



**Théorie bayesienne de la décision statistique et
mesure de la radioactivité**

Thèse de doctorat

présentée à la

Faculté des sciences de
l'Université de Lausanne

par

Jean-Pascal Laedermann

Licencié en Mathématiques
Université de Lausanne

Jury

Prof. Jean-François Valley
Prof. Aurelio Bay
Prof. Max-Olivier Hongler
Dr. François Bochud
Prof. Jean Hernandez

LAUSANNE
2003

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président	Monsieur Prof. Raymond Olivier
Directeur de thèse	Monsieur Prof. Jean-François Valley
Rapporteur	Monsieur Prof. Aurélio Bay
Experts	Monsieur Prof. Max-Olivier Hongler
	Monsieur Dr François Bochud

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Jean-Pascal Laedermann

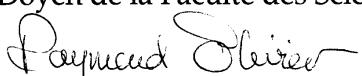
licencié en mathématiques de l'Université de Lausanne

intitulée

**Théorie bayesienne de la décision statistique
et mesure de la radioactivité**

Lausanne, le 5 décembre 2003

pour Le Doyen de la Faculté des Sciences


Prof. Raymond Olivier

A Phyllis

Résumé

L'objet de ce travail est l'évaluation de l'activité de radionucléides présents à l'état de traces, par exemple dans l'environnement. Lorsque les mesures sont de courte durée, ou si les sources sont peu actives, l'analyse statistique standard ne donne plus une estimation fiable de l'activité et de son incertitude. L'introduction du concept bayesien d'a priori permet de modéliser l'information à disposition de l'observateur avant qu'il effectue la mesure. Cette information conduit à une estimation plus cohérente des grandeurs physiques recherchées.

Le cadre de la théorie est tout d'abord présenté, définissant les concepts d'état, d'observation et de décision. La mesure physique est traduite par un modèle statistique qui est une probabilité de transition des états vers les observations. L'information de Fisher et celle de Shannon-Kullback sont introduites dans le but d'obtenir les a priori nécessaires au théorème de Bayes.

Les modèles propres à la mesure de la radioactivité sont ensuite traités. Si l'on peut considérer l'activité comme constante, le modèle est celui de Poisson et conduit à des a priori de type gamma. Pour les grandes activités, ces deux lois se rapprochent des gaussiennes et l'on retrouve l'analyse statistique classique. Lorsque la décroissance du nombre de noyaux n'est plus négligeable, ou lors de l'évaluation de certains temps d'attente, d'autres modèles sont développés.

Quelques applications sont présentées ensuite, notamment une définition cohérente des intervalles de confiance et l'estimation de l'activité de radionucléides à schéma complexe par spectrométrie γ , où l'obtention de tout un spectre permet une analyse multidimensionnelle.

Le paradigme bayesien conduit à une répartition complète et globale pour l'état du système physique mesuré. L'observateur obtient ainsi la meilleure estimation possible de l'état basée sur son modèle d'expérience et l'information préalable en sa possession.

Table des matières

I	Introduction	17
II	Formalisme général	25
1	Théorie bayesienne de la décision statistique	29
1.1	Inversion bayesienne	29
1.2	Espaces de la théorie de la décision statistique	31
1.3	Modèle	33
1.4	Théorème de Bayes	33
1.5	Décision statistique	35
1.5.1	Fonction de perte	35
1.5.2	Stratégies et risque fréquentiste	36
1.5.3	Risque bayesien moyen	37
1.5.4	Risque de Bayes	37
1.5.5	Stratégies optimales	38
2	Incertitude	39
2.1	Introduction	39
2.2	Précision et exactitude	40
2.3	Estimateur de Bayes	40
2.4	Propagation des incertitudes	41
2.5	Ensembles de crédibilité	42
2.6	Estimateurs et ensembles de crédibilité dans le cas poissonnien	42
3	Tests d'hypothèses simples	47
3.1	Introduction	47
3.2	Test de Neyman-Pearson	47
3.3	Probabilités d'erreur	48
3.3.1	Erreur de premier type	48
3.3.2	Erreur de second type	48
4	Evaluation bayesienne des courbes ROC	49
4.1	Définition des courbes ROC	49
4.2	Risques et pertes pour les test d'hypothèses	50
4.3	Stratégies optimales	51
4.4	Calcul du risque	52
4.4.1	Cas différentiable	53

4.4.2 Cas non différentiable	53
4.5 Extension aux hypothèses composées	53
III Modèles, information et a priori	55
5 Lois et modèles de base	59
5.1 Modèle binomial	59
5.2 Modèle poissonnien	60
5.3 Modèle gamma	61
5.4 Modèle négatif-binomial	63
5.5 Modèle bêta	65
5.6 Modèles discrets	66
5.6.1 Modèle discret à transition poissonnienne	66
5.6.2 Modèle discret à transition gamma	66
6 Information contenue dans les modèles	69
6.1 Information entropique	69
6.1.1 Divergence de Kullback-Leibler	69
6.1.2 Information moyenne apportée par la mesure	71
6.2 Information de Fisher	73
6.2.1 Construction	73
6.2.2 Lien avec la divergence de Kullback	74
6.2.3 Inégalité de Cramér-Rao	75
6.2.4 Illustrations	76
7 Choix des a priori	79
7.1 Conjugaison entre l'a priori et l'a posteriori	79
7.1.1 Définition	79
7.1.2 Lois gamma et modèle de Poisson	80
7.1.3 Lois gamma et modèle gamma	81
7.1.4 Lois bêta et modèles binomiaux	81
7.2 A priori sans information préalable	82
7.2.1 A priori constants	82
7.2.2 A priori de Jeffreys	84
7.2.3 A priori maximisant l'information	86
7.3 A priori avec information préalable sur les états	87
7.3.1 Stationnarité de l'état	87
7.3.2 Cas stationnaire, itération du théorème de Bayes	88
7.3.3 Cas non stationnaire, Bayes appliqué à des états probabilistes	91
7.4 Conclusion	96
IV Application de la TBDS à la mesure de la radioactivité	99
8 Introduction	103
9 Information, biais et précision dans un modèle poissonnien simple	105

10 Mesures monocanales à activité constante	107
10.1 Limites de détection	107
10.1.1 Processus	107
10.1.2 Théorie	107
10.1.3 Définitions des limites L_C et L_D	108
10.1.4 Valeurs des limites	108
10.2 Intervalles de confiance	110
10.2.1 Définition	110
10.2.2 Problèmes	110
10.2.3 Intervalles de Feldman-Cousins	111
10.2.4 Intervalles de Giunti	112
10.2.5 Solution bayesienne	112
10.3 Introduction de l'incertitude sur la durée de mesure et l'efficacité dans un modèle poissonnien	113
10.3.1 Construction du modèle	113
10.3.2 Etude du modèle généralisé, moments a posteriori	114
10.3.3 Comparaison avec le calcul classique	115
11 Mesures multicanales à activité constante	117
11.1 Etablissement du modèle	117
11.1.1 Comptages produits par la source	117
11.1.2 Comptages produits par le bruit de fond	118
11.1.3 Modèle global	118
11.1.4 A priori et a posteriori de Jeffreys	119
11.2 Application à la mesure d'une source d'europium-152	121
11.2.1 Bases de la simulation	121
11.2.2 Détermination des plages et des sensibilités	121
11.2.3 A priori sur les taux de bruit de fond	121
11.2.4 A priori sur l'activité	122
11.2.5 Calcul des a posteriori	122
11.2.6 Effet de la prise en compte de plusieurs plages sur les courbes ROC	123
11.2.7 Conclusion	123
11.3 Mesure d'uranifères dans l'environnement	125
11.3.1 Détermination de l'activité de l'uranium-235	125
11.3.2 Calcul de l'activité du thorium-230 utilisant un a priori issu du radium-226	127
12 Temps d'attente et modèles à activité décroissante	129
12.1 Estimation d'un temps d'attente à répartition exponentielle	129
12.1.1 Délai non aléatoire	130
12.1.2 Délai aléatoire	131
12.1.3 Résumé	133
12.2 Modèles avec activité décroissante	134
12.2.1 Détermination de l'activité par les comptages avec constante de désintégration connue	135
12.2.2 Détermination de la constante de désintégration radioactive par les temps de détection	136

V Conclusions et perspectives	143
--------------------------------------	------------

