

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MOULOUD MAMMARI DE TIZI-OUZOU
FACULTE DES SCIENCES
DEPARTEMENT DE MATHEMATIQUES

MEMOIRE DE MASTER

SPECIALITE : STATISTIQUE

Présenté par
Larbi Lydia ép.Hebib

THEME

Sur la décision statistique dans le contexte Bayésien

Devant le jury d'examen composé de :

M. Hamadouche Djamel	Professeur	UMMTO	Président
M. Fellag Hocine	Professeur	UMMTO	Rapporteur
Melle Atil Lynda	Maître de conférence B	UMMTO	Examinatrice
Mme Belkacem Cherifa	Maître assistant	UMMTO	Examinatrice

Soutenu le : .. /.. /2011

Remerciements

*Je tiens à exprimer toute ma reconnaissance à **M.Fellag Hocine** pour l'honneur qu'il m'a fait en assurant la direction et le suivi scientifique et technique du présent mémoire. Je le remercie pour sa grande contribution à l'aboutissement de ce travail, et pour sa disponibilité.*

*Je remercie vivement **M.Hamadouche Djamel** pour l'honneur qu'il me fait en acceptant de présider le jury de ce mémoire.*

*Mes remerciements chaleureux s'adressent également à **Melle Atil Lynda** et **Mme Belkacem Cherifa** pour avoir accepté d'examiner ce travail.*

Je suis également reconnaissante à tous les enseignants qui m'ont formé en général et particulièrement à ceux de l'Université.

*Mes remerciements chaleureux s'adressent également à mes parents, mon mari, mes soeurs spécialement **Bia** et mes frères, ainsi qu'à mes amies : **Nadia** et **Henia**.*

*Un petit coucou aussi à mon futur trésor, et ma petite nièce **Amélia**.*

Je remercie tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste mémoire.

Nous dédions ce modeste travail à toute personne
qui contribu à l'avancement de la science.

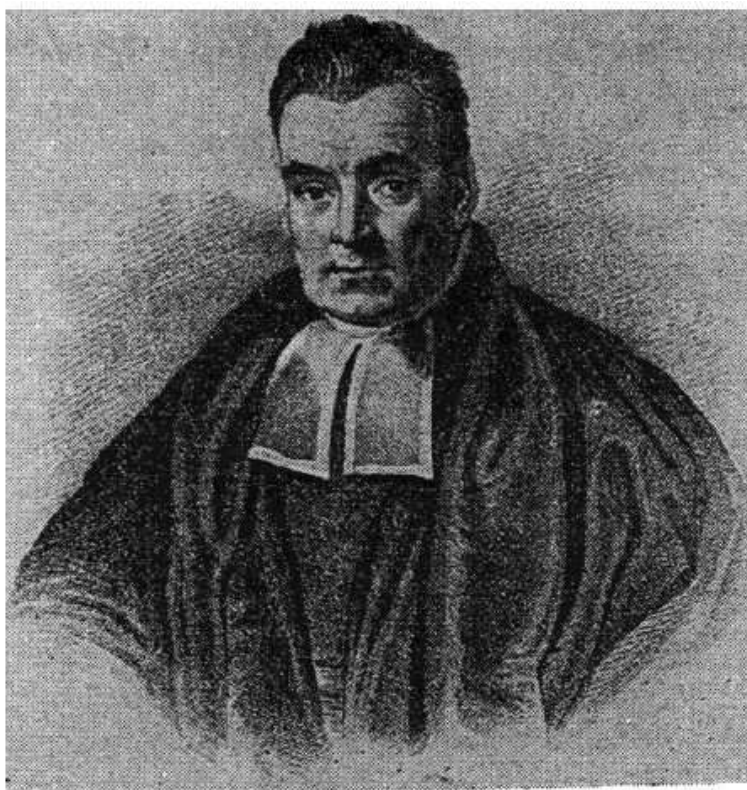
Table des matières

Introduction générale	3
1 Fondement de la Statistique bayésienne	13
1.1 Notions de base	14
1.1.1 Le choix bayésien	14
1.1.2 Espace de la théorie de la décision de statistique	15
1.1.3 Théorème de Bayes	16
1.2 Lois a priori	21
1.2.1 Approximations de la loi a priori	22
1.2.2 Approximations paramétriques	23
1.3 Choix des lois a priori	23
1.3.1 Approche partiellement informative	24
1.3.2 Approche non informative	29
2 Théorie de la décision statistique bayésienne	35
2.1 Estimateurs de Bayes	35
2.1.1 Quelques estimateurs usuelles	36
2.2 Fonction perte et risque	37
2.3 Fonctions perte usuelles	39
2.3.1 Perte quadratique	39
2.3.2 L'erreur de perte absolu	41
2.3.3 Le coût 0 - 1	42
2.3.4 Coût intrinsèque	42
2.4 Quelques alternatives	44
2.5 Admissibilité et minimaxité	44
2.5.1 Minimaxité	44
2.5.2 Admissibilité	46

2.6	Prise de décision bayésienne	47
2.6.1	Ensemble A des décisions	47
2.6.2	Ensemble X des informations	47
2.6.3	Recueil d'information	47
2.6.4	Forme de décision	48
2.7	Représentation probabiliste des connaissances	48
2.8	Etapes de la construction d'un modèle bayésien	49
2.8.1	Représentation probabiliste	49
2.8.2	Résultats expérimentaux	49
2.8.3	Les inconnus θ	50
2.9	Modèle statistique paramétrique	50
3	Applications numériques des méthodes bayésiennes	52
3.1	Choix de l'environnement R pour l'analyse numérique	52
3.2	Exemple introductif	54
3.3	L'avalanche de Montroc (Boreux et al 2010)	57
3.3.1	Présentation du problème de décision à traiter	57
3.3.2	Ensemble des décisions	58
3.3.3	Modélisation probabiliste	59
3.3.4	Inférence bayésienne	60
3.4	Les débits d'une rivière (Parent et Bernier 2007)	67
3.4.1	Présentation du problème	67
3.4.2	Ensemble A des décisions	67
3.4.3	Ensemble X des informations	68
3.4.4	Recueil d'information	68
3.4.5	Forme de décision	69
3.4.6	Représentation probabiliste des connaissances	70
3.4.7	Modèle statistique paramétrique	71
3.4.8	La méthode par introspection	73
	Conclusion générale	80
	Bibliographie	80

Il est bon de suivre sa pente, pourvu que ce soit en montant.
Andre Gide (1869-1951), prix Nobel de Litterature (1947),
Les faux-monnayeurs (1925)

Introduction générale



REV. T. BAYES

Aperçu historique de la probabilité

Les notions de hasard, de flou, d'incertitude renvoient toutes à un manque d'information. La réalité se présente à l'homme du troisième millénaire comme un ensemble d'évènements dont il n'a qu'une vision limitée. Sa compréhension des phénomènes doit donc tenir compte d'une partie cachée. Cette partie se manifeste par le fait que la même situation apparente peut avoir des conséquences fort différentes. Il y a plusieurs réalités possibles qui ont la même apparence. Comme l'évolution temporelle se fait sur la réalité, le futur semble imprédictible. Néanmoins, ces réalités potentielles ne sont pas équivalentes, et l'idée de les différencier en leur donnant des poids différents, émerge sous la forme du calcul des probabilités. Si le concept de hasard est très ancien, sa mesure est très récente dans les sciences. Les premiers pas sont fait par Pascal et Fermat (1650). En 1764 sont publiés les écrits du révérend Thomas Bayes, deux ans après sa mort. Cet essai, qui sera repris et généralisé successivement par Laplace, Condorcet et Boole, donne essentiellement une manière de remonter des effets aux causes, dans un contexte probabiliste. On y retrouve les notions de modèle sous la forme de vraisemblance, d'état sous la forme de cause et d'observation sous la forme d'effet. Cette théorie sera reprise systématiquement deux siècles plus tard par Jeffreys (1930) et Jaynes (1980). La probabilité chez Bayes apparaît plutôt comme mesure de l'incertitude interne à l'observateur relative aux causes possibles d'un évènement particulier. Elle peut ainsi être qualifiée de subjective. Utilisée par Laplace (1749-1827) dans son calcul des erreurs, la probabilité acquiert un statut objectif. La possibilité de répéter une expérience donnera un moyen concret de la mesurer par le traditionnel quotient du nombre de cas favorables sur le nombre de cas possibles. Elle sera alors définie comme limite de cette fréquence et deviendra fréquentiste. C'est dans cette optique que Fisher (1890-1962) développera ses stratégies de maximum de vraisemblance. Axiomatisée par Kolomogorov (1903-1987), la probabilité se fondera dans l'imposante Théorie de la Mesure et de l'Intégration, lui donnant une orientation analytique fortement influencée par les bourbakistes du côté européen de l'Atlantique. De l'autre, le développement de la théorie de l'information lors de la deuxième Guerre mondiale aux Etats-Unis aboutira à un tandem Probabilité-Information.

Statistique Classique versus Statistique Bayésienne

Si par exemple nous sommes confronté à prendre une décision (acheter les ouvrages ou non) et les conséquences de cette décision dépendent d'un paramètre : la proportion p qui

représente ceux qui vont acheter l'ouvrage.

Un statisticien classique nous dira : *"le paramètre p est un paramètre inconnu mais certain"*.

Pour estimer ce paramètre, il nous suggérera de tirer un échantillon au hasard. Supposons que l'on retienne un échantillon de 100 personnes et que sur ces 100, 49 se déclarent prêts à acheter l'ouvrage. Le statisticien classique nous dira alors que l'estimation ponctuelle de p est 49/100. Son raisonnement est fondé sur toute une théorie, dite théorie de l'estimation, et sur les propriétés particulières de l'estimateur du maximum de vraisemblance. De fait, ce statisticien nous dira que le nombre de personnes favorables dans un échantillon de 100 personnes suit une loi Binomiale de paramètres 100 et p et que la valeur de p qui rend maximum la probabilité d'obtenir ce que l'on a obtenu (c'est-à-dire 49 personnes favorables sur 100) est précisément 49/100. Pour le démontrer il suffit d'écrire que la probabilité d'obtenir ce que l'on a obtenu n'est autre que :

$$\frac{100!}{49!51!}p^{49}(1-p)^{51}$$

Pour aller plus loin le statisticien classique nous proposera de réaliser une "estimation par intervalle de confiance". Pour cela, il suffit de se rappeler que dès que la taille de l'échantillon devient suffisamment grande, la loi Binomiale converge vers la loi Normale. Dans notre cas, tout cours de statistique nous dira que la proportion de personnes favorables dans notre échantillon suit approximativement une loi Normale de moyenne p et de variance $p(1-p)/n$. Si nous appelons \bar{X} cette proportion dans l'échantillon (attention : avant de connaître les résultats précis de l'échantillon \bar{X} est une variable aléatoire) on a donc :

$$p(-1.96 < \frac{\bar{X} - p}{\sqrt{p(1-p)/n}} < +1.96) = 0.95$$

la valeur 1,96 étant lue dans les tables de la loi Normale centrée réduite.

On a alors :

$$p(\bar{X} - 1.96\sqrt{p(1-p)/n} < p < \bar{X} + 1.96\sqrt{p(1-p)/n}) = 0.95$$

Arrivé là le statisticien classique sera généralement un peu embarrassé pour nous dire ce dont il s'agit vraiment. Il nous dira d'abord que l'on peut estimer $p(1-p)$ soit par 0,25

(sa valeur maximale) soit par $\bar{X}(1 - \bar{X})$ où \bar{X} est la valeur de la proportion constatée dans l'échantillon. Il nous dira ensuite qu'il a bâti un intervalle de confiance à 95% et aurait obtenu comme réalisation de l'intervalle de confiance (en remplaçant $p(1 - p)$ par 0,25) :

$$p = 0.49 \pm 0.098$$

L'ennui avec cet intervalle, c'est son interprétation. Puisque p est un paramètre certain, il ne fait pas sens d'affirmer que la probabilité pour que p soit compris entre 39,2% et 58,8% est de 95%. Si nous demandons au statisticien classique d'interpréter son intervalle de confiance, il nous fera probablement un long discours pour nous dire que si l'on avait tiré 100 échantillons et si on avait construit 100 intervalles de confiance sur la base de ces échantillons, alors la "vraie" valeur de p aurait été comprise dans 95 de ces intervalles (en moyenne). Tout ceci, nous en conviendrons, est bien difficile à comprendre, d'autant plus que l'on a rarement le temps de tirer 100 échantillons.

Le statisticien bayésien ne nous dira pas du tout la même chose, même si, en bout de course, les deux démarches peuvent parfois se rejoindre.

Le statisticien bayésien nous dira : "Il existe une vraie valeur de p dans la nature. Cette valeur vous est inconnue, mais nous avons à son sujet un certain nombre d'informations. Par exemple nous savons qu'il est très improbable que p soit très proche de 1 ou de 0. Cette information nous vient de notre connaissance des acheteurs et de notre intuition. Nous allons essayer par un certain nombre de questions simples d'encoder nos croyances sous la forme d'une distribution de probabilité subjective sur " p ". Ces questions sont du type "Désirez-vous parier 100 F sur le fait que p est inférieur à 0,4 ou préférez-vous pariez ces 100 F à Pile ou Face, etc.". Nous avons déjà l'habitude de ce type de question. À la suite de cet "encodage" on obtiendra une distribution de probabilité subjective sur p dont on peut représenter graphiquement la densité.

À partir de cette distribution de probabilité subjective, les choses deviennent beaucoup plus simples. Si on veut, pour une raison ou pour une autre, faire de l'estimation ponctuelle, on peut retenir un indicateur de tendance centrale de cette distribution de probabilité : mode, médiane ou moyenne.

Si on décide alors qu'il est rentable pour nous d'acquérir plus d'information (et cette

rentabilité sera appréciée en tenant compte des conséquences de nos décisions) il ne nous reste plus qu'à réviser nos croyances à la lumière de l'information apportée par l'échantillon en utilisant le théorème de Bayes. Dans le cas continu, celui-ci nous dit que :

$$f(p/x) = \frac{f(p)f(x/p)}{\int f(p)f(x/p)}$$

si x est les résultats apportés par l'échantillon.

Si nous le souhaitons, nous pouvons refaire de l'estimation ponctuelle et/ou par intervalle de confiance en utilisant cette nouvelle distribution de probabilité de densité $f(p/x)$.

Les démarches classiques et bayésiennes sont donc fort différentes, avec un net avantage à la seconde du point de vue de la cohérence et de la prise en compte des conséquences du problème, lorsque la statistique a pour but de nous aider à décider.

Ces deux démarches peuvent cependant coïncider au niveau des résultats, en particulier lorsque l'on dispose de très peu d'information a priori et que $f(p)$ tend vers une loi Uniforme. Ceci est relativement naturel puisque, dans l'optique classique, on fait précisément l'hypothèse, implicite, que l'on ne sait rien a priori et que toute l'information est apportée par l'échantillon. Dans le cas où l'on ne sait rien a priori, on a $f(p) = 1, \forall p \in [0, 1]$. Dans la formule de Bayes dans le cas continu on peut remarquer que le dénominateur ne dépend pas de p puisque que l'on intègre sur toute les valeurs possibles de p . On peut donc écrire :

$$f(p/x) = k \times f(x/p)$$

où k est un terme ne dépendant pas de p .

Dans ces conditions faire de l'estimation ponctuelle après révision des croyances revient à chercher une caractéristique de tendance centrale la distribution révisée. Si on choisit d'utiliser le mode de cette distribution, on va retenir la valeur de p maximisant $f(p/x)$. Or cette valeur est maximale lorsque $f(x/p)$ est maximale c'est-à-dire lorsque p maximise la vraisemblance de l'information apportée par l'échantillon. On se retrouve donc, dans ce cas très particulier, dans les conditions de l'estimation ponctuelle classique avec l'estimateur du maximum de vraisemblance.

Débats

Fisher versus Bayes

L'utilisation de la probabilité a suscité de nombreuses étincelles entre les tenants de Bayes et ceux de Fisher ; chose qui est incensé, car il s'agit de la même notion, mais utilisée dans des contextes différents. Pour le physicien par exemple, la répétition de l'expérience est le plus souvent réalisable et les théorèmes limites (loi de grand nombres, théorème central limite) lui permettront d'utiliser une approche fishérienne. Pour l'historien, la répétition n'est pas réalisable. Ce dernier devra donc intégrer des sources d'information provenant d'ailleurs et sera intéressé par l'approche bayésienne.

L'inconvénient majeur des théorèmes limites est qu'ils supposent un grand nombre de répétitions. Si ce nombre est trop petit, le physicien lui-même va se trouver dans la situation de l'historien. Lors de l'analyse d'évènements rares, comme les stunami ou tremblement de terre, l'observateur ne pourra plus utiliser les approximations que lui fournit la théorie fréquentiste.

Particularité de la théorie de Bayes

L' a priori

Le seul concept propre à la théorie de Bayes vue dans le formalisme de la théorie de la décision statistique est l'a priori. L'a priori est une répartition des états qui traduit la connaissance qu'a l'observateur de l'état du système avant de procéder à une mesure. L'obtention d'un résultat de mesure modifie cette répartition en un a posteriori. Dans ce cadre, l'information apportée par l'observation peut se voir comme modification d'une loi de probabilité. Lors de multiples mesures, cette répartition évolue vers une loi qui se rapproche des lois limites utilisées dans l'optique fishérienne.

Un expérimentateur convaincu de l'absolu de la mesure objective aura tendance à remettre en cause le bien fondé de la notion même d'a priori. On fera remarquer qu'un observateur a toujours un a priori avant de faire une mesure, car il utilise un instrument ne fonctionnant que dans certaines conditions. Un thermomètre sera totalement inutilisable dans un four. Cette simple considération limite considérablement l'espace des états.

*”La vérité, ce n’est pas le certain et l’incertain,
ce n’est pas l’ignorance.”*

Ilya Prigogine (1917-2003), prix Nobel de chimie (1977)

Resumé

Le monde qui nous entoure nous place souvent dans une perspective de décision en univers incertain, cependant, il est nécessaire de prendre des décisions en avenir risqué car, bien souvent, les informations sont insuffisantes pour lever complètement les incertitudes. Dans ce cas, ce qui intéresse un décideur c'est de savoir si, ayant adopté une décision "a", son incertitude sur un paramètre inconnu ne risque pas d'entraîner des conséquences désagréables et dans ce cas s'il importe de choisir une autre décision mieux appropriée. Cette complexité est prise en charge par l'approche Bayésienne qui intègre préoccupation au centre de sa démarche. Nous entrons alors dans le monde de la statistique par la porte de la décision sous informations. La statistique décisionnelle porte un langage de représentation du monde et de ses incertitudes.

