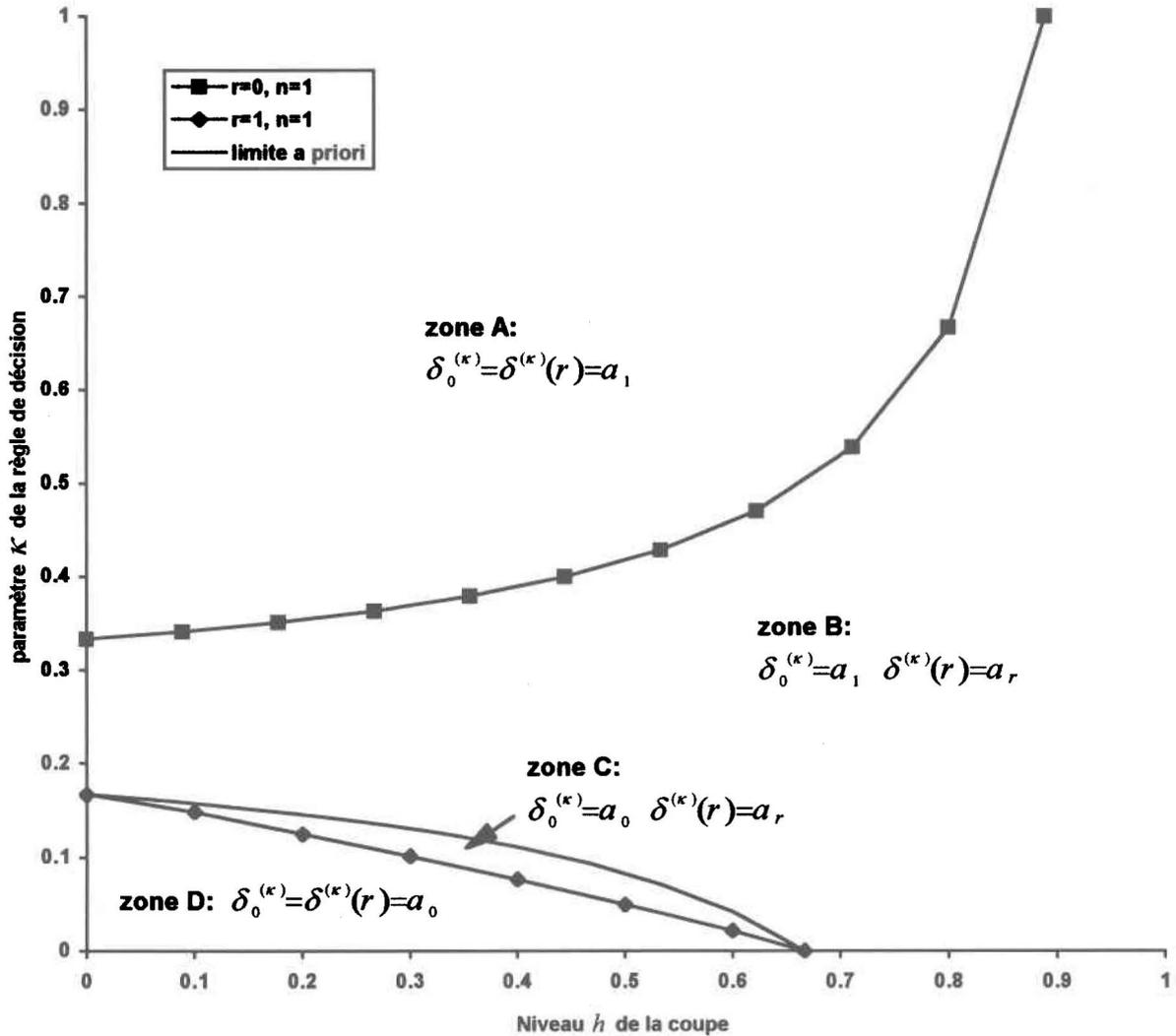


Figure 2.13: Valeur du paramètre κ de la règle de décision à partir de laquelle l'action a_1 est sélectionnée a posteriori, en fonction du niveau h de la coupe du risque flou de l'action a_0

On peut donc distinguer quatre cas possibles, correspondant à chacune des quatre zones identifiées par les lettres A à D sur la Figure 2.13. Les caractéristiques principales de ces quatre zones sont détaillées par le Tableau 2.1. Sachant à quelle zone appartient le couple (h, κ) , on peut ainsi connaître $\delta_0^{(\kappa)}$, $\delta^{(\kappa)}(r=0)$ et $\delta^{(\kappa)}(r=1)$.

Tableau 2.1: Décision a priori et a posteriori correspondant à chacune des quatre zones identifiées sur la Figure 2.13

Zone	$\delta_0^{(\kappa)}$	$\delta^{(\kappa)}(r=0)$	$\delta^{(\kappa)}(r=1)$	bornes pour h et κ
A	a_1	a_1	a_1	$h \leq 8/9, \kappa \geq 1/3$
B	a_1	a_0	a_1	$1/6 \leq \kappa \leq 1/3$
C	a_0	a_0	a_1	$h \leq 2/3, \kappa \leq 1/6$
D	a_0	a_0	a_0	$h \leq 2/3, \kappa \leq 1/6$

2.4.3 Valeur a posteriori de l'information

Les analyses de risque a priori et a posteriori permettent de sélectionner, pour κ fixé, une action a priori $\delta_0^{(\kappa)}$ et une action a posteriori $\delta^{(\kappa)}(r)$ pour $r=0$ et $r=1$. Ceci permet d'établir une prévision imprécise a posteriori de la valeur de l'information. En appliquant la définition proposée de la valeur imprécise de l'information [eq. (2.26)], on obtient pour la prévision inférieure:

$$\underline{VSI}(r=0, \delta^{(\kappa)}) = \begin{cases} 1/6 - \bar{E}[\theta | r=0, n=1] & \text{si } \delta_0^{(\kappa)} = a_1 \text{ et } \delta^{(\kappa)}(0) = a_0 \\ 0 & \text{si } \delta_0^{(\kappa)} = \delta^{(\kappa)}(0) \end{cases} \quad (2.42)$$

$$\underline{VSI}(r=1, \delta^{(\kappa)}) = \begin{cases} \underline{E}[\theta | r=1, n=1] - 1/6 & \text{si } \delta_0^{(\kappa)} = a_0 \text{ et } \delta^{(\kappa)}(1) = a_1 \\ 0 & \text{si } \delta_0^{(\kappa)} = \delta^{(\kappa)}(1) \end{cases} \quad (2.43)$$

On détermine de la même façon la prévision supérieure. Puisque l'on a déjà déterminé une prévision floue a posteriori pour θ à la section précédente, on calcule facilement la valeur floue de l'information pour r et κ fixés.

2.4.4 Prévision floue de la valeur de l'information

Selon la définition proposée [eq. (2.30)], $\underline{XVSI}(\delta^{(\kappa)})$ est donné par la prévision inférieure suivante:

$$\underline{XVSI}(\delta^{(\kappa)}) = \underline{E} \left\{ E^{f(r, n=1|\theta)} [\underline{VSI}(r, \delta^{(\kappa)})] \right\} = \underline{E} [(1-\theta) \cdot \underline{VSI}(r=0, \delta^{(\kappa)}) + \theta \underline{VSI}(r=1, \delta^{(\kappa)})] \quad (2.44)$$

Cette expression se simplifie, car si $\delta_0^{(\kappa)} = a_0$ alors $\underline{VSI}(r=0, \delta^{(\kappa)}) = 0$, et si $\delta_0^{(\kappa)} = a_1$ alors $\underline{VSI}(r=1, \delta^{(\kappa)}) = 0$. Conséquemment, on peut écrire:

$$\underline{XVSI}(\delta^{(\kappa)}) = \begin{cases} \underline{E}[\theta \cdot \underline{VSI}(r=1, \delta^{(\kappa)})] & \text{si } \delta_0^{(\kappa)} = a_0 \text{ et } \delta^{(\kappa)}(1) = a_1 \\ \underline{E}[(1-\theta) \cdot \underline{VSI}(r=0, \delta^{(\kappa)})] & \text{si } \delta_0^{(\kappa)} = a_1 \text{ et } \delta^{(\kappa)}(0) = a_0 \\ 0 & \text{sinon} \end{cases} \quad (45)$$

En remplaçant \underline{VSI} par sa définition, et après quelques manipulations algébriques, on obtient:

$$\underline{XVSI}(\delta^{(\kappa)}) = \begin{cases} \underline{E}(\theta) \cdot \left[\underline{E}(\theta|r=1, n=1) - \frac{1}{6} \right] & \text{si } \delta_0^{(\kappa)} = a_0 \text{ et } \delta^{(\kappa)}(1) = a_1 \\ [1 - \underline{E}(\theta)] \cdot \left[\frac{1}{6} - \underline{E}(\theta|r=0, n=1) \right] & \text{si } \delta_0^{(\kappa)} = a_1 \text{ et } \delta^{(\kappa)}(0) = a_0 \text{ et } h \leq \frac{8}{9} \\ [1 - \underline{E}(\theta)] \cdot \left[\frac{1}{6} - \underline{E}(\theta|r=0, n=1) \right] & \text{si } \delta_0^{(\kappa)} = a_1 \text{ et } \delta^{(\kappa)}(0) = a_0 \text{ et } h > \frac{8}{9} \\ 0 & \text{sinon} \end{cases} \quad (46)$$

De la même façon, on peut déterminer la prévision supérieure de la valeur de l'information:

$$\overline{XVSI}(\delta^{(\kappa)}) = \begin{cases} \overline{E}(\theta) \cdot \left[\overline{E}(\theta|r=1, n=1) - \frac{1}{6} \right] & \text{si } \delta_0^{(\kappa)} = a_0 \text{ et } \delta^{(\kappa)}(1) = a_1 \\ [1 - \overline{E}(\theta)] \cdot \left[\frac{1}{6} - \overline{E}(\theta|r=0, n=1) \right] & \text{si } \delta_0^{(\kappa)} = a_1 \text{ et } \delta^{(\kappa)}(0) = a_0 \\ 0 & \text{sinon} \end{cases} \quad (47)$$

A partir de ces équations, on peut obtenir, pour κ et h fixés, une prévision imprécise de la valeur de l'information. Par exemple, la Figure 2.14 présente la prévision floue de la valeur de l'information obtenue pour une règle de décision minimin ($\kappa=0$), ce qui permet d'observer comment évolue l'imprécision sur la valeur de l'information lorsque l'on fait varier la valeur de h . On observe qu'il ne s'agit pas d'un nombre flou au sens strict, puisque la largeur des intervalles ne décroît pas avec h . En fait, pour $h < 2/3$, la valeur de l'information est nulle, car dans ce cas l'imprécision sur la prévision de θ est telle qu'une information supplémentaire ne permet pas de modifier la décision a priori.

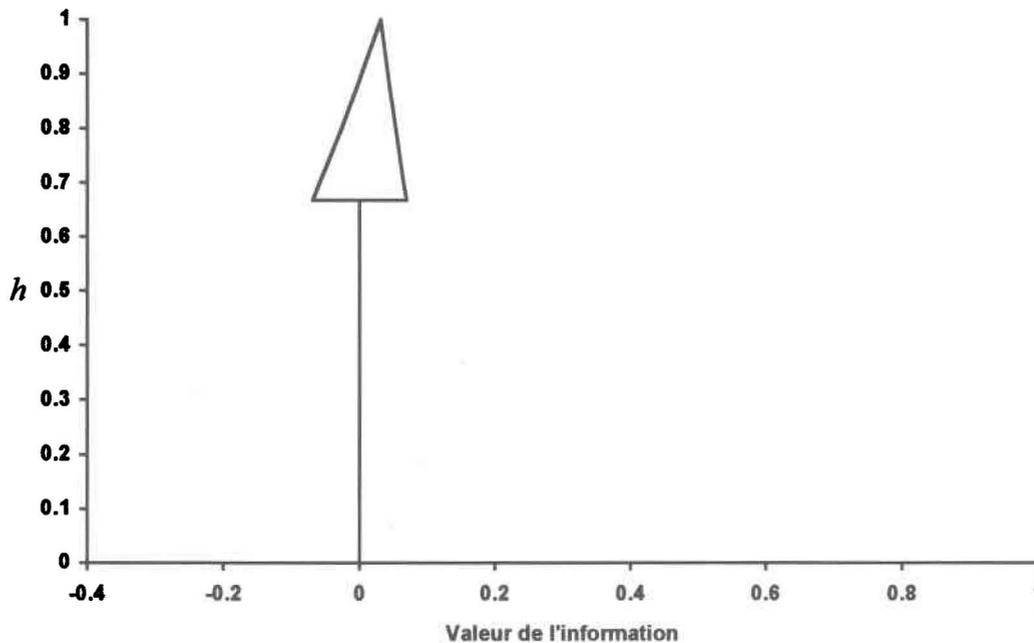


Figure 2.14: Prédiction floue de la valeur de l'information pour une règle de décision minimin ($\kappa=0$)

Pour $1/6 \leq \kappa \leq 1/3$, la Figure 2.13 montre que $\delta_0^{(\kappa)} = a_1$ et $\delta^{(\kappa)}(0) = a_0$ quel que soit h . Dans ce cas, comme le montre la Figure 2.15, la valeur de l'information n'est jamais précisément zéro, puisqu'une information supplémentaire peut toujours modifier la décision a priori. On voit aussi, comme pour la règle de décision minimin, que la prévision inférieure de la valeur de l'information peut être négative. Par contre, la prévision supérieure est toujours positive. Nous donnons à ce phénomène l'interprétation suivante.

L'estimation de la valeur de l'information que nous avons réalisé est basée sur un ensemble limité de contraintes qui ne permettent pas nécessairement d'identifier a priori ou a posteriori une action de risque minimal, et il faut introduire une règle de décision arbitraire. Il n'est pas nécessaire que l'information modélisée par ces contraintes représente la totalité de l'information disponible. Il se pourrait que d'autres informations soient disponibles, qui permettraient de préciser les prévisions, et éventuellement d'identifier une action optimale a priori et a posteriori. Négliger ces informations et utiliser une règle de décision arbitraire peut mener à une action sous-optimale, et donc à une valeur négative de l'information.

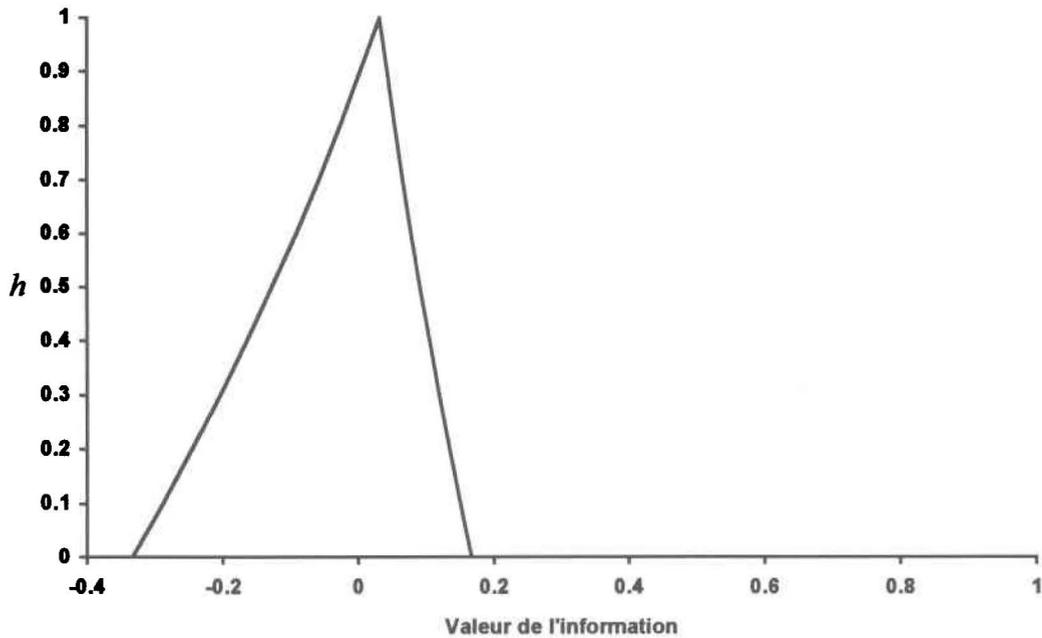


Figure 2.15: Prévision floue de la valeur de l'information lorsque $1/6 \leq \kappa \leq 1/3$

Il est intéressant d'observer la valeur de l'information lorsque $\delta_0^{(\kappa)} = a_0$ et $\delta^{(\kappa)}(1) = a_1$. La Figure 2.13 montre que cette combinaison ne se produit que pour une gamme assez restreinte de valeurs de κ et de h , et qu'elle est impossible lorsque $h > 2/3$. Par contre, la valeur de l'information peut être assez élevée dans ces situations, comme le montre la Figure 2.16. Si la prévision inférieure n'est jamais positive, la prévision supérieure atteint $5/6$.

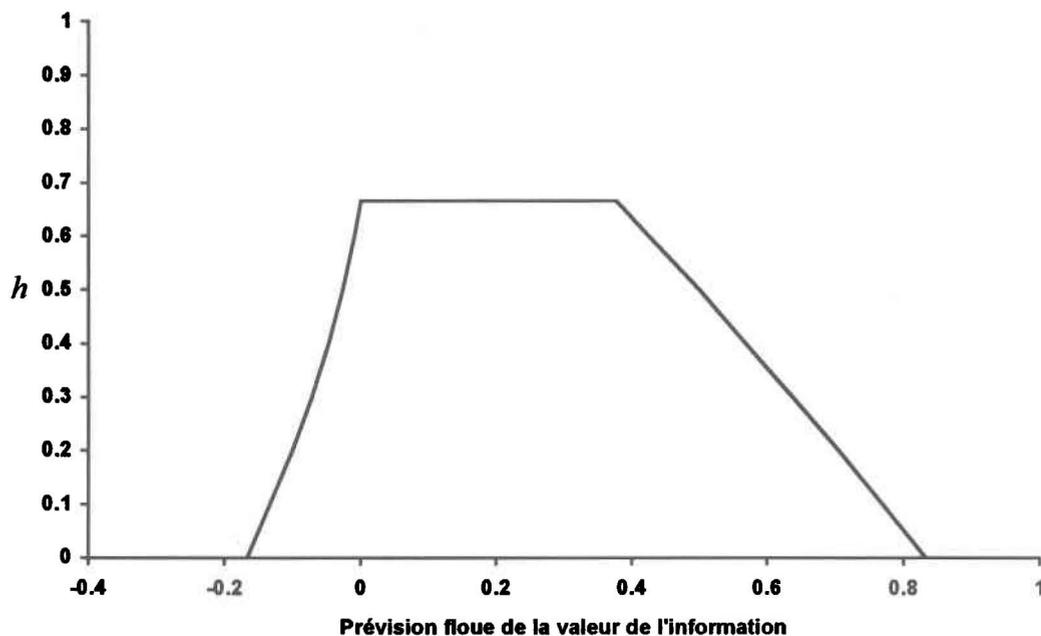


Figure 2.16: Prévion imprécise de la valeur de l'information en fonction de h , lorsque $\delta_0^{(\kappa)} = a_0$ et $\delta^{(\kappa)}(1) = a_1$

Il peut être difficile de déterminer quelle règle de décision choisir. On peut préférer effectuer une prévision conservatrice de la valeur de l'information, en déterminant le minimum de \underline{XVSI} et le maximum de \overline{XVSI} pour toutes les valeurs de κ entre 0 et 1. Pour déterminer si la valeur de l'information est suffisante pour justifier de retarder le projet, il faut la comparer au coût de l'échantillonnage XCS , qui correspond dans ce cas au risque supplémentaire de débordement relié à l'attente $R(a_0)$, duquel on peut retrancher le risque de l'action optimale a priori $R(\hat{a})$. Puisque $R(a_0)$ et $R(\hat{a})$ sont des prévisions floues, on obtient un coût d'échantillonnage flou. La Figure 2.17 montre les deux nombres flous ainsi obtenus pour la valeur de l'information et le coût de l'échantillonnage. Cette dernière prévision peut être interprétée comme le coût minimal d'acquisition d'une information supplémentaire. En effet, elle mesure le coût espéré des dommages qu'une protection supplémentaire aurait permis d'éviter, sans compter le coût relié à la mesure du débit maximum annuel.

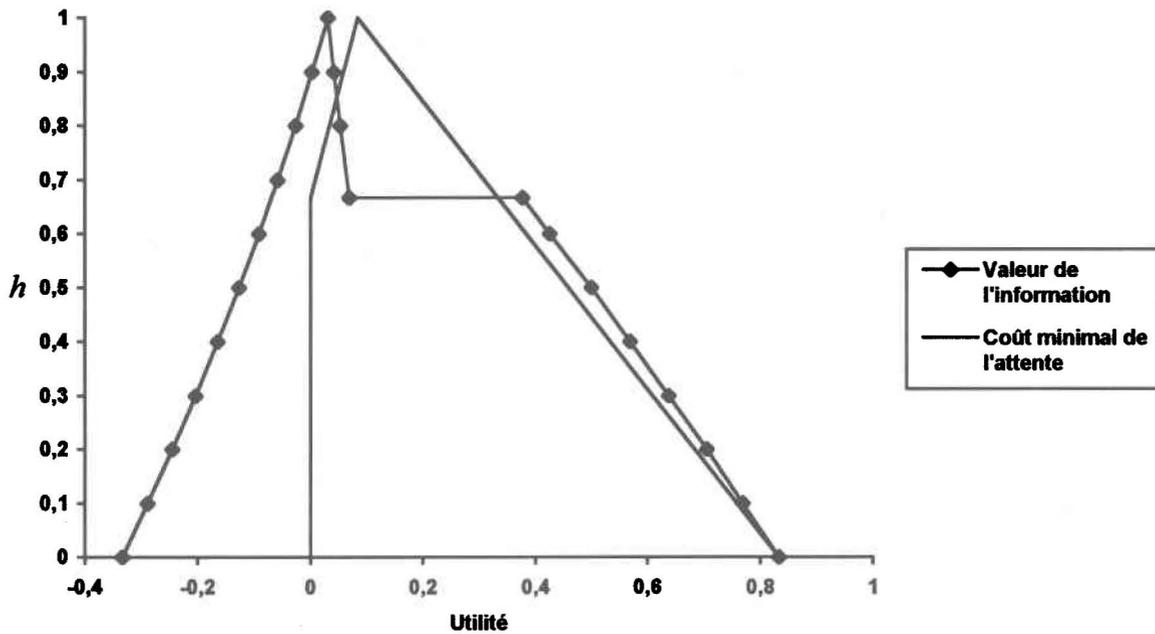
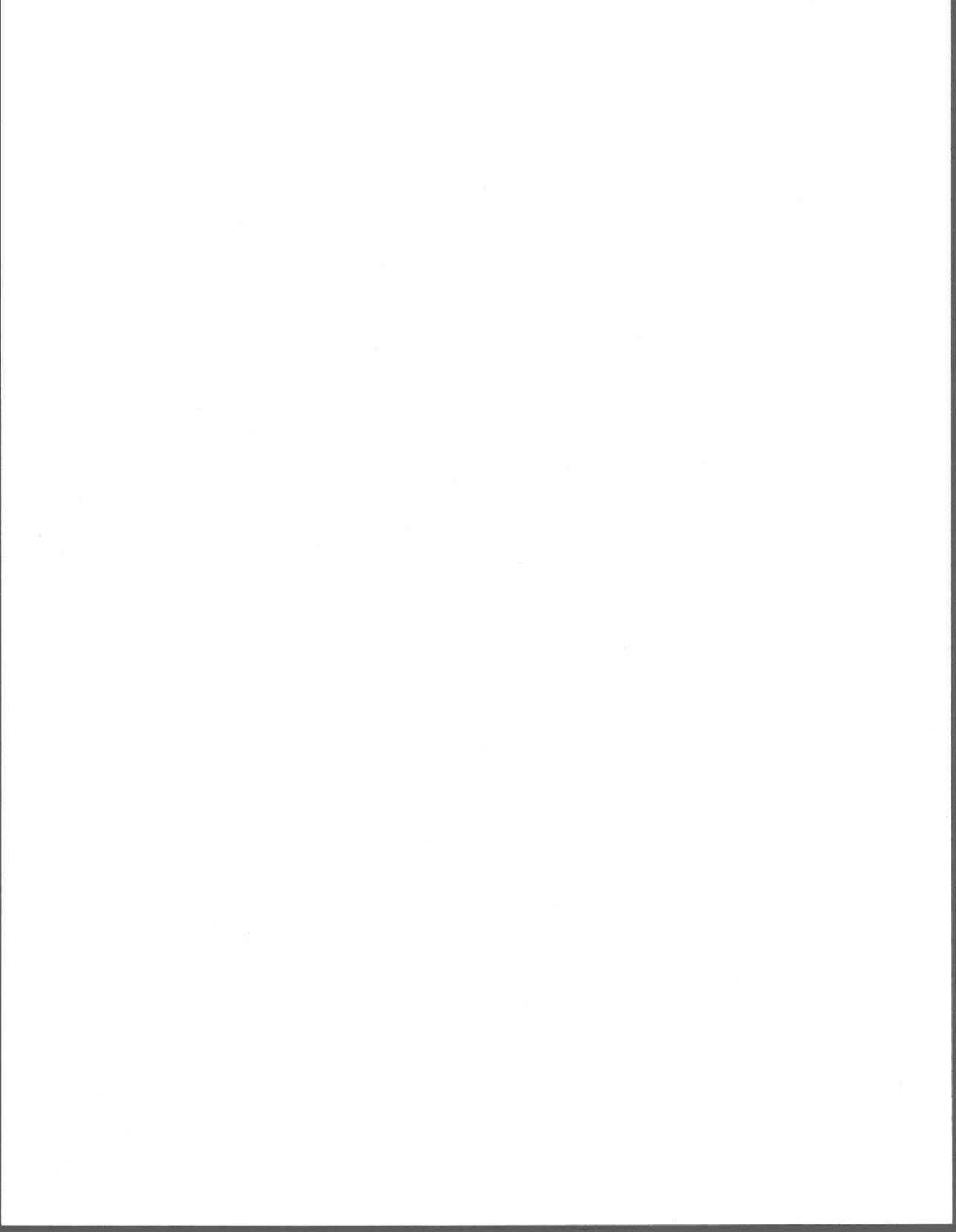


Figure 2.17: Comparaison de la valeur de l'information et du coût lié à son acquisition

On remarque que $\overline{XVSI}_h < \underline{XCS}_h(\theta)$ pour $h > 8/9$. Ainsi, si l'hydrologue accepte de fixer pour θ une prévision au moins aussi précise que $\tilde{E}_{8/9}(\theta) = [2/9, 3/9]$, alors il peut conclure que la valeur de l'information est inférieure à son coût d'acquisition, de sorte qu'il est préférable de prendre immédiatement la décision de relever le niveau de la digue. Autrement, l'information a priori est insuffisante pour permettre de déterminer s'il est préférable d'échantillonner encore une année ou d'agir immédiatement.