VI Remerciements	149
-------------------------	------------

VII Annexe	153
-------------------	------------

A Technique	155
A.1 Semi-groupe binomial	155
A.2 Composition d'une loi gamma et d'une loi de Poisson	156
A.3 Composition de deux lois gamma	157
A.4 Perte d'information par changement de variable	157
A.5 Divergence logarithmique et géodésiques	158
A.6 Divergence logarithmique de deux lois multinormales	159
A.7 Divergence logarithmique de deux lois de Poisson	160
A.8 Divergence logarithmique de deux lois gamma	160
A.9 Divergence logarithmique de deux lois bêta	162
A.10 Information moyenne apportée par une mesure poissonnienne d'a priori gamma	163
A.11 Information moyenne apportée par une mesure de temps d'attente d'a priori gamma	164
A.12 Matrice de Fisher pour le problème à deux voies	165
A.13 Calcul de l'information de Fisher pour le modèle normal à un seul paramètre	165
A.14 Détermination du modèle pour l'estimation de la constante radioactive	166
A.15 Calcul de l'a posteriori pour la constante radioactive à l'aide de l'approximation de Laplace	167
A.16 Elimination du bruit de fond par itération pour la détermination de la constante radioactive	168
A.17 Modèle des comptages	169
A.18 Modèle des temps d'attente	169
A.19 Calcul des limites L_C et L_D	170
B Notes	173
B.1 Modèle des clés	173
B.1.1 Un traitement plus complet	173
B.1.2 Une extension introduisant la durée de la recherche	174
B.2 Divergence de Kullback-Leibler et entropie de Shannon	174
B.2.1 Entropie	174
B.2.2 Bayes, Shannon et Kullback	175
B.3 A priori	177
B.3.1 A priori MDIP	177
B.3.2 Maximisation asymptotique de l'information	177
B.4 A priori maximisant l'information apportée par la mesure	180
B.4.1 Le paradoxe du cas fini	180
B.4.2 Application au modèle poissonnien	180
B.4.3 Application au modèle gamma	181
B.5 A priori de Jeffreys avec données préalables, extension continue Poisson-gamma	182

Table des figures

1	Sous-ensembles relatifs au problème des clés	21
1.1	Etats, observations et modèle	30
1.2	Maximum de vraisemblance	31
1.3	Concepts de la théorie de la décision statistique	32
1.4	Concepts de la théorie bayesienne de la décision statistique	34
1.5	Illustration du modèle poissonnien	36
2.1	Densités a posteriori	43
2.2	Variation des intervalles de crédibilité	44
2.3	Variation des intervalles de crédibilité, détections nombreuses	44
2.4	Variation des intervalles de crédibilité, détections rares	45
4.1	Exemples de courbes ROC	51
4.2	Courbes ROC et axe des risques	54
5.1	Exemples de lois binomiales	60
5.2	Exemples de lois de Poisson	62
5.3	Exemples de lois gamma	63
5.4	Plan gamma	64
5.5	Exemples de lois négatives-binomiales	66
5.6	Exemples de lois bêta	67
7.1	Estimation de la probabilité de lever du soleil	84
7.2	Simplexe probabiliste de dimension 2	92
7.3	Evolution de l'a priori de Jeffreys avec données préalables	96
7.4	A priori pour un modèle à deux états	97
10.1	Limites de détection	109
10.2	Bandes de confiance	111
10.3	Intervalles de confiance	113
10.4	Fréquences des écarts à l'activité effective	116
11.1	Courbes ROC relatives à la détection de l'europium-152	124
11.2	Spectre brut pour la détermination de l'activité d'un échantillon d'uranium-235	125
11.3	Comparaison des activités a posteriori du thorium-230	128
12.1	Nucléide avec un état métastable	129
12.2	Estimation bayesienne du délai	131

12.3	Densité a posteriori pour le paramètre du délai	133
12.4	Evolution de l'a posteriori sur le paramètre du délai	134
12.5	Estimation bayesienne de l'activité initiale d'une source de fluor-18	137
12.6	Estimation de la constanre radioactice par approximation de Laplace	139
12.7	Estimation du λ de l'yttrium-90	141
12.8	Nombres d'impulsions cumulées	141
12.9	Validation statistique de l'estimation	142
A.1	Divergence de deux lois de Poisson	160
B.1	Partition définissant une question	176
B.2	A priori MDIP pour le modèle poissonnien	178

Liste des tableaux

3.1	Probabilités d'erreur pour le test à deux issues	48
4.1	Fonction de perte $L(d \theta)$ pour le test à deux issues	50
6.1	Divergences logarithmiques de quelques lois de base	71
7.1	Lois et modèles de base concernant la détermination de l'activité des sources et de la sensibilité des compteurs	82
7.2	Paramètres de la simulation de l'a priori de Jeffreys avec données préalables	95
11.1	Emissions photoniques dominantes de l'europium-152	122
11.2	Taux de bruit de fond étendus aux plages	122
11.3	Comptages simulés pour le modèle (11.11)	122
11.4	Estimations a posteriori	123
11.5	Qualifications des courbes ROC	123
11.6	Données pour une estimation de l'activité de l'uranium-235	126
11.7	Activité a posteriori	126
11.8	Activités détectables minimales	126
11.9	Données de mesure pour le thorium-230	127
11.10	Estimation de l'activité du thorium	127
12.1	Paramètres de simulation pour une estimation de λ sans bruit de fond	140

Première partie

Introduction

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

Anonyme, XX^e

*Et puis le physicien, alors ?
Oh, eh bien, le physicien aussi.*

Paysan, XXI^e

Clés, professeurs et révérends

L'histoire se passe dans la petite ville d'East Finchley près de Londres. Le temps est clair et la lune à son premier quartier. Aux environs de minuit, le professeur Ronald F. s'apprête à entrer dans son *home sweet home* lorsqu'il s'aperçoit qu'il a perdu ses clés. Un moment plus tard, le révérend Thomas B., passant par là, le trouve à quatre pattes sous un réverbère.

- Mais, qu'est-ce que vous fichez-là, dans cette absurde position, plutôt qu'à bailler tranquillement dans votre lit ? questionne le révérend.

- Je cherche mes clés ! répond le professeur.

- Mais où les avez-vous donc perdues ?

- Par là ... fait le professeur en indiquant une zone obscure à une douzaine de yards du réverbère.

- Mais pourquoi donc cherchez-vous ici ?

- Parce qu'il y a de la lumière !

Cette histoire très connue, adaptée aux besoins de la cause, n'est peut-être pas si idiote qu'elle en a l'air. En effet, nous allons voir que si la connaissance qu'a le professeur du lieu où sont ses clés n'est pas tranchée, et que sa recherche est plus efficace dans la lumière, il se peut tout à fait qu'il ait raison. Par contre, si sa recherche dure trop longtemps sans succès, il aura avantage à aller voir dans l'obscurité. La décision initiale de chercher dans la lumière ou dans l'obscurité se fait en optimisant la probabilité de trouver les clés. Ensuite, l'intégration du fait qu'il ne les trouve pas se fait grâce au théorème de Bayes.

Pour traiter ce problème, définissons tout d'abord l'efficacité de recherche du professeur dans la lumière, S_L , par la probabilité qu'il a de les y trouver, si elles y sont, dans le quart d'heure qui suit. Pour fixer les idées, supposons que $S_L = 90\%$. Similairement, son efficacité de recherche dans l'obscurité sera S_O , que nous fixerons à une valeur évidemment beaucoup plus basse, mettons 1%.

Le professeur a aussi une certaine idée du lieu où sont ses clés. Ce sentiment se traduit par une probabilité a priori qu'elles soient dans la lumière $P(L)$, qui vaudra par exemple 5%. La probabilité qu'elles soient dans l'obscurité est la complémentaire $P(O) = 95\%$. Cette dernière, nettement plus grande, rend compte du fait que le professeur privilégie une zone éloignée du réverbère.

S'il choisit les alentours du réverbère comme lieu de recherche, la probabilité effective de trouver est le produit

$$M_L = S_L P(L) = 4.50\%$$

par contre, s'il choisit la zone obscure

$$M_O = S_O P(O) = 0.95\%$$

Une bonne stratégie consiste à chercher tout d'abord à l'endroit qui donne la probabilité maximale de succès, c'est-à-dire sous le réverbère !

Supposons maintenant qu'après un quart d'heure, le professeur n'ait toujours pas trouvé. Pour tenir compte de cette information, il doit modifier son a priori. Le théorème de Bayes, sous sa forme originale, exprime la probabilité qu'une cause C soit à l'origine d'un effet observé E par

$$P(C | E) = \frac{P(E | C)P(C)}{P(E)}$$

où $P(E | C)$ est la probabilité de l'effet pour une cause donnée, $P(C)$ la probabilité a priori d'occurrence de la cause, et $P(E)$ la probabilité globale d'occurrence de l'effet.