C'est aujourd'hui une science mathématique dont l'objectif est de décrire ce qui s'est produit et de faire des projections quant à ce qu'il peut advenir dans le futur. L'approche Bayésienne occupe une place de choix dans le monde scientifique. Ce qui nous éloigne considérablement des anciennes réticences à employer l'approche bayésienne dûes essentiellement à la conception subjectiviste des probabilités que l'on associe à la démarche Bayésienne. De nos jours, il n'y a aucune ambiguïté à prendre en compte, dans une prévision ou une estimation, les informations que l'on peut avoir a priori.

Ce mémoire présente l'essentiel de ce qui est nécessaire à tout décideur souhaitant utiliser l'approche Bayésienne dans sa gestion. Après une présentation de la formalisation mathématique de cette approche étalée sur deux chapitres, quelques exemples concrets viendront illustrer l'intérêt pratique de la décision statistique dans le contexte Bayésien.

Chapitre 1

Fondement de la Statistique bayésienne

Introduction

”Lorsque un mathématicien se trouve devant une contradiction irréductible, il ajoute une dimension à l’espace du problème.”

Anonyme, XX^e

La Statistique doit être considérée comme l’interprétation d’un phénomène naturel, plutôt que son explication. En effet, l’inférence statistique s’accompagne d’une modélisation probabiliste du phénomène observé et implique nécessairement une étape de formalisation réductrice. Sans cette base probabiliste, aucune conclusion utile ne pourra être tirée.

L’objet principal de la Statistique est de faire une inférence sur la distribution probabiliste à l’origine de ce phénomène, grâce à l’observation d’un phénomène aléatoire, c’est-à-dire faire une analyse à fin de donner une description d’un phénomène passé, ou faire une prédiction d’un phénomène arrivant avec une nature similaire.

La Statistique est la plupart du temps motivée par un objectif tel que :

- Une usine devrait-elle recruter plus de travailleurs ?
- Un laboratoire pharmaceutique devrait-il arrêter la production d’un médicament à cause de quelques critiques ?
- Un pays en faillite devrait-il prendre des prêts d’un autre pays ou pas ?

Le troisième millénaire sera, dit-on, celui de l'information.

Aussi la statistique y sera-telle appelée à jouer un rôle important et le paradigme bayésien plus que tout autre, puisqu'il offre un cadre de raisonnement bien adapté à l'intégration des opinions et des faits de toutes provenances qui interviennent dans la gestion des risques et la prise de décision en contexte d'incertitude. De la collecte de données à la prévision, l'analyse statistique pose plusieurs défis. L'élaboration du modèle représente sans doute la phase la plus délicate de l'exercice, car elle doit répondre à un double impératif de réalisme et de parcimonie. Hormis quelques cas de figure, une démarche bayésienne n'est envisageable qu'à charge de disposer d'outils efficaces pour la quantification et la mise à jour de l'information.

Nous présenterons donc dans ce chapitre les fondements de la théorie bayésienne de décision.

1.1 Notions de base

1.1.1 Le choix bayésien

La statistique est un art interdisciplinaire de la quantification sous incertitudes utilisé par les physiciens, les économistes, les ingénieurs, les biologistes, les assureurs, les psychologues, les météorologues, etc. Tous les praticiens soucieux de bâtir, sur des fondations solides, un pont entre théorie et données expérimentales. Depuis un siècle, la statistique s'est considérablement développée, initiant une révolution dans les modes de pensée, car elle porte un langage de représentation du monde et de ses incertitudes.

C'est aujourd'hui une science mathématique dont l'objectif est de décrire ce qui s'est produit et de faire des projections quant à ce qu'il peut advenir dans le futur. Parfois, la situation peut être simplement décrite par quelques représentations graphiques d'analyse élémentaire des données. Bien souvent, le problème est beaucoup plus compliqué car de multiples facteurs d'influence doivent être pris en compte.

Schématiquement, on construit deux ensembles avec ces facteurs. Un premier paquet contient les facteurs dits explicatifs, bien identifiés, ceux dont on souhaite étudier l'influence en détail. En ce qui concerne le second paquet de facteurs, on ne sait, ou on ne veut pas, représenter leurs effets perturbateurs au cas par cas et, de ce fait, le jargon des modélisateurs le baptise sous le terme bruit, décrit alors de façon plus grossière par ses caractéristiques statistiques générales. Dans tous les cas, l'étude de la variabilité est au

centre des débats : il s'agit d'abord de caractériser l'influence des facteurs identifiés et ensuite de représenter et d'évaluer le bruit résiduel dû à ces autres facteurs non pris en compte dans l'analyse de façon explicite. Dans une telle situation, le statisticien classique utilise à la fois un raisonnement déterministe par l'absurde, afin de proposer des valeurs acceptables pour les paramètres décrivant les effets des facteurs explicatifs et un raisonnement probabiliste, pour traduire la variabilité des résultats observés due au bruit. Ce mode de pensée s'appuie sur l'hypothèse de la réalité objective des paramètres ainsi que sur l'interprétation de la probabilité comme limite des fréquences des résultats observés.

Par contre, le statisticien bayésien utilise le même cadre de pensée pour traiter par le pari probabiliste l'interaction de ces deux niveaux d'incertitudes : ignorance quant aux valeurs possibles des paramètres et l'aléa des bruits entachant les résultats expérimentaux. Choisir la piste bayésienne paraîtra à certains inutilement trop sophistiqué si on se limite aux modèles élémentaires (binomial, normal, etc.), pour ces cas d'école simples, l'approche fréquentiste est facile (nombreux logiciels), et offre au praticien des résultats souvent très proches de ceux que donnerait une analyse bayésienne avec une distribution a priori peu informative. Mais pour peu que l'analyste souhaite prendre à bras le corps des problèmes plus proches de son réel quotidien, apparaissent variables multiples, données manquantes, effets aléatoires, grandeurs latentes..., la structure des modèles de la vie scientifique moderne se présente sous une forme où des couches successives de conditionnement s'emboîtent, et pour lesquels l'approche bayésienne affirme sa véritable pertinence.

1.1.2 Espace de la théorie de la décision de statistique

Nous abrègerons ici, la théorie bayésienne de la décision statistique en TBDS, qui distingue trois espaces.

Le premier espace est celui des états , c'est-à-dire les valeurs vraies ou mesurandes relatives au système. Ces valeurs ne sont généralement pas connues directement, mais à travers le processus de mesure. Les états seront notés θ et leur espace Θ .

Le deuxième espace est celui des observations ou des résultats de mesure. Ces résultats seront notés x et leur espace \mathcal{X} .

Le troisième espace est celui des stratégies.

L'observateur est souvent amené à prendre des décisions. Ces décisions sont envisagées d'une manière très générale. Par exemple, un test d'hypothèses conduisant au choix d'une alternative est associé à un espace discret de décisions, qui n'est autre que celui de l'ensemble des différentes hypothèses envisagées. S'il existe une règle s associant une décision d à une observation x , la fonction $d = s(x)$ est appelée une stratégie.

1.1.3 Théorème de Bayes

Le but étant de trouver une probabilité sur les états, alors que l'on ne dispose au départ que d'une famille de probabilités sur les observations nous conduit à raisonner dans l'espace produit des observations par les états

$$\mathcal{X} \times \Theta$$

La théorie générale de l'intégration montre que la donnée d'une probabilité de transition $P(x/\theta)$ de Θ vers \mathcal{X} ne suffit pas à obtenir une probabilité sur le produit. Il faut encore une probabilité dite a priori $\pi(\theta)$. Dans ce cas, la probabilité conjointe sur le produit est définie par

$$P(x, \theta) = P(x/\theta)\pi(\theta)$$

La probabilité conjointe ainsi construite induit une marginale. En particulier, la marginale sur l'espace des observations sera nommée simplement marginale et sera définie par

$$M(x) = \int_{\theta} P(x/\theta)\pi(\theta)$$

Supposons maintenant que la mesure donne l'observation x . Il est possible de considérer la partie de $\mathcal{X} \times \Theta$, et considérer la probabilité conditionnelle induite par la conjointe sur cette partie.

Alors, cette probabilité conditionnelle s'écrit

$$\pi(\theta/x) = \frac{P(x, \theta)}{M(x)} = \frac{P(x/\theta)\pi(\theta)}{M(x)} = \frac{P(x/\theta)\pi(\theta)}{\int_{\theta} P(x/\theta)\pi(\theta)}$$

Cette probabilité conditionnelle est la probabilité a posteriori sur les états étant donné l'observation x .

Si le modèle est donné par une fonction de vraisemblance $\mathcal{L}(\theta/x)$ et $\pi(\theta)$ par une densité $f(\theta)$, on obtient la formule classique

$$f(\theta/x) = \frac{\mathcal{L}(\theta/x)f(\theta)}{\int_{\theta} \mathcal{L}(\theta/x)f(\theta)}$$

Pour remonter de l'observation à l'état, la tentation est grande de lire cette fonction avec x constant (le résultat effectif de la mesure) et θ variable. Selon l'approche fréquentiste de Fisher, le meilleur choix de l'état consiste à prendre celui qui maximise la fonction \mathcal{L} . C'est le principe *du maximum de vraisemblance*

$$\hat{\theta} = \arg \max \mathcal{L}(\theta/x)$$

Cet estimateur de l'état a un caractère local. En effet, le maximum est souvent obtenu par annulation de la dérivée de \mathcal{L} . Le reste de la fonction est ignoré. En particulier, il est difficile de donner une estimation de l'incertitude pour cet estimateur. Il serait alors possible d'obtenir une moyenne et une variance pour l'état.

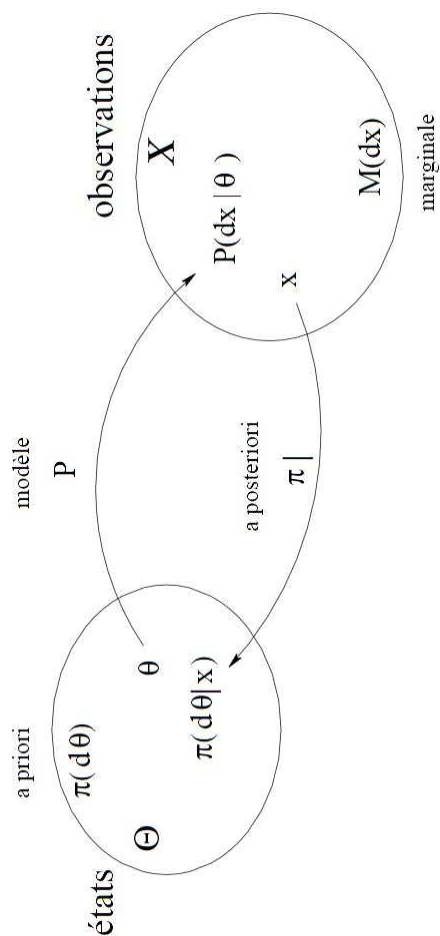


FIG. 1.1 – Concepts de la théorie bayésienne de la décision statistique.

Exemple 1.1. Soit notre paramètre $\theta = p$, alors si $x \sim \mathcal{B}(n, p)$ et $p \sim \mathcal{Be}(\alpha, \beta)$ avec $\alpha = \beta = 1$ dans le cas particulier de Bayes,

$$f(x/p) = C_n^p p^x (1-p)^{n-x}$$

$$\pi(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

La distribution jointe de (x, p) est alors

$$\varphi(x, p) = \frac{C_n^p}{B(\alpha, \beta)} p^{\alpha+x-1} (1-p)^{n-x+\beta-1}$$

et la distribution marginale de x est

$$\begin{aligned}
 M(x) &= \frac{C_n^p}{B(\alpha, \beta)} B(\alpha + x, n - x + \beta) \\
 &= C_n^x \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + x)\Gamma(n - x + \beta)}{\Gamma(\alpha + \beta + n)}
 \end{aligned}$$

puisque la distribution a posteriori de p est

$$\pi(p/x) = \frac{p^{\alpha+x-1}(1-p)\beta + n - x - 1}{B(\alpha + x, \beta + n - x)}$$

qui est une loi bêta $Be(\alpha + x, \beta + n - x)$.

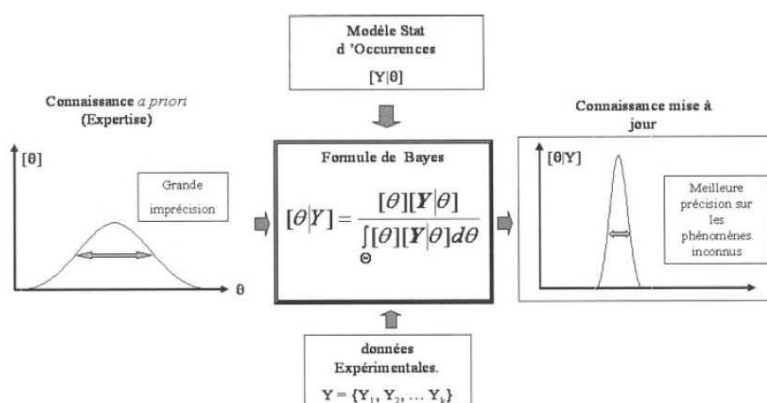


FIG. 1.2 – Formule de Bayes.

Exemple 1.2. Numérique traité avec R

Une étude est réalisée sur un échantillon de petits garçons, le but de cet étude est d'analyser les effets de l'exposition de petits garçons au plomb (rencontrés dans plusieurs types de nourriture), dans le développement cérébral de ses petits gamins.

Les chercheurs analysent la teneur en plomb des dents de lait des petits après les avoir perdus. On prend un échantillon de 29 petits garçons, qui ont une teneur en plomb supérieur à 22.22 ppm(part par million), 22 parmi ses petits ont terminé leur éducation secondaire,

et les 7 autres non.

Supposons une distribution a priori $p \sim \mathcal{Be}(1,1)$ pour les élèves qui ont terminé leurs études secondaires.

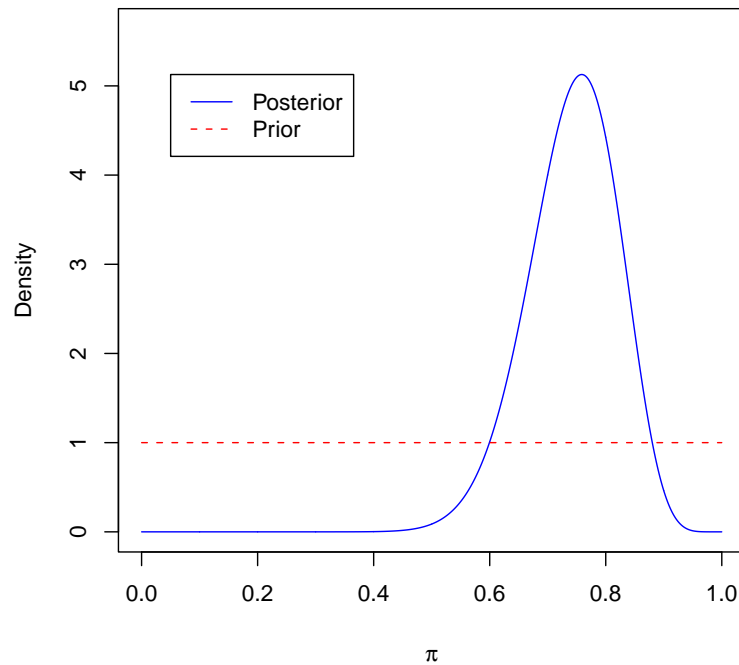
Analyse du problème :

1. Notre échantillon suit une loi de bernouilli $P(x/\theta) = B(p)$,
2. Nous prenons comme loi a priori une bêta, $\pi(\theta) \sim \mathcal{Be}(1,1)$
3. En faisant les mêmes calculs que dans l'exemple théorique passé, on obtient la distribution a posteriori qui est une $\pi(\theta/x) = \mathcal{Be}(1 + 22, 1 + 29 - 22)$

Représentation graphique de la densité a priori et a posteriori :

Nous commencerons par télécharger le package *Bolstad*, nous utiliserons la fonction "binogcp" qui évalue et dessine la densité a posteriori et a priori qui est utilisée seulement dans le cas d'un échantillon binomial et une distribution a priori continue.

```
results <- binogcp(22, 29, density = "beta", params = c(1, 1), n : pi = 1e4)
```

FIG. 1.3 – Distribution a priori $\mathcal{B}e(2, 4)$ et a posteriori $\mathcal{B}e(8, 13)$ *Caractéristiques de la loi a posteriori :*

- . *densite* = `dbeta(x, 23, 8)`
- . *summary(densite)*

<i>Min.</i>	<i>1st Qu.</i>	<i>Median</i>	<i>Mean</i>	<i>3rd Qu.</i>	<i>Max.</i>
0.06994	2.86800	4.10500	3.69400	4.88500	5.12800

1.2 Lois a priori

le point le plus significatif dans l'analyse bayésienne est le choix de la loi a priori, sa détermination est donc l'étape la plus importante dans la mise en oeuvre de cette inférence.

En statistiques bayésiennes, on considère, en plus des données récoltées dans le cadre d'une expérience, un a priori sur le paramètre θ que l'on cherche à estimer. C'est le terme

$p(\theta)$.

Cela peut permettre d'inclure dans les analyses des résultats précédents, formels ou non. En pratique, toute la difficulté consiste à estimer de manière correcte nos a priori.

En statistiques bayésiennes, on va chercher non pas la "meilleure valeur" d'un paramètre, comme dans le maximum de vraisemblance, mais on va estimer la distribution de probabilité de ce paramètre.

On peut imaginer que l'on ne cherche pas à estimer la valeur du paramètre sur un cas, mais sur un grand ensemble de cas pour lesquelles les valeurs divergent. Si on fait suffisamment d'expériences, la distribution de probabilité du paramètre peut être tellement pointue qu'en pratique on considèrera une unique valeur.

Les a priori sont divisés en deux grandes parties.