Maintenant que nous avons illustré la théorie des prévisions inférieures par un exemple simple de prévision de la valeur de l'information, nous pouvons présenter une application concrète pour laquelle le langage des prévisions inférieures permet une réelle prise en compte d'une information a priori subjective.



3. VALEUR D'UNE INFORMATION HYDROLOGIQUE AU SITE DE CONSTRUCTION D'UN OUVRAGE HYDROELECTRIQUE

Ce troisième chapitre sera consacré à une estimation imprécise de la valeur de l'information pour un problème sinon réel, du moins réaliste. Il s'agit d'estimer la valeur d'une information hydrologique locale sur un bassin versant où l'on envisage l'aménagement d'un complexe hydroélectrique. Pour montrer l'intérêt pratique des outils développés au chapitre précédent, nous nous inspirerons largement d'un problème réel auquel l'entreprise d'électricité Hydro-Québec a dû faire face à la fin des années 1980.

Pour suivre l'évolution de la demande québécoise en électricité et pour honorer des contrats d'exportation, cette société devait planifier et réaliser de nouveaux aménagements hydroélectriques (Hydro-Québec, 1991). Différents projets étaient envisagés, entre autres un aménagement hydroélectrique sur la rivière Sainte-Marguerite (SM3). Nous montrerons comment la théorie des prévisions inférieures peut aider à formaliser le processus d'ordonnancement des projets, et peut permettre de prévoir l'utilité économique de l'information hydrologique dans ce type de décision.

L'objectif de cette application est d'abord d'illustrer par un exemple plus concret la théorie développée au chapitre précédent. Les résultats que nous présentons ne sont pas directement utiles sur le plan décisionnel, puisque Hydro-Québec a déjà décidé de réaliser le projet SM3. Les travaux progressent, et la mise en service est prévue pour 2001 (Hydro-Québec, 1997). Cependant, la méthodologie est transférable à des problèmes similaires.

3.1 Problématique générale

Hydro-Québec (1991) prévoyait pour la période 1990-2003 une augmentation de la demande québécoise en électricité de l'ordre de 33 TWh par an, passant de 157 TWh à environ 190 TWh. L'entreprise disposait à l'époque d'un surplus de l'ordre de 20 TWh, et pouvait combler les besoins additionnels en continuant l'aménagement du complexe La Grande. Les centrales La Grande-1, Laforge-1 et Brisay étaient déjà en construction et pourraient produire à elles seules 14 TWh (Hydro-Québec, 1991), ce qui serait suffisant pour la société québécoise. Cependant, l'entreprise avait aussi comme objectif d'exporter de l'électricité, en particulier

vers le Vermont et l'état de New York. Des contrats étaient déjà signés pour 12.6 TWh, et on s'attendait à pouvoir vendre 12.4 TWh supplémentaires d'ici 2003, pour un total de 25 TWh. Simplement pour respecter les engagements déjà pris, il fallait pouvoir produire environ 12 TWh supplémentaires. Les deux dernières centrales prévues dans le cadre du complexe La Grande (Laforge-2 et Eastmain-1) ne pourraient produire que 4.5 TWh. Il fallait entreprendre de nouveaux chantiers, pour produire au moins 7 TWh supplémentaires.

Différents projets étaient envisagés, dont un troisième aménagement hydroélectrique sur la rivière Sainte-Marguerite (projet SM3), ainsi que les complexes hydroélectriques Grande-Baleine et Nottaway-Broadback-Rupert (projet NBR). Le Tableau 3.1 présente quelques caractéristiques de ces trois projets.

Tableau 3.1: Caractéristiques technico-économiques des projets étudiés pour réaliser les objectifs d'exportation d'électricité (adapté de Hydro-Québec, 1991)

Nom du projet	Puissance à la pointe (MW)	Production annuelle estimée (TWh)	Prix de revient estimé (¢/kWh)
Sainte-Marguerite-3 (SM3)	790	4	4.0
Grande-Baleine (GB)	3 160	16	4.4
Nottaway-Broadback-Rupert (NBR)	8 440	45	4.6

En étudiant la rentabilité de ces trois projets, l'entreprise décida de produire l'électricité destinée à l'exportation grâce à Grande-Baleine et SM3, et de reporter après 2003 le projet NBR. En mars 1994, suite à une diminution de la demande en électricité dans l'état de New York et aux protestations du peuple Cri face au projet Grande-Baleine, un contrat de 3 TWh avec la compagnie Consolidated Edison fut annulé (DOE, 1995). Hydro-Québec ne devait plus trouver que 4 TWh pour combler les besoins à venir. L'entreprise délaissa le projet Grande-Baleine et entreprit d'aménager un réservoir et une nouvelle centrale sur la rivière Sainte-Marguerite. Nous présentons dans cette section les grandes lignes de ce projet, les informations hydrologiques disponibles, et nous discutons de certaines des difficultés rencontrées par les hydrologues pour justifier le projet et dimensionner les ouvrages.

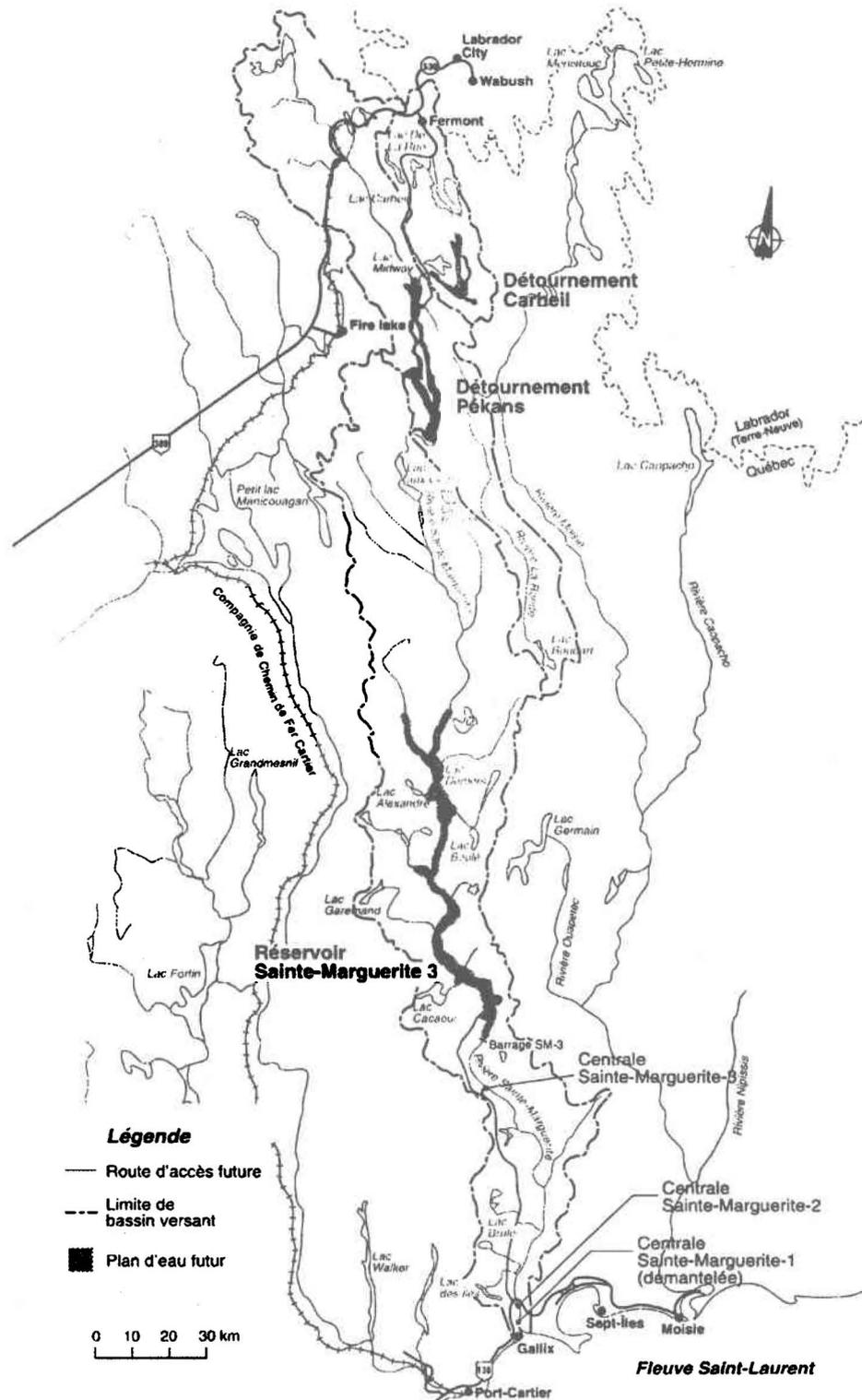


Figure 3.1: Plan général de l'aménagement, projet SM3 (tiré de Hydro-Québec, 1991)

3.1.1 Aménagement hydroélectrique SM3

Comme son nom l'indique, le projet SM3 constitue le troisième aménagement hydroélectrique sur la rivière Sainte-Marguerite, dont l'embouchure est située à mi-chemin entre les villes de Port-Cartier et de Sept-Iles sur la Côte-Nord, à environ 700 km au nord-est de Montréal. La Figure 3.1 montre le plan général de l'aménagement envisagé dans le rapport d'avant-projet (Hydro-Québec, 1991). On remarque l'emplacement des centrales Sainte-Marguerite-1 (aujourd'hui démantelée) et Sainte-Marguerite-2 (SM2) à 10 km de l'embouchure, ainsi que l'emplacement projeté de la centrale Sainte-Marguerite-3 (SM3), environ 90 km en amont de l'embouchure. Alors que les centrales précédentes étaient au fil de l'eau, le projet SM3 comprend la mise en eau d'un ensemble de trois réservoirs permettant de constituer une réserve utile de 4090 hm³ destinée à la production énergétique.

De façon à augmenter la production hydroélectrique à la centrale, il avait été décidé de détourner la rivière Carheil et la rivière aux Pékans, deux affluents du bassin versant de la rivière Moisie, située à l'est de la rivière Sainte-Marguerite. Ceci permettait d'obtenir un bassin versant de 7330 km² au site SM3, ajoutant ainsi 2530 km² au bassin naturel. Pour minimiser les impacts environnementaux sur le bassin de la rivière Moisie, en particulier pour préserver le saumon atlantique, un ouvrage régulateur permettant de restituer une partie suffisante des eaux des rivières Carheil et aux Pékans dans la rivière Moisie était prévu. Notons que suite à la controverse qu'a suscité le projet dans la population, Hydro-Québec n'a pas encore obtenu l'autorisation du Gouvernement du Québec pour effectuer ce détournement (Hydro-Québec, 1997).

3.1.2 Mesures du débit disponibles

Au moment où furent effectuées les études d'avant-projet (1986-1990), on ne disposait d'aucune mesure du débit au site SM3. Une station de mesure du débit journalier (à partir de la hauteur d'eau) avait été installée en 1987, mais les données ne furent disponibles qu'en 1990, entre autres à cause des délais reliés à la validation de la courbe niveau-débit. Par contre, on avait certaines informations au site SM2. Avant la mise en service de cette centrale (1954), il existait à cet endroit une station de mesure hebdomadaire du débit, mise en service en 1937. A partir de la production des turbines et de l'ouverture des vannes de l'évacuateur à SM2, il a

aussi été possible d'estimer le débit journalier après 1954. Finalement, le débit de la portion détournée du bassin de la rivière Moisie pouvait être estimé à partir des données journalières d'une station située sur la rivière aux Pékans, à quelques kilomètres de l'emplacement du détournement envisagé. Cette station fut en opération de 1965 à 1982.

3.1.3 Transfert de l'information régionale

Pour justifier le projet, étudier les variantes d'aménagement et effectuer les études technico-économiques requises, il fallait arriver à estimer le débit au site SM3 à partir des données disponibles. Si l'on pouvait prétendre connaître assez bien le débit détourné du bassin de la rivière Moisie, il en était autrement du débit naturel à SM3. On ne connaissait qu'une estimation du débit 80 km en aval (SM2). Comme cela est expliqué dans CHS (1995), les hydrologues d'Hydro-Québec ont procédé par rapport de superficie de bassin versant pour estimer le débit à SM3, tout en étant conscients de l'imprécision des données reconstituées à SM2 et des limites du procédé de transfert de l'information utilisé.

La confiance limitée qu'avaient les hydrologues dans l'estimation ainsi obtenue du débit à SM3 est reflétée par leur comportement ultérieur. Après avoir observé en 1990 un écart de 20% entre les débits journaliers reconstitués et les débits journaliers mesurés à SM3, on a pensé que ce problème pouvait être lié à l'imprécision de la courbe de rendement des turbines et de la courbe de tarage de l'évacuateur à SM2. Une série de quatorze jaugeages en différentes conditions d'ouverture des vannes d'évacuateur et de turbinage a confirmé ce biais, et les apports à SM2 ont été réévalués à la hausse (CHS, 1995).

Toutefois, toujours selon CHS (1995), cet exercice de correction s'est avéré erroné et a conduit à une surévaluation de la production et de la rentabilité. En 1994, on observait une surestimation de 10% des valeurs transférées. On suppose que cet écart provient de l'utilisation du rapport des superficies de bassin versant pour transférer l'information, en négligeant la variabilité spatiale des précipitations.

La Figure 3.2 présente la série de débits annuels obtenue à SM2 à partir de la production et de l'évacuation, ainsi que les débits estimés à SM3 par rapport de bassin versant. Pendant la période utilisée pour les études d'avant-projet (1958-82), le débit moyen à SM2 fut de $139 \text{ m}^3/\text{s}$, avec un écart-type d'environ $20 \text{ m}^3/\text{s}$. Étant donné que le bassin versant à SM2

fait 6130 km² contre 4800 km² à SM3, on estimait à 109 m³/s le débit moyen pour la même période à SM3, par rapport de superficies de bassin versant (avant la correction de 1992).

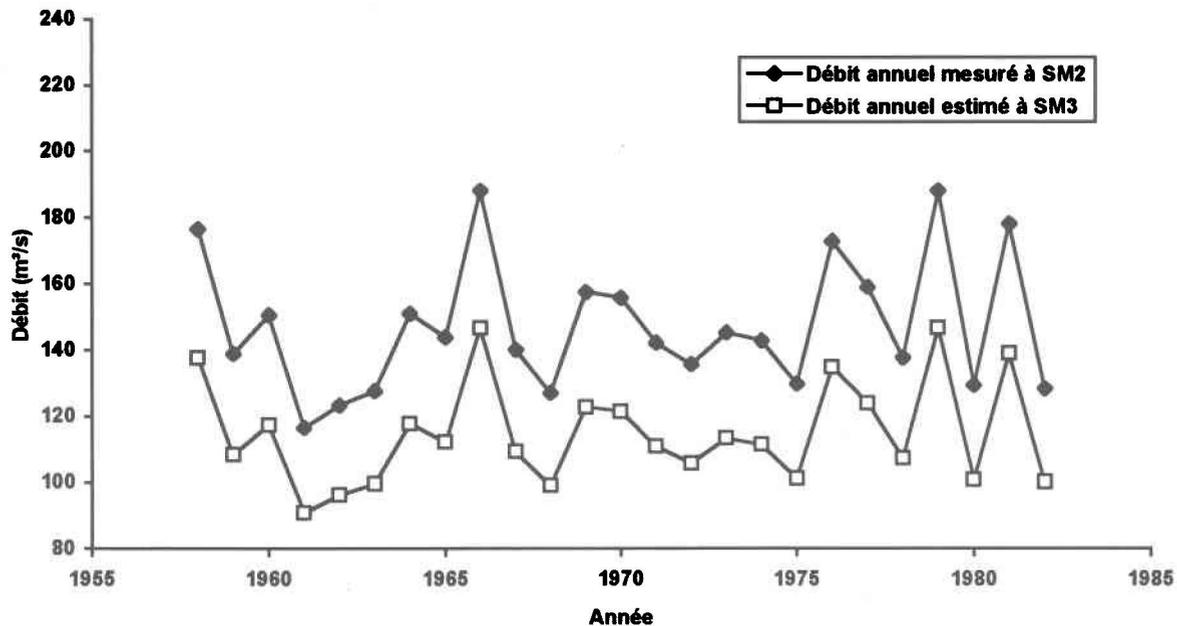


Figure 3.2: Débit annuel à SM2 et à SM3 pour la période 1958-1982 (adapté de Hydro-Québec, 1992)

3.1.4 Caractéristiques hydrologiques et technico-économiques du projet Sainte-Marguerite-3

Dans l'étude d'avant-projet de SM3, trois variantes d'aménagement ont été considérées: une première variante sans détournement des affluents de la rivière Moisie (SM3-A), une seconde variante avec détournement complet des rivières Carheil et aux Pékans (SM3-B), ainsi qu'une troisième variante comprenant un détournement partiel de ces deux rivières (SM3-C), de façon à minimiser l'impact sur l'environnement de la rivière Moisie (Hydro-Québec, 1991). En fonction de l'information disponible et des techniques de transfert d'information utilisées, le Tableau 3.2 présente, pour ces trois variantes, les estimations du module total à SM3, c'est-à-dire le débit moyen annuel naturel au site, auquel s'ajoutent les apports du bassin de la rivière Moisie. La production annuelle, le prix de revient et le coût de construction sont aussi indiqués.

Tableau 3.2: Caractéristiques des variantes étudiées du projet SM3 (adapté de Hydro-Québec, 1991)

Variante	Module (m ³ /s)	Production (TWh/an)	Prix de revient (¢/kWh 1991)	Coût de construction (10 ⁹ \$ 1990)
A: sans détournement	109	2.8	5.0	1.1
B: avec détournement	165	4.4	3.7	1.3
C: détournement et débit réservé	155	4.1	4.0	1.3

Pour faire un compromis entre la rentabilité et la protection de l'environnement, Hydro-Québec (1991) proposait le projet C. Cependant, rappelons qu'Hydro-Québec n'a toujours pas obtenu les autorisations gouvernementales nécessaires pour effectuer ce détournement (Hydro-Québec, 1997). En conséquence, il se pourrait que le prix de revient de l'électricité soit 25% plus élevé que prévu. Dans ces conditions, le projet SM3 serait moins rentable que les projets Grande-Baleine et NBR.

3.1.5 Sensibilité de la rentabilité et des coûts de construction au débit

Le débit à SM3 a bien sûr des répercussions sur le dimensionnement des ouvrages, mais aussi sur la rentabilité du projet. Une étude réalisée par Hydro-Québec suite à la réévaluation des débits à SM2 a conclu, pour la variante C, qu'une augmentation de 6% du module à SM3 (le débit moyen annuel passerait de 155 à 164 m³/s) n'aurait pas d'impact significatif sur le coût de construction mais améliorerait la rentabilité de 5%, en faisant passer le prix de revient de 4.0 ¢/kWh à 3.8 ¢/kWh (Hydro-Québec, 1992). Ceci est en parti dû à une réduction importante du temps de remplissage du réservoir, qui passe de 46 à 39 mois, soit une réduction de 15%. Une méconnaissance du débit peut donc mener à une mauvaise évaluation de la rentabilité du projet, sans avoir d'influence sur les coûts liés au dimensionnement des ouvrages.

3.2 Problématique étudiée

Le projet SM3 est complexe, et l'étude des conséquences des différentes incertitudes sur le dimensionnement des ouvrages et la rentabilité l'est aussi. Dans la suite de ce chapitre, nous nous concentrerons sur le problème d'estimation de la rentabilité et sur l'ordonnement des

projets, en relation avec l'incertitude sur le débit à SM3. En particulier, nous tenterons de répondre aux questions suivantes:

- En fonction de l'information disponible au moment de l'étude d'avant-projet, était-il préférable de lancer la construction de SM3 malgré l'incertitude sur la rentabilité, ou aurait-il été plus judicieux de choisir un autre projet?
- Quelle aurait été l'utilité, au moment de choisir le projet à réaliser, d'une observation du débit moyen annuel?
- Aurait-il été avantageux de retarder la décision d'une année pour obtenir cette information?

Nous utiliserons la théorie des prévisions inférieures cohérentes pour répondre à ces questions, ainsi que la définition de la valeur de l'information proposée au chapitre précédent. Il faudra entre autres établir une fonction de coût de façon à relier le débit à SM3 à la rentabilité, modéliser l'information a priori à l'aide de prévisions imprécises, choisir une distribution d'échantillonnage, appliquer la théorie, et quantifier le coût d'acquisition de l'information.

L'analyse sera basée sur les informations disponibles au moment de l'étude d'avant-projet, en négligeant les développements ultérieurs intervenus dans ce dossier, l'objectif étant de voir ce qu'auraient pu apporter à ce moment une analyse plus formelle de la décision, basée sur la théorie des prévisions inférieures cohérentes.

Notons que les résultats sont présentés avec un nombre limité de chiffres significatifs, de façon à respecter le niveau de précision des données. Bien sûr, les calculs ont été effectués avec beaucoup plus de précision. Un désavantage de cette présentation est que le lecteur qui voudra refaire les calculs obtiendra de légères erreurs d'arrondi.

3.3 Espace des actions

Dans le rapport d'avant-projet, Hydro-Québec (1991) discute de la nécessité d'augmenter sa capacité de production hydroélectrique, en reconnaissant cependant que SM3 n'est pas le seul projet qui puisse être envisagé pour répondre aux besoins futurs en électricité:

Selon le scénario de croissance moyenne, incluant l'objectif d'exportation, de nouveaux équipements seront nécessaires après l'an 2000 pour répondre aux besoins énergétiques. (p. 10)

La solution de remplacement à l'aménagement Sainte-Marguerite-3 consisterait à devancer un autre projet hydroélectrique dont la mise en service serait elle aussi possible pour 2001. En l'occurrence, il s'agirait du complexe Nottaway-Broadback-Rupert... (p. 19)

Par ailleurs, si on considère les besoins en énergie du Québec seulement, sans l'objectif d'exportation de 3 500 MW, on peut envisager deux scénarios:

- soit réaliser les trois centrales du complexe Grande-Baleine de 2002 à 2009 et ensuite Sainte-Marguerite-3 en 2010;
- soit réaliser Sainte-Marguerite-3 dès 2001 et réaliser par la suite soit Grande-Baleine, soit d'autres projets de moindre envergure et de moindre coût. (p. 15)

Pour tenir compte de l'incertitude reliée à la réalisation des objectifs d'exportation, nous effectuerons deux analyses distinctes. Dans la première, nous ferons l'hypothèse que le projet Grande-Baleine (GB) sera entrepris pour produire une partie de l'énergie destinée à l'exportation, et qu'il faut choisir ensuite entre les projets SM3 et NBR. Dans un deuxième temps, nous ne tiendrons pas compte de l'objectif d'exportation et comparerons les projets SM3 et Grande-Baleine. Soit $S_1 = \{SM3, NBR\}$ et $S_2 = \{SM3, GB\}$ les ensembles d'actions possibles correspondant à ces deux scénarios.

3.4 Fonction de coût

Pour effectuer une analyse de la décision, il faut mesurer l'utilité de chaque action, c'est-à-dire établir une échelle de préférences permettant de comparer les actions. En ce qui concerne le choix d'un projet, on retient cette citation du rapport d'avant-projet (Hydro-Québec, 1991):

Pour établir l'ordonnancement des projets, le prix de revient n'est pas le seul critère utilisé. En effet, il faut aussi tenir compte de l'état d'avancement des études et de la production énergétique des équipements. (p. 10)

Puisque nous avons construit l'espace des actions en fonction de l'état d'avancement des études et de la production énergétique des équipements, nous pouvons comparer les actions sur la base du prix de revient de l'électricité. En consultant le Tableau 3.1, on voit que le projet SM3 est plus utile que Grande-Baleine et NBR. Cependant, l'estimation du prix de revient de ce projet dépend entre autres de deux variables, sur lesquelles subsiste une incertitude importante:

- le module naturel à SM3, estimé par rapport de superficie de bassins versants (Q);
- l'autorisation de prélever un débit Q_M dans le bassin versant de la rivière Moisie.

Dans cette section, nous développons une fonction de coût $g(\text{SM3}, Q, Q_M, \theta)$, permettant d'évaluer le prix de revient du projet SM3 en fonction du module naturel à SM3 et de la probabilité θ que le Gouvernement du Québec accepte qu'Hydro-Québec prélève un débit Q_M dans le bassin versant de la rivière Moisie. Puisque la rentabilité des projets Grande-Baleine et NBR n'est vraisemblablement pas influencée par les caractéristiques hydrologiques du projet SM3, nous fixerons $g(\text{GB}, Q, Q_M, \theta) = 4.4 \text{ ¢/kWh}$ et $g(\text{NBR}, Q, Q_M, \theta) = 4.6 \text{ ¢/kWh}$.

Tableau 3.3: Prix de revient de l'électricité à SM3, en fonction du module naturel à SM3, du débit détourné et du débit réservé (adapté de Hydro-Québec, 1991 et 1992)

Valeurs numériques utilisées dans chacune des études

Variante	Module naturel (Q , m ³ /s)	+ Débit détourné (Q_D , m ³ /s)	- Débit réservé (Q_R , m ³ /s)	= Module total (Q_T , m ³ /s)	Prix de revient (¢/kWh 1991)
A	109	0	0	109	5.0
B	109	56	0	165	3.7
C	109	56	10	155	4.0
D	122	52	10	164	3.8

Les études effectuées pour calculer la rentabilité des différentes variantes du projet SM3 peuvent être utilisées pour relier la rentabilité au module naturel à SM3. On dispose de quatre études de ce type: celles réalisées pour les variantes A, B et C, dont les caractéristiques principales sont résumées par le Tableau 3.2, et l'étude effectuée après révision du débit à SM2 (Hydro-Québec, 1992), que nous nommerons pour simplifier la variante D. Ces études évaluent le prix de revient de l'électricité en fonction, entre autres, d'estimations du module naturel à SM3 (Q), du débit détourné du bassin de la rivière Moisie (Q_D), et du débit réservé (Q_R), qui est retourné à la rivière Moisie. Les valeurs de ces paramètres utilisées pour les quatre études sont résumées par le Tableau 3.3. Notons que le débit détourné est différent pour les variantes C et D car ce débit a été réévalué à la baisse en 1992 suite à une période de plus faible hydraulité.