Ici, une cause est l'un des deux lieux de présence possibles des clés, la région lumineuse L ou la région obscure O . L'effet est le fait de les trouver T , ou non $nonT$. L'expérience consiste à *chercher dans la zone lumineuse*. La figure 1 montre les sous-ensembles concernés. Reprenons le théorème de Bayes dans le cas de la recherche infructueuse du professeur.

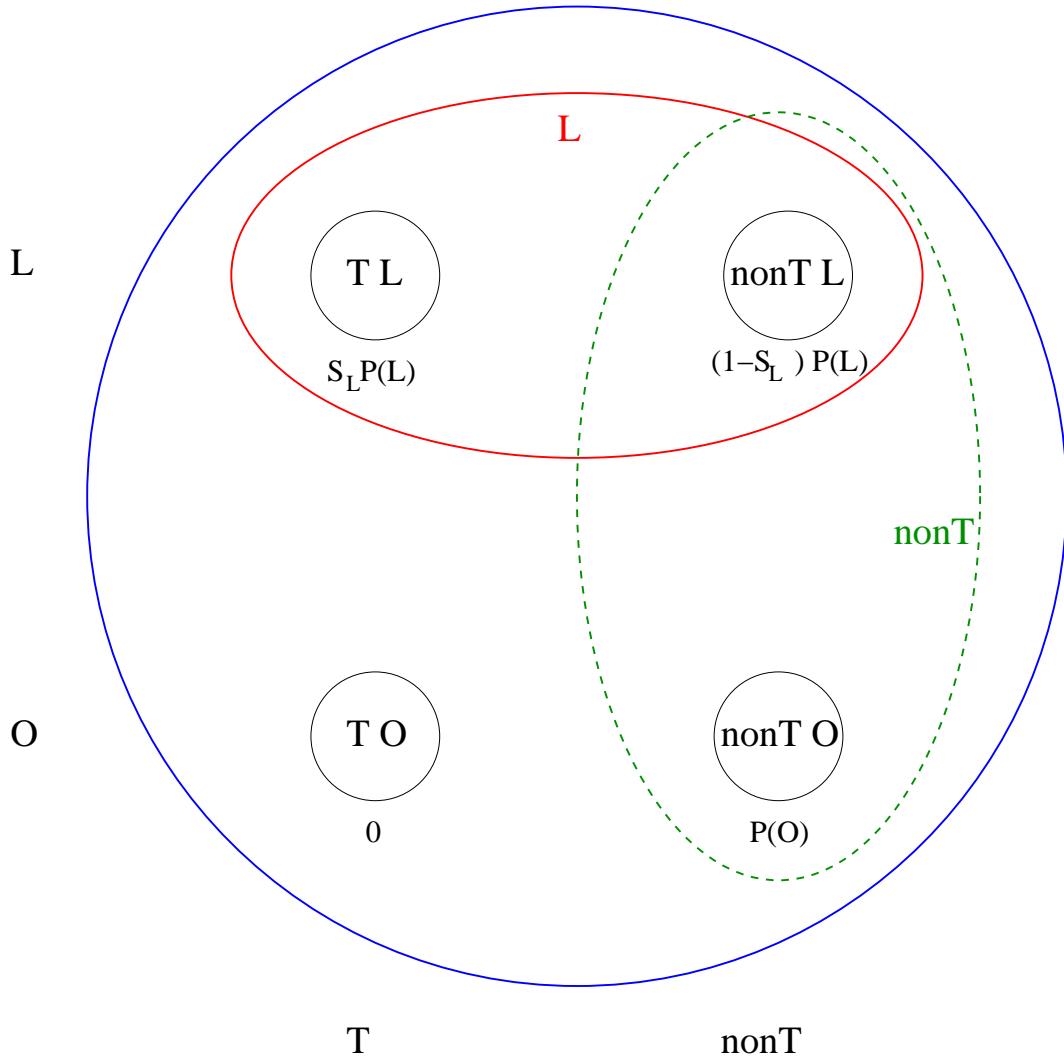


FIG. 1 – Sous-ensembles relatifs au problème des clés. L'espace total est formé de quatre éléments. C'est le produit $\{T, nonT\} \times \{L, O\}$. Les probabilités des différents éléments sont exprimées dans le cas où la recherche se fait dans la lumière.

$$P(L | nonT) = \frac{P(nonT | L)P(L)}{P(nonT)} = \frac{P(nonT \cap L)}{P(nonT)}$$

$P(L | nonT)$ est la nouvelle probabilité de présence des clés dans la lumière

$P(nonT \cap L)$ est la probabilité de l'élément $nonTL$, qui vaut $(1 - S_L)P(L)$

$P(nonT)$ est la probabilité globale de ne pas trouver les clés, c'est-à-dire la somme des probabilités des deux éléments $nonTL$ et $nonTO$ c'est-à-dire $(1 - S_L)P(L) + P(O)$

On obtient ainsi une nouvelle probabilité de présence des clés dans la lumière P' , sachant qu'elles n'ont pas encore été trouvées :

$$P'(L) = P(L \mid \text{non}T) = \frac{(1 - S_L)P(L)}{(1 - S_L)P(L) + P(O)}$$

Ainsi, après une recherche infructueuse de quinze minutes sous le réverbère, le calcul donne une nouvelle estimation des probabilités de présence des clés : $P'(L) = 0.52\%$ dans la lumière et sa complémentaire $P'(O) = 99.48\%$ dans l'obscurité. Les nouvelles probabilités de trouver les clés sont alors :

$$\begin{aligned} M'_L &= S_L P'(L) = 0.47\% \text{ en cherchant dans la lumière} \\ M'_O &= S_O P'(O) = 0.99\% \text{ en cherchant dans l'obscurité.} \end{aligned}$$

Le professeur doit donc changer de tactique¹.

Il ne reste plus qu'à souhaiter que le révérend ne passe pas trop tard ...

Survol 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 suites 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 leur donner des poids différents émerge sous la forme du *calcul des probabilités*.

Si le concept de hasard est très ancien, sa mesure arrive tardivement dans l'évolution des sciences. Les premiers pas sont fait par Pascal et Fermat (1650). En 1764 sont publiés les écrits du révérend Thomas Bayes [1], 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) [2]. 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) [3], la probabilité se fondera dans l'imposante *Théorie de la Mesure et de l'Intégration*, lui donnant une orientation analytique fortement influencée par les bourbakiens [4] 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 [5].

¹Le lecteur intéressé par une discussion plus générale pourra consulter la section B.1

Fisher contre Bayes, faux débats

L'utilisation de la probabilité a suscité de nombreux échanges d'invectives entre les tenants de Bayes et ceux de Fisher ; en pure perte, 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 bayesienne.

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'événements rares, comme la détection de neutrinos ou de bosons de Higgs, l'observateur ne pourra plus utiliser les approximations que lui fournit la théorie fréquentiste.

Le même problème apparaît lors de la mesure de sources radioactives de faible intensité. Plus le temps à disposition est limité, moins l'analyse statistique habituelle est fiable. Dans le but de surmonter cette difficulté, la théorie bayesienne a refait une apparition en Allemagne notamment sous l'impulsion de deux physiciens du Physikalisch-Technische Bundesanstalt (PTB) : K. Weise et W. Woeger [6].

Dans le cadre de cette analyse, la distinction entre état et observation est fondamentale. Ce que le physicien cherche, c'est la valeur d'une ou plusieurs variables associées à un système dans un état donné. Or la mesure ne donne pas directement ces variables, mais un ensemble de grandeurs associées qui constituent l'observation. Par exemple, on n'observe jamais une activité, mais un nombre d'impulsions ou un courant électrique ...

Cette distinction conceptuelle permet de comprendre pourquoi la théorie de Bayes a également sa place en physique. En effet, cette dernière fournit une relation entre l'état et les observations qui est aléatoire, car l'état physique n'est que l'expression de la partie connaissable du phénomène. Pour un état, il peut y avoir plusieurs observations possibles, qui se répartissent selon une certaine loi. Mais lorsque la mesure est terminée, l'observation est unique, et l'état ne peut s'exprimer que par le biais d'une incertitude. Il faut donc remonter des effets aux causes et utiliser le théorème de Bayes. Les répartitions sur les états sont des probabilités au sens de Bayes, alors que celles sur les observations sont des probabilités au sens de Fisher.