Nous considérerons deux approches usuelles, celle des "*objectivistes*", ou on suppose une probabilité a priori dite "Distribution a priori non informative" puisque l'observateur n'a initialement rien à sa disposition, mis à part son instrument de mesure, et celle des "*subjectivistes*", ou on suppose une probabilité dite "Distribution a priori informative", qui permet d'inclure des connaissances sur l'état préexistantes à la mesure.

1.2.1 Approximations de la loi a priori

Ici, nous avons deux cas possibles, le cas où l'espace des états Θ est fini, et le cas où il est infini.

Θ fini

Dans ce cas, souvent on est amené à se référer aux expériences précédentes du même type.

Θ infini

Quand l'espace des paramètres Θ n'est pas dénombrable, par exemple, lorsqu'il s'agit d'un intervalle, la détermination subjective de la loi a priori π est évidemment beaucoup plus compliquée. En général, une première approximation de π est obtenue par le parti-

tionnement de Θ en différents ensembles (par exemple des intervalles) et la détermination de la probabilité de chaque ensemble, $\pi(\theta)$ est alors approchée par un histogramme.

1.2.2 Approximations paramétriques

Une alternative très souvent utilisée pour déterminer un a priori continu consiste à restreindre arbitrairement le choix de π à une famille de densités paramétrées et à déterminer les paramètres correspondants en utilisant les moments ou en utilisant les quantiles, cette seconde option étant plus robuste. Par exemple, des évaluations subjectives de la médiane et du quantile à 75% sont suffisantes pour identifier les deux paramètres d'une distribution normale.

Exemple 1.3. Soit $X_i \sim \mathcal{B}e(n_i, p_i)$ le nombre d'étudiants réussissant l'examen d'introduction à l'analyse, dans une classe de n_i étudiants. Les années précédentes, la moyenne des p_i a été de 0.70, avec une variance de 0.1. Si nous supposons que les p_i sont tous générés selon la même distribution bêta " $\mathcal{B}e(\alpha, \beta)$ ", les paramètres α et β sont estimés par

$$\begin{aligned} -\text{Moyenne} : \frac{\alpha}{\alpha + \beta} &= 0.7 \\ -\text{Variance} : \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} &= 0.1 \end{aligned}$$

Soit $\alpha = 0.77$ et $\beta = 0.33$, ce qui conduit à la distribution a priori

$$p \sim \mathcal{B}e(0.77, 0.33)$$

1.3 Choix des lois a priori

Un grand intérêt de la théorie bayésienne est sa grande cohérence et sa méthodologie unifiée. Ainsi, donner les lois a priori et a posteriori, ainsi que donner la fonction de perte suffit pour déterminer, entre autres, un estimateur optimal. Le choix de la loi a priori π est donc crucial. Avec beaucoup d'observations, le comportement asymptotique peut guider ce choix mais sinon il est nécessaire de le justifier avec précision.

1.3.1 Approche partiellement informative

Quand on dispose de peu d'information a priori, ou quand l'information dont on dispose est trop vague, alors souvent le statisticien ne peut faire une construction subjective complète de l'a priori.

De telles situations peuvent obliger le statisticien à avoir recours à des méthodes d'estimation fréquentiste comme : Estimateur du maximum de vraisemblance, estimateur sans biais optimaux, etc.

Cependant, tout en gardant à l'esprit les fondements bayésiens des critères fréquentistes d'optimalité, il paraît donc préférable de suivre l'approche bayésienne, en utilisant un a priori dit objectif, c'est-à-dire construit à partir du modèle d'échantillonnage.

Lorsque aucune information a priori n'est disponible, ces a priori sont dits *non informatifs*.

1.3.1.1 Maximum d'entropie

Si l'on possède des informations partielles du type $\mathbb{E}^\pi[g_k(\theta)] = \mu_k$ où pour chaque $k = 1, \dots, n$, g_k est une fonction donnée.

Pour $\theta \in \{1, \dots, n\}$ et $\pi(\theta) = (\pi_1, \dots, \pi_n)$ tel que $\pi_i > 0$ et $\sum_{i=1}^n \pi_i = 1$, l'entropie de la loi est définie par

$$Ent(\pi) = - \sum_{i=1}^n \pi_i \log(\pi_i) \leq - \sum_{i=1}^n \frac{1}{n} \log\left(\frac{1}{n}\right) = \log n \quad (1.1)$$

Ce dernier terme correspond à une répartition uniforme. Pour la masse de Dirac $\delta(j)$ (telle que $\pi_j = 1$ et $\forall i \neq j, \pi_i = 0$), $Ent(\delta(j)) = 0$ ce qui correspond à l'intuition puisqu'alors il n'y a plus d'incertitude et l'information est totale. Une entropie petite s'interprète comme une loi concentrée et informative. La maximisation de l'entropie sous les contraintes permet de chercher la loi qui apporte le moins d'information. Le principe à la base de cette méthode est donc de chercher à calculer :

$$\arg \max_{\pi} Ent(\pi) \quad \text{sous la contrainte} \quad \mathbb{E}^\pi[g_k(\theta)] = \mu_k$$

La solution de ce problème est alors donnée par :

$$\pi^* \propto e^{\sum_{k=1}^n \lambda_k g_k(\theta)}$$

où les λ_k sont les multiplicateurs de Lagrange associés. Dans la pratique, on détermine ces valeurs λ à partir des contraintes (systèmes d'équations) comme l'indique l'exemple à suivre.

Exemple 1.4. *Un cas dénombrable*

Ici, $\Theta = \mathbb{N}$ et $\mathbb{E}^\pi[\theta] = x > 1$, c'est-à-dire qu'ici $g(\theta) = \theta$ et $\mu = x$. On sait que $\pi^* \propto e^{\lambda\theta}$ et que λ est déterminé par :

$$\frac{\sum_{\theta \in \mathbb{N}} \theta e^{\lambda\theta}}{\sum_{\theta \in \mathbb{N}} e^{\lambda\theta}} = x$$

Cela conduit à résoudre :

$$\begin{aligned} \frac{x}{1 - e^\lambda} &= \frac{1}{e^\lambda} \frac{e^\lambda}{(1 - e^\lambda)^2} \\ e^\lambda &= \frac{x - 1}{x} \end{aligned}$$

Par exemple si $x = \frac{12}{11}$ alors $\lambda = -\log(12)$

En continu, il n'est pas possible de définir l'entropie comme ci-dessus puisqu'on ne peut dénombrer les états (pas de mesure de comptage) en l'absence de mesure de référence. Dans le cas continu, on définit alors l'équivalent de l'entropie par rapport à une mesure π_0 :

$$Ent(\pi | \pi_0) = \int_{\Theta} \pi(\theta) \log\left(\frac{\pi(\theta)}{\pi_0(\theta)}\right) d\theta$$

C'est en fait la divergence de Kullback. Dans l'idée π_0 est la plus plate possible, la plus proche de la répartition uniforme. L'objectif est donc de maximiser $Ent(\pi | \pi_0)$ sous les contraintes $\mathbb{E}^\pi[g_k(\theta)] = \mu_k$. Là encore, la solution générale est connue :

$$\pi^*(\theta) \propto e^{\sum_{k=1}^n \lambda_k g_k(\theta)} \pi_0(\theta)$$

1.3.1.2 Lois a priori conjuguées

Ce type de lois a priori est utilisé quand l'information a priori disponible sur le modèle est trop vague ou peu faible. Dans ce cas l'analyste regarde la forme de la fonction de vraisemblance et choisit une famille de lois qui se marie bien avec elle.

L'avantage des familles conjuguées est avant tout la simplicité des calculs. Avant l'essor du calcul numérique, ces familles étaient pratiquement les seules qui permettaient de faire aboutir des calculs. L'intérêt principal du caractère conjugué se manifeste quand \mathcal{F} est paramétrée.

Effectivement le passage de distribution a priori à la distribution a posteriori ce n'est dans ce cas qu'une mise à jour des paramètres correspondants. Et par conséquence, les distributions a posteriori sont toujours calculables dans ce cas.

L'approche a priori conjuguée, introduite par Raiffa et Schlaifer (1961), peut être considérée comme un point de départ pour l'élaboration de distributions a priori fondées sur une information a priori limitée.

On considère une variable x suivant une fonction de densité $f(x | \theta)$

Définition 1.1. Famille conjuguée

Une famille F de distributions de probabilité sur Θ est dite conjuguée (ou fermée par échantillonnage) par une fonction de vraisemblance $f(x|\theta)$ si, pour tout $\pi \in F$, la distribution a posteriori $\pi(\cdot|x)$ appartient également à F .

Un exemple trivial d'une famille conjuguée est l'ensemble F_0 de toutes les lois de probabilité sur Θ . L'avantage des familles conjuguées est avant tout de simplifier les calculs. Avant l'essor du calcul numérique, ces familles étaient pratiquement les seules qui permettaient de faire aboutir des calculs.

Les lois a priori conjuguées sont généralement associées à un type particulier de lois d'échantillonnage qui permet toujours leur obtention ; il est même caractéristique des lois a priori conjuguées comme nous le verrons ci-dessous. Ces lois constituent ce qu'on appelle des *familles exponentielles*.

Définition 1.2. Familles exponentielles

Soient μ une mesure σ -finie sur χ , Θ l'espace des paramètres, C et h des fonctions respectivement de χ et Θ dans \mathbb{R}_+ , et R et T des fonctions de Θ et χ dans \mathbb{R}^k . La famille des distributions de densité (par rapport à μ)

$$f(x|\theta) = C(\theta)h(x) \exp\{R(\theta).T(x)\} \quad (1.2)$$

est dite famille exponentielle de dimension k . Dans le cas particulier où $\Theta \subset \mathbb{R}^k$, $\chi \subset \mathbb{R}^k$ et

$$f(x|\theta) = C(\theta)h(x) \exp\{\theta.x\} \quad (1.3)$$

la famille est dite *naturelle*.

D'un point de vue analytique, les familles exponentielles ont certaines caractéristiques intéressantes, pour tout échantillon de (1.8), en particulier, il existe une statistique exhaustive de dimension constante, en effet, si $x_1, \dots, x_n \sim f(x|\theta)$, avec f satisfaisant (1,9),

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \in \mathbb{R}^k$$

est exhaustive pour tout n . La réciproque de ce résultat a été aussi établie par Koopman (1936) et Pitman (1936).

Théorème 1.1. (*Lemme de Pitman-Koopman*) Si une famille de lois $f(\cdot|x)$ à support constant est telle que, à partir d'une taille d'échantillon suffisamment grande, il existe une statistique exhaustive de taille fixe, la famille est exponentielle.

Exemple 1.5. Soit $x \sim N_p(\theta, \sigma^2 I_p)$ alors,

$$\begin{aligned} f(x|\theta) &= \frac{1}{\sigma^p} \frac{1}{(2\pi)^{p/2}} \exp\left\{-\sum_{i=1}^p (x_i - \theta_i)^2 / 2\sigma^2\right\} \\ &= C(\theta, \sigma) h(x) \exp\{x \cdot (\theta/\sigma^2) + \|x\|^2 (-1/2\sigma^2)\} \end{aligned}$$

et la distribution normale appartient à une famille exponentielle de paramètres naturels θ/σ^2 et $-1/2\sigma^2$. De la même façon, si $x_1, \dots, x_n \sim N_p(\theta, \sigma^2 I_p)$, la distribution jointe satisfait

$$f(x_1, \dots, x_n) = C'(\theta, \sigma) h'(x_1, \dots, x_n) \times \exp\{n\bar{x} \cdot (\theta/\sigma^2) + \sum_{i=1}^n \|x_i - \bar{x}\|^2 (-1/2\sigma^2)\}$$

et la statistique $\bar{x}, \sum_i \|x_i - \bar{x}\|^2$ est exhaustive pour tout $n \geq 2$.

Définition 1.3. Soit $f(x|\theta) = C(\theta)h(x) \exp\{\theta \cdot x\}$, une famille exponentielle naturelle. L'espace naturel des paramètres est

$$N = \left\{ \theta; \int_{\mathcal{X}} e^{\theta \cdot x} h(x) d\mu(x) < +\infty \right\}$$

La famille est dite régulière si N est un ensemble ouvert et minimale si $\dim(N) = \dim(K) = k$, où K est la clôture de l'enveloppe convexe du support de μ .

Remarque 1.1. Les familles exponentielles naturelles peuvent aussi être réécrites sous la forme

$$f(x|\theta) = h(x) e^{\theta \cdot x - \psi(\theta)} \tag{1.4}$$

et $\psi(\theta)$ est dite *fonction cumulée des moments*.

Lois conjuguées des familles exponentielles

Soit $f(x|\theta) = h(x)e^{\theta \cdot x - \psi(\theta)}$, loi générique d’une famille exponentielle. Cette loi admet alors une famille conjuguée.

Proposition 1.1. Une famille conjuguée pour $f(x|\theta)$ est donnée par

$$\pi(\theta|\mu, \lambda) = K(\mu, \lambda)e^{\theta \cdot \mu - \lambda \psi(\theta)} \tag{1.5}$$

où $K(\mu, \lambda)$ est la constante de normalisation de la densité. La loi a posteriori correspondante est $\pi(\theta|\mu + x, \lambda + 1)$.

Remarque 1.2. Seuls les modèles à structure exponentielle admettent une famille conjuguée.

Exemple 1.6. Le tableau suivant contient quelques lois a priori conjuguées usuelles.

$f(x \theta)$	$\pi(\theta)$	$\pi(\theta x)$
Normale $N(\theta, \sigma^2)$	Normale $N(\mu, \tau^2)$	$N(\varrho(\sigma^2\mu + \tau^2x), \varrho\sigma^2\tau^2)$ $\varrho^{-1} = \sigma^2 + \tau^2$
Poisson $P(\theta)$	Gamma $\Gamma(\alpha, \beta)$	$\Gamma(\alpha + x, \beta + 1)$
Gamma $\Gamma(\nu, \theta)$	Gamma $\Gamma(\alpha, \beta)$	$\Gamma(\alpha + \nu, \beta + x)$
Binomiale $B(n, \theta)$	Beta $Be(\alpha, \beta)$	$Be(\alpha + x, \beta + n - x)$
Binomiale Négative $Neg(m, \theta)$	Bêta $Be(\alpha, \beta)$	$Be(\alpha + m, \beta + x)$
Multinomiale $M_k(\theta_1, \dots, \theta_k)$	Dirichlet $D(\alpha_1, \dots, \alpha_k)$	$D(\alpha_1 + x_1, \dots, \alpha_k + x_k)$
Normale $N(\mu, 1/\theta)$	Gamma $\Gamma(\alpha, \beta)$	$\Gamma(\alpha + 0.5, \beta + (\mu - x)^2/2)$

Tab 1.1-Lois a priori conjuguées usuelles.

1.3.1.3 Lois a priori subjectives

Précisons tout d’abord que cette démarche n’est pas forcément facile dans la pratique. L’idée est d’utiliser les données antérieures. Par exemple, dans un cadre paramétrique, cela revient à choisir une valeur particulière du paramètre.

Dans un cas concret, il peut être judicieux de baser son raisonnement sur les dires d’experts, notamment à l’aide de questionnaires. Il est alors nécessaire de veiller à ce que

les questions soient compréhensibles 1, par exemple en prenant comme base les quantiles plutôt que les moments. Pour plusieurs experts, il peut être utile de pondérer leurs réponses et d'utiliser des modèles hiérarchiques.

Ainsi, la difficulté ici n'est pas mathématique mais plus psychométrique pour réduire les biais sur les réponses fournies. Nous allons nous concentrer sur le second aspect de la détermination.

1.3.2 Approche non informative

Lorsque aucune information a priori n'est disponible, le choix de la loi a priori est analytique, puisqu'elles donnent des expressions exactes pour quelques quantités a posteriori. Dans de telles situations, il est impossible de justifier le choix d'une loi a priori sur des bases subjectives. Plutôt que de revenir aux alternatives classiques, comme l'estimation par maximum de vraisemblance, ou d'utiliser les données pour approcher ces hyperparamètres, comme dans une analyse bayésienne empirique, il est préférable de faire appel à des techniques bayésiennes, ne serait-ce que parce qu'elles sont à la base des critères classiques d'optimalité. Dans un tel cas, ces lois a priori particulières doivent être construites à partir de la distribution d'échantillonnage, puisque c'est la seule information disponible. Pour des raisons évidentes, de telles lois sont dites *non informatives*.

Les lois de probabilités non informatives nous amènent souvent à des résultats qui sont des mesures et non des probabilités qu'on appelle des lois impropres c'est-à-dire :

$$\int_{\mathbb{R}} \pi(\theta) d\theta = +\infty$$

l'ensemble des lois a priori impropres constituent un prolongement des lois a priori propres. En effet, elle permettent une bonne description du manque d'information a priori.

Voici quelques approches pour déterminer des lois non informatives :

1.3.2.1 Les lois a priori de Laplace

Laplace fut le premier à utiliser des techniques non informatives puisque, bien que ne disposant pas d'information, il munit ces paramètres d'une loi a priori qui prend en compte

son ignorance en donnant la même vraisemblance à chaque valeur du paramètre, soit donc en utilisant une loi uniforme. Son raisonnement, appelé plus tard principe de la raison insuffisante, se fondait sur l'équiprobabilité des événements élémentaires.

Trois critiques ont été plus tard avancées sur ce choix. Premièrement, les lois résultantes sont impropres quand l'espace des paramètres n'est pas compact et certains statisticiens se refusent à utiliser de telles lois, car elles mènent à des difficultés comme le paradoxe de marginalisation

Deuxièmement, le principe des événements équiprobables de Laplace n'est pas cohérent en termes de partitionnement : si $\Theta = \{\theta_1, \theta_2\}$, la règle de Laplace donne $\pi(\theta_1) = \pi(\theta_2) = 1/2$ mais, si la définition de Θ est plus détaillée, avec $\Theta = \{\theta_1, \theta_2, \theta_3\}$, la règle de Laplace mène à $\pi(\theta_1) = 1/3$, ce qui évidemment n'est pas cohérent avec la première formulation, cette cohérence n'est pas un problème important : il peut être évacué en argumentant que le niveau de partitionnement doit être fixé à un certain stade de l'analyse et que l'introduction d'un degré plus fin dans le partitionnement modifie le problème d'inférence.