Vraisemblablement, la rentabilité du projet dépend peu de la provenance de l'eau, mais plus de son volume. Ainsi, on peut supposer que les deux dernières colonnes du Tableau 3.3

donnent une bonne indication de la rentabilité du projet SM3 en fonction du module total $Q_T=Q+Q_M$, quelle que soit la répartition de cette valeur entre le module naturel (Q) et le débit prélevé dans le bassin versant de la rivière Moisie ($Q_M=Q_D-Q_R$). Comme le montre la Figure 3.3, la relation ainsi obtenue entre le module total (Q_T) et le prix de revient semble linéaire. La droite reliant les points, obtenue par moindres carrés, permet d'évaluer le prix de revient par $7.4-0.0224 \cdot (Q+Q_M)$.

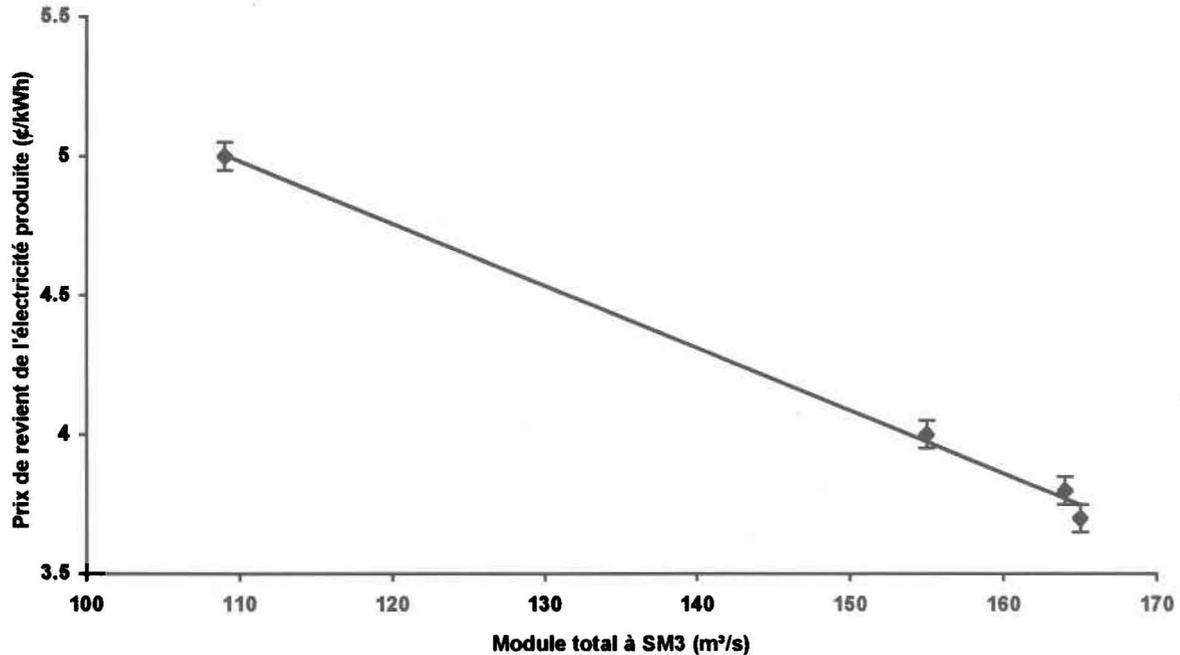


Figure 3.3: Relation entre le prix de revient de l'électricité et le module total à SM3

Si Hydro-Québec estime qu'une requête visant à prélever un débit Q_M dans le bassin de la rivière Moisie sera acceptée avec probabilité θ , alors le prix de revient espéré de l'action SM3 peut être estimé par:

$$g(\text{SM3}, Q, Q_M, \theta) = 7.4 - 0.0224 \cdot (Q + Q_M \cdot \theta) \quad (3.1)$$

Au moment de l'étude d'avant-projet, le débit pouvant être prélevé dans le bassin de la rivière Moisie était mieux connu que le module naturel à SM3, car ce dernier était obtenu par transformation d'une estimation du débit à SM2 plutôt que par des mesures au site. Dans le reste de ce chapitre, nous négligerons l'incertitude sur le débit pouvant être prélevé, en insistant plutôt sur l'incertitude liée à la décision du Gouvernement du Québec d'autoriser ou non les prélèvements dans le bassin de la rivière Moisie. Hydro-Québec (1991) prévoyait

prélever environ 46 m³/s dans le bassin de la rivière Moisie. En fixant pour simplifier $Q_M=46$ m³/s, on obtient la fonction de coût suivante, qui servira de base à notre analyse de la décision:

$$\begin{aligned} g(\text{SM3}, Q, \theta) &= 7.4 - 0.0224 \cdot Q - \theta \\ g(\text{GB}, Q, \theta) &= 4.4 \\ g(\text{NBR}, Q, \theta) &= 4.6 \end{aligned} \quad (3.2)$$

Cette fonction de coût a l'avantage de décrire l'effet combiné de l'incertitude hydrologique à l'aide du paramètre Q et de l'incertitude politique à l'aide du paramètre θ . Il s'agit d'une représentation simplifiée du problème, qui ne tient en particulier pas compte de façon explicite des complications politiques, hydrauliques et géotechniques reliées aux projets Grande-Baleine et NBR. Selon nous, le projet SM3 est assez complexe en lui-même pour permettre d'illustrer de façon convaincante la méthodologie développée au chapitre précédent.

3.5 Analyse du risque a priori

La décision optimale a priori dépend de l'information disponible sur le module naturel à SM3 et de la probabilité d'obtenir l'accord du Gouvernement du Québec pour détourner une partie du bassin versant de la rivière Moisie. Il se peut aussi que cette information soit trop tenue pour permettre d'identifier l'action de risque minimal. Etant donné la forme de la fonction de coût, des prévisions précises de Q et θ sont suffisantes (mais ne sont pas nécessaires) pour départager les actions. Hydro-Québec (1991) propose 109 m³/s comme estimation du module naturel à SM3. Nous formaliserons cette information sous la forme de la prévision précise $E(Q)=109$. Ceci permet d'obtenir le coût de chaque action en fonction de θ uniquement:

$$\begin{aligned} g(\text{SM3}, \theta) &= E[g(\text{SM3}, Q, \theta) | \theta] = 5.0 - \theta \\ g(\text{GB}, \theta) &= 4.4 \\ g(\text{NBR}, \theta) &= 4.6 \end{aligned} \quad (3.3)$$

Ainsi, pour préférer le projet SM3 au projet Grande-Baleine, il faut être prêt à parier au moins avec probabilité $\underline{E}(\theta)=0.6$ sur une décision favorable du Gouvernement du Québec, alors que $\underline{E}(\theta)=0.4$ est suffisant pour rejeter le projet NBR. S'il faut choisir un projet immédiatement, s'il n'est pas possible d'attendre d'obtenir de meilleures informations sur le débit à SM3, alors le projet le plus rentable dépend de l'opinion d'Hydro-Québec sur la

vraisemblance d'une décision favorable du Gouvernement du Québec concernant le détournement du bassin de la rivière Moisie.

S'il est possible de proposer une prévision précise pour Q , cela est plus difficile pour θ . Puisque Hydro-Québec (1991) préférerait le projet SM3 au projet NBR, on pourrait dire que l'entreprise s'est comportée comme si $\underline{E}(\theta) \geq 0.4$. Cependant, une modélisation de l'information a priori basée sur l'observation du comportement du décideur demanderait une interprétation descriptive de la théorie des prévisions inférieures. On peut quand même imaginer que l'entreprise imaginait, avec un certain optimisme, pouvoir détourner une partie des eaux du bassin de la rivière Moisie. Autrement, d'autres projets plus rentables auraient été envisagés à ce stade.

Etant donné le comportement du décideur dans le dossier du projet SM3, nous pourrions proposer une prévision imprécise $[\underline{E}(\theta), 1]$ pour θ . On obtiendrait un risque imprécis $[4.0, 5.0 - \underline{E}(\theta)]$ pour l'action SM3. Cependant, $\bar{R}(\text{SM3}) = 5.0 - \underline{E}(\theta)$ pourrait s'avérer supérieur au risque d'un autre projet, et il faudrait construire une règle de décision arbitraire δ pour choisir une action a priori. Par exemple, en utilisant le type de règle de décision $\delta^{(\kappa)}$, on montre que SM3 est préféré à Grande-Baleine uniquement si $\kappa[1 - \underline{E}(\theta)] < 0.4$. Pour simplifier la notation et la compréhension, on peut proposer l'emploi d'une variable $p = 1 - \kappa[1 - \underline{E}(\theta)]$ qui intégrerait à la fois l'information imprécise sur θ et la règle de décision arbitraire. Le paramètre p peut être interprété comme la probabilité que le Gouvernement du Québec accepte le détournement et qu'Hydro-Québec décide alors de construire SM3 malgré l'indécision qui pourrait subsister. Dans la suite de ce chapitre, nous supposons que le décideur est prêt à préciser cette valeur de p , et nous étudierons la sensibilité des résultats à la valeur de ce paramètre.

Pour $p \in [0, 1]$ connu, le projet SM3 est donc préféré à Grande-Baleine si et seulement si $p > 0.6$, et préféré à NBR si et seulement si $p > 0.4$. De façon générale, si l'on compare SM3 à un projet alternatif PA dont le prix de revient est c , alors le risque a priori des deux actions est donné par:

$$\begin{aligned} R(\text{SM3}) &= 7.4 - 0.0224 \cdot E(Q) - p = 5.0 - p \\ R(\text{PA}) &= c \end{aligned} \tag{3.4}$$

Conséquemment, le risque de l'action optimale a priori est donnée par:

$$R(\hat{a}) = \min\{c, 5.0 - p\} \quad (3.5)$$

3.6 Valeur de l'information parfaite

On peut, a priori, déterminer la valeur d'une information parfaite sur le module naturel à SM3 et sur la décision que prendra le Gouvernement du Québec concernant le détournement du bassin de la rivière Moisie. Cette valeur de l'information parfaite dépend de la quantité d'information a priori et de la façon dont elle est formalisée.

3.6.1 Modélisation de l'information a priori

En premier lieu, nous utiliserons l'information régionale pour limiter le domaine de variation du module naturel Q . Ceci est nécessaire pour éviter une extrapolation abusive de la fonction de coût, qui fut déterminée de façon empirique. Si l'on peut accepter l'hypothèse de linéarité pour l'interpolation, rien ne nous assure que cette hypothèse tient aussi à l'extrapolation. On sait d'ailleurs que la relation ne saurait être linéaire pour des valeurs extrêmes de Q . Par exemple, si $Q > 290 \text{ m}^3/\text{s}$ le prix de revient évalué par la fonction de coût proposée devient négatif! Même en tenant compte de l'incertitude du module estimé à SM2, nous croyons qu'un hydrologue connaissant le bassin versant de la Sainte-Marguerite accorderait un poids négligeable à l'hypothèse que le module réel à SM3 soit supérieur à l'estimation disponible à SM2, et obtenue à partir d'un échantillon de 45 années de mesure. Ainsi, nous supposons que $Q \leq 139 \text{ m}^3/\text{s}$. De façon analogue, nous considérerons négligeable la possibilité que Q soit inférieur au débit estimé de la portion détournée des rivières Carheil et aux Pékans. Rappelons que cette portion du bassin versant de la rivière Moisie fait 2530 km^2 , à peine plus de la moitié du bassin versant naturel à SM3. Il semble donc très peu probable que Q soit inférieur à $Q_M = 46 \text{ m}^3/\text{s}$. Rappelons que Q constitue la moyenne (ou la prévision) du débit annuel. Si l'on note x le débit observé une année quelconque à SM3, alors notre hypothèse sur le domaine de Q est équivalente à la prévision imprécise $\underline{E}(x) = 46 \text{ m}^3/\text{s}$, $\bar{E}(x) = 139 \text{ m}^3/\text{s}$.

Nous utiliserons aussi les informations employées pour identifier une action optimale a priori, c'est-à-dire que nous considérerons que l'information a priori permet d'effectuer la prévision précise $E(Q) = 109 \text{ m}^3/\text{s}$ et de préciser la valeur du paramètre p , qui intègre

l'incertitude politique et l'aspect arbitraire de la décision d'Hydro-Québec. Nous supposons que p peut être fixé par le décideur, mais que sa valeur nous est inconnue pour le moment.

En résumé, l'évaluation de la prévision supérieure du regret \overline{XOL} proposée dans cette section repose sur les informations suivantes:

- le module naturel à SM3 est supérieur à celui des portions détournées des rivières Carheil et aux Pékans, mais inférieur au module naturel à SM2, i.e. $46 \text{ m}^3/\text{s} \leq Q \leq 139 \text{ m}^3/\text{s}$;
- en considérant l'information régionale, on peut estimer le module naturel à $109 \text{ m}^3/\text{s}$, information que nous avons modélisée par la prévision précise $E(Q) = 109 \text{ m}^3/\text{s}$;
- le décideur est capable de fixer la probabilité p que le Gouvernement du Québec accepte qu'Hydro-Québec détourne sur une base annuelle $46 \text{ m}^3/\text{s}$ du bassin de la rivière Moisie, et que l'entreprise choisisse ensuite de construire SM3.

3.6.2 Calcul de la valeur de l'information parfaite à l'aide du théorème de l'enveloppe inférieure

Pour calculer la valeur maximale de l'information parfaite, il est possible d'utiliser le principe d'extension des prévisions inférieures ou d'appliquer le théorème de l'enveloppe inférieure. Les deux approches mènent aux mêmes résultats, mais nous utiliserons la seconde car elle permet d'obtenir plus facilement une solution analytique. Rappelons que, pour Q et p connus et pour un ensemble d'actions S fixé, le regret d'une décision $\delta_0 \in S$ est donné par:

$$OL(\delta_0, Q, p) = g(\delta_0, Q, p) - \inf_{a \in S} g(a, Q, p) = \sup_{a \in S} [g(\delta_0, Q, p) - g(a, Q, p)] \quad (3.6)$$

Au lieu d'étudier séparément les ensembles d'actions $S_1 = \{\text{SM3}, \text{NBR}\}$ et $S_2 = \{\text{SM3}, \text{GB}\}$, nous comparerons le projet SM3 à un projet alternatif PA de coût fixe c . On retrouvera S_1 pour $c = 4.6 \text{ ¢/kWh}$, et S_2 pour $c = 4.4 \text{ ¢/kWh}$. Cela simplifie la présentation, et permet d'établir facilement la valeur de l'information parfaite pour un autre ensemble d'actions.

La comparaison du coût des actions SM3 et PA montre que, pour Q et p connus, le projet SM3 est préféré au projet alternatif PA si et seulement si $g(\text{SM3}, Q, p) = 7.4 - 0.0224 \cdot Q - p < c$. Il est utile de définir le débit critique $Q(c, p)$ en dessous duquel le projet alternatif est plus rentable:

$$Q(c, p) = \frac{7.4 - c - p}{0.0224} = 330 - 45 \cdot (c + p) \quad (3.7)$$

On observe que le débit critique fait partie du domaine de définition du module Q seulement si $4.3 \text{ ¢/kWh} < c+p < 6.4 \text{ ¢/kWh}$. Si ce n'est pas le cas, alors la décision n'est pas influencée par la valeur de Q , et la valeur de l'information est nulle. Nous supposons donc par la suite que cette condition est vérifiée, puisque la solution est triviale dans les autres cas. On remarque que la condition est vérifiée quel que soit p pour les projets Grande-Baleine ($c=4.4 \text{ ¢/kWh}$) et NBR ($c=4.6 \text{ ¢/kWh}$).

Pour l'ensemble d'actions $S=\{\text{SM3, PA}\}$, on obtient les expressions suivantes pour le regret en appliquant la définition (3.6):

$$\begin{aligned} OL(\text{SM3}, Q, p) &= \max[0 ; 7.4 - c - 0.0224 \cdot Q - p] \\ OL(\text{PA}, Q, p) &= \max[0 ; c - 7.4 + 0.0224 \cdot Q + p] \end{aligned} \quad (3.8)$$

La valeur maximale de l'information parfaite est obtenue à partir de la prévision supérieure de ce regret:

$$\overline{XOL} = \inf_{\delta_0 \in S} \overline{E}[OL(\delta_0, Q, p)] \quad (3.9)$$

Pour utiliser le théorème de l'enveloppe inférieure, il faut trouver la mesure de probabilité P qui maximise $XOL^P = \inf_{\delta_0 \in S} XOL^P(\delta_0) = \inf_{\delta_0 \in S} E^P[OL(\delta_0, Q, p)]$, tout en étant compatible avec les contraintes imposées par l'information a priori. Rappelons que le paramètre p est supposé connu, de telle sorte que P est une mesure de probabilité sur le domaine de Q uniquement. Pour P connu, la prévision précise du regret $XOL^P(\delta_0)$ est obtenue en prenant l'espérance de (3.8):

$$\begin{aligned} XOL^P(\text{SM3}) &= E^P[7.4 - c - 0.0224 \cdot Q - p | Q \leq Q(c, p)] \cdot P[Q \leq Q(c, p)] \\ XOL^P(\text{PA}) &= E^P[c - 7.4 + 0.0224 \cdot Q + p | Q > Q(c, p)] \cdot P[Q > Q(c, p)] \end{aligned} \quad (3.10)$$

En remplaçant l'espérance conditionnelle par l'intégrale correspondante, on obtient:

$$\begin{aligned} XOL^P(\text{SM3}) &= (7.4 - c - p) \cdot F(c, p) - 0.0224 \cdot \int_{46}^{Q(c, p)} Q dP \\ XOL^P(\text{PA}) &= (7.4 - c - p) \cdot \{F(c, p) - 1\} + 0.0224 \cdot \int_{Q(c, p)}^{139} Q dP \end{aligned} \quad (3.11)$$

où $F(c,p)=P[Q \leq Q(c,p)]$ correspond à la probabilité de non-dépassement du débit critique. On peut donc établir la valeur de l'information parfaite pour P connu, qui correspond au minimum de $XOL^P(\text{SM3})$ et $XOL^P(\text{PA})$.

$$XOL^P = (7.4 - c - p) \cdot F(c, p) + \min \left\{ -0.0224 \cdot \int_{46}^{Q(c,p)} Q dP ; (p + c - 7.4) + 0.0224 \cdot \int_{Q(c,p)}^{139} Q dP \right\} \quad (3.12)$$

La valeur de l'information parfaite quantifie en unités d'utilité l'influence d'une connaissance parfaite de Q sur la décision. Elle sera donc maximale pour les mesures de probabilités les plus sensibles à une information supplémentaire, et minimale pour des mesures de probabilités qui seraient peu ou pas modifiées par une information supplémentaire. Par exemple, la valeur de l'information parfaite est nulle pour une mesure de probabilité concentrée en un point, puisque dans ce cas on connaît déjà parfaitement la valeur de Q . A l'opposé, la valeur de l'information est maximale pour des distributions concentrées en deux points Q_1 et Q_2 de l'espace, puisque toute information supplémentaire permettra d'établir précisément laquelle des deux hypothèses est la bonne, à condition que ces points soient situés de part et d'autre du débit critique $Q(c,p)$ de façon qu'une information supplémentaire puisse modifier la décision. Pour de telles distributions, le calcul de la valeur de l'information parfaite est considérablement simplifié:

$$XOL^P = (7.4 - c - p) \cdot F(c, p) + \min \left\{ -0.0224 \cdot Q_1 \cdot F(c, p) ; (p + c - 7.4) + 0.0224 \cdot Q_2 \cdot [1 - F(c, p)] \right\} \quad (3.13)$$

Seules les distributions concentrées en deux points ayant pour espérance $E(Q)=109 \text{ m}^3/\text{s}$ sont compatibles avec l'information a priori. Pour Q_1 et Q_2 connus, on peut donc préciser $F(c,p)$:

$$F(c, p) = \frac{Q_2 - 109}{Q_2 - Q_1} \quad (3.14)$$

De plus on observe que, pour $F(c,p)$ fixé, la valeur de l'information parfaite est maximale lorsque Q_1 est minimal et Q_2 maximal. Le maximum de XOL^P est donc atteint pour la distribution concentrée en $Q_1=\inf Q=46 \text{ m}^3/\text{s}$ et $Q_2=\sup Q=139 \text{ m}^3/\text{s}$ ayant pour espérance $E(Q)=109 \text{ m}^3/\text{s}$. La valeur maximale de l'information parfaite est donc donnée par l'équation suivante:

$$\overline{XOL} = \min \left\{ (7.4 - c - p - 0.0224 \cdot \inf Q) \cdot \frac{\sup Q - E(Q)}{\sup Q - \inf Q}; \right. \\ \left. (p + c + 0.0224 \cdot \sup Q - 7.4) \cdot \frac{E(Q) - \inf Q}{\sup Q - \inf Q} \right\} \quad (3.15)$$

$$\overline{XOL} = \min \left\{ (6.4 - c - p) \cdot \frac{30}{93}; (p + c - 4.3) \cdot \frac{63}{93} \right\} \quad (3.16)$$

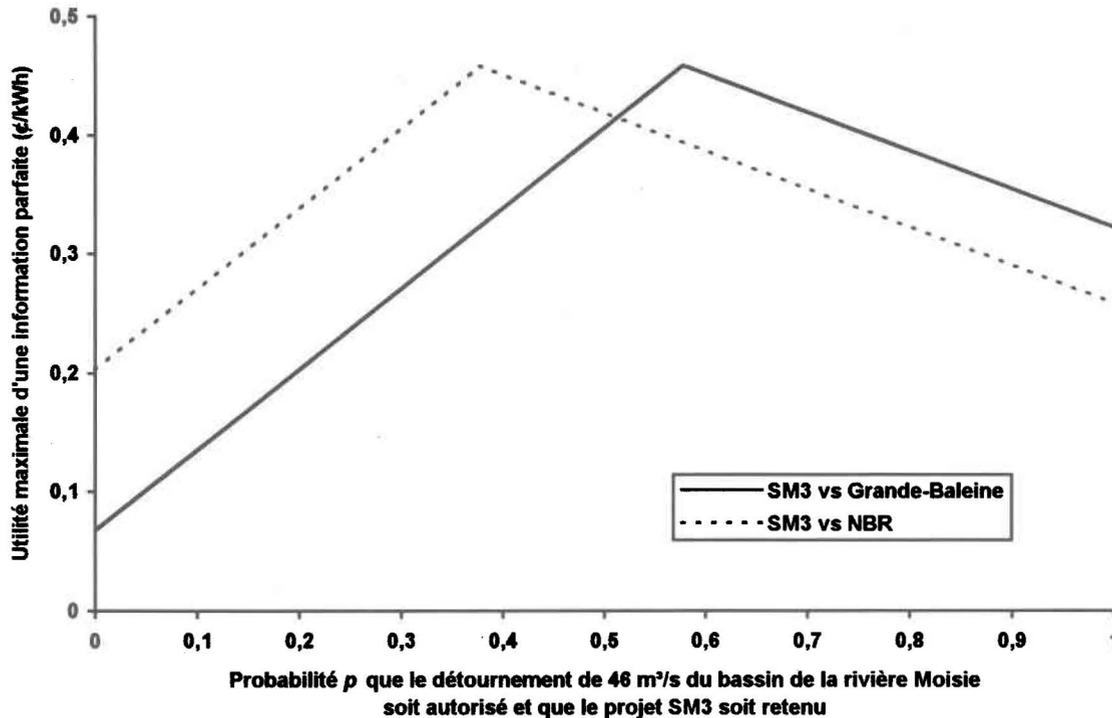


Figure 3.4: Valeur maximale d'une connaissance parfaite du module à SM3

On peut tracer \overline{XOL} en fonction de p pour c fixé. A titre d'exemple, la Figure 3.4 présente la valeur maximale de l'information parfaite pour les projets Grande-Baleine ($c=4.4$ €/kWh) et NBR ($c=4.6$ €/kWh). On observe que le maximum de cette fonction, bien qu'il ne soit pas atteint pour la même valeur de p , est identique pour les deux projets. En fait, on peut montrer que \overline{XOL} est maximal pour $p+c=5.0$. Sa valeur est alors donnée par $[1-F(c,p)]^2=0.46$ €/kWh.

La valeur d'une information hydrologique parfaite est importante, puisqu'elle est du même ordre de grandeur que l'écart de rentabilité entre les projets. Si l'on doit choisir entre

construire SM3 ou entreprendre le complexe NBR, la valeur d'une connaissance parfaite du module à SM3 est équivalente à une réduction du prix de revient de l'électricité d'au moins 0.2 ¢/kWh, ce qui n'est pas négligeable étant donné la production annuelle des projets à l'étude. De plus, en fonction du niveau d'incertitude politique, la valeur d'une information hydrologique parfaite peut facilement doubler.

Il faut comprendre qu'une information additionnelle sur Q peut bien sûr provenir de mesures du débit au site, mais aussi d'une analyse plus détaillée de l'information régionale disponible. Le fait que la mesure P , qui maximise XOL^P et reste compatible avec le modèle de l'information a priori, soit dégénérée peut indiquer que des contraintes supplémentaires pourraient être obtenues à partir des connaissances du décideur. On pourrait par exemple imposer une variance minimale $V(Q) = E\{[(Q-E(Q))^2]\}$ pour Q . Reste à savoir quelle valeur donner à $V(Q)$. Avant de se lancer dans une modélisation plus complexe de l'information a priori, on peut déjà comparer \overline{XOL} au coût d'acquisition d'une information supplémentaire. En effet si la valeur de l'information parfaite, même ainsi surestimée, ne dépasse pas le coût d'échantillonnage on saura déjà qu'il ne vaut pas la peine de retarder la prise de décision.