Evolution des 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 physique 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*. A cela, 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 d'appartement sera totalement inutilisable dans un four. Cette simple considération limite considérablement l'espace des états. La demi-droite $] -273.15^\circ, \infty^\circ[$ devient l'intervalle borné $[-40^\circ, 50^\circ]$. De plus, l'application du théorème de Bayes est souvent robuste lors du choix d'*a priori*

larges, à condition de se souvenir qu'un état exclu dès le départ n'a aucune chance de revenir par la suite. Si l'on avait pris $[-50^\circ, 80^\circ]$ au lieu de $[-40^\circ, 50^\circ]$, le résultat serait peu différent. Cette robustesse permet un choix d'a priori initial assez libre. En revanche, si l'on avait pris la réunion de intervalles $[-50^\circ, 10^\circ]$ et $[20^\circ, 80^\circ]$, aucune température entre 10° et 20° ne pourrait être obtenue, quelle que soit l'observation. Ce genre d'a priori doit évidemment être évité.

Plan de la thèse

Le cadre physique de cette étude est la mesure de la radioactivité. Ce phénomène fait intervenir dès le départ une composante aléatoire, avant même d'introduire les incertitudes relatives aux instruments. De plus, la grandeur de base, l'activité, est par définition réelle positive. Cette positivité pose d'emblée un problème lors de l'utilisation des lois normales, qui sont définies sur tout l'axe réel. Ces considérations ont motivé une approche plus générale, dans laquelle le théorème de Bayes permet un calcul rigoureux de l'activité de sources faibles.

Le formalisme général de la théorie bayesienne de la décision statistique sera présenté dans le cadre de la théorie des probabilités, ainsi que ses implications sur la notion d'incertitude et des tests d'hypothèses. Propres aux modèles régissant la désintégration des noyaux, la loi de Poisson et les répartitions gamma seront utilisées en lieu et place des habituelles gaussiennes.

Ensuite, une discussion plus approfondie au sujet des modèles et des a priori basée sur les notions d'information sera développée. Le concept d'information dans le contexte probabiliste fait l'objet de deux approches. L'une, celle de Shannon, mesure globalement la capacité de transmission d'un canal. Cette notion peut être reprise en physique par simple substitution de termes : l'émetteur devient l'état, le récepteur l'observation et le canal de transmission l'instrument de mesure. C'est une notion globale. L'autre, celle de Fisher, permet d'évaluer la sensibilité d'un système de mesure dans un état donné. C'est une notion locale. Nous montrerons comment les deux définitions de l'information sont liées et peuvent servir à construire des a priori.

Enfin, une série d'applications de cette théorie est présentée.

- La procédure de détermination des limites de détection sera reprise. Ces limites sont actuellement définies uniquement dans l'espace des observations, créant des paradoxes qui sont facilement élucidés dans un cadre bayésien.
- La généralité des espaces d'observation et d'état sera mise à profit pour le traitement simultané de plusieurs informations. En particulier la détermination de l'activité de nucléides complexes par spectrométrie γ sera améliorée en considérant plusieurs régions du même spectre.
- L'activité d'un ensemble de noyaux radioactifs varie exponentiellement dans le temps. Les premières applications seront développées en faisant l'hypothèse que les durées de mesures sont suffisamment petites pour admettre que l'activité est constante. Un modèle tenant compte de cette décroissance sera proposé ensuite.

Deuxième partie

Formalisme général

N'essayez jamais de convaincre un physicien qu'il est bayesien sans le savoir

G. D'Agostini [7]

Chapitre 1

Théorie bayesienne de la décision statistique

1.1 Inversion bayesienne

La physique fonctionne en isolant certaines parties de l'univers pour en faire des modèles. Un système physique est caractérisé par un certain nombre de variables définissant son état. Les valeurs de ces variables, les *mesurandes* [8], sont obtenues par le biais d'un système de mesure. Les *résultats de la mesure*, que nous appellerons aussi *observations*, ne permettent qu'une estimation de l'état, car interviennent dans le processus des phénomènes de nature aléatoire non maîtrisés par l'observateur. On obtient une correspondance entre l'état et l'observation qui est de nature statistique, associant à un état fixé une répartition des observations (gaussienne, poissonnienne ou autre). Le but de la mesure d'une grandeur physique est alors d'inverser cette relation, en ce sens que l'observateur ayant obtenu un résultat, doit en inférer l'état.

La relation donnant la répartition des observations pour un état donné se nomme le *modèle*. Elle est *objective*, extérieure à l'observateur, car elle n'est dépendante que de l'objet mesuré et de l'instrument utilisé. En répétant l'expérience en face du même état, l'observateur verra ses résultats se distribuer selon le modèle. On trouve une probabilité au sens fréquentiel.

Comme un état peut donner des résultats de mesure différents, une observation peut très bien correspondre à plusieurs états. Quel est le bon ? Ou mieux, quelle répartition peut-on imaginer sur les états ayant en main cette observation ? L'incertitude passe du domaine des résultats dans celui des états. Cette incertitude est *subjective*, en ce sens qu'elle est propre à l'observateur. La répartition des états n'existe que dans la description interne qu'a l'observateur du système à un moment donné, mais nullement dans la réalité qui lui est extérieure.

En statistique classique, le modèle est donné par une fonction dite *de vraisemblance* $\mathcal{L}(\theta | x)$, où θ désigne l'état du système physique et x l'observation. Cette fonction, lue avec x variable et θ constant, est normalisée et considérée comme densité de probabilité. Elle donne le modèle par (cf fig 1.1)

$$P(dx | \theta) = \mathcal{L}(\theta | x)dx$$

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

états

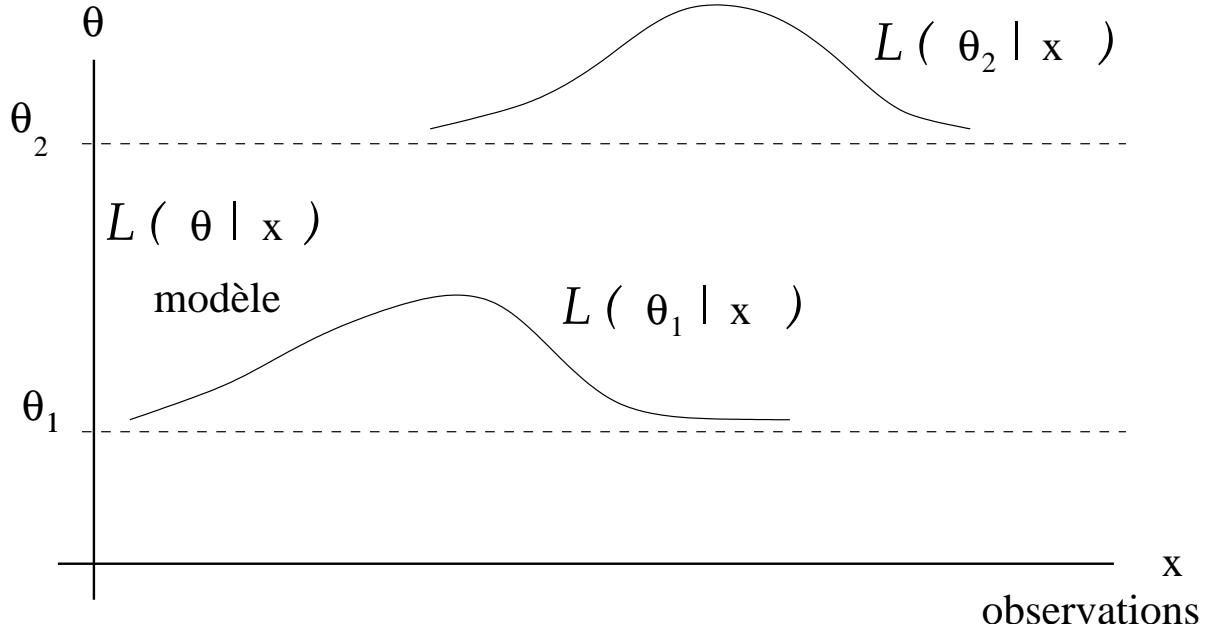


FIG. 1.1 – Etats, observations et modèle. *Le modèle est donné par une fonction des deux variables x et θ dite de vraisemblance. C'est une densité de probabilité en x avec θ fixe, qui n'est généralement pas normalisée en θ avec x constant.*

de l'état consiste à prendre celui qui maximise la fonction \mathcal{L} . C'est le principe de maximum de vraisemblance (cf figure 1.2)

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(\theta | x)$$

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. Pour retrouver un caractère global, il faut considérer la courbe en θ dans son ensemble et essayer d'en faire une probabilité. Il serait alors possible d'obtenir une moyenne et une variance pour l'état. La première idée consiste à poser comme répartition des états étant donné x

$$\pi(d\theta | x) = C \cdot \mathcal{L}(\theta | x) d\theta$$

avec une constante C de normalisation. Cette manière de faire consiste à considérer les états comme initialement équiprobables. La généralisation de l'inversion ci-dessus se fait simplement en remplaçant $d\theta$ par une répartition dite a priori $\pi(d\theta)$. Cet a priori traduit la connaissance que l'observateur a du système avant de commencer sa mesure. Le passage de $P(dx | \theta)$ à $\pi(d\theta | x)$ est dit *inversion bayésienne* (cf section 1.4).