La troisième critique est plus fondamentale, car elle concerne le problème de l'invariance par reparamétrisation. Si on passe de $\theta \in \Theta$ à $\eta = g(\theta)$ par une transformation bijective g , l'information a priori reste totalement inexistante et ne devrait pas être modifiée. Cependant, si $\pi(\theta) = 1$, la loi a priori sur η est :

$$\pi^*(\eta) = \left| \frac{d}{d\eta} g^{-1}(\eta) \right|$$

1.3.2.2 Loi a priori de Jeffreys

les lois a priori non informatives de Jeffreys sont fondées sur l'information de Fisher, donnée par

$$I(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 \right]$$

dans le cas unidimensionnel.

Sous certaines conditions, cette information est aussi égale à

$$I(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right]$$

La loi a priori de Jeffreys est

$$\pi_j \propto \sqrt{|I(\theta)|}$$

Ainsi la loi a priori de Jeffreys est invariante par reparamétrisation. Malgré l'immense intérêt d'une telle propriété, il faut savoir que la loi a priori de Jeffreys n'a de bonnes propriétés que dans le cas des petites dimensions et en particulier de la dimension 1. Vérifions que c'est effectivement une loi de probabilité invariante par reparamétrisation : si $\eta = g(\theta)$ avec $g \in \mathcal{C}^1$, alors $\pi_j(\eta) = \sqrt{|I(\eta)|}$. En outre, $\tilde{I}(\eta) = \nabla(\theta)'I(\theta)\nabla g(\theta)$, il s'en suit que dans le cas de la dimension 1, on a :

$$\pi_j(\eta) = \sqrt{|I(\eta)|} = \sqrt{|I(\theta)|} \left| \frac{d\theta}{d\eta} \right| = \pi_j(\theta) \left| \frac{d\theta}{d\eta} \right|$$

Ainsi la loi a priori de Jeffreys est invariante par reparamétrisation. Malgré l'immense intérêt d'une telle propriété, il faut savoir que la loi a priori de Jeffreys n'a de bonnes propriétés que dans le cas des petites dimensions et en particulier de la dimension 1.

Il existe une variante de l'approche précédente : la loi de référence de Bernardo dont l'idée est de reposer sur un critère d'objectivité. On regarde ici la loi a priori qui amène le moins d'information par rapport à ce que pouvait fournir les données. On s'intéresse donc à la fonction de π , la loi a priori, suivante :

$$I_n(\pi) = \int K(\pi(\theta|x^n); \pi(\theta)) m_\pi(x^n) dx^n$$

En outre, on peut vérifier que :

$$I_n(\pi) \sim_{n \rightarrow \infty} \frac{d}{2} \log \frac{n}{2\pi} - \int \pi(\theta) \log \frac{\pi}{\sqrt{|I(\theta)|}} d\theta + o_p(1)$$

Dans le cas d'un modèle régulier $I_n(\pi)$ est donc maximal lorsque n tend vers $+\infty$ si $\pi(\theta) \propto \sqrt{|I(\theta)|}$.

Bernardo introduit la stratégie suivante avec un paramètre θ de la forme $\theta = (\theta_1, \theta_2)$ où θ_1 est un paramètre d'intérêt et θ_2 est un paramètre de nuisance. On raisonne alors séquentiellement en écrivant : $\pi(\theta_1, \cdot) = \pi(\cdot|\theta_1)\pi(\theta_1)$ et en prenant pour θ_1 la loi de Jeffreys. Cette manière de faire peut se généraliser si $\theta = (\theta_1, \dots, \theta_n)$ où l'on a ordonné sans perte de généralité les θ_i par intérêt croissant. Un algorithme consiste à prendre pour θ_1 la loi de Jeffreys, puis de calculer $\pi(\theta_2|\theta_1)$ la plus objective puis $\pi(\theta_3|\theta_1, \theta_2)$ la plus objective, etc. Cependant, il est clair ce raisonnement n'est pas purement objectif parce que donner plus d'importance à un paramètre qu'à un autre relève une fois encore d'un choix.

Exemple 1.7. Soit $x \sim \mathcal{B}(n, p)$,

$$\begin{aligned} f(x/p) &= C_n^x p^x (1-p)^{n-x} \\ \frac{\partial^2 f(x/p)}{\partial^2 p^2} &= \frac{x}{p^2} + \frac{n-x}{(1-p)^2} \\ I(p) &= n \left[\frac{1}{p} + \frac{1}{p-1} \right] = \frac{n}{p(1-p)} \end{aligned}$$

Donc la loi de Jeffreys pour ce modèle est

$$\pi(p) \propto [p(1-p)]^{-1/2}$$

et est alors propre, car il s'agit de la distribution $\mathcal{Be}(1/2, 1/2)$.

1.3.2.3 Loi de référence

Cette technique a été mis au point par Bernardo (1979), c'est une modification de l'approche de Jeffreys dans le cas multidimensionnel, appelée approche de la loi de référence.

La différence qui caractérise cette méthode est que la loi a priori résultante par la méthode de référence ne dépend pas seulement de la loi d'échantillonnage, mais aussi du problème inférentiel considéré.

Quand $x \sim f(x/\theta)$ et $\theta = (\theta_1, \theta_2)$, où θ_1 est le paramètre d'intérêt, la loi de référence est obtenue en définissant d'abord $\pi(\theta_2/\theta_1)$ comme la loi de Jeffreys associée à $f(x/\theta)$ pour θ_1 fixé, puis en calculant la loi marginale

$$\tilde{f}(x/\theta_1) = \int f(x/\theta_1, \theta_2) \pi(\theta_2/\theta_1) d\theta_2$$

et la loi de Jeffreys $\pi(\theta_1)$ associée à $\tilde{f}(x/\theta_1)$.

1.3.2.4 Loi a priori de concordance (matching priors)

Le but est de trouver une loi a priori concernant le paramètre θ qui se rapproche le plus possible de la méthode de choix fréquentiste, cela revient à faire en sorte que le tirage x n'influence pas le résultat.

On rappelle dans un premier temps des exemples d'une région de confiance ou α -crédible.

Elle peut être par exemple un intervalle unilatéral $\{\theta \leq \theta_d^{(x)}\}$ ou bien bilatéral $\{\theta_{\alpha,1} \leq \theta \leq \theta_{\alpha,2}\}$. Il peut s'agir aussi de région HPD, $\theta \in \mathcal{C}_\alpha^\pi$ avec par exemple $\{\log(\hat{\theta}) - \log(\theta) \leq h_\alpha\}$ tel que $P^\pi(\theta \in \mathcal{C}|x) = 1 - \alpha$.

On cherche π tel que $\forall \theta; P_\theta(\theta \in \mathcal{C}) = 1 - \alpha$, appelé la parfaite concordance. C'est en général impossible. On va alors chercher r_n le plus petit possible tel que :

$$\forall \theta \in \Theta; \forall \alpha \in]0, 1[, P_\theta(\theta \in \mathcal{C}) = 1 - \alpha + o(r_n)$$

La loi a priori est alors dite concordante à l'ordre r_n .

1.3.2.5 Lois a priori invariantes impropres

Définition 1.4. Une loi impropre M sur Θ est invariante par transformation sur le paramètre h de Θ dans Θ si M est identique à son image par h définie par Moh^{-1} .

Remarque 1.3. Si M est caractérisée par sa densité π et h est bijective bidérivable, la densité de Moh^{-1} est égale à $\|\rho^{-1}\|(Moh^{-1})$, la condition d'invariance s'exprime par l'égalité :

$$\pi = \|\partial h^{-1}\|(\pi oh^{-1})$$

où $\|\partial h^{-1}\|$ est la valeur absolue du déterminant de la matrice des dérivées partielles.

Exemple 1.8. (*Berger et yang (1995)*).

Dans le modèle $AR(1)$, en prenant la transformation :

$$h : \rho \longmapsto h(\rho) = \frac{1}{\rho}, \quad |\rho| > 1$$

et la loi a priori :

$$\begin{cases} 1/(2\pi\sqrt{(1-\rho^2)}), & \text{si } |\rho| < 1; \\ 1/(2\pi|\rho|\sqrt{(\rho^2-1)}), & \text{si } |\rho| > 1. \end{cases}$$

Dans ce cas, la condition d'invariance par la transformation h sur le paramètre ρ s'exprime par l'égalité

$$\pi(h(\rho)) = \pi(\rho) \left| \frac{\partial h(\rho)}{\partial \rho} \right|^{-1}$$

Exemple 1.9. (*Le choix Bayésien (2002)*).

La famille de lois $f(x - \theta)$ est invariante par translation, car $y = x - x_0$ a une loi de la

même famille pour tout x_0 , $f(y - (\theta - x_0))$, θ est alors dit paramètre de position et une exigence d'invariance est que la loi a priori soit invariante par translation, donc satisfasse

$$\pi(\theta) = \pi(\theta - \theta_0)$$

pour tout θ_0 . La solution est $\pi(\theta) = c$ la loi uniforme sur Θ .

Chapitre 2

Théorie de la décision statistique bayésienne

Introduction

”Prévoir consiste à projeter dans l’avenir ce qu’on a perçu dans le passé.”

Henri Bergson, le possible et le réel (1930)

Dans ce paragraphe, nous allons nous intéresser à la statistique vu d’un autre angle, celui de la prise de décision bayésienne en présence d’informations. On parle aussi de décision en avenir risqué car, bien souvent, les informations sont insuffisantes pour lever complètement les incertitudes.

Nous allons, nous intéresser à la façon dont la décision sera prise, et ceci en l’étudiant sur un exemple illustratif avant de l’appliquer sur des données réelles dans le chapitre suivant.

On se propose de repérer les éléments-clés qui assurent une même expression formelle : la décision, les informations et leurs modes de collecte, la façon d’associer une décision à une information.

2.1 Estimateurs de Bayes

Soit une fonction de coût $L(\theta, \delta)$, et une loi de probabilité a priori (ou une loi impropre) π , pour trouver l’estimateur de Bayes $\delta^\pi(x)$, on applique la règle suivante :

$$\delta^\pi(x) = \min_{\delta} \mathbb{E}^\pi[L(\theta, \delta)/x]$$

L'estimateur $\delta^\pi(x)$ sera déterminé analytiquement ou numériquement ceci dépendra de la fonction de perte, de sa nature et complexité.

Généralement, les solutions associées à des coûts classiques sont formellement connues et correspondent aux caractéristiques usuelles d'une distribution (moyenne, médiane, fractiles, etc.).

Par exemple, l'estimateur de Bayes associé au coût quadratique est la moyenne a posteriori. Cette construction formelle des estimateurs de Bayes classiques n'évite pas toujours le recours à une approximation numérique, particulièrement dans des cas multidimensionnels.

Lemme 2.1. *Soit $f(x/\theta)$ une distribution de probabilité appartenant à une famille exponentielle. Pour toute loi a priori π la moyenne a posteriori de θ est donnée par*

$$\delta^\pi(x) = \nabla \log m_\pi(x) \nabla \log h(x)$$

où ∇ est l'opérateur gradient et m_π est la loi marginale associée à π .

Preuve. *L'espérance a posteriori est donnée par*

$$\begin{aligned} \mathbb{E}^\pi[\theta_i/x] &= \frac{\int_{\Theta} \theta_i h(x) e^{\theta \cdot x - \psi(\theta)} \pi(\theta) d\theta}{m_\pi(x)} \\ &= \left(\frac{\partial}{\partial x_i} \int_{\Theta} h(x) e^{\theta \cdot x - \psi(\theta)} \pi(\theta) d\theta \right) \frac{1}{m_\pi(x)} - \left(\frac{\partial}{\partial x_i} \right) \frac{1}{h(x)} \\ &= \frac{\partial}{\partial x_i} [\log m_\pi(x) - \log h(x)]. \end{aligned}$$

2.1.1 Quelques estimateurs usuelles

Estimateurs de Bayes du paramètre θ sous coût quadratique pour les lois a priori conjuguées des familles exponentielles usuelles.

Loi de x	Loi conjuguée	Moyenne a posteriori
Normale $N(\theta, \sigma^2)$	Normale $N(\mu, \tau^2)$	$\frac{\mu\sigma^2 + \tau^2x}{\sigma^2 + \tau^2}$
Poisson $P(\theta)$	Gamma $\Gamma(\alpha, \beta)$	$\frac{\alpha + x}{\beta + 1}$
Gamma $\Gamma(\nu, \theta)$	Gamma $\Gamma(\alpha, \beta)$	$\frac{\alpha + \nu}{\beta + x}$
Binomiale $B(n, \theta)$	Beta $Be(\alpha, \beta)$	$\frac{\alpha + x}{\alpha + \beta + n}$
Binomiale Négative $Neg(m, \theta)$	Bêta $Be(\alpha, \beta)$	$\frac{\alpha + n}{\alpha + \beta + x + n}$
Multinomiale $M_k(\theta_1, \dots, \theta_k)$	Dirichlet $D(\alpha_1, \dots, \alpha_k)$	$\frac{\alpha_i + x_i}{(\sum_j \alpha_j) + n}$
Normale $N(\mu, 1/\theta)$	Gamma $\Gamma(\alpha, \beta)$	$\frac{\alpha + 1}{\beta + (\mu - x)}$ ²

Tab 2.1-Quelques estimateurs de Bayes usuels.

2.2 Fonction perte et risque

Pour le modèle $X \in \{\chi, \beta, \{P_\theta, \theta \in \Theta\}\}$, on définit D , l'ensemble des décisions possibles. C'est-à-dire l'ensemble des fonctions de Θ dans $g(\Theta)$, où g dépend du contexte :

- si le but est d'estimer θ alors $D = \Theta$
- pour un test, $D = \{0, 1\}$

La fonction de perte est une fonction mesurable de $(\Theta \times D)$ à valeurs réelles positives : $L : \Theta \times D \rightarrow \mathbb{R}_+$. Elle est définie selon le problème étudié et constitue l'armature du problème statistique.

Définition 2.1. Risque fréquentiste

Pour (θ, δ) , le risque fréquentiste est défini par :

$$\begin{aligned} R(\theta, \delta) &= E_\theta[L(\theta, \delta(X))] \\ &= \int_X L(\theta, \delta(X))f(X|\theta)dX \end{aligned}$$

C'est une fonction de θ et ne définit donc pas un ordre total sur D et ne permet donc pas de comparer toutes les décisions et estimateurs. Il n'existe donc pas de meilleur estimateur dans un sens absolu. Ainsi, l'approche fréquentiste restreint l'espace d'estimation en préférant la classe des estimateurs sans biais dans laquelle il existe des estimateurs de risque uniformément minimal; l'école bayésienne ne perd pas en généralité en définissant un risque a posteriori. L'idée est d'intégrer sur l'espace des paramètres pour pallier cette difficulté.

Définition 2.2. Risque a posteriori

Une fois donnée la loi a priori sur le paramètre et la fonction de perte, le risque a posteriori est défini par :

$$\rho(\pi, \delta|X) = E^\pi[L(\theta, \delta)|X] \\ \int_{\Theta} L(\theta, \delta)\pi(\theta|X)d\theta$$

Ainsi, le problème change selon les données; ceci est dû à la non existence d'un ordre total sur les estimateurs.

Définition 2.3. Risque intégré

A fonction de perte donnée, le risque intégré est défini par :

$$r(\pi, \delta) = E^\pi[R(\theta, \delta)] \\ \int_{\Theta} R(\theta, \delta)d\pi(\theta)$$

Définition 2.4. Estimateur bayésien

Un estimateur bayésien est un estimateur vérifiant :

$$r(\pi, \delta^\pi) = \inf_{\delta \in D} r(\pi, \delta) < \infty \quad (2.1)$$

La valeur $r(\pi, \delta^\pi)$ est alors appelée risque de bayes.

Pour obtenir la valeur de l'infimum du risque intégré il faut donc en théorie minimiser une intégrale double δ . L'introduction du risque intégré se justifie par le théorème suivant. Il suffira de minimiser une grandeur qui ne dépend plus que des données, ceci permet donc d'arriver à des estimateurs satisfaisants.

Théorème 2.1. *Méthode de calcul*

Si $\exists \delta \in D, r(\pi, \delta) < \infty$ et $\forall X \in \chi \delta^\pi(X) = \arg \min_{\delta} \rho(\pi, \delta|X)$, alors $\delta^\pi(X)$ est un estimateur bayésien.

2.3 Fonctions perte usuelles

La théorie générale de la décision statistique définit un certain nombre de fonctions de coût dont nous allons toute suite donner un aperçu.

2.3.1 Perte quadratique

On a souvent recours à cette fonction dite "perte quadratique" ou "coût quadratique", elle est donnée par Legendre (1805) et Gauss (1810), son expression est la suivante

$$L(\theta, d) = (\theta - d)^2$$

elle estime le coût consécutif à la décision "d" si le système est dans l'état θ .

Exemple 2.1. *Nous cherchons à évaluer la performance de l'estimateur*

$$\delta(x_1, x_2) = \begin{cases} \frac{x_1+x_2}{2}, & \text{si } x_1 \neq x_2; \\ x_1 + 1, & \text{sinon.} \end{cases}$$

par $\alpha(x_1, x_2)$ sous le critère quadratique

$$[\mathbb{I}_\theta(\delta(x_1, x_2)) - \alpha(x_1, x_2)]^2$$

où

$$\mathbb{I}_\theta = \begin{cases} 1, & \text{si } v = 0; \\ 0, & \text{sinon.} \end{cases}$$

La fonction α estime donc d'une certaine façon la probabilité que δ prenne la vraie valeur θ . Deux estimateurs peuvent être considérés :

(i) $\alpha_0(x_1, x_2) = 0.75$, qui donne l'esperance de $\mathbb{I}_\theta(\delta(x_1, x_2))$; et

(ii)

$$\alpha_1(x_1, x_2) = \begin{cases} 1, & \text{si } x_1 \neq x_2; \\ 0.5, & \text{sinon.} \end{cases}$$

Le risque des deux estimateurs est alors

$$\begin{aligned} R(\theta, \alpha_0) &= \mathbb{E}_\theta[\mathbb{I}_\theta(\delta(x_1, x_2)) - 0.75]^2 \\ &= 0.75 - (0.75)^2 \\ &= 0.1875 \end{aligned}$$

et

$$\begin{aligned} R(\theta, \alpha_1) &= \mathbb{E}_\theta[\mathbb{I}_\theta(\delta(x_1, x_2)) - \alpha_1(x_1, x_2)]^2 \\ &= (0.5)^2 \frac{1}{2} \\ &= 0.125 \end{aligned}$$

Par conséquent, α_1 est un meilleur estimateur des performances de δ que α_0 .