3.7 Evaluation du coût relié à l'observation du débit annuel

Le coût d'acquisition d'une observation du débit annuel est relié principalement aux conséquences financières du report du projet. Parce que l'énergie produite était destinée à l'exportation, on peut évaluer le coût d'attente en fonction du bénéfice attendu de la vente de l'électricité.

Hydro-Québec (1991) mentionne que le prix de vente de l'énergie, pour les contrats de 12.6 TWh déjà signés au moment de l'étude d'avant-projet, était de 40% supérieur au prix de revient. Les livraisons, débutant plusieurs années avant la mise en service de la centrale SM3, devaient être assurées par la mise en service progressive du complexe Grande-Baleine, dont le prix de revient est estimé à 4.4 ¢/kWh. On en déduit que le prix de vente de l'énergie se situait autour de 6.2 ¢/kWh.

En faisant l'hypothèse d'un prix de vente similaire pour l'énergie produite par SM3 et en considérant un prix de revient espéré donné par le risque de l'action optimale a priori [eq. (3.5)] $R(\hat{a}) = \min\{c, 5.0 - p\}$, retarder la mise en service de la centrale SM3 de 2001 à 2002,

de façon à recueillir une observation du débit annuel, coûterait donc $\max\{6.2 - c, 1.2 + p\}$ ¢ pour chaque kWh ainsi perdu. En étudiant les prévisions présentées dans le rapport d'avant-projet (Hydro-Québec, 1991), on observe qu'il était prévu d'utiliser seulement une partie de l'énergie produite à SM3 la première année, soit environ 2 TWh. On pouvait donc prévoir un coût d'attente de l'ordre de $\max\{124 - 20c, 24 + 20p\}$ millions de dollars, payable une seule fois. Le taux d'actualisation était à ce moment de 11% (Hydro-Québec, 1992), de sorte que l'entreprise pouvait assimiler le coût d'attente à un paiement annuel de $\max\{14 - 2.2c, 2.6 + 2.2p\}$ millions de dollars. Pour les projets alternatifs qui nous intéressent (NBR, avec $c=4.6$ ¢/kWh et Grande-Baleine, avec $c=4.4$ ¢/kWh), cela représente de 4 à 5 millions de dollars par année.

Le coût d'acquisition de l'information est relativement élevé. Mais, étant donné la production hydroélectrique prévue entre 3 et 4 TWh à SM3 (selon que l'on réalise ou non le détournement), une économie de l'ordre de 0.1 ¢/kWh serait suffisante pour justifier l'acquisition d'une information supplémentaire (en effet, $0.1 \text{ ¢/kWh} \cdot 1 \text{ TWh} = 1 \text{ millions \$}$). Remarquons que 0.1 ¢/kWh correspond justement à la précision utilisée par Hydro-Québec pour spécifier le prix de revient d'un projet. Conséquemment, la valeur d'une observation du débit annuel à SM3 dépasse son coût d'acquisition lorsqu'elle devient significative, relativement à la précision des mesures du prix de revient effectuées par Hydro-Québec.

La valeur d'une connaissance parfaite du module naturel à SM3 [Figure 3.4], est au moins 0.1 ¢/kWh. Par contre, rien ne permet d'affirmer qu'une seule observation du débit a une telle valeur; cela dépend de l'idée a priori que l'on a de sa précision.

3.8 Analyse du risque a posteriori

Pour p fixé, le coût du projet SM3 est une fonction linéaire du module Q , ce qui implique que le risque a posteriori dépend de la prévision a posteriori de ce module, sur laquelle nous n'avons fait jusqu'à présent aucune hypothèse. Le principe d'extension des prévisions inférieures n'est pas très utile pour déterminer une prévision imprécise pour le pari $Q|x$, parce que des mesures de probabilité trop concentrées en quelques points de l'espace sont compatibles avec les prévisions utilisées jusqu'à présent pour représenter l'information a priori.

Ainsi, une distribution a priori dégénérée en $E(Q)$ donnera $E(Q|x)=E(Q)$. A l'opposé, une distribution a priori concentrée sur deux points (x,y) donnera $E(Q|x)=x$.

Pour éviter que des distributions a priori trop concentrées soient compatibles avec les contraintes imposées par l'information a priori, nous proposons l'ajout d'une prévision précise $E(Q|x)$ au modèle de l'information a priori. Selon que le poids relatif de l'information a priori est faible ou élevé en comparaison de l'information apportée par x , il semble naturel d'accorder à $E(Q|x)$ une valeur plus ou moins proche de $E(Q)$. On pourrait par exemple proposer une moyenne pondérée pour $E(Q|x)$:

$$E(Q|x) = \frac{n_0 \cdot E(Q) + x}{n_0 + 1} \tag{3.17}$$

où n_0 quantifie le poids de l'information a priori, dans une unité de mesure analogue à une taille d'échantillon. Si $n_0 \ll 1$, on obtient $E(Q|x) \approx x$, ce qui signifie que l'information a priori est tellement vague que l'on préfère effectuer une prévision sur Q basée principalement sur une seule observation du débit. Au contraire, si $n_0 \gg 1$, alors une seule observation du débit ne modifie pas beaucoup l'idée que l'on se faisait a priori de la valeur de Q .

Il est difficile de préciser une valeur pour n_0 , et on pourrait envisager de représenter ce paramètre par un intervalle, ce qui rendrait imprécise la prévision du pari conditionnel $Q|x$. Cependant, pour des raisons techniques liées à la forme de la fonction de coût, nous supposons que le décideur accepte de fixer une valeur précise pour n_0 . Ainsi, il est possible de déterminer l'action optimale a posteriori sans avoir à définir une règle de décision arbitraire, et l'on peut calculer la valeur a posteriori de l'information de façon précise, sans recourir à l'emploi de la règle de Bayes généralisée.

Pour p fixé, le risque a posteriori du projet SM3 est donné par:

$$\begin{aligned} R(\text{SM3}|x) &= 7.4 - 0.0224 \cdot E(Q|x) - p \\ &= \frac{7.4(n_0 + 1) - 0.0224 \cdot [n_0 E(Q) + x]}{n_0 + 1} - p \\ &= \frac{2.4 - 0.0224 \cdot x}{n_0 + 1} + (5.0 - p) \\ &= \frac{2.4 - 0.0224 \cdot x}{n_0 + 1} + R(\text{SM3}) \end{aligned} \tag{3.18}$$

On observe que le risque du projet SM3 diminue ou augmente selon que l'observation x est inférieure ou supérieure à la prévision a priori $E(Q)$. Cette modification du risque est d'autant plus faible que le poids n_0 de l'information a priori est faible.

Si l'on compare a posteriori le projet SM3 à un projet alternatif de coût fixe c , on trouve que le projet SM3 est préféré dans la mesure où cette observation est supérieure à un certain seuil z :

$$\hat{a}_{(x)} = \text{SM3} \quad \text{ssi} \quad x \geq z, \quad \text{où} \quad z = z(c, p, n_0) = 109 - \frac{(c + p - 5.0)(n_0 + 1)}{0.0224} \quad (3.19)$$

Ce seuil est plus facile à atteindre lorsque l'on a confiance que le projet de détournement pourra être réalisé, i.e. lorsque p est élevé.

3.9 Valeur a posteriori d'une observation du débit annuel

Etant donné que l'information a priori permet d'identifier les actions optimales a priori et a posteriori, on peut obtenir facilement la valeur a posteriori d'une observation du débit annuel. Si l'on compare le projet SM3 à un projet alternatif PA de coût c , la valeur a posteriori de l'information est, par définition, nulle si $\hat{a} = \hat{a}_{(x)}$, égale à $R(\text{SM3}|x) - c$ si $\hat{a} = \text{SM3}$ et $\hat{a}_{(x)} = \text{PA}$, et égale à $c - R(\text{SM3}|x)$ si $\hat{a} = \text{PA}$ et $\hat{a}_{(x)} = \text{SM3}$. En combinant les équations (3.18) et (3.19), on obtient:

$$VSI(x) = \begin{cases} \frac{2.4 - 0.0224 \cdot x}{n_0 + 1} - (c + p) + 5.0 & \text{si} \quad c + p > 5.0 \text{ et } x < z \\ \frac{0.0224 \cdot x - 2.4}{n_0 + 1} + (c + p) - 5.0 & \text{si} \quad c + p \leq 5.0 \text{ et } x > z \\ 0 & \text{sinon} \end{cases} \quad (3.20)$$

On remarque que $VSI(x)$ ne dépend pas des valeurs individuelles de c et p , seulement de leur somme. Ainsi, le paramètre p agit comme une pénalité sur le prix de revient du projet alternatif.

3.10 Prévision imprécise de la valeur de l'information basée sur une hypothèse de normalité du débit annuel

Pour obtenir une prévision de la valeur de l'information, on peut procéder en deux étapes:

1. Calculer l'espérance $XVSI(\theta)$ de $VSI(x)$ par rapport à la distribution d'échantillonnage $f(x|\theta)$ de l'observation x . Il ne s'agit pas d'un intervalle, étant donné que la valeur a posteriori de l'information est précise.
2. Construire une prévision imprécise pour $XVSI(\theta)$, après avoir modélisé l'information a priori sur les paramètres θ . Pour que l'information x soit utile, il faut évidemment que le vecteur θ de paramètres comprennent au moins le module Q , mais x peut aussi dépendre d'autres paramètres, que l'on nomme usuellement des paramètres nuisibles.

3.10.1 Modèle d'échantillonnage

Le débit annuel est constitué de la moyenne des débits journaliers. On peut s'attendre à ce que cette distribution soit approximativement normale, par application du théorème de la limite centrale, même si les débits journaliers ne sont pas indépendants ni identiquement distribués. Pour montrer que cette hypothèse est compatible avec les données, nous avons porté sur papier normal la série chronologique du débit annuel à SM3 [Figure 3.5]. En plus des données de la période 1958-1982, utilisées lors de l'étude d'avant-projet, nous avons inclus les données plus récentes de la période 1983-1993. Il reste que l'hypothèse de normalité est introduite principalement pour simplifier les calculs, même si elle peut sembler moins gênante pour des débits annuels que pour d'autres variables. Notons tout de même que Perreault et al. (1996) ont conduit une étude des apports énergétiques (i.e. une transformation en kWh des apports en eau) de huit complexes hydroélectriques gérés par Hydro-Québec, et n'ont rejeté l'hypothèse de normalité que pour le complexe Bersimis (test de Shapiro-Wilk (1965) au niveau de signification de 5%).

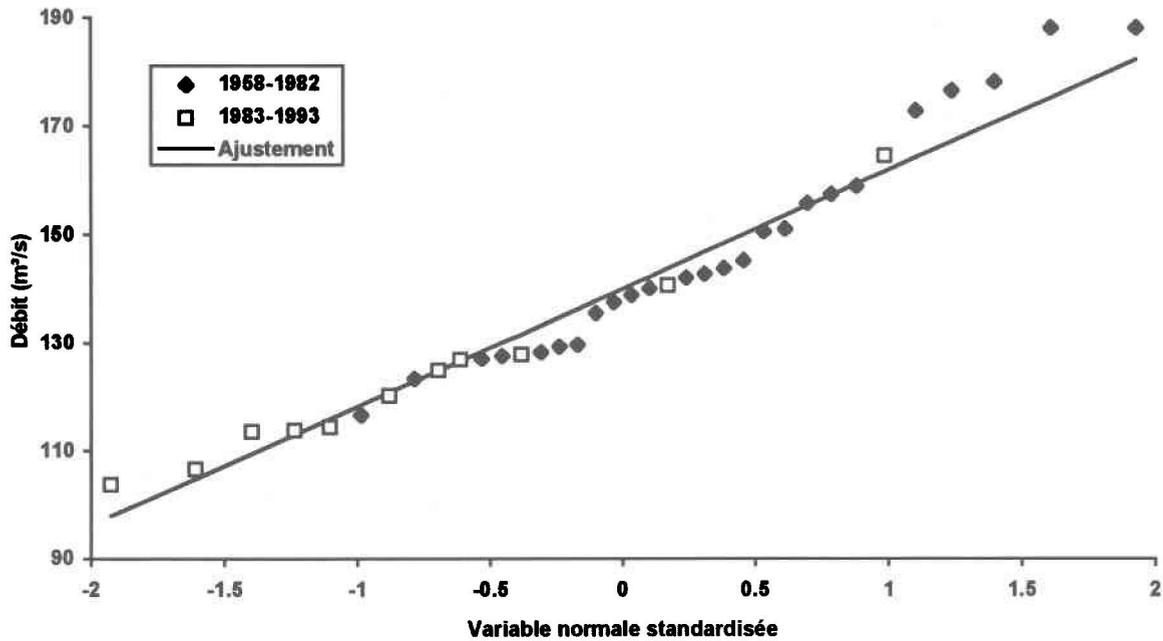


Figure 3.5: Représentation du débit annuel à SM2 sur papier normal pour la période 1958-1993

On remarque que plusieurs des valeurs de la période 1983-1993 sont parmi les plus faibles de la série. Ceci est dû à une sécheresse qui a débuté au milieu des années 1980. On peut d'ailleurs observer une tendance à la baisse des apports énergétiques sur l'ensemble du réseau d'Hydro-Québec depuis 1985 (Perreault et al., 1996). Au moment où l'étude d'avant-projet a été réalisée à SM3, aucune information n'était disponible sur ce changement de moyenne apparent. Nous continuerons donc l'analyse sans en tenir compte, étant donné qu'une étude du problème de la non-stationarité est un problème complexe qui mérite d'être étudié séparément.

Nous supposons donc que le débit annuel x a une distribution normale de moyenne Q et d'écart-type σ . Puisque σ n'apparaît pas dans la fonction de coût, il s'agit d'un paramètre nuisible. La vraisemblance de (Q, σ) pour x fixé est donc donnée par l'équation suivante:

$$L_x(Q, \sigma) = f(x|Q, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-Q)^2}{2\sigma^2}\right\} \quad (3.21)$$

3.10.2 Prédiction de la valeur de l'information en fonction du module et de l'écart-type du débit annuel

Pour Q et σ connus, on peut déterminer une prédiction précise $XVSI(Q, \sigma) = E^{f(x|Q, \sigma)} [VSI(x)]$ de la valeur de l'information. En combinant (3.20) et (3.21) avec la définition de l'espérance conditionnelle, on obtient:

si $c + p > 5.0$

$$XVSI(Q, \sigma) = E^{f(x|Q, \sigma)} \left[\frac{2.4 - 0.0224 \cdot x}{n_0 + 1} - (c + p) + 5.0 \mid x < z \right] \cdot F(z|Q, \sigma) \quad (3.22)$$

si $c + p \leq 5.0$

$$XVSI(Q, \sigma) = E^{f(x|Q, \sigma)} \left[\frac{0.0224 \cdot x - 2.4}{n_0 + 1} + (c + p) - 5.0 \mid x > z \right] \{1 - F(z|Q, \sigma)\}$$

où $F(z|Q, \sigma) = \int_{-\infty}^z f(x|Q, \sigma) dx$ correspond à la fonction de répartition de la loi normale (probabilité de non-dépassement). On sait que la distribution de x conditionnelle à un intervalle correspond à une distribution normale tronquée (Patel et Read, 1996). On peut donc écrire:

si $c + p > 5.0$

$$XVSI(Q, \sigma) = \left\{ \frac{2.4 - 0.0224 \cdot E^{f^*(y|Q, \sigma, z)}(y)}{n_0 + 1} - (c + p) + 5.0 \right\} F(z|Q, \sigma) \quad (3.23)$$

si $c + p \leq 5.0$

$$XVSI(Q, \sigma) = \left\{ \frac{0.0224 \cdot E^{f_*(y|Q, \sigma, z)}(y) - 2.4}{n_0 + 1} + (c + p) - 5.0 \right\} \{1 - F(z|Q, \sigma)\}$$

où $f_*(y|Q, \sigma, z)$ est la loi normale tronquée inférieurement en z et $f^*(y|Q, \sigma, z)$ est la loi normale tronquée supérieurement en z . On connaît l'espérance de la loi normale tronquée (Patel et Read, 1996):

$$E^{f^*(y|Q, \sigma, z)}(y) = Q - \frac{f(z|Q, \sigma)}{F(z|Q, \sigma)} \sigma^2$$

$$E^{f_*(y|Q, \sigma, z)}(y) = Q + \frac{f(z|Q, \sigma)}{1 - F(z|Q, \sigma)} \sigma^2 \quad (3.24)$$

En combinant (3.23) et (3.24) on obtient:

si $c + p > 5.0$

$$XVSI(Q, \sigma) = \left\{ \frac{2.4 - 0.0224 \cdot Q}{n_0 + 1} - (c + p) + 5.0 \right\} F(z|Q, \sigma) + \frac{0.0224\sigma^2}{n_0 + 1} f(z|Q, \sigma) \quad (3.25)$$

si $c + p \leq 5.0$

$$XVSI(Q, \sigma) = \left\{ \frac{0.0224 \cdot Q - 2.4}{n_0 + 1} + (c + p) - 5.0 \right\} \{1 - F(z|Q, \sigma)\} + \frac{0.0224\sigma^2}{n_0 + 1} f(z|Q, \sigma)$$

En utilisant une approximation numérique pour $F(z|Q, \sigma)$, on arrive donc à calculer assez facilement une prévision de la valeur de l'information pour Q et σ connus. Évidemment, ce résultat n'est pas intéressant en soi. C'est la prévision imprécise de cette quantité, en fonction de l'information disponible sur Q et sur σ , qui est utile.

3.10.3 Modélisation de l'information sur l'écart-type du débit annuel

Pour obtenir une prévision imprécise de la valeur de l'information, il faut d'abord compléter la modélisation de l'information a priori. Si celle-ci peut être suffisamment détaillée pour Q , il reste à représenter l'information sur le paramètre nuisible σ . Comme nous l'avons fait pour Q , nous proposons de limiter le domaine de définition de σ et d'effectuer une prévision de sa valeur.

Au lieu de modéliser l'information sur σ directement, il est plus facile de travailler avec le coefficient de variation $CV = \sigma/Q$. En effet, ce paramètre est souvent au centre des études régionales en hydrologie. Une hypothèse de base de plusieurs méthodes d'estimation régionale du débit consiste à supposer qu'il existe des «régions homogènes» à l'intérieur desquelles le coefficient de variation varie peu d'un site à l'autre. En fait, en reconstituant le débit à SM3 par rapport de superficie de bassin, les hydrologues d'Hydro-Québec ont conservé le coefficient de variation de SM2, faisant l'hypothèse implicite que les deux sites appartenaient à une même région homogène. Pour modéliser cette information régionale, nous proposons d'ajouter aux contraintes une prévision sur CV égale au coefficient de variation observé à SM2, c'est-à-dire 14% pour la période 1958-1982.

Le climat québécois fournit aussi une information permettant de borner le domaine de définition du coefficient de variation. En effet, la valeur de CV est un indicateur du type de régime hydrologique. Selon Vogel (communication personnelle, 1997), aux Etats-Unis les

valeurs de $CV > 1$ sont typiques des régions désertiques. Puisque le climat du bassin de la Sainte-Marguerite n'est pas désertique, nous pourrions poser $\sup CV = 1$. Il est plus difficile d'établir une borne inférieure pour CV , même si l'on sait que le débit annuel à SM3 n'est pas constant. Etant donné la difficulté de fixer des valeurs précises pour les bornes, nous proposons de modéliser ces bornes par un nombre flou, de façon à étudier la sensibilité de la valeur de l'information à la largeur de l'intervalle. Parce que $E(CV) = 0.14$ doit nécessairement faire partie de cet intervalle, nous utiliserons un nombre flou triangulaire atteignant son maximum en 0.14 [Figure 3.6].

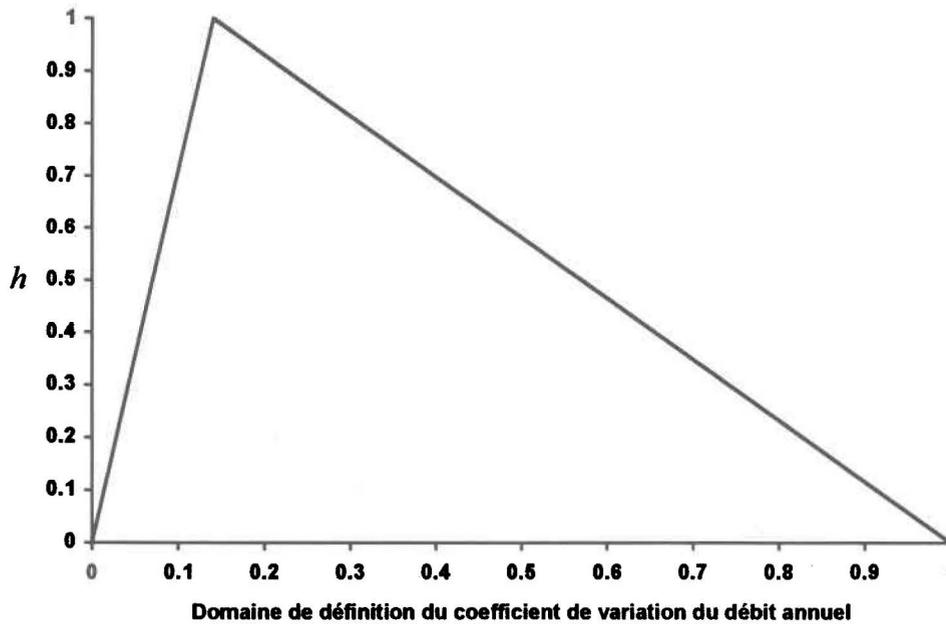


Figure 3.6: Nombre flou utilisé pour effectuer une analyse de sensibilité de la valeur d'une observation du débit annuel à SM3 au domaine de définition du coefficient de variation de ce débit annuel

Puisque l'on s'intéresse plus au coefficient de variation qu'à l'écart-type, on peut réécrire le modèle d'échantillonnage de même que la prévision de la valeur de l'information en fonction de CV plutôt qu'en fonction de σ (en tenant compte de la relation $CV = \sigma/Q$):

$$L_x(Q, CV) = f(x|Q, CV) = \frac{1}{CV \cdot Q \cdot \sqrt{2\pi}} \exp\left\{-\frac{(x/Q - 1)^2}{2 \cdot CV^2}\right\} \quad (3.26)$$

$$F(z|Q, CV) = \int_{-\infty}^z f(x|Q, CV) dx \quad (3.27)$$

si $c + p > 5.0$

$$XVSI(Q, CV) = \left\{ \frac{2.4 - 0.0224 \cdot Q}{n_0 + 1} - (c + p) + 5.0 \right\} F(z|Q, CV) + \frac{0.0224 \cdot (CV \cdot Q)^2}{n_0 + 1} f(z|Q, CV) \quad (3.28)$$

si $c + p \leq 5.0$

$$XVSI(Q, CV) = \left\{ \frac{0.0224 \cdot Q - 2.4}{n_0 + 1} + (c + p) - 5.0 \right\} \{1 - F(z|Q, CV)\} + \frac{0.0224 \cdot (CV \cdot Q)^2}{n_0 + 1} f(z|Q, CV)$$

3.10.4 Application du principe d'extension des prévisions inférieures

On peut appliquer le principe d'extension des prévisions inférieures pour obtenir une prévision de la quantité $XVSI(Q, \sigma)$ qui soit cohérente avec les contraintes modélisant l'information a priori:

1. $Q \in [46, 139] \text{ m}^3/\text{s}$;
2. $E(Q) = 109 \text{ m}^3/\text{s}$;
3. $E(Q|x) = \frac{109 \cdot n_0 + x}{n_0 + 1}, \quad \forall x$
4. $CV \in [0.14 \cdot h, 1 - 0.86 \cdot h]$
5. $E(CV) = 0.14$

Notons que, pour calculer une prévision imprécise de $XVSI(Q, CV)$, il faudra fixer les paramètres n_0 , c , p et h . Remarquons aussi que la prévision du pari $Q|x$ représente en fait une famille non dénombrable de prévisions, puisque l'on accepte cette contrainte pour toutes les valeurs de x . Ceci complique l'application du principe d'extension des prévisions inférieures. Selon ce principe, une prévision inférieure $\underline{XVSI} = E[XVSI(Q, CV)]$ cohérente avec l'information a priori peut être obtenue en solutionnant l'équation suivante:

$$\underline{XVSI} = \sup_{\mu_1, \mu_2, (\lambda_n), (x_n)} \left\{ \inf_{Q, \sigma} [XVSI(Q, CV) + \mu_1 \{Q - E(Q)\} + \mu_2 \{CV - E(CV)\}] \right. \\ \left. + \sum_{n=1}^k \lambda_n \cdot L_{x_n}(Q, CV) \cdot \{Q - E(Q|x_n)\} \right\} \quad (3.29)$$

où le supremum doit être pris par rapport à toutes les valeurs de μ_1, μ_2, λ_n réelles, et pour toutes les suites réelles finies $(x_n, n=1, 2, \dots, k)$. De la même façon, on obtient une prévision supérieure $\overline{XVSI} = \overline{E}[XVSI(Q, CV)]$ de la valeur de l'information en solutionnant l'équation suivante:

$$\overline{XVSI} = \inf_{\mu_1, \mu_2, (\lambda_n), (x_n)} \left\{ \sup_{Q, \sigma} [XVSI(Q, CV) + \mu_1 \{Q - E(Q)\} + \mu_2 \{CV - E(CV)\}] \right. \\ \left. + \sum_{n=1}^k \lambda_n \cdot L_{x_n}(Q, CV) \cdot \{Q - E(Q|x_n)\} \right\} \quad (3.30)$$

Il est difficile de voir comment l'on pourrait résoudre ces deux équations, même par des méthodes numériques. Nous proposons une méthode d'estimation approximative de cette prévision imprécise en deux étapes. Définissons d'abord les fonctions $\underline{w}[(x_n)]$ et $\overline{w}[(x_n)]$:

$$\underline{w}[(x_n)] = \underline{w}[(x_n, n=1, 2, \dots, k)] = \sup_{\mu_1, \mu_2, (\lambda_n)} \left\{ \inf_{Q, \sigma} [XVSI(Q, CV) + \mu_1 \{Q - E(Q)\}] \right. \\ \left. + \mu_2 \{CV - E(CV)\} + \sum_{n=1}^k \lambda_n \cdot L_{x_n}(Q, CV) \cdot \{Q - E(Q|x_n)\} \right\} \quad (3.31)$$

$$\overline{w}[(x_n)] = \overline{w}[(x_n, n=1, 2, \dots, k)] = \inf_{\mu_1, \mu_2, (\lambda_n)} \left\{ \sup_{Q, \sigma} [XVSI(Q, CV) + \mu_1 \{Q - E(Q)\}] \right. \\ \left. + \mu_2 \{CV - E(CV)\} + \sum_{n=1}^k \lambda_n \cdot L_{x_n}(Q, CV) \cdot \{Q - E(Q|x_n)\} \right\} \quad (3.32)$$

Pour une suite $(x_n, n=1, 2, \dots, k)$ fixée, les valeurs de $\underline{w}[(x_n)]$ et $\overline{w}[(x_n)]$ peuvent être déterminées en optimisant numériquement des fonctions de $k+4$ variables. Si cela n'est pas facile, c'est au moins possible. On voit que \underline{XVSI} et \overline{XVSI} peuvent être définis en fonction de $\underline{w}[(x_n)]$ et $\overline{w}[(x_n)]$:

$$\underline{XVSI} = \sup_{(x_n)} \underline{w}[(x_n)] \\ \overline{XVSI} = \inf_{(x_n)} \overline{w}[(x_n)] \quad (3.33)$$

On ne peut pas, en pratique, utiliser ces deux équations puisqu'il faudrait pouvoir calculer $\underline{w}[(x_n)]$ et $\overline{w}[(x_n)]$ pour toutes les suites réelles finies. Par contre, on peut obtenir la valeur de ces deux fonctions pour un ensemble quelconque de suites $\{(x_n)_1, (x_n)_2, \dots, (x_n)_N\}$. On comprend que le maximum des valeurs de $\underline{w}[(x_n)]$ et le minimum des valeurs de $\overline{w}[(x_n)]$ obtenues pour ces suites bornent la prévision imprécise de la valeur de l'information:

$$\begin{aligned} \max\{\underline{w}[(x_n)_1], \underline{w}[(x_n)_2], \dots, \underline{w}[(x_n)_N]\} &\leq \sup_{(x_n)} \underline{w}[(x_n)] = \underline{XVSI} \\ \min\{\overline{w}[(x_n)_1], \overline{w}[(x_n)_2], \dots, \overline{w}[(x_n)_N]\} &\geq \inf_{(x_n)} \overline{w}[(x_n)] = \overline{XVSI} \end{aligned} \quad (3.34)$$

Ce n'est pas la seule façon de résoudre le problème. En fait, pour k fixé, on pourrait déterminer de façon numérique $\sup \underline{w}[(x_1, x_2, \dots, x_k)]$ et $\inf \overline{w}[(x_1, x_2, \dots, x_k)]$. Il faudrait appliquer une méthode capable de maximiser ces fonctions de k variables réelles. Cependant, parce que l'évaluation des fonctions $\underline{w}[(x_1, x_2, \dots, x_k)]$ et $\overline{w}[(x_1, x_2, \dots, x_k)]$ demande aussi une optimisation numérique, ce procédé devient trop lourd.