Le but la mesure est de permettre à l'observateur de prendre des décisions. Or ces décisions devraient, pour être optimales, ne dépendre que de l'état du système physique en question. La théorie de la décision statistique propose un formalisme qui inclut la recherche des meilleures décisions basées uniquement sur les observations.

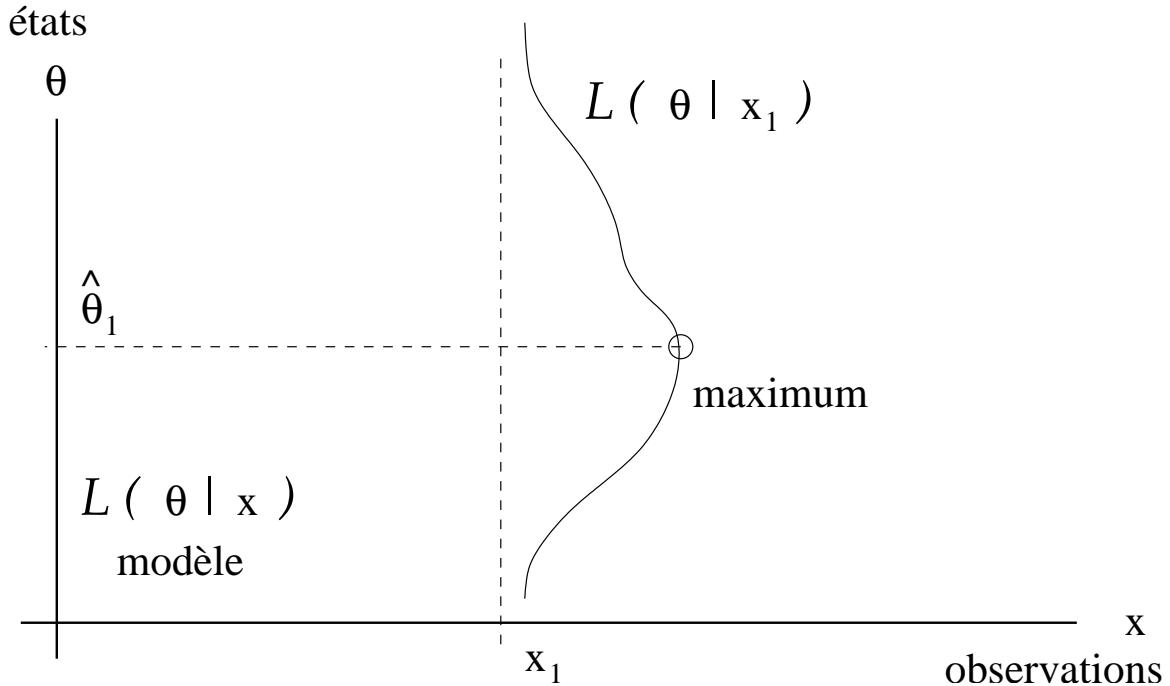


FIG. 1.2 – Maximum de vraisemblance. La fonction de vraisemblance $L(\theta | x)$ peut être considérée à x constant. L'interprétation de cette fonction comme densité de probabilité en θ implique le théorème de Bayes.

1.2 Espaces de la théorie de la décision statistique

La théorie bayésienne de la décision statistique, que nous abrégerons en *TBDS*, distingue trois espaces. Nous suivrons ici les notations de Berger-Bernardo [9] [10], qui tendent à devenir un standard.

Le premier espace est celui des *états*, c'est-à-dire les valeurs vraies ou mesurandes relatives au système physique. 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} .

Au vu de certains résultats, l'observateur est 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. L'estimation d'un état constitue aussi une décision, souvent élément d'un espace continu qui s'identifie à celui des états. L'ensemble des *décisions* forme le troisième espace \mathcal{D} , dont les éléments seront notés d . S'il existe une règle s associant une décision d à une observation x , la fonction $d = s(x)$ est appelée une *stratégie*¹. Dans le cas particulier où la décision est un état, les stratégies deviennent des *estimateurs*. Ces espaces ainsi que les relations qui les lient sont représentées à la figure 1.3.

¹Nous ne considérerons ici que des stratégies déterministes.

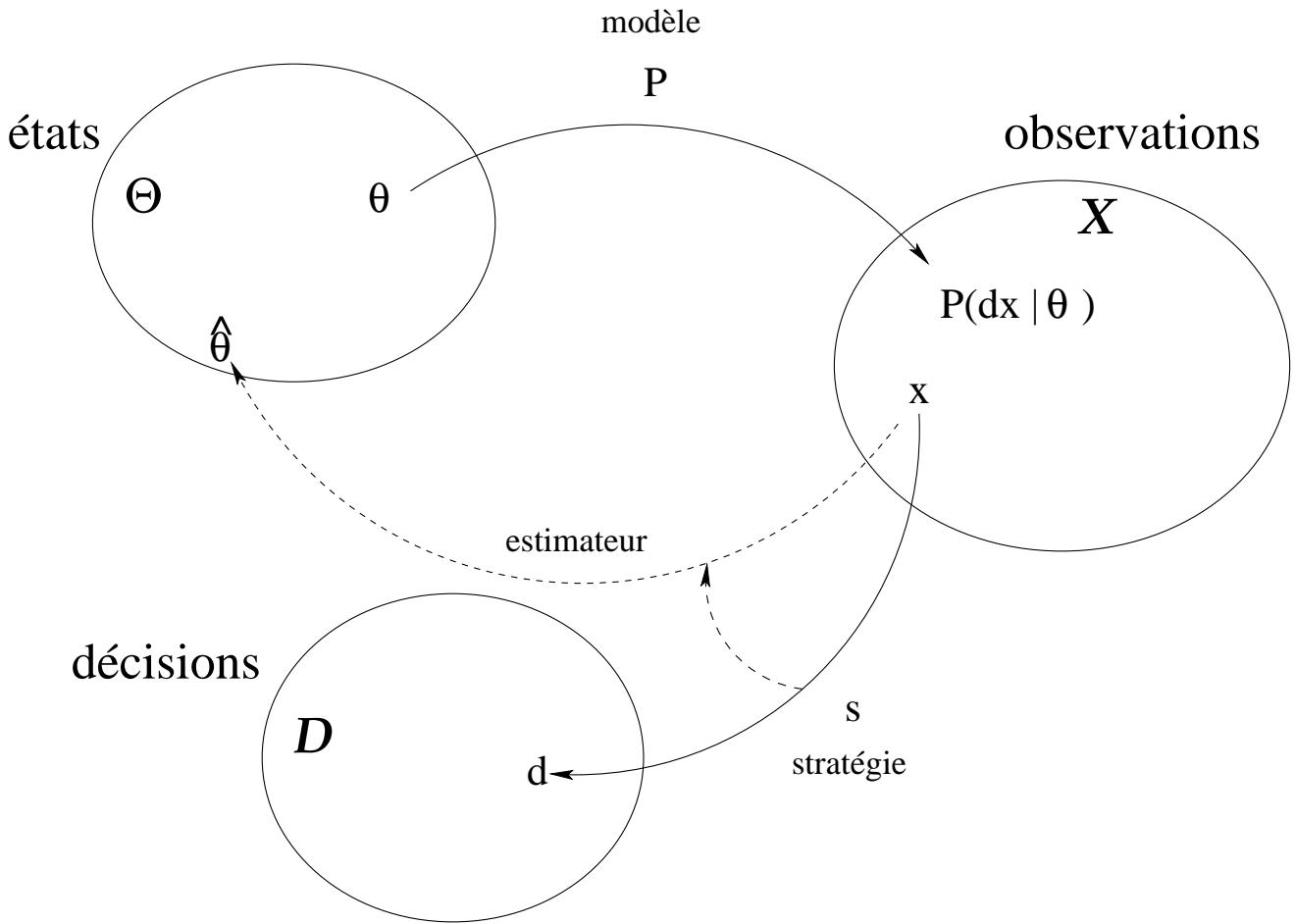


FIG. 1.3 – Concepts de la théorie de la décision statistique. *Le modèle donne pour un état, une répartition sur les observations. Une stratégie associe une décision à une observation. Le pointillé montre qu'un estimateur est une stratégie lorsque les décisions consistent en le choix d'un état, identifiant \mathcal{D} à Θ .*

Illustration Considérons la mesure d'une source radioactive en présence d'un rayonnement ambiant. Si les durées de mesure envisagées ne sont pas trop longues, l'activité de la source peut être considérée comme constante de valeur a [Bq]. Supposons aussi le rayonnement ambiant stationnaire et caractérisé par un taux moyen de b impulsions par seconde dans le détecteur. Le système physique se réduit ici à ces deux variables : $\theta = (a, b)$ et l'espace des états est ici $\Theta = \mathbb{R}_+^2$.