Proposition 2.1. *L'estimateur de Bayes δ^Π associé à la loi a priori π et au coût quadratique est la moyenne a posteriori*

$$\delta^\pi(x) = \mathbb{E}^\pi[\theta|x] = \frac{\int_\theta \theta f(x|\theta)\pi(\theta)d\theta}{\int_\theta f(x|\theta)\pi(\theta)d\theta}$$

Preuve. *Comme*

$$\mathbb{E}^\pi[(\theta - \delta)^2/x] = \mathbb{E}^\pi[(\theta^2/x) - 2\delta\mathbb{E}^\pi[(\theta/x)] + \delta^2]$$

le minimum du coût a posteriori est effectivement atteint par $\delta^\pi(x) = \mathbb{E}^\pi[(\theta/x)]$

Les corollaires suivants s'ensuivent de manières directes.

Corollaire 2.1. *L'estimateur de Bayes δ^π associé à π et au coût quadratique ponderé*

$$L(\theta, \delta) = \omega(\theta)(\theta - \delta)^2$$

où $\omega(\theta)$ est une fonction positive, est

$$\delta^\pi(x) = \frac{\mathbb{E}^\pi[\omega(\theta)\theta/x]}{\mathbb{E}^\pi[\omega(\theta)/x]}$$

Corollaire 2.2. *Quand $\Theta \in \mathbb{R}^p$, l'estimateur de Bayes δ^π associé à π et au coût quadratique*

$$L(\theta, \delta) = (\theta\theta - \delta)^t Q(\theta - \delta),$$

est la moyenne a posteriori, $\delta^\pi = \mathbb{E}^\pi[(\theta/x)]$, pour toute matrice $Q_{(p \times p)}$ symétrique définie positive.

La fonction de perte quadratique est particulièrement intéressante lorsque l'espace des paramètres est borné et le choix d'un coût plus subjectif est impossible. En effet, cette fonction est assez simple d'utilisation et l'erreur d'approximation est alors de faible importance. L'indétermination de la fonction de perte (et son remplacement par une approximation quadratique) est fréquent en évaluation de la précision, qui inclut par exemple l'estimation du coût.

2.3.2 L'erreur de perte absolu

Une autre technique d'estimer la perte est d'utiliser la fonction de perte absolue, son expression s'écrit comme suit :

$$L(\theta, d) = |\theta - d|$$

Coût linéaire par fraction

$$L_{k_1, k_2}(\theta, d) = \begin{cases} k_2(\theta - d), & \text{si } \theta > d; \\ k_1(d - \theta), & \text{sinon.} \end{cases}$$

De telles fonctions croissent plus lentement que le coût quadratique.

Proposition 2.2. *L'estimateur de Bayes associé à la loi a priori π et à la fonction de coût linéaire par morceaux est le fractile $(k_2/(k_1 + k_2))$ de la distribution a posteriori $\pi(\theta/x)$.*

Preuve. *Nous avons*

$$\begin{aligned} \mathbb{E}^\pi[L_{k_1, k_2}(\theta, d)/x] &= k_1 \int_{-\infty}^d (d - \theta)\pi(\theta/x)d\theta + k_2 \int_d^{+\infty} (\theta - d)\pi(\theta/x)d\theta \\ &= k_1 \int_{-\infty}^d P^\pi(\theta < y/x)dy + k_2 \int_d^{+\infty} P^\pi(\theta > y/x)dy \end{aligned}$$

est obtenue par une intégration par parties.

Si on dérive par rapport à d on obtient :

$$k_1 P^\pi(\theta < d/x) - k_2 P^\pi(\theta > d/x)$$

Et donc,

$$P^\pi(\theta < d/x) = \frac{k_2}{k_1 + k_2}$$

Remarque 2.1. Si $k_1 = k_2$, l'estimateur de Bayes est la médiane a posteriori.

2.3.3 Le coût 0 - 1

Proposé par Neyman et Pearson, ce coût est surtout utilisé dans l'approche classique des tests d'hypothèses. En effet, pour ce coût, la pénalité associée à un estimateur δ est 0 si la réponse est correcte et 1 sinon.

Exemple 2.2. Soit le test de

$$\begin{cases} H_0 : \theta \in \Theta, & ; \\ H_1 : \theta \notin \Theta, & . \end{cases}$$

Alors $\mathcal{D} = 0, 1$, où 1 représente l'acceptation de H_0 et 0 son rejet. Pour la fonction de coût 0 - 1, qui vaut

$$L(\theta, d) = \begin{cases} 1 - d, & \text{si } \theta \in \Theta_0; \\ d, & \text{sinon.} \end{cases}$$

le risque associé est

$$R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(\theta))] = \begin{cases} P_\theta(\delta(\theta) = 0), & \text{si } \theta \in \Theta_0; \\ P_\theta(\delta(\theta) = 1), & \text{sinon.} \end{cases}$$

Ce qui donne exactement les erreurs de première et deuxième espèce qui sous-tendent la Théorie de Neyman-Pearson.

Proposition 2.3. L'estimateur de Bayes associé à π et au coût est

$$\delta^\pi(x) = \begin{cases} 1, & \text{si } P(\theta \in \Theta_0/x) > P(\theta \notin \Theta_0/x); \\ 0, & \text{sinon.} \end{cases}$$

donc $\delta^\pi(x)$ vaut 1 si et seulement si $P(\theta \in \Theta_0/x) > 1/2$.

2.3.4 Coût intrinsèque

Nous rencontrons certains cas pratiques, où il y a absence totale d'informations, d'où l'impossibilité d'une paramétrisation d'un modèle pour ces cas, et de même une fonction de coût inconnue.

Dans un tel contexte non informatif, il semble naturel d'utiliser des coûts comparant directement les distributions $f(\cdot/\theta)$ et $f(\cdot/\delta)$ associées au vrai paramètre θ et l'estimateur

δ . De telles fonctions de coût,

$$L(\theta, \delta) = d(f(\cdot/\theta), f(\cdot/\delta)),$$

sont effectivement indépendantes de la paramétrisation. Deux distances standard pour les distributions sont :

1. La distance entropique

$$L_e(\theta, \delta) = \mathbb{E}_\theta[\log(\frac{f(x/\theta)}{f(x/\delta)})]$$

Dite aussi divergence de Kullback-Leibler et qui n'est pas une distance au sens mathématique à cause de son asymétrie ;

2. La distance de Hellinger

$$L_H(\theta, \delta) = \frac{1}{2} \mathbb{E}_\theta[(\sqrt{\frac{f(x/\theta)}{f(x/\delta)}} - 1)^2]$$

Exemple 2.3. Soit $x \sim \mathcal{N}(0, 1)$. On a alors

$$L_e(\theta, \delta) = \frac{1}{2} \mathbb{E}_\theta[-(x - \theta)^2 + (x - \delta)^2] = \frac{1}{2}(\delta - \theta)^2 ;$$

$$L_H(\theta, \delta) = 1 - \exp\{-(\delta - \theta)^2/8\}.$$

Dans le cas normal où $\pi(\theta/x)$ est une loi $\mathcal{N}(\mu(x), \sigma^2)$, il est trivial de démontrer que l'estimateur de Bayes est $\mu(x)$ dans les deux cas.

Remarque 2.2. Le coût de Hellinger est le plus souvent utilisé dans les cas des distributions usuelles où il mène à des expressions explicites de $L_H(\theta, \delta)$, malheureusement il ne permet pas le calcul explicite des estimateurs de Bayes. En revanche, pour les familles exponentielles, le coût entropique fournit un estimateur explicite qui est la moyenne a posteriori du paramètre.

2.4 Quelques alternatives

Lorsque la fonction de coût n'a pu être entièrement déterminée, on peut supposer qu'elle appartient à une famille paramétrique de fonctions de coût, le décideur choisissant le paramètre le plus approprié. Mis à part les coûts L_p , deux autres possibilités sont

(i)

$$L_1(\theta, \delta) = \log(\alpha\|\theta - \delta\|^2 + 1),$$

(ii)

$$L_2(\theta, \delta) = 1 - \exp\{-c\|\theta - \delta\|^2\}.$$

2.5 Admissibilité et minimaxité

Définition 2.5. Estimateur randomisé

Pour le modèle $X \in \{\chi, \beta, \{P_\theta, \theta \in \Theta\}\}$, un ensemble de décisions D , on définit D^* comme l'ensemble des probabilités sur D . $\delta^* \in D^*$ est appelé estimateur randomisé.

L'idée à l'origine de cette notion est de rendre D convexe pour pouvoir maximiser facilement.

Théorème 2.2. *Pour toute distribution a priori π sur Θ , le risque de Bayes pour l'ensemble des estimateurs randomisés est le même que celui pour l'ensemble des estimateurs non randomisés, soit*

$$\inf_{\delta \in D} r(\pi, \delta) = \inf_{\delta^* \in D^*} r(\pi, \delta^*) = r(\pi)$$

2.5.1 Minimaxité

Le critère de minimaxité apparaît comme une assurance contre le pire, car il vise à minimiser le coût moyen dans le cas le moins favorable. Il représente aussi un effort fréquentiste pour éviter de recourir au paradigme bayésien, tout en engendrant un ordre (faible) sur D^* .

Définition 2.6. On appelle risque minimax

$$\bar{R} = \inf_{\delta \in D^*} \sup_{\theta} R(\theta, \delta) = \inf_{\delta \in D^*} \sup_{\theta} \mathbb{E}_{\theta}[L(\theta, \delta(x))]$$

et estimateur minimax tout estimateur δ_0 tel que

$$\bar{R} = \sup_{\theta} R(\theta, \delta_0)$$

L'estimateur minimax correspond au point de vue de faire le mieux dans le pire des cas, c'est-à-dire à s'assurer contre le pire. Il est utile dans des cadres complexes mais trop conservateur dans certains cas où le pire est très peu probable. Il peut être judicieux de voir l'estimation comme un jeu entre le statisticien (choix de δ) et la Nature (choix de θ), l'estimation minimax rejoint alors celle de la Théorie des Jeux.

règle minimax et stratégie maximin

Une difficulté importante liée à la notion de minimaxité est que les estimateurs minimax n'existent pas nécessairement. En particulier, il existe une stratégie minimax quand Θ est fini et la fonction de coût est continue. Plus généralement, Brown (1976) (voir aussi Le Cam, 1986, et Strasser, 1985) considère l'espace de décision D comme plongé dans un autre espace de telle manière que l'ensemble des fonctions de risque sur D soit compact dans ce grand espace. Dans cette perspective et sous des hypothèses supplémentaires, il est alors possible de construire des estimateurs minimax lorsque la fonction de coût est continue.

Théorème 2.3. *Si $D \subset \mathbb{R}^k$ est convexe et compact et si $L(\theta, d)$ est continue et convexe en tant que fonction de d , pour chaque $\theta \in \Theta$, il existe un estimateur minimax non randomisé.*

Lemme 2.2. *Le risque de Bayes est toujours plus petit que le risque minimax,*

$$\underline{R} = \sup_{\pi} r(\pi) = \sup_{\pi} \inf_{\delta \in D} r(\pi, \delta) \leq \bar{R} = \inf_{\delta \in D^*} \sup_{\theta} R(\theta, \delta)$$

La première valeur est dite *risque maximin* et une distribution π^* telle que $r(\pi^*) = \underline{R}$ est appelée *distribution a priori la moins favorable*, quand de telles distributions existent. En général, la borne supérieure $r(\pi^*)$ est atteinte plutôt par une distribution impropre pouvant s'exprimer comme une limite de distributions a priori propres π_n . Mais ce phénomène n'empêche pas nécessairement la construction d'estimateurs minimax. Quand elles existent, les distributions les moins favorables sont celles qui ont le risque de Bayes le plus grand, donc aussi les moins intéressantes en terme de coût lorsqu'elles ne sont pas suggérées par l'information a priori disponible. Le résultat ci-dessus est assez logique au sens où l'information a priori ne peut qu'améliorer l'erreur d'estimation, même dans le pire des cas.

Définition 2.7. Un problème d'estimation est dit *admettre une valeur* si $\underline{R} = \overline{R}$, c'est-à-dire quand

$$\sup_{\pi} \inf_{\delta \in D} r(\pi, \delta) = \inf_{\delta \in D^*} \sup_{\theta} R(\theta, \delta)$$

Quand le problème admet une valeur, certains estimateurs minimax sont des estimateurs de Bayes correspondant aux lois a priori les moins favorables. Cependant, ils peuvent être randomisés. Par conséquent le principe minimax ne fournit pas toujours des estimateurs acceptables.

Lemme 2.3. Si δ_0 est un estimateur de Bayes pour π_0 et si $R(\theta, \delta_0) \leq r(\pi_0)$ pour tout θ dans le support de π_0 , δ_0 est minimax et π_0 est la distribution la moins favorable.

2.5.2 Admissibilité

Définition 2.8. Estimateur admissible

Soit $X \in \{\chi, \beta, \{P_{\theta}, \theta \in \Theta\}\}$ un modèle paramétrique et L une fonction de perte sur $\Theta \times D$ où D est l'ensemble des décisions. On dit que $\delta \in \Theta$ est inadmissible si et seulement si $\exists \delta_0 \in D, \forall \theta \in \Theta, R(\theta, \delta) \geq R(\theta, \delta_0)$ et $\exists \theta_0 \in \Theta, R(\theta_0, \delta) > R(\theta_0, \delta_0)$. Dans le cas contraire, δ est admissible.

Proposition 2.4. S'il existe un unique estimateur minimax, cet estimateur est admissible.

Notons que la réciproque de ce résultat est fautive, car il peut exister plusieurs estimateurs minimax admissibles. Par exemple, dans le cas $N_p(\theta, I_p)$, il existe des estimateurs de Bayes réguliers minimax pour $p \geq 5$. Quand la fonction de coût L est absolument convexe (en d), la caractérisation suivante est aussi possible.

Proposition 2.5. Si δ_0 est admissible de risque constant, δ_0 est l'unique estimateur minimax.

Théorème 2.4. Estimateurs bayésiens admissibles

Si l'estimateur bayésien δ^{π} associé à une fonction de perte L et une loi a priori π est unique, alors il est admissible.

Proposition 2.6. Si un estimateur de Bayes, δ^{π} , associé à une loi a priori (propre ou impropre) π , est tel que le risque de Bayes,

$$r(\pi) = \int_{\Theta} R(\theta, \delta^{\pi}) \pi(\theta) d\theta$$

soit fini, δ^{π} est admissible.

Définition 2.9. π -admissibilité

Un estimateur δ_0 est π -admissible si et seulement si

$$\forall(\delta, \theta), R(\theta, \delta) \leq R(\theta, \delta_0) \Rightarrow \pi(\{\theta \in \Theta, R(\theta, \delta) < R(\theta, \delta_0)\}) = 0$$

Propriété 2.1. Tout estimateur bayésien tel que $r(\pi) < \infty$ est π -admissible.

Théorème 2.5. *Continuité et π -admissibilité*

Si $\pi > 0$ sur Θ , $r(\pi) < \infty$ pour une fonction de perte L donnée, si δ^π estimateur bayésien correspondant existe et si $\theta \mapsto R(\theta, \delta)$ est continu, alors δ^π est admissible.

2.6 Prise de décision bayésienne

2.6.1 Ensemble A des décisions

On l'appelle aussi l'ensemble des alternatives. L'ensemble A explicite toutes les actions applicables sur l'étude. On note "a" une alternative, et l'ensemble A des alternatives "a" peut-être discret et, s'il est de cardinal "n" fini, on énumérera alors les décisions possibles :

$$A = \{a_j, j = 1, 2, \dots, n\}$$

La détermination de l'ensemble A est un point essentiel de notre étude, puisqu'il faut précisément inclure toutes les décisions possibles et raisonnables à prendre pour le problème posé.

2.6.2 Ensemble X des informations

Il s'agit de la détermination des informations, encore appelées données, observations ou résultats expérimentaux, les notations x, y sont réservées à ces quantités.

2.6.3 Recueil d'information

L'information x est souvent obtenue après une expérimentation ou une campagne de collecte d'information, notée $e \in E$ dont l'espace même des observations dépend, et, par la suite, nous noterons l'information X_e pour souligner cette dépendance.

2.6.4 Forme de décision

Le processus de décision peut prendre deux formes selon la disponibilité des informations, ces deux formes sont :

- **Analyse extensive** (ou a posteriori) Ici, les données ont été observées : le décideur est supposé connaître l'information x . Pour choisir la décision optimale $a \in A$ dont il a inventorié les conséquences, il doit s'intéresser aux incertitudes associées. Ces incertitudes sont en partie (mais en partie seulement) levées par l'information x .
- **Analyse normale** (ou prédictive) Ici, la donnée x n'a pas encore été observée, et le décideur est dans une situation plus complexe, car ne disposant pas de l'information x , il doit d'abord choisir le type d'expérimentation "e" qui lui fournira x_e . Souvent il dispose déjà d'une certaine information mais s'interroge sur l'opportunité de la compléter.

Ainsi doit-il utiliser une procédure de décision séquentielle :

$$\text{choix de } e \in E \rightarrow \text{obtenir } x_e \in X_e \rightarrow \text{choix de } a \in A$$

Au début de cette procédure, le décideur doit choisir une "e" mais les conséquences de son choix dépendront de sa décision "a" finale, qui est non seulement soumise aux incertitudes sur les conséquences mais aussi - et c'est nouveau par rapport à la situation précédente de l'analyse extensive - aux incertitudes sur l'information x_e que pourrait générer l'expérimentation "e".

2.7 Représentation probabiliste des connaissances

Dans le but d'obtenir un bon résultat, ainsi qu'une bonne règle de décision, il faut savoir traduire le caractère stochastique de chaque situation, qui se présente à nous. L'objet de la théorie de la décision statistique est de rechercher un comportement décisionnel cohérent minimisant les chances d'occurrences de situation où une décision censée mène à un résultat indésirable.

Toutes les informations récoltées quantitatives et qualitatives que peut recevoir un décideur sont formalisées à travers un couple comprenant d'une part le modèle statistique

de représentation du phénomène étudié, et d'autre part, le modèle de représentation des connaissances a priori. Ici, l'outil mathématique essentiel est le modèle. Un modèle est une construction mentale qui a pour but la traduction opérationnelle d'un ensemble de connaissances à des fins de déduction.