Autrement dit, on sait borner inférieurement \underline{XVSI} et borner supérieurement \overline{XVSI} : une approximation de la valeur imprécise de l'information basée sur l'étude d'un ensemble de suites en fournira toujours une estimation conservatrice. Nous proposons d'étudier les suites de la forme suivante, de taille $k=2^{m-1}+1$:

$$(x_n)_m = \left(\inf Q + (\sup Q - \inf Q) \cdot \frac{n-1}{2^{m-1}} \right) = \left(46 + 93 \cdot \frac{n-1}{2^{m-1}} \right) \quad (3.35)$$

Ces suites divisent le domaine de définition de Q en 2^{m-1} intervalles de même longueur. Elles ont aussi l'avantage de former une famille croissante d'ensembles, i.e. $(x_n)_2 \subset (x_n)_3 \subset \dots \subset (x_n)_N$. Ainsi, on peut montrer que le maximum des valeurs de $\underline{w}[(x_n)_m]$ et le minimum des valeurs de $\overline{w}[(x_n)_m]$ est atteint pour $(x_n)_N$:

$$\begin{aligned} \max\{\underline{w}[(x_n)_1], \underline{w}[(x_n)_2], \dots, \underline{w}[(x_n)_N]\} &= \underline{w}[(x_n)_N] \leq \underline{XVSI} \\ \min\{\overline{w}[(x_n)_1], \overline{w}[(x_n)_2], \dots, \overline{w}[(x_n)_N]\} &= \overline{w}[(x_n)_N] \geq \overline{XVSI} \end{aligned} \quad (3.36)$$

On peut tracer $\underline{w}[(x_n)_m]$ et $\overline{w}[(x_n)_m]$ en fonction de m pour voir comment évolue cette approximation conservatrice de la valeur imprécise de l'information. Supposons que le poids de l'information a priori soit analogue à une année d'observation ($n_0=1$), que l'on compare les projets SM3 et NBR ($c=4.6$ €/kWh) et que le risque a priori des projets SM3 et

NBR soit semblable ($p=0.4$). Dans ces conditions particulières, la Figure 3.7 montre l'évolution de $\underline{w}[(x_n)_m]$ et $\bar{w}[(x_n)_m]$ en fonction de $k=2^{m-1}+1$. On voit que les courbes se stabilisent à partir de $k=5$. Nous avons observé un effet semblable pour d'autres valeurs de n_0 , c et p . Ceci signifie que les premières contraintes sur la valeur de l'espérance conditionnelle ont un effet plus important sur le résultat de l'application du principe d'extension de prévisions inférieures que les contraintes suivantes. On gagne beaucoup plus en précision de $k=0$ à $k=5$ que de $k=5$ à $k=9$.

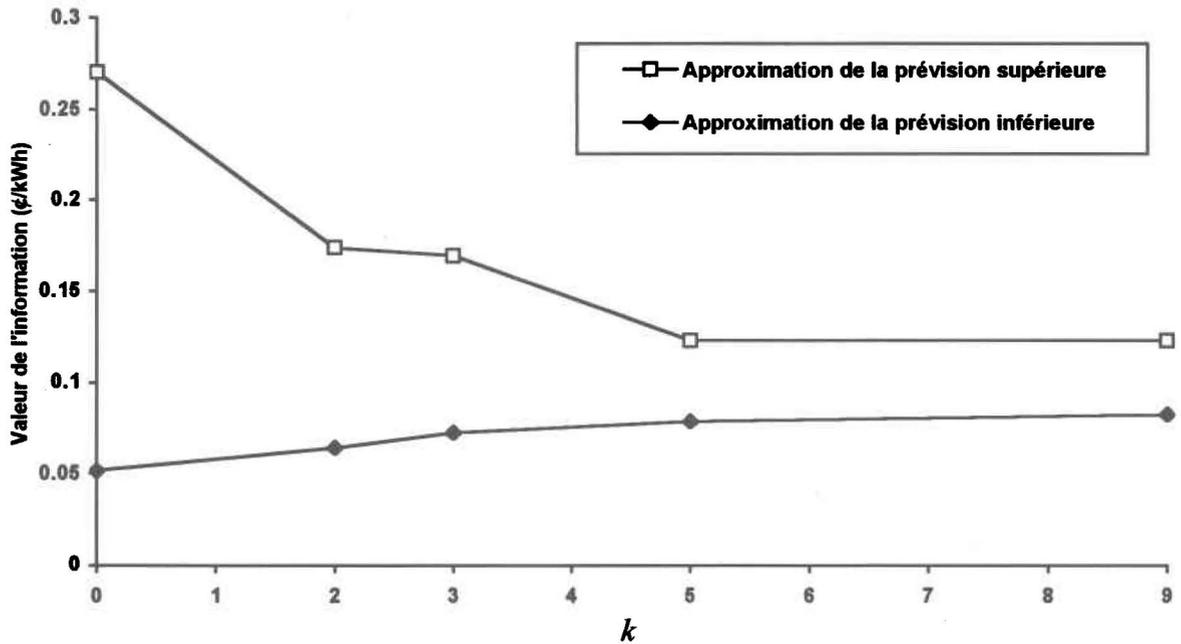


Figure 3.7: Approximation de la prévision imprécise de la valeur de l'information, pour $n_0=1$, $c=4.6$ €/kWh et $p=0.4$, en fonction de la longueur $k=2^{m-1}+1$ de la suite $(x_n)_m$

La valeur de k a une influence importante sur le temps nécessaire pour appliquer numériquement le principe d'extension. Il importe donc de choisir la plus petite valeur possible. Selon le degré de précision voulu, nous fixerons k entre 3 et 9, ce qui correspond à des valeurs de m entre 2 et 4.

3.10.5 Résultats numériques

La prévision floue de la valeur de l'information dépend du paramètre n_0 et de la somme des paramètres c et p . Etudions d'abord la situation où l'on est certain que le détournement d'une

partie du bassin de la rivière Moisie sera accepté ($p=1$) et où l'on considère que l'information a priori a un poids analogue à une année d'observation du débit annuel ($n_0=1$). La Figure 3.8 présente le nombre flou obtenu lorsque le projet alternatif est NBR ($c+p=5.6$), alors que la Figure 3.9 présente le résultat obtenu pour la comparaison de SM3 et Grande-Baleine ($c+p=5.4$). Dans les deux cas, la valeur de l'information est inférieure à son coût d'acquisition (≤ 0.1 €/kWh), quel que soit le domaine de définition du coefficient de variation. Si l'on considère que l'information régionale doit avoir un poids au moins équivalent à une seule observation du débit annuel, et si l'on a une confiance absolue d'obtenir l'autorisation de détourner une partie du bassin versant de la rivière Moisie, alors il n'est pas utile d'attendre d'avoir observé le débit au site pour débiter la réalisation du projet SM3.

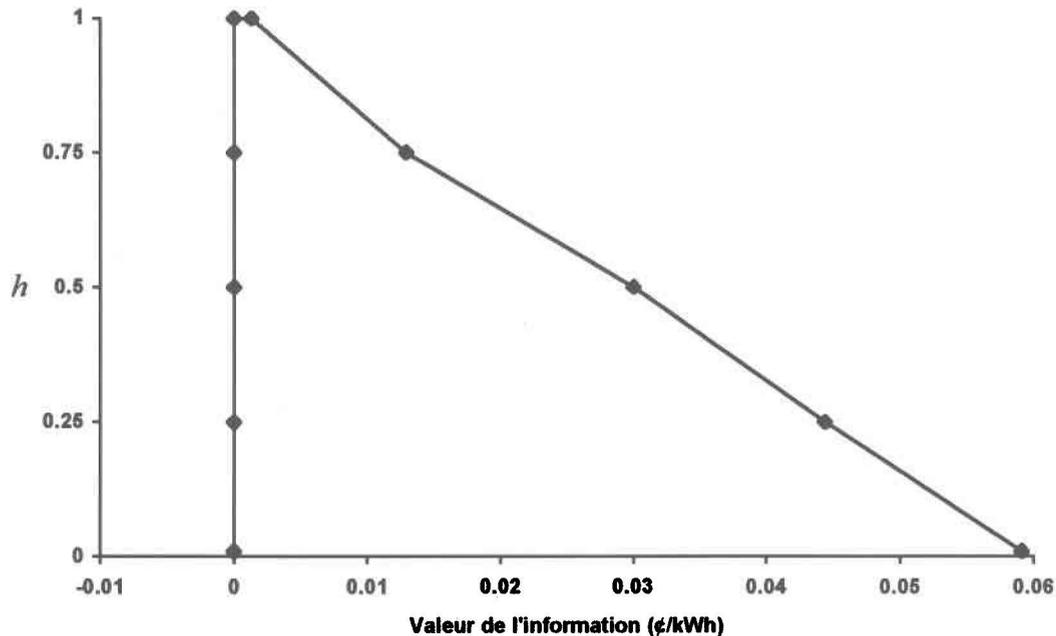


Figure 3.8: Prévion floue de la valeur de l'information hydrologique à SM3 pour $n_0=1$ et $c+p=5.6$ ($m=4$), par exemple pour le projet alternatif NBR avec $p=1$

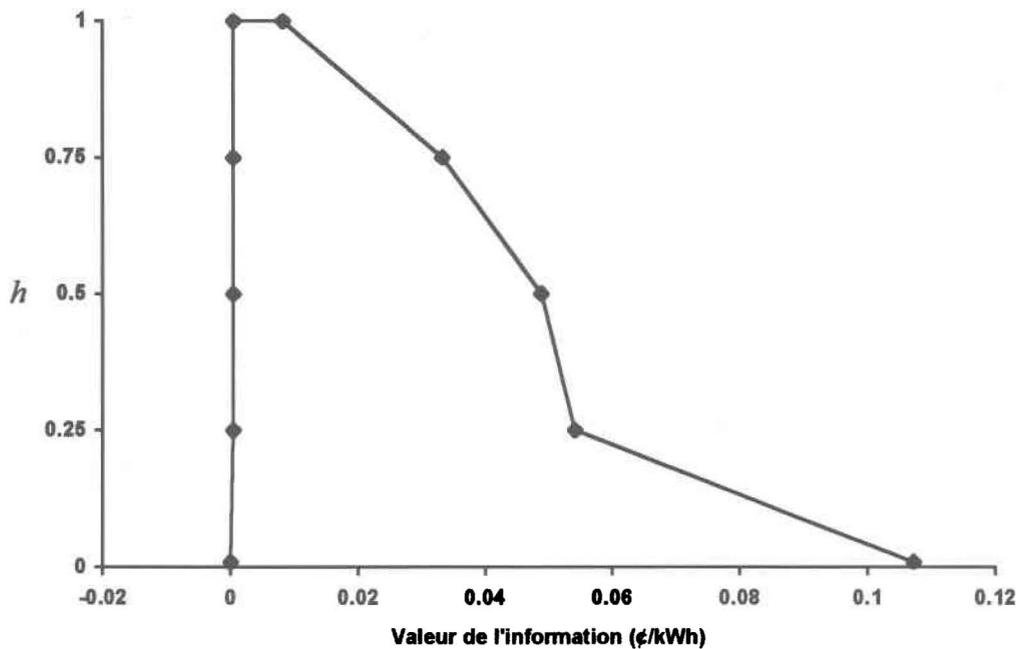


Figure 3.9: Prévision floue de la valeur de l'information hydrologique à SM3 pour $n_0=1$ et $c+p=5.4$ ($m=4$), par exemple pour le projet alternatif Grande-Baleine avec $p=1$

La situation est différente si l'on n'est pas certain de pouvoir effectuer le détournement ($p < 1$). La Figure 3.10 présente une prévision floue de la valeur de l'information dans le cas où cette incertitude est telle que le risque a priori du projet alternatif est équivalent au risque a priori du projet SM3 ($c+p=5.0$). Si l'on compare SM3 et NBR, ceci correspond à $p=0.4$, alors que si l'on compare SM3 et Grande-Baleine, c'est plutôt $p=0.6$. Dans ce cas, la valeur de l'information est nettement plus élevée (≥ 0.1 €/kWh). S'il est impossible de mobiliser plus d'information régionale, il serait prudent dans cette situation d'observer le débit à SM3 avant de choisir le projet hydroélectrique à réaliser.

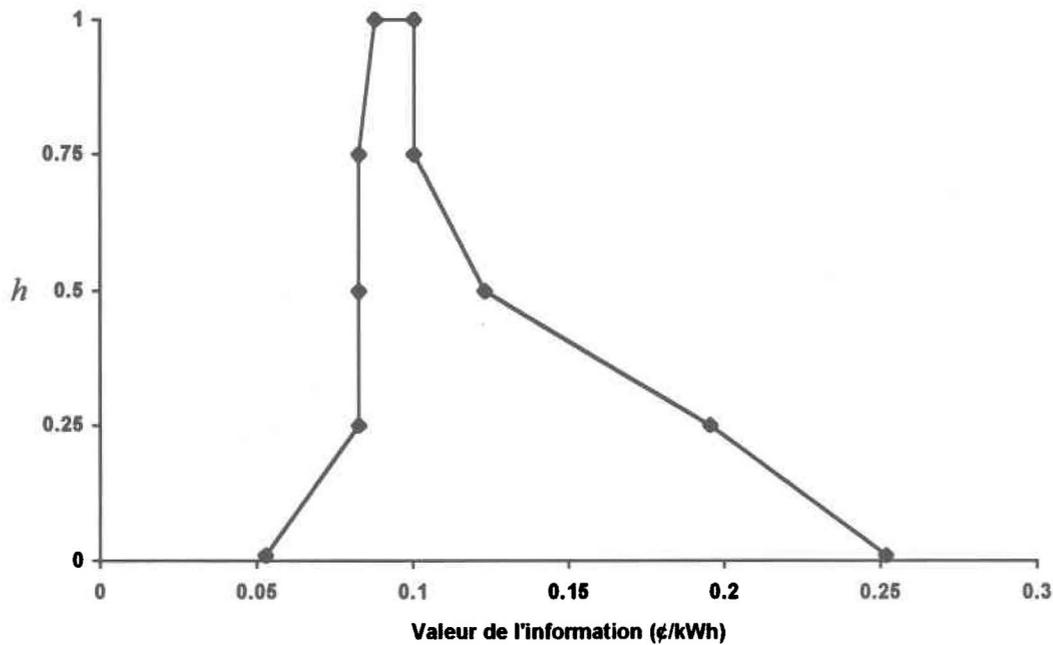


Figure 3.10: Prévision floue de la valeur de l'information hydrologique à SM3 pour $n_0=1$ et $c+p=5.0$ ($m=4$)

Les résultats présentés jusqu'à maintenant montrent que, pour $n_0=1$, la valeur de l'information hydrologique est influencée par l'incertitude politique relative au détournement d'une partie du bassin de la rivière Moisie. Si l'on accorde plus ou moins de poids à l'information a priori, on peut s'attendre à une réponse différente. Par exemple, la Figure 3.11 présente une prévision supérieure de la valeur de l'information obtenue en fixant $h=0$ pour différentes valeurs de n_0 . On voit que la valeur de l'information culmine lorsque $c+p=5.0$, ce qui correspond à la situation où le risque a priori des deux projets à l'étude est similaire. Pour ce niveau d'incertitude sur le coefficient de variation, on voit que la prévision supérieure de la valeur de l'information dépasse toujours 0.1 €/kWh pour $n_0=0.5$. Par contre, lorsque le poids de l'information a priori est plus élevé, cette prévision supérieure dépasse le coût d'acquisition de l'information seulement pour un intervalle de valeurs de $c+p$.

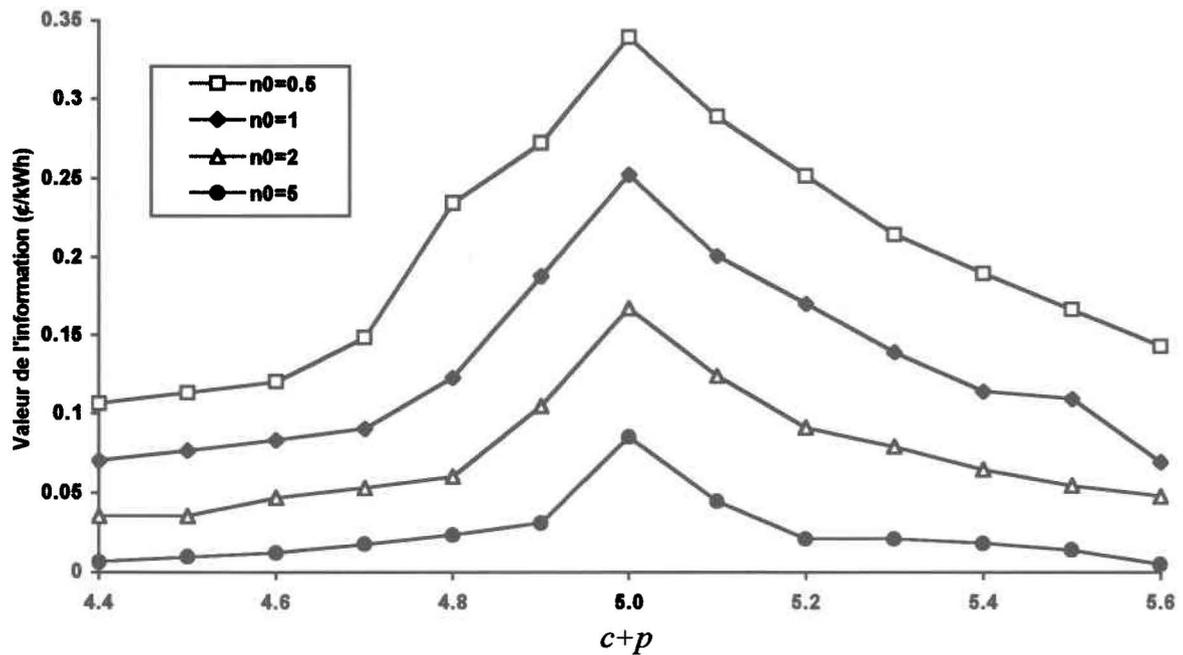


Figure 3.11: Pr evision sup erieure de la valeur de l'information pour $h=0$ et diff erentes valeurs de n_0 , en fonction de $c+p$ ($m=2$)

Les figures suivantes pr esentent une pr evision sup erieure de la valeur de l'information pour $h=0.5$ (Figure 3.12) et $h=1$ (Figure 3.13). On observe qu'une meilleure connaissance du domaine de d efinition du coefficient de variation, refl et ee par une plus grande valeur de h , a pour effet de diminuer la valeur de l'information d' echantillonnage, de sorte que m eme pour $n_0=0.5$ la valeur de l'information peut ˆtre inf erieure   son cˆot d'acquisition.

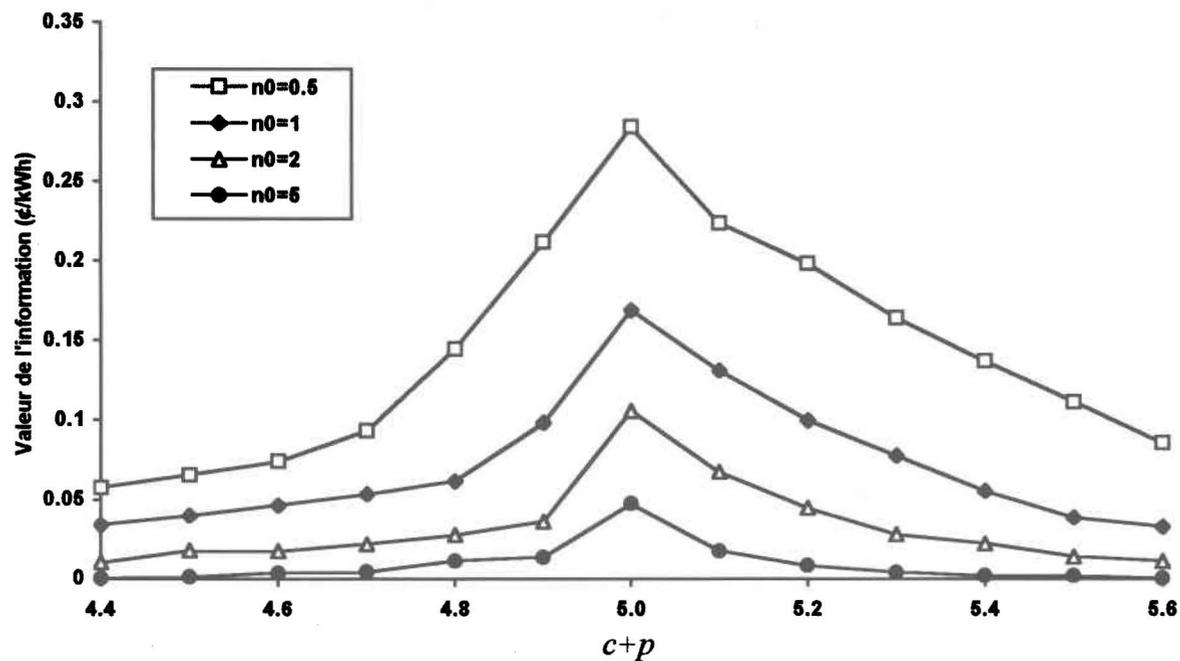


Figure 3.12: Prédiction supérieure de la valeur de l'information pour $h=0.5$ et différentes valeurs de n_0 , en fonction de $c+p$ ($m=2$)

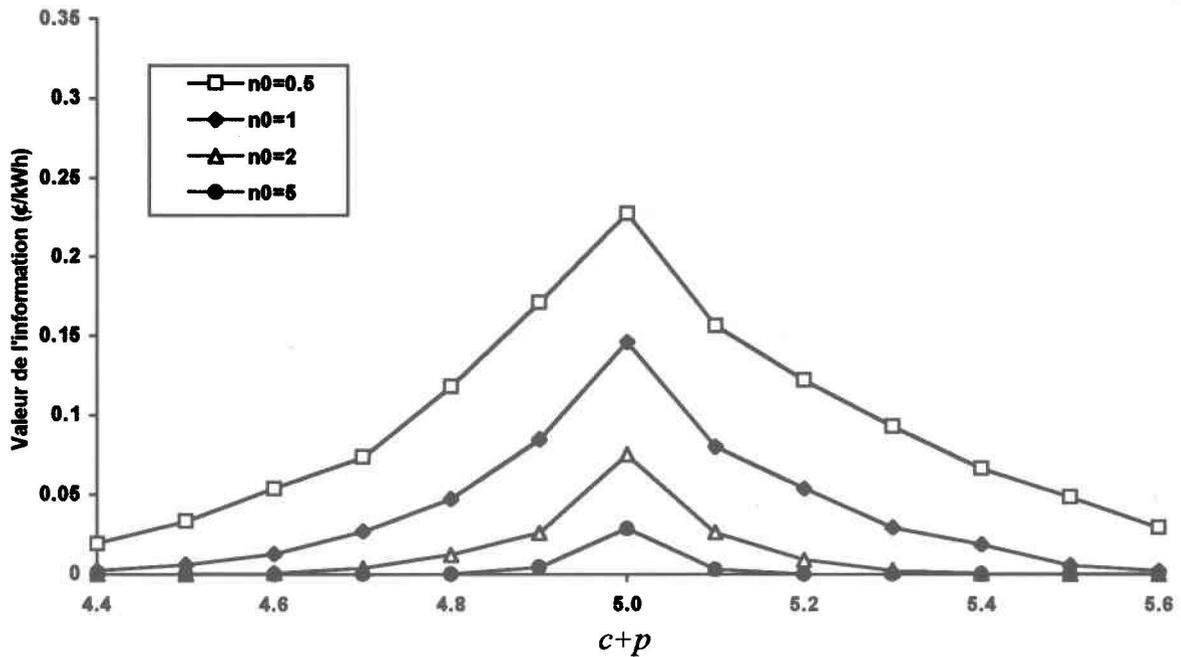


Figure 3.13: Pr evision sup erieure de la valeur de l'information pour $h=1$ et diff erentes valeurs de n_0 , en fonction de $c+p$ ($m=2$)

Il est utile de d eterminer en fonction de h et de n_0 l'intervalle des valeurs de $c+p$ pour lesquelles la pr evision sup erieure de la valeur de l'information d epasse 0.1 €/kWh. On peut repr esenter le r esultat  a l'aide de nombres flous (Figure 3.14). On observe que pour $n_0=2$, la pr evision sup erieure de la valeur de l'information est toujours inf erieure  a 0.1 €/kWh pour $h>0.5$. Lorsque l'on consid ere que l'information a priori a un poids  equivalent  a $n_0=5$ observations du d ebit annuel, alors la valeur de l'information hydrologique est inf erieure  a son c ot  d'acquisition pour toutes les valeurs de h et de $c+p$, de sorte qu'il n'y a pas de courbe correspondant  a $n_0=5$ sur la Figure 3.14.