Les grandeurs a et b sont inconnues. Leur détermination se fait par deux mesures, l'une sans source, de durée t_0 , l'autre avec source, de durée t_1 , le taux de bruit de fond étant supposé le même. Si le comptage sans la source donne N_0 impulsions, et celui avec source N_1 , le couple de ces comptages formera une observation $x = (N_1, N_0)$ et l'espace des observations sera $\mathcal{X} = \mathbb{N}^2$.

Deux types de décisions apparaissent ici. Si l'observateur est intéressé par l'absence ou la présence d'un certaine activité a connue, son problème est de choisir entre les hypothèses \mathcal{H}_0 : il n'y a pas de source, et \mathcal{H}_1 : il y a une source. L'espace des décisions devient un ensemble à deux éléments $\mathcal{D} = \{\mathcal{H}_0, \mathcal{H}_1\}$. Si, au contraire, celui-ci est intéressé par la valeur de l'activité elle-même, sa décision sera un élément de l'axe a des états et l'espace \mathcal{D} s'identifie à la composante a de Θ .

1.3 Modèle

La répartition aléatoire des observations pour un état donné est une *probabilité de transition* (cf par exemple [11]) de l'espace des états vers celui des observations

$$P : \Theta \rightsquigarrow \mathcal{X} \quad (1.1)$$

Cette probabilité de transition se nomme le *modèle*. Elle associe à tout état une répartition des observations. On notera par $P(dx | \theta)$, la probabilité d'obtenir un résultat dans dx étant donné que l'état est θ . Nous supposerons toujours que ces répartitions sont soit discrètes, soit densitaires dans le cas continu. Si $f(x)$ est une fonction sur l'espace des observations, et que θ est un état fixé, son espérance sera souvent écrite

$$E_\theta f = Ef | \theta = \int_x f(x)P(dx | \theta) \quad (1.2)$$

Illustration Reprenons le cas de la source radioactive d'activité a en présence d'un bruit de fond de taux b . Les répartitions des comptages suivent des lois de Poisson. Introduisons l'efficacité du détecteur ε , et notons $\mathcal{P}(n | \lambda) = \frac{\lambda^n}{n!} e^{-\lambda}$ la répartition de Poisson de paramètre λ . Le comptage issu de la mesure sans la source donne la répartition suivante pour le nombre d'impulsions N_0

$$P(N_0 | b) = \mathcal{P}(N_0 | bt_0) \quad (1.3)$$

et celui pour la mesure avec la source

$$P(N_1 | a, b) = \mathcal{P}(N_1 | (a\varepsilon + b)t_1) \quad (1.4)$$

Dans le cas où ces mesures sont indépendantes le modèle s'exprime alors par

$$P(N_1, N_0 | a, b) = P(N_1 | a, b)P(N_0 | b) = \mathcal{P}(N_1 | (a\varepsilon + b)t_1)\mathcal{P}(N_0 | bt_0) \quad (1.5)$$

1.4 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 $P(dx | \theta)$ de Θ vers \mathcal{X} ne suffit pas à obtenir une probabilité sur le produit. Il faut encore une probabilité dite *de départ* sur Θ : $\pi(d\theta)$. Dans ce cas, la probabilité *conjointe* sur le produit est définie par

$$P(dx, d\theta) = P(dx | \theta)\pi(d\theta) \quad (1.6)$$

Dans le contexte bayésien, $\pi(d\theta)$ reçoit le nom d'*a priori*. La probabilité conjointe ainsi construite induit une marginale sur chacun des facteurs du produit. En particulier, la marginale sur l'espace des observations sera nommée simplement *marginale* et sera définie par

$$M(dx) = \int_\Theta P(dx | \theta)\pi(d\theta) \quad (1.7)$$

Lorsque les états sont pris aléatoirement en suivant π , $M(dx)$ donne la répartition des observations. $M(dx)$ sera aussi notée $P(dx | \pi)$.

Supposons maintenant que la mesure ait donné l'observation x . Il est possible de considérer la tranche de $\mathcal{X} \times \Theta$ à x constant, et considérer la probabilité conditionnelle induite par la conjointe sur cette tranche. Soit donc un infinitésimal dx au voisinage de x , cette probabilité conditionnelle s'écrit

$$\pi(d\theta | x) = \frac{P(dx, d\theta)}{M(dx)} = \frac{P(dx | \theta)\pi(d\theta)}{M(dx)} = \frac{P(dx | \theta)\pi(d\theta)}{\int_{\theta'} P(dx | \theta')\pi(d\theta')} \quad (1.8)$$

Cette probabilité conditionnelle est la probabilité *a posteriori* sur les états étant donné l'observation x . L'a posteriori donne donc une solution au problème de l'inversion du modèle (cf. figure 1.4). C'est la version générale du théorème de Bayes.

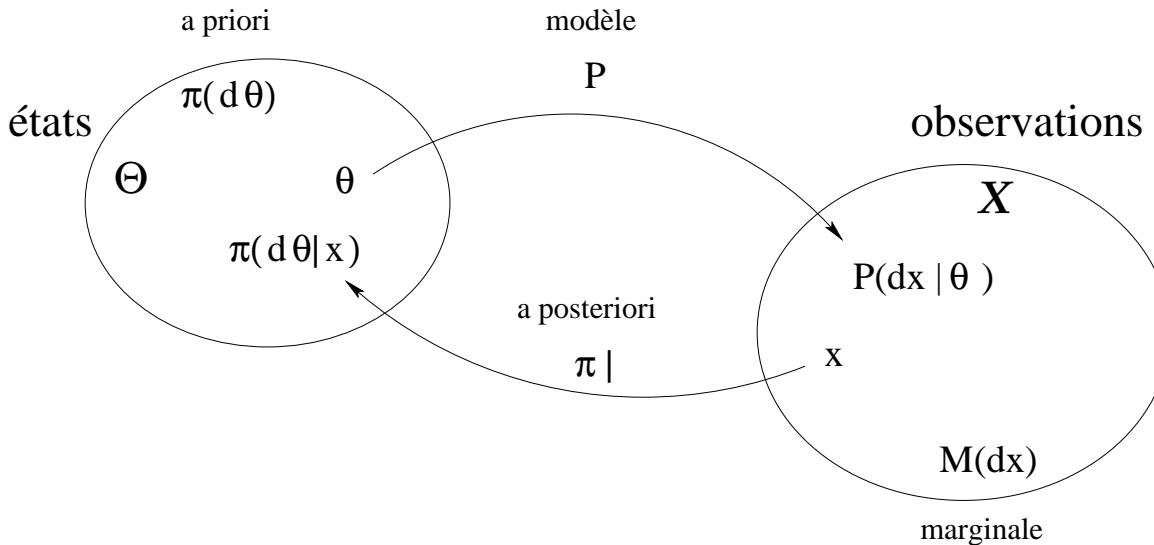


FIG. 1.4 – Concepts de la théorie bayésienne de la décision statistique. La donnée d'un *a priori* $\pi(\theta)$ permet d'inverser le modèle et de passer de x à θ par l'*a posteriori* $\pi(d\theta | x)$ qui associe à toute observation une répartition sur les états.

Remarquons la dualité entre les deux espaces

$$P(dx, d\theta) = P(dx | \theta)\pi(d\theta) = \pi(d\theta | x)M(dx) \quad (1.9)$$

L'*a priori* devient la *marginale sur les états* et l'*a posteriori* un modèle inversé, c'est-à-dire une probabilité de transition des observations vers les états.

Remarque L'*a priori* et le modèle donnent directement la conjointe par simple produit. Le problème de construire un *a priori* qui corresponde à une marginale et un modèle donnés n'est pas aussi immédiat, car ce sont des répartitions sur le même espace (les observations). Ce problème sera abordé plus bas dans cette étude, lorsque nous chercherons à déterminer un *a priori* compatible avec des observations préalables donnant une approximation de la marginale.

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

$$f(\theta | x) = \frac{\mathcal{L}(\theta | x)f(\theta)}{\int \mathcal{L}(\theta' | x)f(\theta')d\theta'} \quad (1.10)$$

Si l'on admet une équiprobabilité a priori des états, ce qui serait une manière de dire que l'on ne sait rien, la fonction $f(\theta)$ est constante, et la méthode du maximum de vraisemblance consiste à travailler avec le mode de l'a posteriori.