2.8 Etapes de la construction d'un modèle bayésien

2.8.1 Représentation probabiliste

Les connaissances que l'on doit représenter (et quantifier grâce au calcul des probabilités) concernent les observables et les inconnues.

Remarque 2.3. Dans la suite du document, la notation $[\]$ désigne une loi de probabilité : il s'agira de la fonction de densité de probabilité pour les variables aléatoires continues et de la fonction de répartition pour les variables discrètes.

Remarque 2.4. On utilisera les notations suivantes :

- La loi des probabilités conditionnelles s'écrira $[A, B] = [A/B][B]$.
- Les paramètres sont notés conventionnellement par une lettre grecque θ et les observations par des lettres latines minuscules (généralement x ou y).

2.8.2 Résultats expérimentaux

En général, un statisticien est intéressé par la régénération des résultats expérimentaux en utilisant un modèle statistique paramétrique de représentation du phénomène. Tous les paramètres inconnus dans ce modèle forment *l'état de la nature*.

Le modèle permettra de quantifier le risque généré par chacune des décisions possibles applicables. Un modèle comprend généralement trois étapes importantes :

- Exprime des hypothèses fondatrices qui organisent une représentation simplifiée du phénomène,

- Contient une formulation quantifiée des relations possibles entre grandeurs d'intérêt,
- comprend une partie probabiliste qui sert à représenter les influences échappant au savoir actuel.

Les écarts entre réalité et le modèle dit écarts aléatoires seront décrits par des distributions de probabilité. La conséquence importante de cette opération est de doter l'ensemble X des informations d'une structure probabiliste, décrite par une famille de lois de probabilités indexée par le paramètre θ . Par abus de notations, on notera encore X l'observable.

2.8.3 Les inconnus θ

D'autres connaissances, en général, de nature plus qualitative et provenant d'une expertise, portent sur des grandeurs non observables. Pour qu'en pratique ces connaissances aient trait au problème, il faut bien sûr qu'elles soient liées à l'état de la nature θ . Une distribution de probabilité $[\theta]$, dite loi a priori ou prior, notée aussi $\pi(\theta)$, traduit la plus ou moins grande incertitude associée aux grandeurs non observables θ : la probabilité, dite alors subjective, est une mesure de l'engagement personnel en termes de pari à miser sur telle ou telle valeur de l'évènement incertain. Dans ce cas, il s'agit d'un modèle d'expertise, la probabilité s'interprète comme un degré de crédibilité des valeurs que peut prendre la grandeur incertaine θ . Par conséquent, elle ne possède de sens que si elle est définie conditionnellement à un niveau fixe de connaissance et peut changer quand l'état de connaissance change. Deux individus différents peuvent bien sûr ne pas partager le même prior.

2.9 Modèle statistique paramétrique

Les statisticiens classiques tirent beaucoup moins parti des connaissances a priori et désignent sous le nom de modèle tout court la représentation formelle des seules grandeurs observables, c'est-à-dire le modèle statistique d'occurrence des données. Pour comprendre et utiliser ensuite ce point de vue par lequel nous commençons, il suffit de se mettre dans la situation où l'état de la nature θ prend une valeur unique et précise. Pour mieux cerner cette hypothétique existence d'une telle grandeur inconnue, nous essayerons tant que possible de donner un sens phénoménologique à la grandeur θ . Le résultat de ce travail conceptuel est la loi de probabilité $\pi(x/\theta)$ qui a reçu le nom de modèle statistique (d'occurrence des observables). Cette loi donne, conditionnellement à une valeur fixée du paramètre θ , la

probabilité de l'évènement $X=x$ (où la densité de probabilité de la variable aléatoire X au point x , sachant θ).

Chapitre 3

Applications numériques des méthodes bayésiennes

Introduction

”La pratique, c’est quand tout fonctionne et que personne ne sait pourquoi.”

Albert EINSTEIN

Ni l’étude, ni la comparaison des méthodes de décision et de prévision bayésiennes ne pourrait être complète sans une mise en oeuvre pratique des résultats théoriques établis. Le but de cette partie est l’application de la théorie générale de la décision statistique à des problèmes réels. Ainsi, ce chapitre sera consacré d’une part à la mise en oeuvre pratique de quelques méthodes vu dans les chapitres précédents et l’analyse des résultats trouvés.

3.1 Choix de l’environnement **R** pour l’analyse numérique

R a été initié dans les années 90 par **Robert Gentleman** et **Ross Ihaka** du Département de Statistique de l’ Université d’Auckland, Nouvelle-Zélande.

R est un système qui est communément appelé langage et logiciel. Il permet de réaliser, entre autre, des analyses statistiques.

Plus particulièrement, il comporte des moyens qui rendent possibles la manipulation des données, les calculs et les représentations graphiques. **R** a aussi la possibilité d’exécuter

des programmes stockés dans des fichiers textes.

En effet **R** possède :

- > un système efficace de manipulation et de stockage des données
- > différents opérateurs pour le calcul sur tableaux, en particulier les matrices
- > un grand nombre d'outils pour l'analyse des données et les méthodes statistiques
- > des moyens graphiques pour visualiser les analyses
- > un langage de programmation simple et performant comportant : conditions, boucles, moyens d'entrées sorties, possibilité de définir des fonctions récursives.

R comporte un grand nombre de procédures statistiques.

Parmi elles, nous avons : les modèles linéaires, les modèles linéaires généralisés, la régression non-linéaire, les séries chronologiques, les tests paramétriques et non-paramétriques classiques, etc.

Il y a également un grand nombre de fonctions fournissant un environnement graphique flexible afin de visualiser et créer divers genres de présentations de données.

Ce logiciel étant de domaine public son point fort est représenté par le développement d'applications, de modules qui sont mis à la disposition de tous les utilisateurs et développeurs.

Son réseau international de développement est en perpétuelle évolution. L'intérêt majeur de **R** est qu'il est ouvert à tous. De ce fait, tout le monde peut "apporter sa pierre à l'édifice".

Cela laisse envisager de phénoménales extensions au système. Le potentiel de **R** semble donc énorme.

Dans ce chapitre, nous allons nous intéresser à la statistique vu d'un autre angle, celui de la prise de décision bayésienne en présence d'informations. On parle aussi de décision en avenir risqué car, bien souvent, les informations sont insuffisantes pour lever complètement les incertitudes.

Nous allons, nous intéresser à la façon dont la décision sera prise, et ceci en l'étudiant sur deux exemples illustratifs.

On se propose de repérer les éléments-clés qui assurent une même expression formelle : la décision, les informations et leurs modes de collecte, la façon d'associer une décision à une information.

3.2 Exemple introductif

Comme introduction, nous allons présenter un cas réel, qui est un article d'un journal, et grâce à la formule de Bayes démontrer que ce qui est annoncé dans ce journal est loin d'être la réalité!!!

Jean-Michel Claverie, Professeur de génomique et bioinformatique médicale, faculté de médecine, Université de la Méditerranée, Marseille et directeur du laboratoire "information génomique et structurale", a développé une analyse intitulée

"La formule de Bayes, ou comment résister à la xénophobie"

Et dans un journal à grand tirage, a publié l'étude suivante :

« 50% des voleurs à la tire sont des immigrés, alors qu'ils ne représentent que 10% de la population. Devons-nous continuer d'accueillir de nouveaux immigrants ou fermer nos frontières ? »

Traduisons ces chiffres en terme probabiliste, et tachons d'étayer rationnellement notre conclusion.

L'énoncé nous donne :

- $p(\text{imm}/\text{vol})=1/2$
- $p(\text{imm})=1/10$

Ce qui nous intéresse, c'est bien évidemment d'estimer la fraction de la population d'immigrés qui –théoriquement- est malhonnête. Il nous manque une donnée : la fraction

de voleurs dans la population générale en France ; disons $p(\text{vol}) = 1/10000$.

Nous pouvons donc utiliser le retournement bayésien :

$$P(\text{vol}/\text{imm}) = \frac{P(\text{imm}/\text{vol})P(\text{vol})}{P(\text{imm})}$$

Application numérique :

$$P(\text{vol}/\text{imm}) = \frac{0.5 * 0.0001}{0.10} = \frac{1}{2000}$$

La fraction de voleur dans la population immigrée est donc de $1/2000$. Renvoyer tous les immigrés chez eux revient donc à déporter 1999 personnes « innocentes » pour une malhonnête, une mesure que l'on peut considérer comme (relativement) injuste !

Certes, la proportion de délinquant reste plus 5 fois plus grande que dans la population générale (ce qui peut simplement traduire une répartition socio-économique défavorable), mais on est loin des chiffres (50% de délinquants) suggérés par la Une du journal !

Pour montrer que ces formules correspondent à une réalité, voici une population respectant ces valeurs :

Total population	Immigrés	Dont voleurs	Natifs	Dont voleurs
100.000	10.000	5	90.000	5

On a bien $P(\text{imm}) = 1/10$, $P(\text{vol}) = 1/10000$, $P(\text{imm}/\text{vol}) = 1/2$, et $P(\text{vol}/\text{imm}) = 1/2000$

Total honnêtes	Dont Immigrés	Dont natifs
99.990	9995	89995

Les immigrés rencontrant des difficultés d'intégration économique, leur fréquence relative tend à être plus grande dans les couches sociales à faible revenu. Le dénuement favorisant le vol (sans l'excuser bien sûr au plan moral), on peut facilement trouver des distributions pour lesquelles la proportion de délinquance immigrée est plus faible que pour la population native **à revenu égal** :

Revenu	immigrés	Dont voleur	P(vol/im)	natif	Dont voleur	P(vol/nat)
Bas	6000	5	0.00083	5000	5	0.001
moyen	3000	0	0	55000	0	0
fort	1000	0	0	30000	0	0

Le fait que la population immigrée puisse être plus vertueuse ($0.00083 < 0.001$) que la population native dans chaque tranche de revenu, n'est clairement pas intuitive dans l'énoncé initial.

Dans d'autres cas, la population immigrée peut être tantôt plus ou moins vertueuse que la population native :

Revenu	immigrés	Dont voleur	P(vol/im)	natif	Dont voleur	P(vol/nat)
Bas	6000	3	0.0005	5000	5	0.001
moyen	3000	2	0.00066	55000	0	0
fort	1000	0	0	30000	0	0

Toutes les répartitions sont donc possibles, et l'énoncé initial peut refléter des réalités sociales très différentes.

Conclusion :

1. Il faut se méfier des raisonnements simplistes et des Unes !
2. La frange délinquante d'une minorité ethnique peut avoir une influence négative disproportionnée sur la réputation de l'ethnie toute entière,
3. Il faut pratiquer chaque matin un « retournement bayésien » comme gymnastique mentale, et développer ainsi notre résistance à la manipulation par les médias.

3.3 L'avalanche de Montroc (Boreux et al 2010)

3.3.1 Présentation du problème de décision à traiter

Le 9 février 1999, une avalanche meurtrière (douze décès) a détruit une partie du hameau de Montroc près de Chamonix (France). Cette coulée de neige a englouti vingt-trois chalets dans une zone déclarée «constructible», car considérée comme hors d'atteinte d'après la cartographie des risques établie en 1992. En fait, avant la date fatidique, la dernière avalanche sur ce site avait été observée en 1945. Toutefois, selon Le journal Dauphiné libéré, cinq avalanches survenues entre 1843 et 1945 n'auraient pas été prises en compte, après cet incident Le maire de Chamonix a été condamné à 3 mois de prison avec sursis le 17 juillet 2003.



FIG. 3.1 – Avalanche de Montroc.

3.3.2 Ensemble des décisions

Nous sommes en 1992, c'est-à-dire avant que la catastrophe ne se soit produite, et nous devons trancher sur la question de la constructibilité du site ou non, c'est-à-dire accepter ou refuser de déclarer la zone constructible.

Nous disposons de certain faits, c'est-à-dire certaines informations ou données :

- la dernière avalanche a été observée en 1945,
- cinq autres avalanches ont affecté le site entre 1843 et 1945 (on ignore les années).

Décision possible à prendre :

$d_1 \equiv$ déclarer la zone constructible \Leftrightarrow perdre C_1 Millions d'euros si le site subit une avalanche dans les h prochaines années.

$d_2 \equiv$ refuser le projet \Leftrightarrow perdre C_2 Millions d'euros si le site ne subit aucune avalanche dans les h prochaines années, faute de non-recettes.

Après ceci, nous devons donc fixer le h (nombre d'années), et essayer d'évaluer $\mathbf{p}(h)$, qui représente la probabilité d'observer au moins une avalanche dans la période fixée h .

Nous pouvons donc conclure, d'après ce que nous savons que :

$$E(C_t/d_1, h) = p(h) \times C_1 \quad (3.1)$$

$$E(C_t/d_2, h) = (1 - p(h)) \times C_2 \quad (3.2)$$

Ceci peut se résumer dans le tableau suivant :

Coûts	Etat de la nature θ	
	$p(h)$	$1-p(h)$
d_1	C_1	0
d_2	0	C_2

Tab 3.1 - Montroc : pertes associées aux décisions.

Règle de décision :

On a donc le résultat suivant

$$\frac{E(C_t/d_1, h)}{E(C_t/d_2, h)} = r = \frac{p(h)}{(1-p(h))} \times \frac{C_1}{C_2} \quad (3.3)$$

qui fournit une règle de décision rationnelle

$$r < 1 \Rightarrow d_1, \quad r \geq 1 \Rightarrow d_2 \quad (3.4)$$

Remarque 3.1. il est possible de discuter la valeur du rapport C_1/C_2 , car r augmente avec lui. Ainsi, sous les mêmes hypothèses, dès que le rapport des coûts vaut 1 il faut recommander d_2 . On peut d'ailleurs faire une analyse de sensibilité sur ce rapport.

3.3.3 Modélisation probabiliste

Si on pose que dans une année prise au hasard s'est produite une avalanche, alors on codera cette année avec le chiffre (1), et dans le cas contraire où dans une année prise au hasard, il n'y a eu aucun évènement, celle-ci sera donc codée (0), et nous savons qu'à Montroc, il y a eu six années où il y a eu des incidents donc codée (1).

Dans ce qui suit, nous étudierons le modèle binomial.

Remarque 3.2. D'autres modèles d'échantillonnage sont possibles. En effet, il ne faut pas confondre l'absence d'information avec l'absence réelle d'avalanche, car on peut très bien imaginer que des coulées de neige n'aient pas été enregistrées. La modélisation de ce modèle de données manquantes sort du cadre de ce mémoire.

Le processus de Bernoulli

A chaque année t , on associe une variable aléatoire de Bernoulli, disons y_t , qui prend la valeur 1 avec la probabilité π_t si le site de Montroc subit au moins une avalanche grave et la valeur 0 avec la probabilité complémentaire dans le cas contraire. Si on considère que ces variables aléatoires sont indépendantes et identiquement distribuées ($\forall t : \pi_t = \pi$), la suite y_t constitue un processus de Bernoulli.

$$\forall t : [y_{t+1}/y_t, \pi] = [y_{t+1}/\pi]$$

Le modèle binomial

Puisque chaque année est représentée par une variable aléatoire de Bernoulli, leur somme

$$x = \sum_{t=1}^n y_t$$

est une variable aléatoire binomiale, de paramètres n , π , dont la densité s'écrit :

$$[x/\pi, n] = C_n^x \pi^x (1 - \pi)^{n-x} \quad (3.5)$$

où

$$C_n^x = \frac{n!}{(n-x)!x!}$$

3.3.4 Inférence bayésienne

Ayant imaginé un processus susceptible de générer les observations, c'est-à-dire le processus de Bernoulli qui aboutit au modèle binomial, il faut maintenant estimer son paramètre caractéristique.

La règle de Bayes dit comment réactualiser cette expertise disposant des données : il suffit de multiplier la vraisemblance par le prior. La distribution a posteriori du paramètre, ou posterior, implique la normalisation de ce produit.

La vraisemblance

Pour l'avalanche de Montroc, le modèle d'observation est la loi binomiale. La vraisemblance est donc immédiate

$$[x/\theta, n] \propto \theta^x (1 - \theta)^{n-x} \quad (3.6)$$

Choix du prior et application de la règle de Bayes

Quand on regarde la vraisemblance, on reconnaît immédiatement la signature fonctionnelle d'une densité de probabilité bêta.

On dit qu'un prior bêta est conjugué à une vraisemblance binomiale. La conjugaison a déjà été abordée au chapitre 1.

Bien sûr, il faut préciser les paramètres du prior bêta, disons $a > 0$ et $b > 0$:

$$[\theta/a, b] \propto \theta^{a-1} (1 - \theta)^{b-1} \quad (3.7)$$

Les paramètres des lois a priori découlent de l'expertise reconnue et sont ponctuels, c'est-à-dire sans incertitude.

Comment déterminer les hyperparamètres d'un prior bêta ?

le paramètre θ représente la probabilité qu'une année calendaire, choisie au hasard, voit au moins une avalanche dévaler sur le site de Montroc. Ces années « noires » sont plutôt rares, sinon le problème de décision n'aurait aucun sens.

Biensur, nous ne pouvons nous passer de l'avis d'un spécialiste des avalanches.

Un expert accorde une chance sur dix à θ de dépasser la valeur 0.05 et cinq chances sur cent, d'être inférieure à la valeur 0.01, l'expert justifie ces informations apportées, mais elles témoignent aussi de son savoir et de son expérience.

A partir de ces données, nous pouvons donc utiliser une méthode numérique pour déterminer les hyperparamètres de la loi bêta.

Voici la méthode donnée par l'auteur.

Cas où L'expert donne deux quantiles de π :

Par construction, on a obtenu les quantiles π_q et π_p :

$$[\pi \leq \pi_q] = q, \quad [\pi \leq \pi_p] = p$$

Avec les notations de **R**, pbeta est la fonction de répartition de la distribution bêta. Les hyperparamètres recherchés sont les solutions du système suivant :

$$\begin{cases} p - \text{pbeta}(\pi_p, r, s) = 0, & ; \\ q - \text{pbeta}(\pi_q, r, s) = 0, & . \end{cases}$$

Il faut bien sur disposer d'un solveur numérique, et dans notre cas sa sera le **R**.