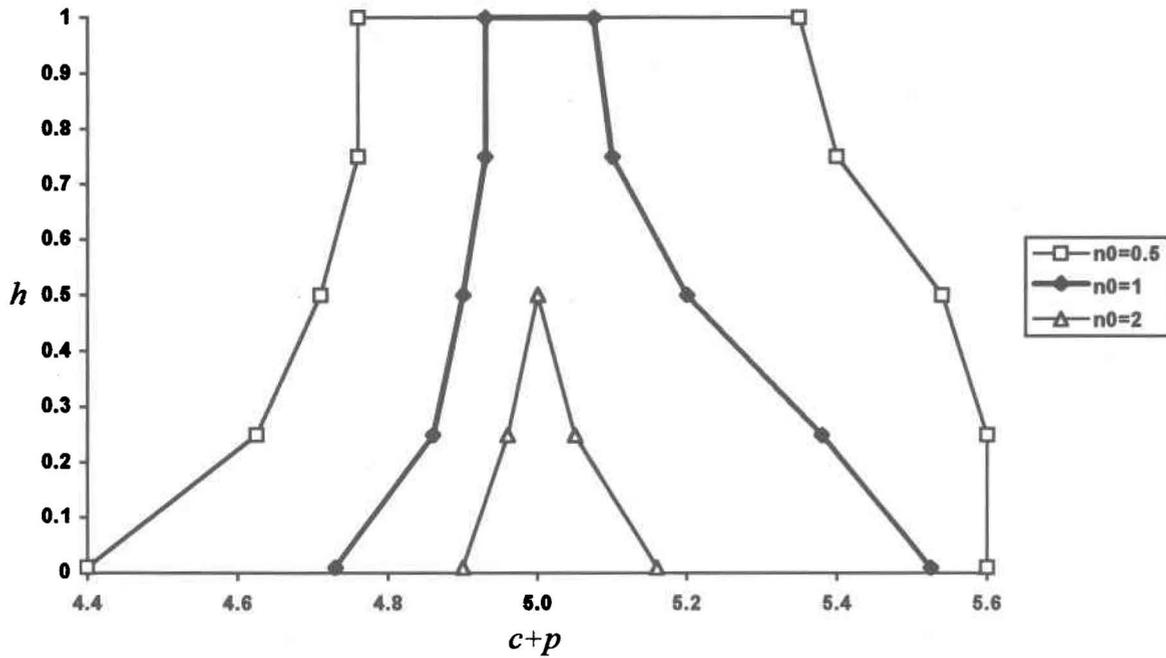


Figure 3.14: Nombres flous présentant les valeurs de $c+p$ pour lesquelles la prévision supérieure de la valeur de l'information hydrologique dépasse 0.1 €/kWh

3.11 Construction d'une prévision supérieure de la valeur de l'information sans l'hypothèse de normalité du débit annuel

L'approche développée à la section précédente pour construire une prévision floue de la valeur de l'information hydrologique à SM3 demande de choisir un modèle d'échantillonnage, de représenter l'incertitude sur les paramètres nuisibles et de résoudre des équations non linéaires complexes. On peut contourner ces difficultés en modélisant directement l'incertitude sur le débit annuel x , ce qui permet de construire une prévision imprécise pour la valeur a posteriori de l'information $VSI(x)$ [eq. (3.20)].

Par définition, le module naturel Q correspond à l'espérance du débit moyen x , i.e. $E^{f(x|Q)}(x) = Q$. Par application du postulat d'indépendance, on obtient $E(x) = E(Q) = 109 \text{ m}^3/\text{s}$. Supposons qu'il soit possible de borner le domaine de définition du débit annuel. On sait que $x \geq 0$, mais il est plus difficile d'établir une borne supérieure. La série d'observations du débit annuel reconstituée à partir des données observées à SM2 a pour maximum $139 \text{ m}^3/\text{s}$, valeur

qui fut atteinte en 1979. Il est évident que le débit annuel pourrait dépasser le maximum observé sur une si courte série. Imaginons que l'hydrologue accepte de fixer une borne supérieure $\sup x \geq 139 \text{ m}^3/\text{s}$ pour le débit annuel.

A partir de ce modèle simple de l'incertitude sur x , on peut construire une prévision supérieure de $VSI(x)$. Des considérations analogues à celles utilisées pour établir la valeur de l'information parfaite permettent de montrer que la mesure de probabilité P qui est compatible avec l'information sur x et qui maximise $E^P[VSI(x)]$ est concentrée en $x=0$ et en $x=\sup x$, avec probabilité $\frac{\sup x - E(Q)}{\sup x}$ en $x=0$, et avec probabilité $\frac{E(Q)}{\sup x}$ en $x=\sup x$. Pour $c+p \geq 5$, on

obtient la prévision supérieure suivante:

$$\begin{aligned} \overline{XVSI} &= E^P[VSI(x)] = \frac{\sup x - E(Q)}{\sup x} \cdot VSI(0) + \frac{E(Q)}{\sup x} \cdot VSI(\sup x) \\ &= \frac{\sup x - 109}{\sup x} \cdot \max\left\{0; \frac{2.4}{n_0 + 1} - c - p + 5.0\right\} \end{aligned} \quad (3.37)$$

Selon la valeur de la borne supérieure du domaine de x (entre 139 et $+\infty$), la prévision supérieure de la valeur de l'information variera entre les bornes suivantes:

$$\frac{30}{139} \cdot \max\left\{0; \frac{2.4}{n_0 + 1} - c - p + 5.0\right\} \leq \overline{XVSI} \leq \max\left\{0; \frac{2.4}{n_0 + 1} - c - p + 5.0\right\} \quad (3.38)$$

Cette façon de résoudre le problème permet de comprendre plus facilement l'influence des paramètres n_0 , c et p sur la valeur de l'information. Par contre, la prévision obtenue est généralement trop imprécise pour permettre de prendre une décision.

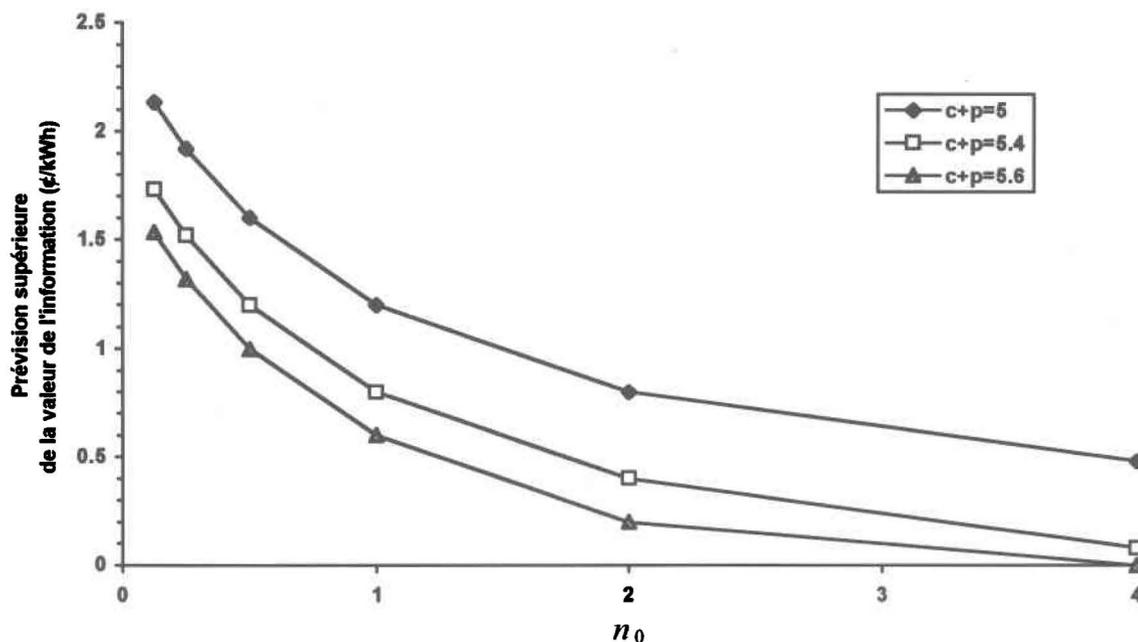


Figure 3.15: Prévision supérieure de la valeur de l'information hydrologique à SM3 en fonction de n_0 et de $c+p$, sans hypothèse sur la distribution d'échantillonnage

Par exemple, la Figure 3.15 présente l'évolution de la prévision supérieure de la valeur de l'information \overline{XVSI} obtenue lorsque $\sup x \rightarrow \infty$ pour différentes valeurs de n_0 et de $c+p$. On observe que \overline{XVSI} est supérieur à 0.1 €/kWh pour une vaste gamme de valeurs de n_0 et de $c+p$.

3.12 Conclusions générales

L'analyse de la décision d'Hydro-Québec d'entreprendre le projet SM3 suggère qu'il aurait peut-être été préférable d'attendre de disposer d'une meilleure information hydrologique. Cependant, cela dépend de plusieurs facteurs:

1. le prix de revient du projet alternatif considéré (c);
2. la probabilité que le Gouvernement du Québec accorde l'autorisation de réaliser le détournement d'une partie du bassin de la rivière Moisie et qu'Hydro-Québec décide par la suite de réaliser le projet SM3 (p);
3. le poids de l'information régionale (n_0);

4. le domaine de définition du coefficient de variation [CV, \overline{CV}].

Même si l'information régionale est considérée au moins aussi utile qu'une seule mesure du débit annuel ($n_0 \geq 1$), la valeur de l'information peut atteindre 0.25 ¢/kWh, ce qui correspond à une économie annuelle de l'ordre de 10 millions de dollars, étant donné la capacité de production à SM3. Cependant, la valeur de l'information hydrologique est surtout importante tant qu'il subsiste une incertitude politique ($c+p \approx 5$). Ainsi, si la décision d'entreprendre immédiatement SM3 a été prise parce que l'on avait une grande confiance d'obtenir les autorisations gouvernementales nécessaires pour réaliser le détournement des eaux de la rivière Moisie ($p \approx 1$), alors il s'agissait vraisemblablement de la bonne décision.

Comme pour plusieurs problèmes d'analyse formelle de la décision, ce ne sont pas les résultats numériques qui sont les plus intéressants, mais la lumière nouvelle qu'apporte une réflexion structurée. Dans ce cas précis, la définition de la fonction de coût a permis de faire ressortir les liens existant entre les incertitudes hydrologiques et politiques. La valeur de l'information hydrologique est en effet influencée de façon importante par le niveau d'incertitude sur la décision que prendra le Gouvernement du Québec concernant le détournement d'une partie du bassin de la rivière Moisie. Ainsi, une étude hydrologique ne permet pas d'arriver à une évaluation éclairée du rendement du projet SM3. Il faut aborder le problème de façon plus globale. Dans ce contexte d'analyse globale, la théorie des prévisions inférieures cohérentes peut être particulièrement utile, puisqu'elle permet d'effectuer une analyse formelle de la décision à partir d'une information incomplète et imprécise. De plus, une fois maîtrisé le langage des prévisions imprécises, la modélisation de l'information disponible est assez facile, et permet en particulier de prendre en compte l'intuition et l'expérience du décideur.

Notre analyse du processus décisionnel entourant le projet SM3 est évidemment incomplète. L'objectif était d'illustrer par un exemple concret et réaliste la méthodologie, et non d'analyser en détail un problème réel. Plusieurs aspects ont été négligés, par exemple les problèmes de dimensionnement des ouvrages: hauteur de la digue, puissance installée, capacité de l'évacuateur de crue, ... Nous n'avons pas non plus tenté de construire un modèle détaillé de l'incertitude politique, représentée par le paramètre p . Pourtant, une année d'attente apporterait certainement aussi une information supplémentaire sur la décision gouvernementale

concernant le bassin de la rivière Moisie. Il se pourrait même qu'il s'agisse d'une information parfaite, si la décision est prise pendant l'année d'attente.

4. DISCUSSION ET CONCLUSION

Nous avons montré dans les chapitres précédents comment la théorie des prévisions inférieures cohérentes pouvait permettre d'évaluer la valeur de l'information hydrologique à partir d'une information a priori imprécise. Ainsi, il n'est pas nécessaire de construire la distribution jointe des variables d'état de la nature pour obtenir une prévision de la valeur de l'information: connaître le domaine de définition et avoir une idée de la tendance centrale peut suffire. Nous proposons un langage plus riche pour représenter l'incertitude, qui généralise le langage probabiliste de la théorie bayésienne, et dont l'utilisation devrait faciliter la modélisation explicite de l'expérience de l'hydrologue.

Pour convaincre les praticiens d'utiliser la théorie des prévisions inférieures cohérentes, il faudra d'abord répondre à un ensemble de questions soulevées dans ce travail:

1. Est-il vraiment nécessaire d'abandonner la théorie bayésienne des probabilités? Une approche bayésienne robuste n'est-elle pas suffisante?
2. S'il faut généraliser la théorie bayésienne, est-ce la meilleure façon de procéder? En particulier, comment se comparent la théorie de Dempster-Shafer, la théorie des possibilités et la théorie des ensembles flous avec la théorie des prévisions inférieures cohérentes?
3. Comment doit-on procéder pour construire des prévisions imprécises représentant fidèlement l'information a priori?
4. Peut-on faire confiance à une résolution numérique du principe d'extension? Peut-on envisager une solution analytique?

Dans le reste de ce chapitre, nous débattons ces questions, pour ensuite proposer un bilan et discuter des perspectives futures de développement et d'application de la théorie proposée.

4.1 Avantages de la théorie des prévisions inférieures cohérentes sur la théorie bayésienne des probabilités

Les fondements axiomatiques de la théorie bayésienne des probabilités et de la théorie des prévisions inférieures cohérentes sont semblables, exception faite du principe de précision. A la fois sur le plan formel et au niveau de l'interprétation, la théorie des prévisions inférieures

cohérentes constitue une généralisation de la théorie bayésienne des probabilités. Un avantage important de la théorie bayésienne des probabilités est que le langage probabiliste s'est imposé dans le domaine scientifique pour la communication des résultats, de sorte qu'un langage différent ne sera adopté que si les gains compensent pour l'effort supplémentaire qu'il faudra faire sur le plan de la communication scientifique.

Le principal avantage de la théorie des prévisions inférieures semble être la facilité avec laquelle peut être modélisée une information a priori subjective. Que ce soit en consultant un expert ou en effectuant une revue de littérature, il est souvent relativement facile de connaître le domaine de définition d'un paramètre et d'effectuer une prévision de sa valeur sous la forme d'un intervalle. Il est beaucoup plus difficile de construire une distribution de probabilité précise qui reflète adéquatement l'information a priori. En hydrologie, nous croyons qu'une méthode d'analyse formelle du risque qui tient compte de l'expérience de l'hydrologue serait la bienvenue. En effet, un hydrologue peut obtenir une connaissance générale d'un bassin versant par l'observation de divers phénomènes hydrologiques, géologiques ou météorologiques (au niveau local comme à une échelle régionale), et on peut gagner à combiner cette expérience aux mesures ponctuelles effectuées sur le terrain, telles que les séries chronologiques de pluie et de débit.

Sur le plan technique, les difficultés de mise en oeuvre de la théorie des prévisions inférieures cohérentes ne dépassent pas celles de l'approche bayésienne robuste, puisque les mêmes outils mathématiques peuvent être utilisés dans les deux cas, selon le théorème de l'enveloppe inférieure. L'approche bayésienne robuste peut même être plus difficile à appliquer, puisqu'il faut s'assurer d'éliminer les distributions a priori qui ne sont pas «raisonnables», par exemple les distributions dégénérées en quelques points de l'espace.

Finalement, rappelons que pour le problème de la prévision de la valeur de l'information, l'hypothèse de base de l'approche bayésienne robuste n'est généralement pas réaliste. En effet, comme nous l'avons mentionné au Chapitre 2, l'approche bayésienne robuste revient à supposer que l'on fera l'effort de fixer une distribution de probabilité précise a priori, et permet d'évaluer la valeur de l'information en fonction des diverses distributions auxquelles on pourrait éventuellement arriver. Ainsi, le problème fondamental de l'indécision est négligé

par l'approche bayésienne robuste, alors qu'il est traité explicitement par la théorie des prévisions inférieures cohérentes.

Evidemment, seule l'application pratique répétée de la théorie des prévisions inférieures cohérentes pourra démontrer qu'il s'agit d'une approche efficace pour la modélisation de l'expérience de l'hydrologue.

4.2 Autres modèles d'incertitude permettant de s'affranchir du principe de précision

Nous avons proposé l'utilisation de la théorie des prévisions inférieures cohérentes pour s'affranchir du principe de précision. Il existe cependant d'autres modèles d'incertitude qui auraient pu être utilisés à cet effet, en particulier la théorie de Dempster-Shafer, la théorie des ensembles flous, et la théorie des possibilités.

4.2.1 Théorie de Dempster-Shafer

La théorie de Dempster-Shafer, issue des travaux de Dempster (1968) et de Shafer (1976), s'intéresse à une classe particulière de probabilités inférieures cohérentes, les capacités de Choquet complètement monotones (Choquet, 1953-54). Ces probabilités inférieures particulières ont l'avantage de pouvoir être générées par une fonction $m: 2^\Theta \rightarrow [0,1]$, où 2^Θ représente l'ensemble des parties du référentiel Θ . Pour simplifier la présentation de cette théorie, nous supposerons dans un premier temps que Θ est un ensemble fini.

4.2.1.1 Modélisation de l'incertitude

Le problème d'une modélisation probabiliste de l'incertitude est la difficulté de répartir entre les singletons θ du référentiel Θ les poids de probabilité de façon à représenter le plus fidèlement possible l'information disponible. En pratique, il arrive que l'on sache seulement qu'un poids $m(A)$ devrait être réparti entre les éléments d'un sous-ensemble A du référentiel. L'idée de base de la théorie de Dempster-Shafer consiste à attribuer aux sous-ensembles, plutôt qu'aux singletons, les poids de probabilité. Ceci permet entre autres de modéliser une ignorance totale en attribuant tout le poids de probabilité au référentiel sans spécifier comment il doit être distribué entre les éléments.

De façon formelle on doit, pour appliquer la théorie de Dempster-Shafer, définir une fonction $m: 2^\Theta \rightarrow [0,1]$ telle que:

$$\sum_{A \in 2^\Theta} m(A) = 1, \quad 0 = m(\emptyset) \leq m(A) \quad (4.1)$$

A partir de cette fonction, que l'on peut nommer la distribution des poids de probabilités, on obtient la probabilité inférieure d'un événement A en faisant la somme de tous les poids de probabilités qui sont attribués aux sous-ensembles de A . En effet, quelle que soit la façon dont on distribue le poids $m(B)$ d'un sous-ensemble $B \subseteq A$ entre les éléments $\theta \in B$, ils contribueront de la même façon à la probabilité totale de l'événement A :

$$\underline{P}(A) = \sum_{B \subseteq A} m(B) \quad (4.2)$$

De la même façon, la probabilité supérieure $\bar{P}(A)$ d'un événement A est obtenue en additionnant tous les poids de probabilité associés à des événements ayant une intersection non vide avec A . En effet, le poids de probabilité $m(B)$ d'un événement B tel que $A \cap B \neq \emptyset$ pourrait être entièrement attribué aux éléments de $A \cap B \neq \emptyset$:

$$\bar{P}(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (4.3)$$

On peut montrer que ces probabilités imprécises respectent la règle $\underline{P}(A) + \bar{P}(A^c) = 1$, et que les probabilités inférieures sont des capacités complètement monotones (Shafer, 1976). Rappelons qu'une fonction \underline{P} est dite k -monotone si elle respecte la condition suivante pour toute suite d'événements $(A_n, n=1,2,\dots,k)$, et qu'une fonction \underline{P} est complètement monotone si elle vérifie cette condition pour tout entier $k > 1$ (Choquet, 1953-54):

$$\underline{P}\left(\bigcup_{n=1}^k A_n\right) \geq \sum_{\emptyset \neq W \subseteq \{1,2,\dots,k\}} (-1)^{|W|+1} \underline{P}\left(\bigcap_{n \in W} A_n\right) \quad (4.4)$$

Les capacités complètement monotones peuvent permettre de modéliser correctement l'incertitude dans un grand nombre de situations. D'ailleurs, Shafer (1976) présente plusieurs problèmes pour lesquels il est naturel de distribuer les poids de probabilité entre des sous-ensembles plutôt qu'entre des singletons. Walley (1991) propose par contre quelques exemples où cette représentation n'est pas assez générale. Il montre aussi que les capacités k -monotones ($k > 1$) sont des prévisions inférieures cohérentes. Il est donc possible d'utiliser à l'intérieur de la

théorie des prévisions inférieures cohérentes un modèle d'incertitude fondé sur une distribution de poids de probabilité $m: 2^{\Theta} \rightarrow [0,1]$. Le traitement des capacités complètement monotone est d'ailleurs assez facile dans le cadre de la théorie des prévisions inférieures cohérentes, puisque les simplifications présentées pour les fonctions 2-monotones s'appliquent.

4.2.1.2 Eléments focaux, support et noyau

Les éléments focaux $\Phi(m)$ d'une distribution de poids de probabilité m sont les sous-ensembles $A \subseteq \Theta$ tels que $m(A) > 0$:

$$\Phi(m) = \{A \in 2^{\Theta} : m(A) > 0\} \quad (4.5)$$

Le support d'une variable aléatoire θ est le sous-ensemble $\Lambda(m)$ de tous les éléments de Θ faisant partie d'au moins un élément focal, c'est-à-dire tous les éléments θ pour lesquels la probabilité inférieure du singleton correspondant $\{\theta\}$ est strictement positive:

$$\Lambda(m) = \{\theta \in \Theta : \underline{P}(\{\theta\}) > 0\} \quad (4.6)$$

Lorsque tous les éléments focaux sont des singletons, les poids de probabilités ne sont alors plus mobiles de sorte que $\underline{P}(A) = \bar{P}(A) = \sum_{\theta \in A} m(\{\theta\})$. La fonction $f(\theta) = m(\{\theta\})$ correspond alors à une distribution de probabilité au sens bayésien; on qualifie donc par extension m de distribution bayésienne de poids de probabilité. Lorsque tous les éléments focaux sont imbriqués, i.e. on peut former une suite croissante d'ensemble à partir de ceux-ci, on dira que m est consonante. Finalement, dans le cas où $m(\Theta) = 1$, on dira que m est ignorante, puisqu'il s'agit d'un modèle approprié pour l'ignorance a priori.

4.2.1.3 Combinaison de l'incertitude à l'aide de la règle de Dempster

Shafer (1976) présente une règle, qu'il nomme la règle de Dempster, pour combiner deux distributions de poids de probabilité m_1 et m_2 obtenues de façon indépendante sur une même variable aléatoire θ . Avant de présenter cette règle, définissons le niveau de discordance entre m_1 et m_2 . Nous dirons que les deux mesures d'incertitude m_1 et m_2 se contredisent partiellement s'il existe des ensembles disjoints A et B tels que $m_1(A) \cdot m_2(B) \neq 0$. Si l'on note \underline{P}_1 et \underline{P}_2 les probabilités inférieures engendrées respectivement par m_1 et m_2 , une discordance partielle signifie en fait qu'il n'existe pas de mesure de probabilité précise P qui domine à la

fois P_1 et P_2 . D'un point de vue d'analyse de sensibilité, la moindre discordance de cette nature rend impossible la combinaison de m_1 et m_2 . On peut mesurer le niveau de discordance total $0 \leq \kappa \leq 1$ en faisant la somme des produits $m_1(A) \cdot m_2(B)$ pour tous les ensembles disjoints A et B :

$$\kappa = \sum_{A \cap B = \emptyset} m_1(A) \cdot m_2(B) \quad (4.7)$$

On remarque qu'un niveau maximal de discordance $\kappa=1$ est atteint lorsque les supports $\Lambda(m_1)$ et $\Lambda(m_2)$ sont disjoints, et qu'un niveau minimal de contradiction $\kappa=0$ est obtenu entre autres lorsque $m_1(\Theta)=1$ ou $m_2(\Theta)=1$.

De façon analogue, on peut construire sur un sous-ensemble C une mesure de concordance $0 \leq \eta(C) \leq 1$ entre m_1 et m_2 en considérant la somme des produits $m_1(A) \cdot m_2(B)$ pour tous les ensembles A et B ayant pour réunion C :

$$\eta(C) = \sum_{A \cup B = C} m_1(A) \cdot m_2(B) \quad (4.8)$$

Remarquons que si $m_1(\Theta)=1$, alors $\eta(C)=m_2(C)$. Si m_1 est bayésienne, on obtient $\eta(C)=0$ sauf si C est un singleton. Shafer (1976) montre que $\eta(C)$ est une distribution de poids de probabilité s'il n'y a pas de discordance entre m_1 et m_2 , i.e. si $\kappa=0$. Autrement, il est possible de normaliser $\eta(C)$ pour obtenir une distribution de poids de probabilité m , qui constitue alors une façon de combiner les fonctions m_1 et m_2 :

$$m(C) = \frac{\eta(C)}{1 - \kappa} = \frac{\sum_{A \cup B = C} m_1(A) \cdot m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A) \cdot m_2(B)} \quad (4.9)$$

On utilisera l'opérateur \otimes pour représenter l'utilisation de la règle de Dempster: $m = m_1 \otimes m_2$. On remarque que cette formule n'est pas applicable si la discordance est totale entre m_1 et m_2 , i.e. $\kappa=1$. Si $m_1(\Theta)=1$, on obtient $m(C)=m_2(C)$, et si m_1 est bayésienne alors m est aussi bayésienne quel que soit m_2 . La règle de Dempster, bien qu'elle possède ces propriétés attrayantes, n'en garde pas moins un parfum empirique.