La probabilité a priori permet de tenir compte de connaissances préalables à la mesure. Par exemple si l'on sait que l'état ne peut sortir d'un sous-ensemble d'états Θ' , il suffit de se donner un $\pi(d\theta)$ qui s'annule en dehors. Mais il faut être extrêmement prudent avec ce type d'a priori, car selon la formule de Bayes (1.8), l'a posteriori s'annulera aussi en dehors de Θ' . Les états en dehors de ce sous-ensemble auront une probabilité nulle quel que soit le résultat de la mesure. Ils sont exclus définitivement. Dans le cas où l'état effectif est en dehors Θ' , la mesure aura tendance à concentrer l'a posteriori vers le bord de ce sous-ensemble. Ceci est d'ailleurs un indice de mauvais choix. D'une manière générale, les a priori pointus (de petite variance) donnent peu de pouvoir à la mesure, alors que les a priori larges lui en donnent plus. Ceci se retrouvera lors de la recherche d'a priori dits non informatifs. Cette structure permet de plus la formalisation d'une répétition de mesures en face du même état, en prenant comme a priori d'une nouvelle mesure l'a posteriori résultant de la précédente.

Illustration, modèle poissonnien simple Lors de la mesure d'une source radioactive, si le rayonnement ambiant est négligeable, la répartition du nombre d'impulsions N sous une activité a est la poissonnienne (cf 5.2)

$$P(N | a) = \mathcal{P}(N | a\epsilon t) \quad (1.11)$$

Où ϵ est la sensibilité du détecteur et t la durée de la mesure. La figure 1.5 synthétise les différentes notions définies ci-dessus pour un a priori basé sur une loi gamma. Nous verrons plus tard l'intérêt de ce type d'a priori. L'a priori se situe à gauche de la figure. Le modèle est représenté en haut pour un état particulier. La zone centrale représente la répartition conjointe obtenue par produit du modèle et de l'a priori. En bas est représentée la marginale qui donne la répartition des impulsions que l'on obtiendrait si l'activité de la source était prise au hasard selon l'a priori. Un a posteriori particulier, issu d'un tranche verticale de la conjointe est placé à droite. Remarquons que l'a priori est aussi la marginale sur l'axe des états.

1.5 Décision statistique

La théorie générale de la décision statistique définit une certain nombre de fonctions que nous allons rapidement passer en revue.

1.5.1 Fonction de perte

La prise de décision ne peut pas s'effectuer en se basant sur l'état du système, qui est inconnu. L'observateur doit donc trouver un moyen d'utiliser les observations de manière optimale, malgré leur caractère non déterministe.

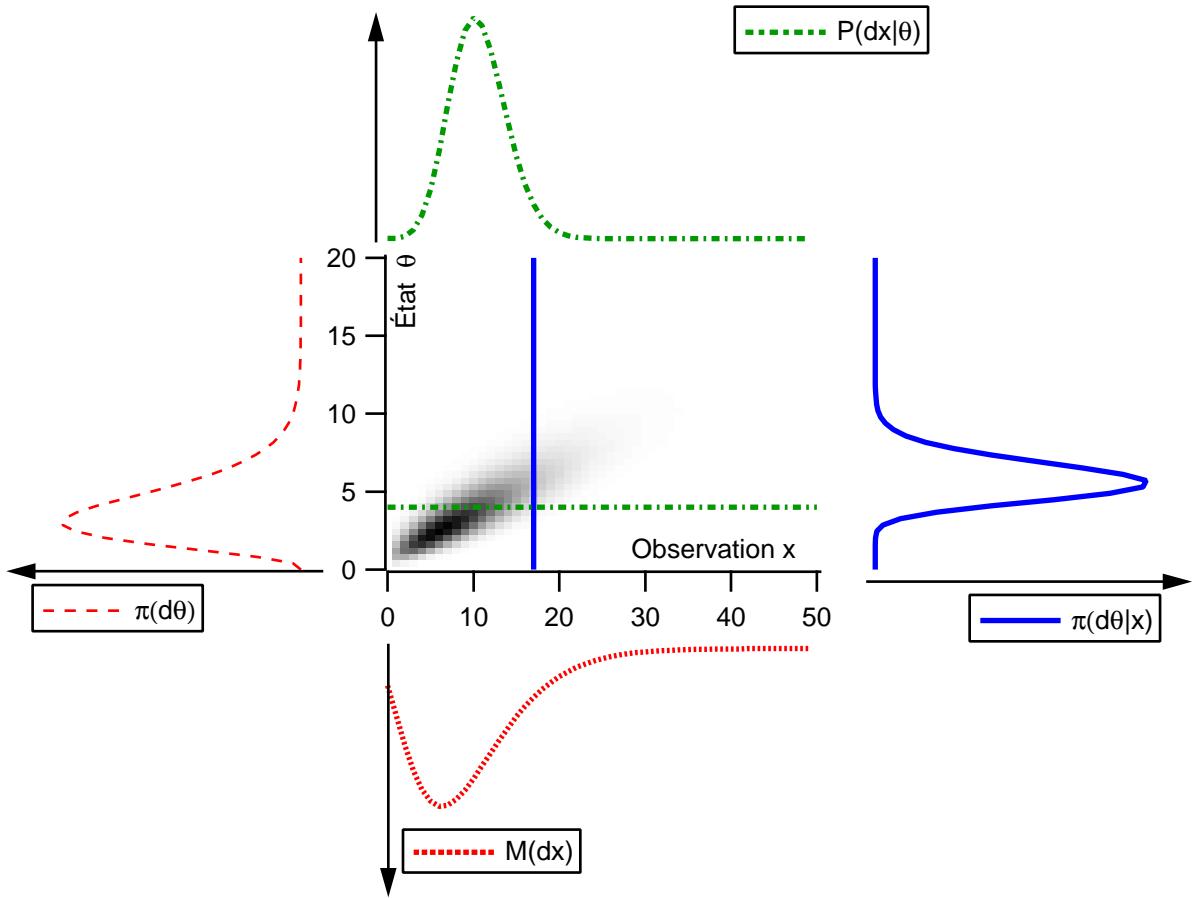


FIG. 1.5 – Illustration du modèle poissonnien. Pour un *a priori* de type gamma et un modèle poissonnien, on a représenté (dans le sens des aiguilles d'une montre depuis neuf heures) l'*a priori* $\pi(d\theta)$, le modèle $P(dx | \theta)$ pour une valeur particulière de θ , l'*a posteriori* $\pi(d\theta | x)$ et la marginale $M(dx)$. Les états sont décrits par un activité $\theta = a$ et les observations sont des nombres d'impulsions $x = N$.

Le risque de prendre une mauvaise décision est quantifié par une *fonction de perte*

$$L(d | \theta)$$

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

Illustration Supposons que l'observateur ait à sa disposition un algorithme lui permettant d'estimer une grandeur physique réelle. Il se trouve dans le cas de figure où l'espace des décisions s'identifie à celui des états : $\mathcal{D} = \Theta$. La fonction de perte utilisée dans ce cas est souvent une fonction croissante de l'écart entre la valeur estimée $\hat{\theta}$ et l'état effectif θ . Les fonctions de perte quadratiques $L(\hat{\theta} | \theta) = (\hat{\theta} - \theta)^2$ sont particulièrement intéressantes.

1.5.2 Stratégies et risque fréquentiste

Soit s une stratégie :

$$s : \mathcal{X} \longrightarrow \mathcal{D}$$

Le modèle étant donné, il est possible d'effectuer une moyenne par intégration de la fonction de perte. Ceci conduit à une évaluation des diverses stratégies. La grandeur obtenue est le *risque fréquentiste* qui s'exprime par

$$R(s | \theta) = \int_x L(s(x) | \theta) P(dx | \theta)$$

qui évalue le coût moyen de la stratégie s si le système est dans l'état θ .

1.5.3 Risque bayesien moyen

Si l'on dispose d'une répartition $\pi(d\theta)$ sur les états, l'intégration de la fonction de perte donne ce qu'il est convenu d'appeler le *risque bayesien moyen*

$$\rho(d | \pi) = \int_{\theta} L(d | \theta) \pi(d\theta)$$

C'est le coût moyen de la décision d . Une décision minimisant ce coût est dit *décision de Bayes* pour la probabilité en question.