En utilisant, le même système proposé par l'auteur et en remplaçant avec les données proposées par l'expert nous obtenons le système d'équations suivant :

$$\begin{cases} p(\theta \geq 0.05) = 1/10, & ; \\ p(\theta \geq 0.01) = 5/10, & . \end{cases}$$

l'expert nous fourni donc les deux quantiles $\theta_{90} \simeq 0.05$ et $\theta_5 \simeq 0.01$.

Voici l'écriture sous R su système à résoudre :

$$\begin{cases} 1/10 - \text{pbeta}(0.9, a, b) = 0, & ; \\ 5/10 - \text{pbeta}(0.05, a, b) = 0, & . \end{cases}$$

A partir de ceux-ci, on détermine les hyperparamètres a et b : $a \simeq 3.82$ et $b \simeq 124.1$.

La règle de Bayes réactualise cette expertise en tenant compte des données : $x=6$ pour $n=150$.

En utilisant ce que nous avons vu dans le chapitre 1, dans les familles conjuguées des familles exponentielles, nous disposons d'un prior $\pi(\theta)=\text{Bêta}(3.82,124.1)$, le posterior $\pi(\theta/x)$ est encore une densité bêta (intérêt de la conjugaison), dont les paramètres intègrent l'expertise et les observations, c'est-à-dire toute l'information disponible :

$$[\theta/n, a, b] \sim \text{dbeta}(\theta/x + a, n - x + b) \quad (3.8)$$

$$[\theta/n, a, b] \sim \text{dbeta}(\theta/6 + 3.82, 150 - 6 + 124.1)$$

$$[\theta/n, a, b] \sim \text{dbeta}(\theta/9.82, 268.1)$$

La figure 3.1 montre le prior et le posterior ainsi obtenus.

Le problème décisionnel, autoriser ou refuser de déclarer la zone constructible, est posé dans une perspective prédictive, c'est-à-dire que cette décision concerne les h années futures.

Nous commencerons par donner une définition de la distribution prédictive a posteriori.

La distribution prédictive a posteriori

Nous voulons déterminer la probabilité d'observer Y années "noires" dans les h prochaines années sachant que, dans le passé, on en a réellement observé x en n années.

Définition 3.1. Soit Y est une variable aléatoire discrete prenant ses valeurs dans l'ensemble $\Omega(Y) = \{0, 1, \dots, n\}$.

La distribution prédictive a posteriori donne les chances de chacune des occurrences $y \in \Omega(Y)$ en impliquant les données connues (a, b, n, x) et l'horizon de prévision envisagé h . On l'obtient en intégrant la distribution jointe de Y et θ sur toutes les valeurs possibles de θ .

$$[Y = y/h, a, b, n, x] = \int_{\Theta} [Y/\theta, h][\theta/n, n, a, b] d\theta \quad (3.9)$$

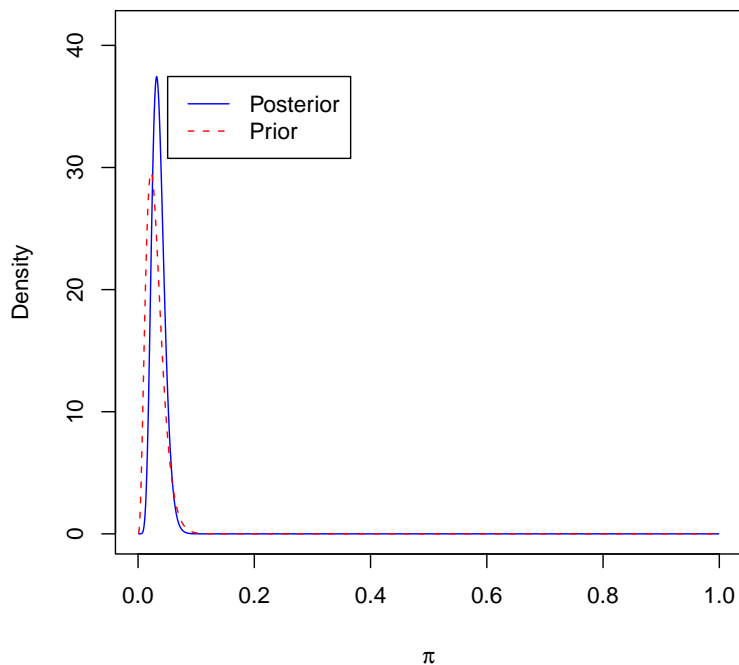


FIG. 3.2 – **Avalanche de Montroc : Distribution a priori $\mathcal{B}e(3.82, 124.1)$ et a posteriori $\mathcal{B}e(9.82, 268.1)$**

- $[Y/\theta, h]$ est la probabilité de y donnée par la loi binomiale équation (3.5).
- $[\theta/n, n, a, b]$ est le posterior bêta obtenu ci-dessus équation (3.8).

Le calcul de la distribution prédictive a posteriori est réalisé en intégrant un produit de distributions de probabilité, de notre modèle précédent c'est-à-dire :

- Vraisemblance \sim binomiale,
- Posterior \sim Bêta.

qu'on appelle aussi le modèle *bêta-binomiale*.

De cette intégration il resultera, une distribution de *Polya* :

$$[Y = y/h, a, b, n, x] = C_h^y \times \frac{B(y + x + a, h - y + n - x + b)}{B(x + a, n - x + b)} \quad (3.10)$$

La figure 3.2 montre la distribution de *Polya* en faisant varier quatre fois le nombre d'années "h" à prévoir.

Exemple 3.1. *La probabilité d'observer au moins une année « noire » à l'horizon h est le complément de n'en observer aucune*

$$\begin{aligned}
 p(h) &= [Y \geq 1/h, a, b, n, x] \\
 &= 1 - [Y = 0/h, a, b, n, x] \\
 &= 1 - \frac{B(x+a, h+n-x+b)}{B(x+a, n-x+b)}
 \end{aligned}$$

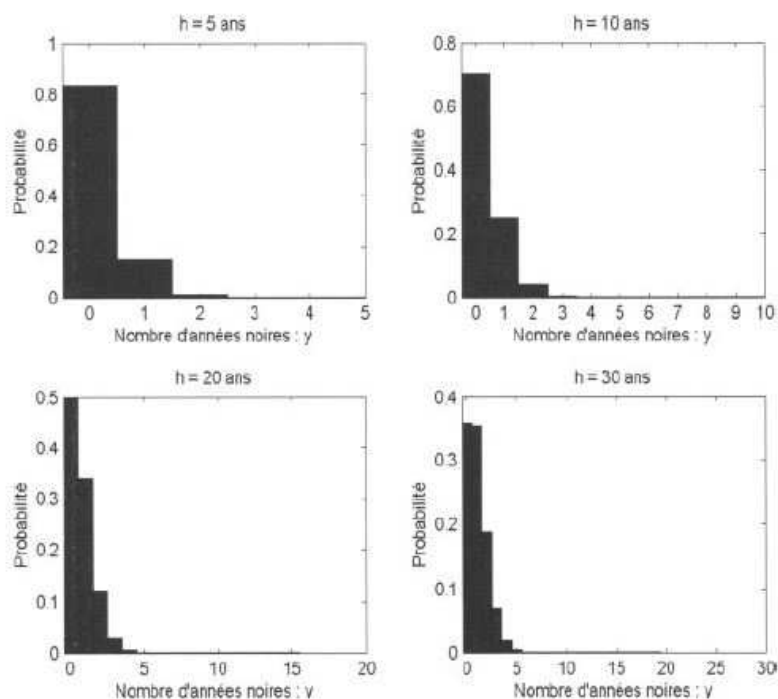


FIG. 3.3 – **Avalanche de Montroc : distribution de Polya pour quatre horizons de prévision.**

Maintenant nous sommes en mesure d'appliquer notre règle de décision (équation 3.3 et équation 3.4), tout en variant C_1/C_2 . En analysant la figure 3.5, on voit que refuser de rendre le site de Montroc constructible (décision d_2) est une décision rationnelle dès que l'on envisage un horizon de prévision compatible avec un projet de lotissement.

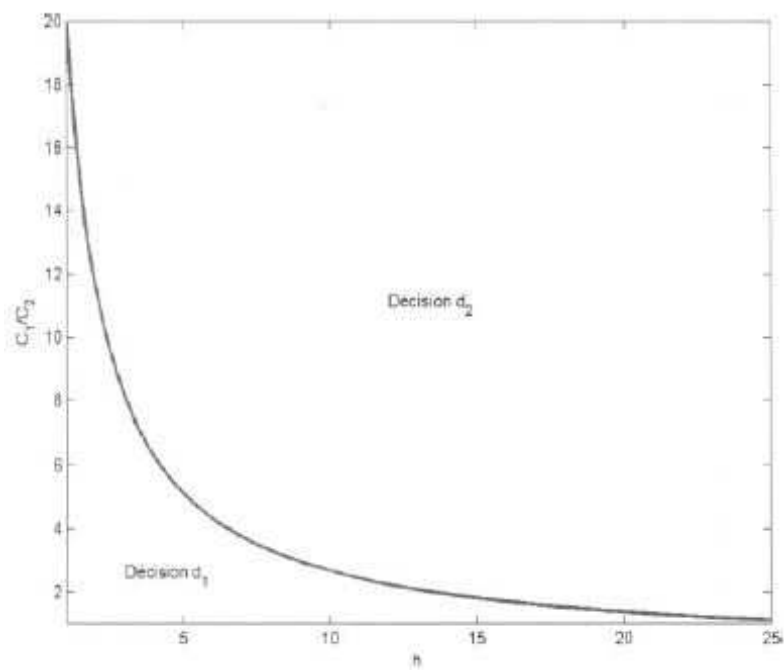


FIG. 3.4 – Avalanche de Montroc : règle de décision.

3.4 Les débits d'une rivière (Parent et Bernier 2007)

3.4.1 Présentation du problème

Les débits d'une rivière comme la Garonne sont très variables. La figure 3.5 présente la série des débits de crue de la Garonne à Mas d'Agenais qui ont dépassé le seuil de $2\,500\text{m}^3/\text{s}$ sur la période 1913-1977. Faut-il construire une digue de protection contre les crues ou considérer que le risque de débordement est acceptable ?

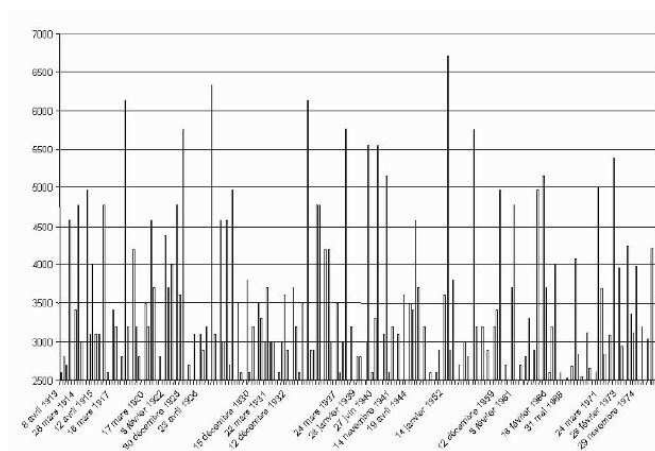


FIG. 3.5 – 151 dépassements des pointes de crue au-delà de $2\,500\text{ m}^3 / \text{s}$ de la Garonne à Mas d'Agenais durant la période 1913 1977

3.4.2 Ensemble A des décisions

on énumérera ici les différentes décisions possibles pour l'exemple qu'on traite :

Construire une digue de hauteur $a \geq 0$ ($a=0$ représente l'alternative ne pas protéger). En termes d'ingénierie hydraulique, a est souvent dénommé crue de projet et sera exprimé dans la même unité que les débits de la rivière (puisque en une section donnée de la rivière, il y a une correspondance biunivoque entre la hauteur et le débit). L'ensemble A est, dans cet exemple, formé des réels positifs.

3.4.3 Ensemble X des informations

L'exemple posé propose les informations suivantes :

Décider de construire ou de ne pas construire un ouvrage destiné à protéger un site contre le débordement d'une rivière repose sur la connaissance de la probabilité d'apparition d'une crue dommageable. Cette probabilité est incertaine mais sur laquelle on dispose néanmoins de quelques enregistrements des débits passés de la rivière. Sur la figure 3.5, on a noté, par exemple, les débits de pointe (en m^3/s) des crues supérieures à $2\,500\,m^3/s$ apparues sur la période 1913-1977. Plusieurs pointes peuvent apparaître chaque année. $X = \{(jour\ de\ la\ mesure\ j, x_j = \text{débit enregistré en } m^3/s)\}$.

3.4.4 Recueil d'information

Informations collectées à partir de l'exemple posé :

L'hydrologue sait bien qu'une crue de la Garonne n'est pas un événement isolé. Chaque crue a une histoire qui débute au moment où, par exemple, des précipitations notables tombent sur le bassin de la rivière. Celles-ci, comme c'est le cas de la Garonne, sont réparties sur un vaste territoire. Prendre ces précipitations en compte, ou une partie d'entre elles, constitue une campagne de collecte d'informations "e" différente de la campagne e que nous avons implicitement supposée : n'utiliser que les débits supérieurs à $2\,500\,m^3/s$. Dans le tableau 3.1 par exemple, au lieu des dépassements de seuil, on a considéré une autre campagne de collecte d'informations : on a enregistré le maximum de débit journalier qui s'est écoulé au cours de chacune des années de 1913 à 1977. Les données x sont formées de la collection (jour de la mesure j , $x_j = \text{débit enregistré en } m^3/s$). L'ensemble X des données possibles est $(R^{+})^{65}$.*

Année	Max.	Année	Max.	Année	Max.	Année	Max.	Année	Max.
1913	4 579	1926	3 200	1939	2 800	1952	6 721	1965	4 968
1914	4 774	1927	6 332	1940	5 553	1953	2 700	1966	5 163
1915	4 968	1928	4 968	1941	5 163	1954	3 000	1967	2 600
1916	4 774	1929	1 950	1942	3 100	1955	5 747	1968	2 530
1917	3 400	1930	7 500	1943	3 600	1956	2 300	1969	4 073
1918	6 137	1931	3 700	1944	4 579	1957	3 200	1970	3 120
1919	4 189	1932	3 600	1945	3 200	1958	2 900	1971	4 696
1920	4 579	1933	2 500	1946	950	1959	4 968	1972	5 377
1921	2 800	1934	3 700	1947	1 850	1960	3 400	1973	3 956
1922	4 384	1935	6 137	1948	2 000	1961	4 774	1974	4 228
1923	5 747	1936	4 189	1949	1 900	1962	2 300	1975	3 200
1924	3 200	1937	5 747	1950	2 600	1963	2 700	1976	4 209
1925	3 100	1938	3 200	1951	2 900	1964	3 300	1977	4 482

Tab 3.1-Débits annuels maximaux (en m³ / s) de la Garonne à Mas d’Agenais sur la période 1913-1977.

En utilisant la fonction `barplot` du **R**, on obtient la figure 3.6, qui représente les dépassements des pointes de crue au-delà de 2 500 m³/s de la Garonne à Mas d’Agenais durant la période 1913-1977.

3.4.5 Forme de décision

Imaginons qu’il y ait “ n ” mesures de débit x_j dépassant 2 500 m³/s. L’ensemble des décisions possibles est le continuum de toutes les hauteurs de digue “ a ” allant de 0 à une hauteur a_{\max} que l’on peut prendre, par commodité, infinie. Dans la suite de cet exemple on admettra une relation entre hauteur et débit de la rivière de telle sorte qu’on peut exprimer “ a ” dans la même unité que les débits, et on parlera de crue de projet associée à la hauteur choisie “ a ”. Comme dans les autres exemples, une règle de décision est une des applications :

$$(x_1, \dots, x_n) \mapsto a = \delta(x_1, \dots, x_n)$$

Une pratique courante de l’ingénierie est de prendre la règle δ^1 :

$$\delta^1(x_1, \dots, x_n) = k \times \max_{i=1 \dots n} (x_1, \dots, x_i, \dots, x_n)$$

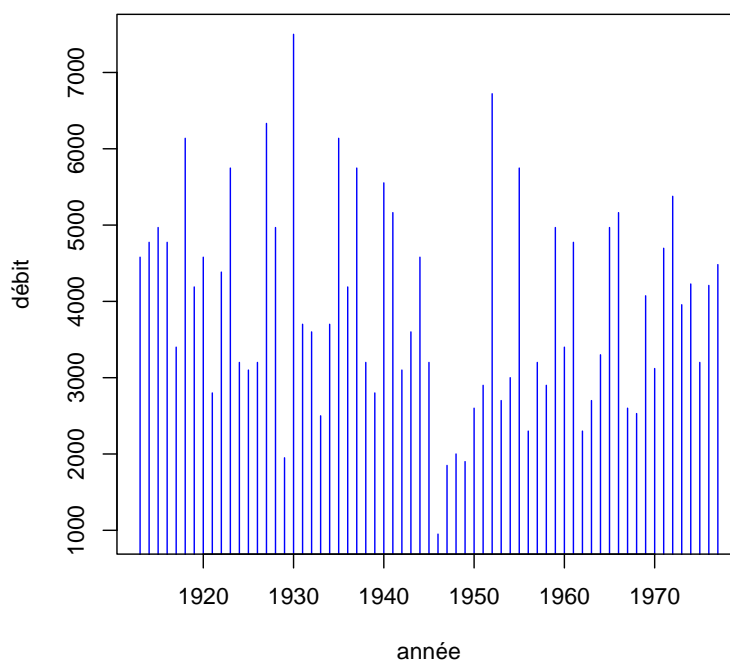


FIG. 3.6 – Les dépassements des pointes de crue.

Est-ce que cette stratégie définit toujours un pari intéressant ? Que se passe-t-il quand l'échantillon de données s'accroît avec le temps ? Que faire lorsque l'on ne dispose pas d'une série de débits, au site où l'on envisage de construire l'ouvrage (site non jaugé) ?

3.4.6 Représentation probabiliste des connaissances

Observons le comportement stochastique de l'exemple posé :

Les rivières sont capricieuses. Même dans le cadre d'un régime hydrologique stationnaire (on supposera ici que le réchauffement climatique global n'affecte pas le cours d'eau que l'on étudie, que l'urbanisation ne change pas les caractéristiques du cours d'eau, etc.), les débits fluctuent naturellement et comme les amplitudes de variations naturelles peuvent être fortes, l'échantillon de débits enregistrés dont on dispose (généralement assez court) peut laisser croire à un régime plus humide (ou plus sec) qu'il n'est en réalité.