Shafer (1976) montre que la règle de Dempster généralise d'une certaine façon le théorème de Bayes. Considérons que m_1 modélise l'information a priori sur θ , et qu'on observe un événement H . On peut modéliser cette observation par une distribution de poids de probabilité $m_2(H)=1$. En combinant m_1 et m_2 à l'aide de la règle de Dempster, on obtient une

nouvelle distribution de poids de probabilité qui correspond à la mise à jour de m_1 après observation de H , que l'on peut noter $m_1(\cdot|H)$. Si l'on note $\bar{P}(\cdot)$ la probabilité supérieure générée par $m_1(\cdot)$ et $\bar{P}(\cdot|H)$ la probabilité supérieure générée par $m_1(\cdot|H)$, on peut alors démontrer à partir de la règle de Dempster la formule suivante pour obtenir une probabilité supérieure conditionnelle:

$$\bar{P}(A|H) = \frac{\bar{P}(A \cap H)}{\bar{P}(H)} \quad (4.10)$$

Démonstration: Pour $m_2(H)=1$, on calcule κ et $\eta(C)$:

$$\kappa = \sum_{A \cap H = \emptyset} m_1(A) = 1 - \sum_{A \cap H \neq \emptyset} m_1(A) = 1 - \bar{P}(H), \quad \eta(C) = \sum_{A \cap H = C} m_1(A)$$

On peut alors obtenir $\bar{P}(A|H)$:

$$\bar{P}(A|H) = \sum_{A \cap B \neq \emptyset} m_1(B|H) = \frac{\sum_{A \cap B \neq \emptyset} \eta(B)}{1 - \kappa} = \frac{\sum_{A \cap B \neq \emptyset} \sum_{D \cap H = B} m_1(D)}{\bar{P}(H)} = \frac{\sum_{(A \cap H) \cap D} m_1(D)}{\bar{P}(H)} = \frac{\bar{P}(A \cap H)}{\bar{P}(H)} \quad \blacksquare$$

Évidemment, il s'agit d'une généralisation du théorème de Bayes, qu'on obtient comme cas limite lorsque $\underline{P}(A) = \bar{P}(A)$. Par contre, on peut observer la différence avec la formule obtenue à l'aide de la théorie des prévisions inférieures pour les probabilités supérieures 2-monotones [eq. (2.20)].

4.2.1.4 Extension de la théorie de Dempster-Shafer aux domaines infinis

L'extension de la théorie de Dempster-Shafer aux domaines infinis pose de véritables problèmes. En effet, les opérations de sommation sur la fonction m nécessitent la construction d'un espace mesuré approprié. En pratique, on doit simplifier cette tâche en supposant que les éléments focaux ont une structure topologique très simple. Lorsque θ est une variable réelle, Caselton et Luo (1992) suggèrent de limiter les éléments focaux aux intervalles. Ainsi, une distribution de poids de probabilité peut être représentée par une fonction bivariable $m(a,b) \geq 0$ dont les arguments spécifient les bornes de l'intervalle $[a,b]$, et définie telle que $\int \int m(a,b) db \cdot da = 1$.

4.2.1.5 Inférence statistique

On peut proposer une méthode d'inférence statistique basée sur la règle de Dempster. Si l'on modélise l'information a priori sur θ par une distribution de poids de probabilité m et si l'on construit une seconde distribution de poids de probabilité m_x à partir d'une réalisation x de vraisemblance $f(x|\theta)$, alors on peut combiner ces deux distributions à l'aide de la règle de Dempster pour obtenir une distribution de poids de probabilité $m(\cdot|x)$ conditionnelle à l'observation x (Shafer, 1986). Il reste cependant à déterminer comment doit être construite m_x à partir de $f(x|\theta)$. Wasserman (1990) propose trois critères pour la construction de m_x :

1. **analogie avec la règle de Bayes:** si m est bayésienne, alors $m(\cdot|x)$ doit être bayésienne et correspondre à la distribution a posteriori obtenue par application de la règle de Bayes;
2. **respect des préférences induites par la vraisemblance:** la probabilité inférieure $\underline{P}_x(\{\theta\})$ associée à tout singleton par m_x ne doit pas décroître si la vraisemblance ne décroît pas: $f(x|\theta_1) \geq f(x|\theta_2) \Rightarrow \underline{P}_x(\{\theta_1\}) \geq \underline{P}_x(\{\theta_2\})$;
3. **principe de vraisemblance généralisé:** toute l'information est contenue dans la vraisemblance, de sorte que si $f(x|\theta_1) = f(x|\theta_2)$, alors $m_x(\{\theta_1, \theta_2\}|\{\theta_1, \theta_2\}) = 1$, i.e. sachant que la variable aléatoire vaut θ_1 ou θ_2 et obtenant la même vraisemblance pour ces deux valeurs, il ne reste aucune information pour distinguer θ_1 et θ_2 , d'où le choix d'une distribution conditionnelle de poids de probabilité qui soit ignorante.

Wasserman (1990) montre que seul le choix d'une distribution consonante des poids de probabilité permet de respecter ces trois critères. Cette distribution consonante doit de plus être choisie de façon que la probabilité supérieure des singletons soit proportionnelle à la vraisemblance:

$$\bar{P}_x(\{\theta\}) = \frac{f(x|\theta)}{\sup_{\theta} f(x|\theta)} \quad (4.11)$$

En pratique, cette méthode d'inférence est très difficile à appliquer lorsque θ est composé d'un vecteur de paramètres (Luo et Caselton, 1997), et d'ailleurs à notre connaissance il n'existe aucune application de la théorie de Dempster-Shafer à des distributions ayant plusieurs paramètres indépendants. Caselton et Luo (1992) présentent une application de

la théorie de Dempster-Shafer pour une distribution binomiale, et Luo (1993) présente des résultats pour les distributions normale et Gumbel avec un des deux paramètres de fixé.

En plus d'être limitée aux distributions n'ayant qu'un seul paramètre, nous croyons que la théorie de Dempster-Shafer apporte une bien maigre contribution au problème de l'inférence statistique. En effet, dans le cas où l'approche fonctionne le mieux, c'est-à-dire pour l'estimation du paramètre d'une loi binomiale, on peut proposer une solution équivalente, beaucoup plus simple, et compatible à la fois avec l'approche bayésienne robuste et avec la théorie des prévisions inférieures.

Considérons en effet la prévision imprécise proposée par Caselton et Luo (1992) pour le paramètre θ d'une distribution binomiale:

$$\underline{E}(\theta|\mathbf{x}) = \frac{r}{n+1}, \quad \bar{E}(\theta|\mathbf{x}) = \frac{r+1}{n+1} \quad (4.12)$$

où \mathbf{x} est un échantillon aléatoire de taille n contenant r succès et $n-r$ échecs. Dans le cadre de la théorie de Dempster-Shafer, il faut à Caselton et Luo (1992) une dizaine de pages de théorie pour obtenir ce résultat.

Si l'on utilise plutôt une distribution bêta(α, β) pour modéliser l'information a priori et que l'on applique la règle de Bayes, on obtient les expressions suivantes pour l'espérance a priori et l'espérance a posteriori:

$$E(\theta) = \frac{\alpha}{\alpha + \beta}, \quad E(\theta|\mathbf{x}) = \frac{r + \alpha}{n + \alpha + \beta} \quad (4.13)$$

Supposons que l'information a priori permette seulement d'établir que $\alpha + \beta = 1$. Ceci revient à dire qu'on a une idée du poids que devrait avoir l'information a priori dans l'analyse, sans avoir d'idée sur l'espérance a priori. Dans ce cas, on peut effectuer une analyse de sensibilité en faisant varier $E(\theta)$ dans l'intervalle $[0,1]$. On voit que ceci permet d'obtenir exactement les mêmes bornes que l'analyse basée sur la théorie de Dempster-Shafer. L'interprétation de la prévision imprécise obtenue pour θ est aussi beaucoup plus claire. Finalement, on peut facilement modifier la contrainte sur la quantité $\alpha + \beta$ et ainsi faire varier le poids de l'information a priori dans l'analyse. Cette façon de traiter l'inférence pour une loi binomiale se généralise facilement à la loi multinomiale (Walley, 1996a), contrairement à la théorie de Dempster-Shafer.

4.2.2 Théorie des ensembles flous

La théorie des sous-ensembles flous a été introduite par Zadeh (1965) comme moyen de formaliser les ambiguïtés langagières. On sait que la signification d'une proposition exprimée en langage courant est souvent ambiguë. En particulier, les limites d'un ensemble spécifié à l'aide du langage usuel sont souvent floues.

Par exemple, si l'on définit un écoulement laminaire à l'aide du langage écrit en affirmant qu'il s'agit d'un écoulement pour lequel les vitesses transversales sont négligeables, alors il sera difficile dans certaines situations de déterminer si un écoulement est laminaire ou turbulent. Zadeh (1965) propose de modéliser ce type d'ambiguïté en quantifiant le degré d'appartenance d'un écoulement particulier à la famille des écoulements laminaires et à la famille des écoulements turbulents. Si x dénote un écoulement particulier, on pourra représenter par $\mu_L(x)$ le degré d'appartenance de x à l'ensemble L des écoulements laminaires et par $\mu_T(x)$ le degré d'appartenance de x à l'ensemble T des écoulements turbulents. Ces nombres sont censés mesurer à quel point l'élément x appartient aux ensembles L et T , dont les limites sont floues.

Pour distinguer les ensembles usuels (non flous) des ensembles flous, on utilisera le symbole \sim . Par exemple, \tilde{A} désigne un ensemble flou. Un ensemble flou \tilde{A} est défini par une fonction d'appartenance $\mu_{\tilde{A}}(x)$ qui spécifie le degré d'appartenance de chaque élément x d'un référentiel Ω cet ensemble. Par convention, on pose $\mu_{\tilde{A}}(x)=1$ pour signifier qu'un élément x appartient de façon certaine à un ensemble flou \tilde{A} , et $\mu_{\tilde{A}}(x)=0$ pour signifier qu'il n'appartient pas à cet ensemble. Si la fonction $\mu_{\tilde{A}}(x)$ ne prend que les valeurs 0 ou 1, alors l'ensemble \tilde{A} est équivalent à un ensemble non flou, et $\mu_{\tilde{A}}(x)$ correspond à la fonction indicatrice de cet ensemble. L'appartenance partielle d'un élément x à un ensemble flou \tilde{A} est modélisée par un nombre $\mu_{\tilde{A}}(x)$ entre 0 et 1.

Lorsque le référentiel x correspond à la droite réelle, et que $\mu_{\tilde{A}}(x)$ comporte une branche croissante, de 0 à 1, et une branche décroissante, de 1 à 0, on qualifie \tilde{A} de nombre flou. Nous avons proposé une utilisation du concept de nombre flou pour effectuer une analyse

de sensibilité sur l'effet de l'imprécision d'une prévision. En dehors de ce contexte particulier, la signification concrète d'un degré d'appartenance $0 < \mu_{\tilde{A}}(x) < 1$ est difficile à saisir.

Plusieurs auteurs, comme Kaufmann (1977), accordent peu d'importance à l'interprétation opérationnelle et à la construction pratique de fonctions d'appartenance, et s'attardent plus sur les propriétés mathématiques des ensembles flous. Cet état de chose est soulevé par de nombreuses critiques de la théorie des ensembles flous. Par exemple, French (1984) fait remarquer qu'il peut être plus utile d'éliminer les ambiguïtés que de les modéliser, en adoptant des définitions plus précises pour les concepts que l'on manipule. Pour l'exemple de caractérisation de l'écoulement que nous avons présenté, c'est l'approche qui est retenue en pratique: au lieu de modéliser le niveau d'appartenance partiel d'un écoulement à l'ensemble L , on mesure son nombre de Reynolds Re . Cette quantité est plus facilement interprétable et permet de comparer différents écoulements.

En plus de permettre la modélisation des ambiguïtés liées au langage, Zadeh (1978) suggère qu'une fonction d'appartenance $\mu(\theta)$ peut aussi être utilisée pour quantifier l'incertitude sur un paramètre inconnu θ . Cette idée a été développée par Dubois et Prade (1988) qui proposent une théorie des possibilités.

4.2.3 Théorie des possibilités

De la même façon qu'une distribution de probabilité $b(\theta)$ permet de générer une mesure de probabilité P , certaines fonctions d'appartenance $\mu(\theta)$ peuvent permettre de générer une mesure de probabilité supérieure \bar{P} . Dubois et Prade (1988) montrent que si $\sup \mu(\theta) = 1$, alors $\bar{P}(A) = \sup_{\theta \in A} \mu(\theta)$ et $\underline{P}(A) = 1 - \bar{P}(A^c)$ sont des probabilités supérieures cohérentes au sens de

Walley (1991). Dans ce cas, on nomme $\mu(\theta)$ une distribution de possibilité. On remarque que la probabilité supérieure d'un singleton est donnée par son niveau d'appartenance: $\bar{P}(\{\theta\}) = \mu(\theta)$. Puisque $\sup \mu(\theta) = 1$, on comprend aussi que pour tout événement A soit $\bar{P}(A) = 1$, soit $\underline{P}(A) = 0$.

Il est possible que l'information a priori puisse être adéquatement modélisée par une distribution de possibilité, mais on ne peut pas espérer que ce soit le cas de façon générale. On peut en effet construire des exemples très simples pour lesquels l'ensemble des distributions de

possibilité n'est pas assez riche pour modéliser l'information disponible. Par exemple, considérons une urne dans laquelle se trouve un nombre fini de balles, dont au moins une noire et une blanche, dans laquelle on tire un balle au hasard. A partir de ces informations, on peut au moins affirmer que $\underline{P}(A) > 0$ et que $\overline{P}(A) < 1$, où A correspond à l'observation d'une balle noire. Ces contraintes ne peuvent pas être modélisées à l'aide d'une distribution de possibilité, puisque l'on doit avoir soit $\overline{P}(A) = 1$, soit $\underline{P}(A) = 0$ pour toute probabilité imprécise générée à partir d'une distribution de possibilité.

Il serait possible de définir le concept de valeur de l'information dans le cadre de la théorie de Dempster-Shafer et de la théorie des possibilités. Cependant, ces deux théories ne nous ont pas semblé pouvoir apporter plus que la théorie des prévisions inférieures cohérentes, puisque les modèles d'incertitude qu'elles proposent correspondent à des prévisions imprécises ayant des structures particulières. Il ne nous paraît pas utile de nous restreindre aux prévisions inférieures cohérentes engendrées par des distributions de poids de probabilité, ou par des distributions de possibilité. Par contre, la combinaison des concepts de prévision imprécise et de nombre flou nous ont permis d'intégrer une analyse de sensibilité au degré d'imprécision des prévisions effectuées a priori dans le calcul d'une prévision imprécise de la valeur de l'information.

4.3 Influence de la représentation de l'information a priori

Nous avons insisté tout au long de ce travail sur la facilité avec laquelle peut être modélisée une information a priori subjective à l'aide de prévisions imprécises. En particulier, nous avons montré qu'il est possible d'étudier le risque et de mesurer la valeur de l'information parfaite lorsque l'information a priori est constituée uniquement d'une prévision imprécise $[\underline{E}(\theta), \overline{E}(\theta)]$ de l'état de la nature θ (connaissant aussi son domaine de variation). Nous croyons que cette approche est intéressante, car un expert est souvent capable de construire un intervalle $[a, b]$ de valeurs de θ qu'il considère vraisemblables, comme le montrent les exemples, tirés de la *Revue des Sciences de l'Eau*, présentés au début du Chapitre 2.

On ne peut cependant modéliser de cette façon tout intervalle de valeurs que peut fournir un expert. Il faut s'assurer de comprendre la signification qu'il donne à cet intervalle. Il peut, dans certains cas, s'agir du domaine de définition de la variable. Si l'individu a une

culture statistique, il peut s'agir d'un intervalle de crédibilité dont il faudra déterminer le seuil. L'expert peut aussi vouloir signifier simplement qu'il préférerait parier sur l'événement $\theta \in [a, b]$ que sur son contraire. Ceci pourrait être modélisé par la probabilité inférieure $\underline{P}(\theta \in [a, b]) = 1/2$.

Le même genre de difficulté se produit lorsque l'expert ne fournit qu'un seul nombre μ pour résumer son information a priori sur un paramètre. On peut utiliser cette information pour fixer une prévision précise $E(\theta) = \mu$, mais on pourrait aussi utiliser cette idée de la tendance centrale pour fixer la médiane plutôt que la moyenne: $P(\theta \leq \mu) = 1/2$. Dans les exemples que nous avons présentés, nous avons préféré utiliser les estimations ponctuelles des paramètres pour fixer la moyenne plutôt que la médiane pour deux raisons pratiques. D'abord, les fonctions d'utilité considérées étaient linéaires, de sorte que le risque a priori des actions pouvait être évalué précisément si l'espérance de l'état de la nature était connu. De plus, cela simplifiait la résolution numérique du principe d'extension des prévisions inférieures. En effet, fixer la probabilité de certains événements introduit des discontinuités dans la fonction qu'il faut optimiser pour appliquer le principe d'extension. Finalement, fixer l'espérance plutôt que la médiane conduit généralement à des prévisions plus précises. Illustrons cela par un exemple.

Soit θ le paramètre d'une loi binomiale et soit μ une estimation fournie par un expert. On cherche à construire une probabilité imprécise pour l'observation de deux succès consécutifs, événement dont la vraisemblance est donnée par θ^2 . Si l'on modélise l'information a priori par la contrainte $E(\theta) = \mu$, on peut obtenir une prévision imprécise du pari θ^2 en appliquant le principe d'extension des prévisions inférieures, ce qui demande de résoudre les équations suivantes:

$$\begin{aligned} \underline{E}_1(\theta^2) &= \sup_{\lambda} \inf_{\theta} \{ \theta^2 - \lambda(\theta - \mu) \} \\ \bar{E}_1(\theta^2) &= \inf_{\lambda} \sup_{\theta} \{ \theta^2 - \lambda(\theta - \mu) \} \end{aligned} \quad (4.14)$$

Quelques manipulations algébriques permettent de montrer que $\underline{E}_1(\theta^2) = \mu^2$ et que $\bar{E}_1(\theta^2) = \mu$. Si l'on modélise plutôt l'information a priori par la contrainte $P(\theta \leq \mu) = 1/2$, les équations à résoudre sont différentes:

$$\begin{aligned} \underline{E}_2(\theta^2) &= \sup_{\lambda} \inf_{\theta} \{ \theta^2 - \lambda(I_{[0,\mu]}(\theta) - 1/2) \} \\ \bar{E}_2(\theta^2) &= \inf_{\lambda} \sup_{\theta} \{ \theta^2 - \lambda(I_{[0,\mu]}(\theta) - 1/2) \} \end{aligned} \quad (4.15)$$

où $I_{[0,\mu]}(\theta)$ représente la fonction indicatrice de l'événement $[0,\mu]$. C'est cette fonction qui introduit une discontinuité et qui rend plus difficile l'application de méthodes numériques. Dans le cas présent, une solution analytique existe cependant: $\underline{E}_2(\theta^2) = \mu^2/2$ et $\bar{E}_2(\theta^2) = (\mu^2 + 1)/2$. La prévision imprécise de deux succès consécutifs est donc influencée par la façon dont est modélisée l'information a priori.

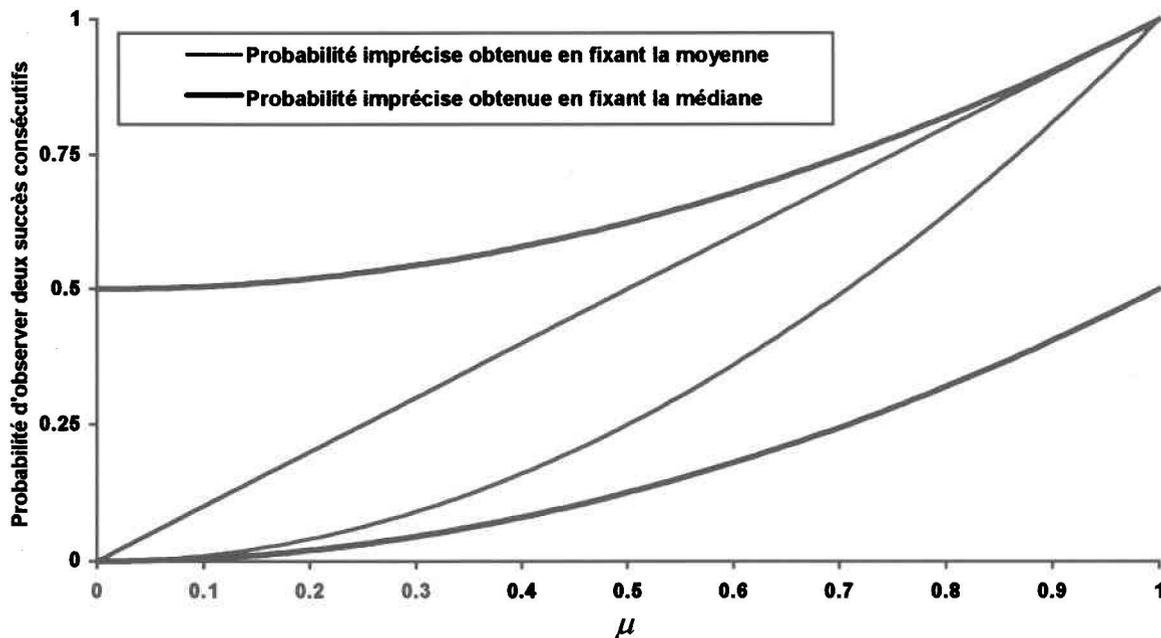


Figure 4.1: Probabilité imprécise d'observer deux succès consécutifs en fonction d'une estimation ponctuelle μ du paramètre de la loi binomiale

La Figure 4.1 permet de visualiser la différence entre les deux probabilités imprécises d'observer deux succès consécutifs. On observe que la probabilité imprécise obtenue en fixant la moyenne est toujours contenue par la probabilité imprécise obtenue en fixant la médiane, i.e. $[\underline{E}_1(\theta^2), \bar{E}_1(\theta^2)] \subseteq [\underline{E}_2(\theta^2), \bar{E}_2(\theta^2)]$. Une façon de savoir comment représenter l'information a priori est de discuter avec l'expert et de lui expliquer le sens d'une prévision imprécise.

Berger (1985) fait remarquer qu'il est difficile de fixer à l'aide d'une information subjective les moments d'une distribution a priori lorsque le domaine de définition est infini, à cause de l'influence du poids des queues sur la valeur des moments. Dans ces situations, il serait préférable de modéliser l'information a priori en fixant des probabilités imprécises pour certains événements, plutôt qu'en effectuant des prévisions imprécises pour des paris de la forme θ^m .

4.4 Application pratique du principe d'extension des prévisions inférieures et de la règle de Bayes généralisée

L'application pratique du principe d'extension des prévisions inférieures est difficile. On peut le formuler ainsi:

$$\underline{E}(Z) = \sup_{\underline{\lambda} > 0} \left\{ \inf_{\theta} W(\theta, \underline{\lambda}) \right\}, \quad W(\theta, \underline{\lambda}) = Z(\theta) - \sum_{n=1}^k \lambda_n G(\theta, X_n) \quad (4.16)$$

où $\underline{\lambda} = \{\lambda_n, n=1, 2, \dots, k\}$. La fonction à optimiser $W(\theta, \underline{\lambda})$ est généralement non linéaire en θ , et peut comporter un grand nombre de variables. Dans le cadre de ce travail, nous avons résolu cette équation à l'aide des outils d'optimisation numérique du logiciel MATLAB, sans trop l'étudier. Nous avons remarqué que les solutions obtenues étaient sensibles au point de départ utilisé pour initialiser l'algorithme, de sorte qu'il fut chaque fois nécessaire d'appliquer le même algorithme à partir d'un grand nombre de points de départ différents pour éviter d'obtenir un optimum local, multipliant ainsi le temps de calcul.

Il serait probablement plus efficace de développer un algorithme de calcul numérique adapté à la forme particulière de l'équation. En effet, bien que $W(\theta, \underline{\lambda})$ soit non linéaire en θ , elle est linéaire en $\underline{\lambda}$. Ainsi, si l'on était capable de résoudre de façon analytique l'équation $Y(\underline{\lambda}) = \inf_{\theta} W(\theta, \underline{\lambda})$ pour toutes les valeurs de $\underline{\lambda}$, on pourrait utiliser des outils de programmation linéaire pour maximiser $Y(\underline{\lambda})$. De façon générale, il est impossible de résoudre analytiquement $Y(\underline{\lambda})$. Il reste que cette fonction est nécessairement linéaire par morceau, même si les équations des hyperplans qui la définissent sont inconnues. On pourrait vraisemblablement utiliser ces particularités de la fonction $Y(\underline{\lambda})$ pour proposer une méthode numérique mieux adaptée et plus efficace.

Lorsque la dimension du vecteur θ est faible, la fonction $W(\theta, \underline{\lambda})$ peut généralement être minimisée avec précision pour $\underline{\lambda}$ fixé, de sorte que l'erreur sur $Y(\underline{\lambda})$ peut être contrôlée. Ce fut entre autres le cas pour les problèmes que nous avons résolus au cours des chapitres précédents. Si, en maximisant par la suite $Y(\underline{\lambda})$, on obtient un maximum local, l'estimation de la prévision inférieure que l'on obtiendra pourra seulement être plus faible que $\underline{E}(Z)$, qui correspond au maximum global de $Y(\underline{\lambda})$. Ainsi, une méthode numérique peu efficace pour maximiser $Y(\underline{\lambda})$ ne pourra que sous-estimer la prévision inférieure. Par des arguments similaires, on montre que la prévision supérieure ne pourra être que surestimée. Conséquemment, s'il s'est glissé des erreurs au niveau de l'optimisation numérique dans les calculs effectués dans ce travail, elles n'ont pu mener qu'à une évaluation plus conservatrice de la valeur de l'information.

Les problèmes numériques liés à l'utilisation du principe d'extension des prévisions inférieures ont des conséquences pour l'emploi de la règle de Bayes généralisée. En effet, pour construire une prévision conditionnelle $\underline{E}(X|H)$, il faut trouver μ tel que $\underline{E}[H \cdot (X - \mu)] = 0$, ce qui nécessite d'appliquer le principe d'extension des prévisions inférieures pour différentes valeurs de μ . La qualité des résultats obtenus à l'aide de la règle de Bayes généralisée dépend donc de la précision de la résolution numérique du principe d'extension des prévisions inférieures. De plus la méthode itérative utilisée pour trouver la valeur de μ qui permet d'annuler $\underline{E}[H \cdot (X - \mu)]$ doit être adaptée pour tenir compte de l'erreur sur $\underline{E}[H \cdot (X - \mu)]$. Nous avons en effet constaté que des méthodes fondées sur la dérivée de $\underline{E}[H \cdot (X - \mu)]$ par rapport à μ , comme la méthode de Newton-Raphson ou la méthode de la descente la plus rapide, sont peu efficaces à cause de l'amplitude de l'erreur sur la dérivée, qui est évaluée de façon numérique. Nous avons eu des problèmes similaires avec la méthode numérique proposée par Walley (1991, note 6.4.1). Il est préférable d'utiliser des algorithmes dont la convergence théorique est moins rapide, mais qui sont moins sensibles à l'erreur d'estimation de $\underline{E}[H \cdot (X - \mu)]$, comme la méthode Regula-Falsi ou la méthode de la bisection (voir Scheid (1968) pour le détail de ces méthodes numériques). Soulignons que dans le cas où le référentiel est fini, il existe des méthodes analytiques permettant d'obtenir des approximations conservatrices des probabilités imprécises a posteriori (White, 1986; Snow, 1991; Salo, 1996).

4.5 Bilan et perspectives

Nous nous étions fixé comme objectif d'explorer le problème de l'estimation de la valeur de l'information hydrologique sans faire appel au principe de précision. Nous avons proposé une façon de prendre en compte l'indécision qui peut résulter de l'imprécision, et nous avons présenté le concept de prévision floue comme façon de visualiser la sensibilité de la valeur de l'information à l'imprécision des prévisions effectuées a priori.

4.5.1 Aspects originaux du travail réalisé

Ce travail constitue à notre connaissance la première tentative effectuée pour définir le concept de valeur de l'information à l'extérieur de la théorie des probabilités. C'est aussi la première fois que la théorie des prévisions inférieures cohérentes est utilisée en hydrologie.

La contribution la plus importante de ce travail consiste peut-être à montrer que l'on peut réellement appliquer les concepts introduits dans un problème concret d'hydrologie opérationnelle. L'estimation de la valeur de l'information hydrologique à SM3 constitue d'ailleurs à notre connaissance le premier exemple où une analyse de l'incertitude utilisant le concept de probabilité imprécise a été effectuée pour un modèle d'échantillonnage ayant plus d'un paramètre, à l'exception de la loi multinomiale.

Nous avons aussi montré que notre approche est plus générale que l'approche bayésienne robuste, plus facile à appliquer, à interpréter et à justifier que la théorie de Dempster-Shafer, et qu'elle peut être généralisée de façon à englober l'arithmétique floue.

4.5.2 Travaux ultérieurs à mener

Un problème qui reste peu étudié dans le cadre de l'analyse de la valeur de l'information est la non stationnarité des processus, causée par exemple par un éventuel changement climatique ou par l'action de l'homme sur le territoire. Krzysztofowicz (1994) et Hobbs (1997) proposent une analyse bayésienne, mais il est difficile selon nous de quantifier par des probabilités précises l'information disponible sur un éventuel changement climatique et sur ses conséquences sur l'hydrologie. Luo et Caselton (1997) montrent d'ailleurs l'intérêt des probabilités imprécises pour modéliser l'information sur le réchauffement global obtenue à

partir d'une revue de littérature. Il faudrait arriver à coupler cette information imprécise au modèle proposé par Krzysztofowicz (1994).

Il faudrait aussi s'attarder aux situations où la fonction d'utilité est imprécise. Lorsque les conséquences des actions sont multiples et complexes, il se peut fort bien qu'un individu soit incapable d'établir une fonction d'utilité précise. S'il croit qu'il serait possible de la construire après une réflexion plus approfondie, alors une analyse de sensibilité est suffisante. Autrement le principe de précision idéale n'est pas respecté, et il faut envisager une approche plus générale pour évaluer la valeur de l'information.

Ce qu'il faut peut-être étudier avant tout, c'est comment l'on peut amener un expert à formaliser ses connaissances a priori à l'aide de prévisions imprécises. Le concept de probabilité subjective étant encore loin d'être compris par plusieurs, on peut s'attendre à ce qu'un certain apprentissage soit nécessaire pour assimiler les concepts de pari, de prévision et de cohérence. De plus, la vérification de la cohérence des prévisions émises par un expert peut aussi être difficile. Le principe de cohérence est d'ailleurs probablement trop fort pour être réaliste, puisqu'il demande que les préférences d'un individu reflètent une compréhension poussée des implications de suites dénombrables de transactions. Il serait intéressant d'étudier le concept de valeur de l'information pour des modèles comportementaux moins contraignant à ce niveau, comme le modèle dichotomique (Quiggin, 1982).

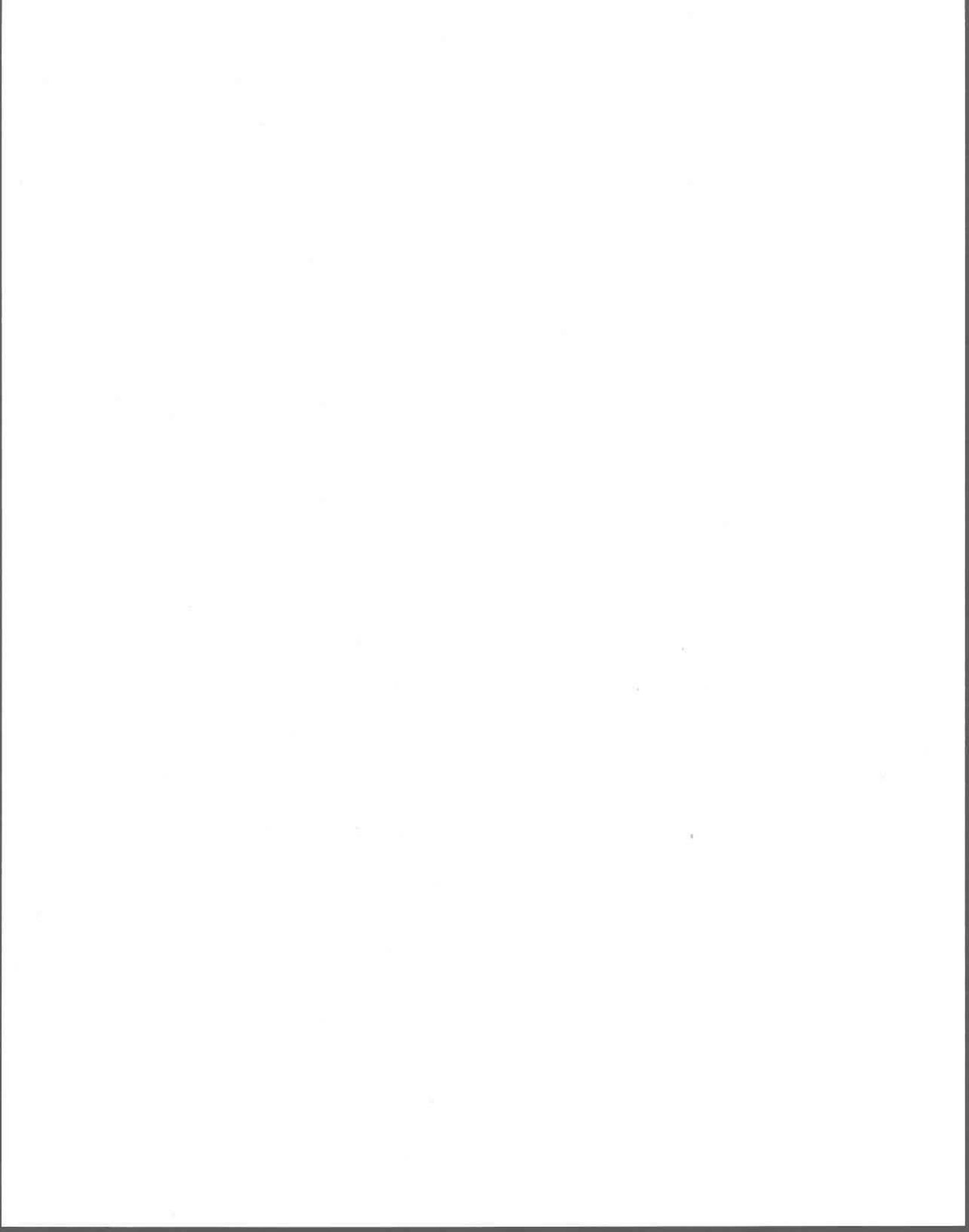
Le rôle que pourra jouer la théorie des prévisions inférieures cohérentes dans les problèmes d'analyse du risque hydrologique dépendra des efforts qui seront faits pour diffuser la théorie, en présentant des exemples réalistes d'application, mais aussi pour continuer les développements théoriques et numériques de façon à faciliter l'utilisation de cette approche et à rassurer les praticiens quant à la précision des résultats obtenus.

4.5.3 Retombées du travail effectué

Nous croyons tout de même que l'état d'avancement des travaux est suffisant pour qu'une entreprise comme Hydro-Québec, qui sait quantifier les conséquences de ses actions et qui doit évoluer dans un environnement incertain, puisse utiliser l'analyse de la valeur de l'information comme outil de gestion permettant de justifier et d'optimiser les programmes d'acquisition

d'information sur l'état de la nature. L'emploi de probabilités et de prévisions imprécises permet alors de faciliter la prise en compte de l'expérience du gestionnaire et de l'ingénieur.

Dans le cas particulier du projet Sainte-Marguerite-3, le fait d'inclure dans les études d'avant-projet une analyse du risque et de la décision aurait peut-être pu permettre de réduire le prix de revient espéré de l'électricité produite, mais aurait surtout permis de mieux justifier par la suite les décisions prises au départ.



RÉFÉRENCES

- Abramowitz, W. and I. Stegun (1970). *Handbook of mathematical functions*, National Bureau of Standards, Washington D.C.
- Agbekodo, K.M., J.P. Croue, S. Dard et B. Legube (1996). Analyse par HPCL et CG/SM des constituants du carbone organique dissous (COD), du COD biodégradable (COBD) et des composés organohalogénés (TOX) d'un perméat de nanofiltration. *Revue des Sciences de l'Eau*, 9(4): 535-555.
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica*, 21(4): 503-546.
- Attanasi, E.D. et M.R. Karlinger (1979). Worth of geophysical data in natural-disaster-insurance rate setting. *Journal of Applied Meteorology*, 21: 453-460.
- Attanasi, E.D. et M.R. Karlinger (1982). Worth of data and natural disaster insurance. *Water Resources Research*, 15(6): 1763-1766.
- Barcelo, A. et J. Coudray (1996). Nouvelle carte des isohyètes annuelles et des maxima pluviométriques sur le massif du Piton de la Fournaise (île de la Réunion). *Revue des Sciences de l'Eau*, 9(4): 457-484.
- Bárdossy, Á. et L. Duckstein (1995). *Fuzzy Rule-Based Modeling with Applications to Geophysical, Biological and Engineering Systems*. CRC Press, Boca Raton, U.S.A.
- Bárdossy, Á., L. Duckstein et I. Bogárdi (1993). Combination of fuzzy numbers representing expert opinions. *Fuzzy Sets and Systems*, 57: 173-181.
- Bartle, R.G. (1995). *The Elements of Integration and Lebesgue Measure*, Wiley Classics Library Edition, New York.
- Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2ème éd., Springer-Verlag, New York.
- Bernardo, J.M. et A.F.M. Smith (1994). *Bayesian Theory*, Wiley, Chichester, Angleterre.
- Bernier, J. (1990). Les incertitudes hydrologiques dans les problèmes de dimensionnement d'ouvrages. Valeur des informations locales et spatiales. *Revue des Sciences de l'Eau*, 3: 37-53.

- Caselton, W.F. et W.B. Luo (1992). Decision making with imprecise probabilities - Dempster-Shafer theory and application. *Water Resources Research*, 28(12): 3071-3083.
- Chateauneuf, A. et J.-Y. Jaffray (1989). Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Math. Social Sci.*, 17:263-283.
- Choquet, G. (1953-54). Theory of capacities. *Annales de l'Institut Fourier*, 5: 131-295.
- CHS (1995). *Projet 2: Analyse statistique des séries d'apports énergétiques, Procès-verbal de la 7ième réunion générale*, Chaire en Hydrologie Statistique Hydro-Québec/ CRSNG, INRS-Eau, Sainte-Foy, Québec, Procès-verbal C2- RG-7 (19-01-1995).
- Davis, D.R., L. Duckstein et R. Krzysztofowicz (1979). The Worth of hydrologic data for nonoptimal decision making. *Water Resources Research*, 15(6): 1733-1742,
- Davis, D.R. et W.M. Dvoranchik (1971). Evaluation of the worth of additional data. *Water Resources Bulletin*, 7(4): 700-707.
- Davis, D.R., C.C. Kisiel et L. Duckstein (1972). Bayesian decision theory applied to design in hydrology. *Water Resources Research*, 8(1): 33-41.
- Davis, D.R. and S. Nnaji (1982). The Information needed to evaluate the worth of uncertain information, predictions and forecasts. *Journal of Applied Meteorology*, 21: 461-470.
- Dawdy, D.R. (1979). The Worth of hydrologic data. *Water Resources Research*, 15(6): 1726-1732.
- Dawdy, D.R., H.E. Kubik et E.R. Close (1970). Value of streamflow data for project design - A Pilot study. *Water Resources Research*, 6(4): 1045-1050.
- Dempster, A.P. (1968). A Generalization of Bayesian inference (with discussion). *Journal of the Royal Statistical Society B*, 30: 205-247.
- DOE (1995). Hydroelectric trade and Canada. *International Energy Outlook 1995*, Energy Information Administration, U.S. Department of Energy.
- Duckstein, L., R. Krzysztofowicz et D. Davis (1978). To build or not to build: A Bayesian analysis. *Journal of Hydrological Sciences*, 5(1): 55-68.
- Dubois, D. et H. Prade (1988). *Théorie des possibilités: Applications à la représentation des connaissances en informatique*, 2nd ed., Masson, Paris.

- Freeze, R.A., B. James, J. Massmann, T. Sperling et Leslie Smith (1992). Hydrogeological decision analysis: 4. The Concept of data worth and its use in the development of site investigation strategies. *Ground Water*, 30(4): 574-588.
- French, S. (1984). Fuzzy decision analysis: Some criticisms. *Studies in the Management Sciences*, 20: 29-44.
- Gallinari, F., S. Elmaleh et R. Ben Aim (1996). Influence de la dissipation énergétique sur l'efficacité de la flottation à air dissous: analogie avec la floculation. *Revue des Sciences de l'Eau*, 9(4): 485-498.
- Gates, J.S. and C.C. Kisiel (1974). Worth of additional data to a digital computer model of a groundwater basin. *Water Resources Research*, 10(5): 1031-1038.
- Good, I.J. (1983). *Good Thinking*, University of Minnesota Press, Minneapolis.
- Hobbs, B.F. (1997). Bayesian methods for analysing climate change and water resource uncertainties. *Journal of Environmental Management*, 49: 53-72.
- Hydro-Québec (1991). *Aménagement hydroélectrique Sainte-Marguerite-3, Rapport d'avant-projet, Parties 1-3: Justification du projet, variantes d'aménagement, études technoéconomiques*, Hydro-Québec, Montréal, Québec.
- Hydro-Québec (1992). *Aménagement hydroélectrique Sainte-Marguerite-3, Mise à jour des données hydrologiques de la rivière Sainte-Marguerite*, Hydro-Québec, Montréal, Québec.
- Hydro-Québec (1997). *Rapport annuel 1996*. Hydro-Québec, Montréal, Québec.
- Jacobi, S. (1975). Economic optimum record length defined in the context of a statistical decision approach. *Nordic Hydrology*, 6:28-42.
- Jaffray, J.-Y. (1989). Généralisation du critère de l'utilité espérée aux choix dans l'incertain régulier. *Recherche Opérationnelle*, 23(3): 237-267.
- Jaffray, J.-Y. et F. Philippe (1997). On the existence of subjective upper and lower probabilities. *Mathematics of Operations Research*, 22(1): 165-185.
- James, B.R. and R.A. Freeze (1993). The Worth of data in predicting aquitard continuity in hydrogeological design. *Water Resources Research*, 29(7): 2049-2065.
- James, B.R. and S.M. Gorelick (1994). When enough is enough: The worth of monitoring data in aquifer remediation design. *Water Resources Research*, 30(12): 3499-3513.

- Jaynes, E.T. (1968). Prior probabilities. *IEEE Trans. Systems Science and Cybernetics*, SSC-4: 227-241.
- Jeffreys, H. (1961). *Theory of Probability*, Macmillan, Londres.
- Kaufmann, A. (1977). *Introduction à la théorie des sous-ensembles flous à l'usage des ingénieurs (Fuzzy Sets Theory) 1. Éléments théoriques de base*, Masson, Paris.
- Kaufmann, A. et M.M. Gupta (1991). *Introduction to fuzzy arithmetic: Theory and applications*, Van Nostrand Rheinhold, New York.
- Klemes, V. (1977). Value of information in reservoir optimization. *Water Resources Research*, 13(5): 837-850.
- Krzysztofowicz, R. (1986). Optimal water supply planning based on seasonal runoff forecasts. *Water Resources Research*, 22(3): 313-321.
- Krzysztofowicz, R. (1994). Strategic decisions under nonstationary conditions: a stopping-control paradigm. Dans: *Engineering in Natural Resources Management (With Special References to Hydrosystems under Changes of Physical or Climatic Environment)*. L. Duckstein et E. Parent édés. NATO ASI Series E: Applied Sciences, Kluwer, Amsterdam, pp. 295-316.
- Krzysztofowicz, R. et L.M. Watada (1986). Stochastic model of seasonal runoff forecasts. *Water Resources Research*, 22(3): 296-302.
- Lane, L.J., D.R. Davis et S. Nnaji (1979). *Water Resources Research*, 15(6): 1851-1858.
- Laplace, P.S. de (1812). *Théorie Analytique des Probabilités*. Courcier, Paris.
- Lehmann, E.L. (1983). *Theory of Point Estimation*. Wiley, New York.
- Lindley, D.V. (1968). Discussion de l'article 'A Generalization of Bayesian Inference' par A.P. Dempster. *Journal of the Royal Statistical Society*, série B, 30(2): 237-238.
- Lindley, D.V. (1994). Foundations. In: *Subjective Probability*, Ayton, P. and G. Wright eds., Wiley, 3-15.
- Loève, M. (1978). *Probability Theory, Vol 2*, 4ème éd., Springer-Verlag, New York.
- Luo, W.B. (1993). *Imprecise probability and decision in civil engineering - Dempster-Shafer theory and application*. Thèse de doctorat, Université de Colombie-Britannique, Canada.
- Luo, W.B. et W.F. Caselton (1997). Using Dempster-Shafer theory to represent climate change uncertainties. *Journal of Environmental Management*, 49: 73-97.

- Moore, R. (1979). *Methods and Applications of Interval Analysis*. SIAM Studies on Applied Mathematics, Vol. 2, Philadelphie.
- Munier, B. (1989). Calcul économique et révision de la théorie de la décision en avenir risque. *Revue d'Économie Politique*, 99(2): 276-306.
- Munier, B. et E. Parent (1995). Le développement récent des sciences de la décision: un regard critique sur la statistique décisionnelle bayésienne. *Méthodes statistiques et approches bayésiennes en hydrologie*. Comptes-rendus de la conférence internationale tenue en l'honneur du Pr. Jacques Bernier par l'UNESCO.
- Nakache, F., A. Deguin et A. Kerneis (1996). Evolution dans un réseau de distribution des micro-organismes et d'un nutriment: le CODB. Incidence du temps de transit. *Revue des Sciences de l'Eau*, 9(4): 499-522.
- O'Hagan, A. and J.O. Berger (1988). Ranges of posterior probabilities for quasiunimodal priors with specified quantiles. *Journal of the American Statistical Association*, Theory and Methods, 83: 503-508.
- Olive, Ph., P. Hubert et S. Ravailleau (1996). Estimation pratique de «l'âge» des eaux souterraines en Europe par le tritium. *Revue des Sciences de l'Eau*, 9(4): 523-534.
- Parent, É., A. Chaouche et Ph. Girard (1995). Sur l'apport des statistiques bayésiennes au contrôle de la qualité par attribut, partie 1: contrôle simple. *Revue de Statistique Appliquée*, 43(4): 5-18.
- Parent, É., G. Lang et Ph. Girard (1996). Sur l'apport des statistiques bayésiennes au contrôle de la qualité par attribut, partie 2: contrôle progressif tronqué. *Revue de Statistique Appliquée*, 44(1): 37-54.
- Patel, J.K. et C.B. Read (1996). *Handbook of the Normal Distribution*. Marcel Dekker, New York.
- Perreault, L., M. Haché et B. Bobée (1996). *Projet 2: Analyse statistique des séries d'apports énergétiques*. Rapport de recherche No R-474, Chaire en Hydrologie Statistique Hydro-Québec/ CRSNG, INRS-Eau, Sainte-Foy, Québec, Procès-verbal C2- RG-7 (19-01-1995).
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organisation*, 3(4): 323-343.

- Raiffa, H. et R. Schlaifer (1961). *Applied Statistical Decision Theory*, Graduate School of Business Administration, Harvard University, Cambridge, Mass.
- Reichard, E.G. et J.S. Evans (1989). Assessing the value of hydrogeologic information for risk-based remedial action decisions. *Water Resources Research*, 25(7): 1451-1460.
- Reuter, V., T. Gaillard, E. Praet et J.L. Vassel (1996). Potentialités des bioélectrodes et des bioréacteurs à biomasse fixée pour l'estimation rapide de la DBO. *Revue des Sciences de l'Eau*, 9(4): 435-456.
- Robert, C. (1992). *L'analyse statistique bayésienne*, Economica, Paris.
- Salo, A.A. (1996). Tighter estimates for the posteriors of imprecise prior and conditional probabilities. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 26(6): 820-825.
- Sarin, R. et P. Wakker (1992). A simple axiomatization of nonadditive expected utility. *Econometrica*, 60(6): 1255-1272.
- Savage, L.J. (1972). *The Foundations of Statistics*, 2ème éd., Dover, New York.
- Scheid, F. (1968). *Schaum's Outline of Theory and Problems of Numerical Analysis*, McGraw-Hill, New York.
- Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, 57: 571-587.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton University Press, New Jersey.
- Shafer, G. (1986). Savage revisited. *Statistical Science*, 1(4): 463-501.
- Shapiro, S.S. et M.B. Wilk (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4): 591-611.
- Slack, J.R., J.R. Wallis et N.C. Matalas (1975). On the value of information to flood frequency analysis. *Water Resources Research*, 11(5): 629-647.
- Slovic, P. et A. Tversky (1974). Who accepts Savage's axiom? *Behavioral Science*, 19: 368-373.
- Smith, C.A.B. (1961). Consistency in statistical inference and decision (avec discussion). *Journal of the Royal Statistical Society, série B*, 23:1-37.

- Snow, P. (1991). Improved posterior probability estimates from prior and conditional linear constraint systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 21: 464-469.
- Szidarovszky, F., I. Bogárdi, L. Duckstein et D.R. Davis (1976). Economic Uncertainties in Water Resources Project Design. *Water Resources Research*, 12(4): 573-580.
- von Neumann, J. et O. Morgenstern (1953). *Theory of Games and Economic Behavior*, 3ème éd., Princeton University Press.
- Wakker, P. (1990). Under stochastic dominance Choquet-expected utility and anticipated utility are identical. *Theory and Decision*, 29: 119-132.
- Walley, P. (1991). *Statistical Reasoning with Imprecise probabilities*, Chapman and Hall, London.
- Walley, P. (1993). *A Bounded Derivative Model for Prior Ignorance About a Real-valued Parameter*. Research report, University of Western Australia.
- Walley, P. (1996a). Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58(1): 3-57.
- Walley, P. (1996b). Measures of uncertainty in expert systems. *Artificial Intelligence*, 83: 1-58.
- Walley, P. et F.M. Campello de Souza (1990). Uncertainty and indeterminacy in assessing the economic viability of energy options: A case study of solar heating systems in Brazil. *Energy Systems and Policy*, 14: 281-304.
- Walley, P., L. Gurrin et P. Burton (1996). Analysis of clinical data using imprecise prior probabilities. *Journal of the Royal Statistical Society, Series D*, 45(4): 457-485.
- Wallis, J.R., N.C. Matalas et J.R. Slack (1976). Effect of sequence length n on the choice of assumed distribution of floods. *Water Resources Research*, 12(3): 457-471.
- Wasserman, L.A. (1990). Belief function and statistical inference. *Canadian Journal of Statistics*, 18: 197-204.
- White, C.C. (1986). A posteriori representations based on linear inequality descriptions of a priori and conditional probabilities. *IEEE Transactions on Systems, Man, and Cybernetics*, 16: 570-573.

- Williams, P.M. (1976). Indeterminate probabilities. Dans: *Formal Methods in the Methodology of Empirical Sciences*, M. Przelecki, K. Szaniawski et R. Wojcicki édés, Reidel, Dordrecht, 229-246.
- Winkler, R.L., A.H. Murphy et R.W. Katz (1983). The value of climate information: A decision-analytic approach. *Journal of Climatology*, 3: 187-197.
- WMO (1990). *Cost-Benefit Assessment Techniques and User Requirements for Hydrological Data*, World Meteorological Organization, Genève, Suisse.
- Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8: 338-353.
- Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1: 3-28.