3.4.7 Modèle statistique paramétrique

Un modèle à deux paramètres Poisson-exponentiel pour décrire les débordements d'une rivière

Le modèle dit de renouvellement-dépassement est couramment utilisé en hydrologie traditionnelle pour représenter l'occurrence des crues (ici, on appellera crue tout événement dont le débit maximum enregistré dépasse, sur notre exemple de la Garonne à Mas d'Age-nais, le seuil de $u=2\,500\text{ m}^3/\text{s}$). Ce modèle est complètement décrit dans Miquel, 1984.

Précisons les notations de la figure 3.7 Dans le passé, n crues indépendantes dommageables d'amplitude x_j ($j = 1..n$) se sont produites durant r années successives. Par commodité dans la suite du traitement de cet exemple, on réalise une translation des échelles ($u=2\,500\text{ m}^3/\text{s}$), de sorte que X mesure l'intensité du dépassement au-delà du seuil u instaurant l'état de crue :

($x=100$) désigne par exemple une crue de : $2\,500+100 = 2600\text{ m}^3/\text{s}$. On note k_i le nombre de crues dommageables ayant survécu au cours de l'année i ($i = 1..r$). Le modèle de renouvellement-dépassement est construit de la façon suivante :

la probabilité du nombre d'évènements dommageables se produisant sur une année suit une loi de Poisson de paramètre θ . La probabilité de voir k crues dommageables sur l'année i est donc :

$$[K_i = k/\mu] = \frac{\mu^k e^{-\mu}}{\Gamma(k+1)}$$

Le paramètre μ , espérance mathématique de la grandeur aléatoire K_i , représente le nombre moyen annuel de crues. On considère que chaque année est indépendante, que le phénomène de débordement est stationnaire, et on suppose que l'amplitude X d'une crue dommageable suit une loi exponentielle de paramètre ρ . Le paramètre ρ s'interprète ici comme l'inverse du dépassement moyen au-delà du seuil de $2\,500\text{ m}^3/\text{s}$. La probabilité qu'une crue dommageable quelconque X_j soit inférieure ou égale à une valeur x s'écrit alors :

$$[X_j < x/\rho] = G(x) = \int_0^x \rho e^{-\rho z} dz = 1 - e^{-\rho x}$$

Si on connaissait les valeurs exactes de μ et de ρ , la probabilité d'observer n crues d'amplitude x_j sur r années consécutives serait :

$$[X_1 = x_1, ..X_{k_1} = x_{k_1}, ..., X_n = x_{k_1+...k_r}/\rho, \mu] = [X = x/r, \rho, \mu]$$

$$[X = x/\rho, \mu] = \frac{\mu^n e^{-r\mu}}{\prod_{i=1}^n \Gamma(k_i + 1)} \rho^N e^{-\rho \sum_{i=1}^n x_i}$$

L'équation précédente est la vraisemblance associée à l'échantillon pour le modèle Poisson-exponentiel. Ce modèle est un modèle très parcimonieux : deux paramètres inconnus μ et ρ suffisent à représenter un phénomène déjà complexe. Avec les notations génériques, $\theta = (\rho, \mu)$. De plus, ce modèle est un cas particulier d'une classe plus vaste dont l'existence et la pertinence sont justifiées par des considérations rigoureuses de comportement asymptotique de l'étude de phénomènes. Néanmoins la fixation du seuil u , choisi ici à 2 500 m³/s, requiert une bonne connaissance pratique du fonctionnement hydraulique de la rivière : en le fixant trop haut, on réduit drastiquement la taille de l'échantillon, tandis qu'en le fixant trop bas on ne garantit plus les conditions théoriques dans lesquelles ce modèle est réaliste, ni d'ailleurs les conditions pratiques permettant l'indépendance de la succession des événements définis comme crues.

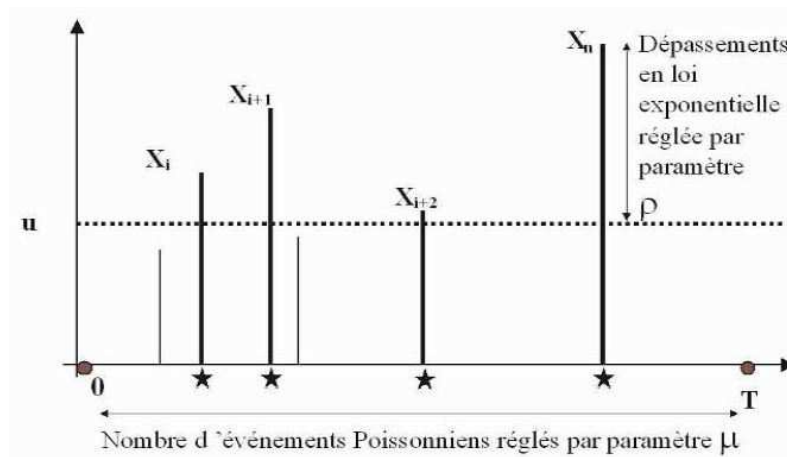


FIG. 3.7 – Modèle de renouvellement poissonnien avec dépassements exponentiels.

3.4.8 La méthode par introspection

Reprenons notre exemple et essayons de représenter l'incertitude a priori sur les états de la nature $\theta = \{\mu, \rho\}$ que pourrait posséder un ingénieur hydrologue. Il interprétera μ comme le nombre moyen de dépassements, ses connaissances hydrologiques lui feront dire, par exemple, qu'il parierait volontiers que μ vaut 3 dépassements par an mais qu'il en est peu sûr, et il donnerait une variabilité ordinaire à ce pari de plus ou moins 2.

Si nous adoptons pour μ une distribution a priori de la famille gamma à deux paramètres c_μ, d_μ :

$$[\mu] = \frac{(c_\mu)^{d_\mu}}{\Gamma(d_\mu)} \mu^{d_\mu-1} e^{-\mu c_\mu}$$

et que nous interprétons l'expertise de l'hydrologue sous la forme $\mathbb{E}[\mu] = 3$ et $\text{Var}[\mu] = 2 \times 2 = 4$ alors les propriétés de la loi gamma permettent de déduire deux équations pour déterminer les deux coefficients :

$$\begin{aligned} \mathbb{E}[\mu] &= \frac{d_\mu}{c_\mu} = 3 \\ \text{Var}[\mu] &= \frac{d_\mu}{(c_\mu)^2} = 4 \end{aligned}$$

Soit $c_\mu = \frac{3}{4}$ et $d_\mu = \frac{9}{4}$.

Une élicitation de la même forme peut être réalisée pour le paramètre ρ qui s'interprète comme l'inverse du dépassement au-delà de 2 500m³/s. Par exemple, ses connaissances du terrain et des rivières analogues feront parier l'ingénieur sur un dépassement moyen de 1 000m³/s que nous interpréterons comme

$$\mathbb{E}[\rho] = 10^{-3}$$

Son expertise peut l'amener à annoncer un écart-type de l'ordre du tiers de sa première estimation de ρ , de sorte que :

$$\text{Var}[\rho] = (0.33)^2 \times (10^{-3})^2 = 10^{-1} \times (10^{-3})^2$$

Si on adopte une fois de plus une loi gamma pour quantifier son expertise

$$[\rho] = \frac{(a_\rho)^{b_\rho}}{\Gamma(b_\rho)} \rho^{b_\rho-1} e^{-\rho a_\rho}$$

les coefficients $a_\rho = 10^4$ et $b_\rho = 10$.

Lois exponentielle et conjuguée

Poursuivons l'exemple en choisissant de représenter l'incertitude a priori sur les états de la nature $\theta = \{\mu, \rho\}$ par deux distributions gamma indépendantes :

$$[\rho] = \frac{(a_\rho)^{b_\rho}}{\Gamma(b_\rho)} \rho^{b_\rho-1} e^{-\rho a_\rho}$$

$$[\mu] = \frac{(c_\mu)^{d_\mu}}{\Gamma(d_\mu)} \mu^{d_\mu-1} e^{-\mu c_\mu}$$

où $a_\rho, b_\rho, c_\mu, d_\mu$ désignent les hyperparamètres qui caractérisent les distributions a priori. μ et ρ sont indépendants a priori et l'élicitation réalisée au paragraphe précédent proposait $a_\rho = 10^4$, $b_\rho = 10$, $c_\mu = \frac{3}{4}$ et $d_\mu = \frac{9}{4}$. En appliquant la formule de Bayes avec la vraisemblance du modèle de renouvellement-dépassement, on trouve qu'a posteriori μ et ρ demeurent indépendants et leurs distributions $[\rho/X]$, $[\mu/X]$ mises à jour grâce aux informations de l'échantillon de crues x sont encore des lois gamma dont les paramètres $a_\rho^\lambda, b_\rho^\lambda, c_\mu^\lambda, d_\mu^\lambda$ s'obtiennent de la façon suivante :

$$a_\rho \rightarrow a_\rho^\lambda = a_\rho + S(n); \quad b_\rho \rightarrow b_\rho^\lambda = b_\rho + n$$

$$c_\mu \rightarrow c_\mu^\lambda = c_\mu + r; \quad d_\mu \rightarrow d_\mu^\lambda = d_\mu + n$$

On voit que l'information contenue dans l'échantillon de crues x n'intervient que par les quantités récapitulatives :

- n : nombre de crues (ici 155)
- r : nombre d'années de mesures (valant ici 65)
- $S(n) = \sum_{i=1}^n x_i$: cumul des dépassements au-delà du seuil de $2\,500\text{m}^3/\text{s}$.

Les figures 3.4 et 3.5 montrent l'effet de cette prise en compte de l'information $n=35$, $r=65$, $S(n)=158\,143$ sur les lois a priori avec les valeurs $a_\rho = 10^4$, $b_\rho = 10$, $c_\mu = \frac{3}{4}$, $d_\mu = \frac{9}{4}$.

Notons que, d'après les propriétés de la loi gamma :

$$\begin{aligned}\mathbb{E}[\rho/n, r, S(n)] &= \frac{b_\rho^\lambda}{a_\rho^\lambda} = \frac{b_\rho + n}{a_\rho + S(n)} \\ \text{Var}[\rho/n, r, S(n)] &= \frac{b_\rho^\lambda}{a_\rho^{2\lambda}} = \frac{b_\rho + n}{(a_\rho + S(n))^2} \\ \mathbb{E}[\mu/n, r, S(n)] &= \frac{d_\mu^\lambda}{c_\mu^\lambda} = \frac{d_\mu + n}{c_\mu + r}\end{aligned}$$

Remarquons enfin que si l'on choisit un prior en faisant tendre $a_\rho, b_\rho, c_\mu,$ et d_μ vers 0, les lois a posteriori restent définies et les moyennes théoriques de μ et de ρ s'identifient naturellement à la valeur moyenne du dépassement et à l'inverse du nombre moyen de crues par an. Un tel prior sera dit **non informatif** et utilisé pour le modèle de renouvellement-dépassement en l'absence d'expertise hydrologique.

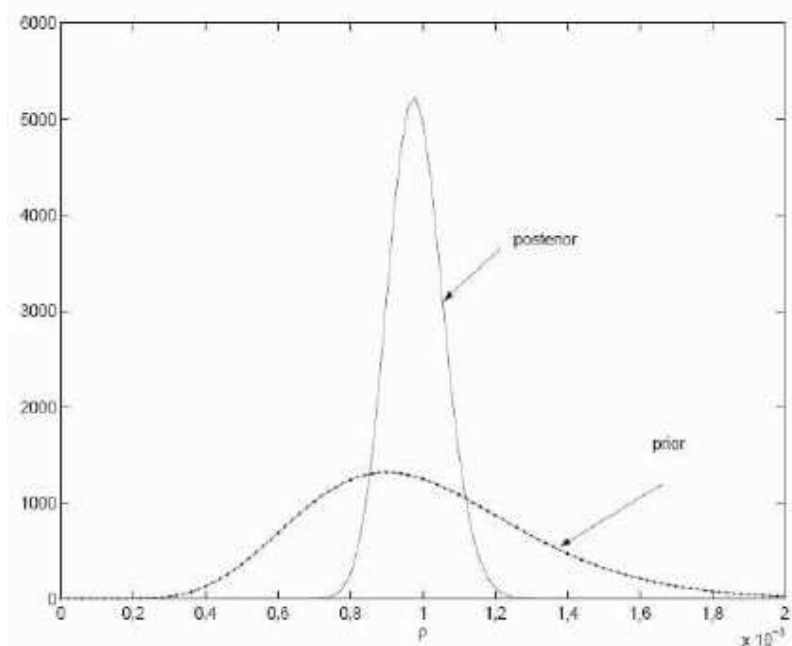
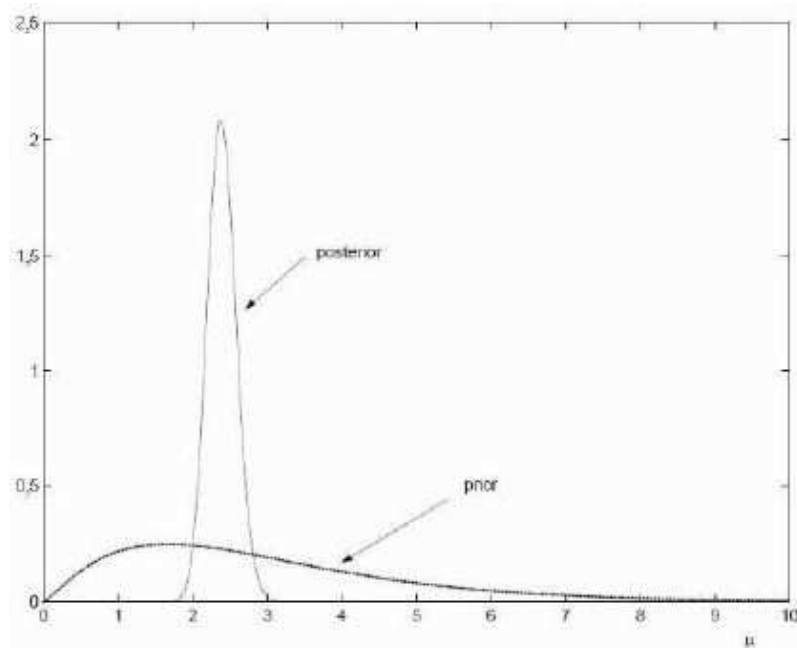


FIG. 3.8 – Prise en compte de l'information pour actualiser la distribution de ρ .

FIG. 3.9 – Mise a jour du parametre μ .

Conclusion générale

"All models are false, some are useful"

Bernardo et Smith, 1994

L'approche bayésienne apporte une plus grande souplesse dans les méthodologies statistiques. D'une part les procédures bayésiennes standard, peuvent maintenant être mises en oeuvre aussi facilement que les tests traditionnels, et elles ont le statut privilégié d'objectivité nécessaire à la communication scientifique. D'autre part différentes distributions a priori, exprimant des résultats antérieurs ou des opinions d'experts, favorables ou défavorables à la conclusion recherchée, peuvent être utilisées pour éprouver la "robustesse" des conclusions et prendre ainsi des décisions "personnelles". En outre les probabilités prédictives permettent d'évaluer les chances d'obtenir une conclusion donnée pour des observations futures, sur la base d'une étude "pilote" ou à partir des résultats partiels d'une étude en cours.

En perspectives, il serait mettre à profit les notions développées dans ce travail pour mettre en oeuvre des applications sur données réelles dans des domaines touchant la vie socio-économique ou environnementale. Par ailleurs, une recherche approfondie de la robustesse des méthodes peut donner lieu à une investigation originale.

Bibliographie

-
- [1] A.Yousfate. Introduction à R. Ateliers de simulation Numérique Tizi- Ouzou du 22 au 26 mars 2009.
- [2] A.Yousfate. Sur l'utilisation du langage fonctionnel R. U.M.M.T.O, 2009.
- [3] ABDOU NIANDOU Daouda, *Etude et comparaison des méthodes de prévision des séries chronologiques*. MEMOIRE D'INGENIEUR D'ETAT en RECHERCHE OPERATIONNELLE. U.M.M.T.O, 2009.
- [4] Éric Parent, Jacques Bernier. *Le raisonnement bayésien :Modélisation et inférence*. Edition Springer, 2007.
- [5] Christian P. Robert, *Le choix bayésien :Principes et pratique*. Edition Springer, 2006.
- [6] Jayanta, K. Ghosh, Mohan Delampady and Tapas Samanta. An introduction to Bayesian analysis : Theory and methods. Springer Science and Business Media, LLC, USA, 2006.
- [7] J.J.Droesbeke, J.Fine, G.Saporta. *Méthodes bayésiennes en statisque*. Editions TECHNIP, 2002.
- [8] Jean-Jacques Boreux, Eric Parent, Jacques Bernier. *Pratique du calcul bayésien*. Edition Springer, 2010.
- [9] Jean-Pascal Laedermann, *Théorie bayésienne de la décision statistique et mesure de la radioactivité*. Mémoire de doctorat en mathématiques. LAUSANNE, Suisse, 2003.
- [10] J.J.Droesbeke, J.Fine, G.Saporta. *Méthodes bayésiennes en statisque*. Editions TECHNIP, 2002.
- [11] J.M.Bernardo, A.F.M.Smith. *Bayesian theory*. Editions wiley series in probability and statistics, 2000.
- [12] J.Roudier, *Une application de la théorie de la décision statistique bayésienne*. Revue de statistique appliquées, tome 22, n°4, 1974, p. 3-28.
- [13] J.Ulmo, *La décision statistique dans le cadre bayésien*. Revue de statistique appliquées, tome 19, n°3, 1971, p. 27-66.
- [14] Hocine Fellag. Introduction aux statistiques bayésiennes et applications. Cours de 2ème année Master, U.M.M.T.O Tizi ouzou, 2011.
- [15] Glen Meeden, glen@stat.umn.edu, *Simulating from the Polya posterior*, March 06.
- [16] Le site "www.wikipedia.com".
- [17] Peter Congdon. Applied Bayesian Modelling. John Wiley et Sons Ltd, England, 2003.

[18] Peter Dalgaard. *Introductory Statistics with R*, Second Edition Springer 2008.

[19] William Feller. *An introduction of probability theory and its applications*, volume I. Wiley, New York, 1968.