Illustration Si la fonction de perte est quadratique, on peut écrire

$$\rho(\hat{\theta} | \pi) = \int_{\theta} (\hat{\theta} - \theta)^2 \pi(d\theta) \quad (1.12)$$

Le risque minimum s'obtient en annulant la dérivée par rapport à $\hat{\theta}$

$$\frac{\partial \rho}{\partial \hat{\theta}} = 0 \quad (1.13)$$

$$2 \int_{\theta} (\hat{\theta} - \theta) \pi(d\theta) = 0 \quad (1.14)$$

$$\hat{\theta} = \mathbf{E}\theta \quad (1.15)$$

$$\rho_{min} = \mathbf{V}\theta \quad (1.16)$$

La décision de Bayes est donc l'espérance mathématique de la répartition $\pi(d\theta)$, et le risque minimum est sa variance.

1.5.4 Risque de Bayes

La donnée simultanée d'une stratégie et d'un a priori permet alors d'envisager une fusion des deux risques précédents. Le *risque de Bayes* sera l'intégration du risque fréquentiste par rapport à l'a priori

$$r(s | \pi) = \int_{\theta} R(s | \theta) \pi(d\theta) = \int_{\theta} \pi(d\theta) \int_x L(s(x) | \theta) P(dx | \theta)$$

Ce risque mesure le coût moyen associé à la stratégie s . La minimisation de ce risque est l'un des buts de la théorie de la décision statistique.

1.5.5 Stratégies optimales

Supposons que nous ayons obtenu une observation x . Partant de l'a priori $\pi(d\theta)$, le théorème de Bayes donne l'a posteriori $\pi(d\theta | x)$. Considérons le risque bayesien moyen pour une décision d associé à cette nouvelle probabilité en indiquant par un indice la dépendance en x

$$\rho_x(d) = \int_{\Theta} L(d | \theta) \pi(d\theta | x)$$

Il existe généralement une décision donnant une valeur minimale pour cette fonction. Associons donc cette décision à l'observation x , nous avons une stratégie

$$s(x) = \arg \min_d \rho_x(d)$$

Evaluons le risque de Bayes associé à cette stratégie

$$\begin{aligned} r(s | \pi) &= \int_{\Theta} \pi(d\theta) \int_x L(s(x) | \theta) P(dx | \theta) \\ &= \int_{\Theta, x} \pi(d\theta) P(dx | \theta) L(s(x) | \theta) \\ &= \int_{\Theta, x} \pi(d\theta | x) M(dx) L(s(x) | \theta) \\ &= \int_x M(dx) \rho_x(s(x)) \end{aligned}$$

Mais par définition, pour n'importe quelle stratégie $s'(x)$

$$\rho_x(s(x)) \leq \rho_x(s'(x))$$

ce qui prouve que $r(s | \pi) \leq r(s'(x) | \pi)$ car $M(dx)$ est positive. La stratégie qui consiste à prendre la décision d qui minimise le risque moyen de Bayes pour l'a posteriori issu de l'observation x est donc optimale. En d'autre termes, la meilleure stratégie est de prendre la décision de Bayes relative à l'a posteriori.

Chapitre 2

Incertitude

2.1 Introduction

Il est paradoxal que l'une des préoccupations majeures des sciences exactes soit l'incertitude. Particulièrement en physique, aucune évaluation d'une grandeur ne peut (en principe) être présentée sans barre d'erreur ou intervalle de confiance. Il existe plusieurs approches pour décrire une connaissance incomplète. La plus connue consiste à attribuer à chaque résultat possible son poids. Au lieu de donner le résultat de mesure par un seul nombre, on donne l'ensemble de ces poids, qui devient une répartition ou une probabilité. La probabilité d'un résultat est ainsi celle de son apparition. Le résultat devient une variable aléatoire. Un cadre plus général que la probabilité est apparu dans le deuxième moitié du XX^e siècle. Il s'agit de la *logique floue* [12] [13] qui associe aussi un nombre à chaque évènement, sa *possibilité*. Cette notion, proche de la probabilité, n'est pas soumise aux mêmes lois. En particulier, on n'exige pas la normalisation, ni l'additivité. Cette théorie n'est pas encore couramment utilisée en physique, mais a un bel avenir en ce qui concerne la représentation des a priori bayesiens.

La description d'une répartition complète nécessite beaucoup d'information. Aussi cherche-t-on à la résumer en quelques nombres déterminants. Les plus importants sont l'*espérance* et la *variance*. Lors de la répétition de la mesure, ces deux grandeurs peuvent être estimées par la moyenne arithmétique des résultats et le carré de la déviation standard de l'échantillon. Deux théorèmes appuient cette manière de faire. La *loi des grands nombres*, qui assure la convergence de la moyenne vers l'espérance, et le *théorème central limite*, qui permet l'estimation de la variance par le carré de la déviation standard. L'incertitude exprimée par la variance est dite statistique. Mais il se peut que la manière d'estimer le résultat soit tributaire d'une erreur systématique.

Le théorème central limite implique aussi que les répartitions pour les expériences répétées convergent vers la *loi de Gauss*, justifiant ainsi le remplacement de toute la courbe par l'espérance et la variance seules. D'autre part, les variances des combinaisons de résultats peuvent alors facilement s'estimer au premier ordre par la *loi de propagation des incertitudes*.

Deux approximations essentielles apparaissent ici. Premièrement, le théorème central limite donne une convergence au sens mathématique, c'est-à-dire lorsque le nombre de répétitions devient infini, ce qui n'est jamais le cas en réalité. Deuxièmement, la loi de propagation des incertitudes n'est valable qu'au premier ordre, si les variances des constituants sont petites. La mesure de la radioactivité de bas niveau se trouve justement confrontée à des nombres de répétitions qui sont faibles et ne permettent plus de justifier ces approximations.

Ces concepts prennent un sens plus précis dans le contexte de la TBDS, qui permet en plus de s'affranchir des deux approximations ci-dessus. Les notions centrales sont ici l'*estimateur* et l'*a posteriori*.

2.2 Précision et exactitude

D'une manière générale donnons-nous un modèle réel multidimensionnel $P(dx | \theta)$ et un estimateur de l'état $\tau(x)$. Pour un état donné θ , l'espérance de cet estimateur vaut

$$E_\theta \tau = \int_x P(dx | \theta) \tau(x) \quad (2.1)$$

Lorsque l'observateur répète son expérience en face du même état, la moyenne de ses estimations converge vers cette espérance. Un estimateur sera donc bon si la vraie valeur est proche de cette espérance. On introduit ainsi la notion de *biais*, mesurant l'exactitude tributaire des erreurs systématiques

$$b_\theta(\tau) = |E_\theta \tau - \theta|^2 \quad (2.2)$$

La variance d'une estimation représente l'incertitude relative à sa *précision*, qui est définie par la diagonale de la matrice de covariance

$$(C_\theta \tau)_{ij} = \int_x P(dx | \theta) (\tau_i(x) - E_\theta \tau_i)(\tau_j(x) - E_\theta \tau_j) \quad (2.3)$$

où les éléments non diagonaux expriment les corrélations entre ses composantes.

Remarque Si nous mesurons la qualité d'une estimateur à l'aide d'une fonction de perte quadratique $L(\theta' | \theta) = (\theta - \theta')^2$, le risque fréquentiste s'exprime

$$R(\tau | \theta) = \int_x (\tau(x) - \theta)^2 P(dx | \theta) \quad (2.4)$$

$$= \int_x (\tau(x) - E_\theta \tau + E_\theta \tau - \theta)^2 P(dx | \theta) \quad (2.5)$$

$$= \int_x (\tau(x) - E_\theta \tau)^2 P(dx | \theta) + \int_x (E_\theta \tau - \theta)^2 P(dx | \theta) \\ + 2 \int_x (\tau(x) - E_\theta \tau)(E_\theta \tau - \theta) P(dx | \theta) \quad (2.6)$$

$$= V_\theta \tau + b_\theta(\tau) \quad (2.7)$$

C'est-à-dire que le risque se mesure comme somme de la variance et du biais.

2.3 Estimateur de Bayes

Dans le cas réel, l'*a posteriori* consécutif à une mesure est une répartition qui admet une espérance (vectorielle) et une matrice de covariance. Soit x l'observation, l'espérance *a posteriori* donne un état qui est l'*estimateur de Bayes*

$$\tau_{Bayes,i}(x) = E\theta_i | x = \int_\theta \theta_i \pi(d\theta | x) \quad (2.8)$$

l'incertitude étant représentée par la matrice de covariance de l'*a posteriori*

$$C_{Bayes,ij}(x) = \int_\theta (\theta_i - \tau_{Bayes,i}(x))(\theta_j - \tau_{Bayes,j}(x)) \pi(d\theta | x) \quad (2.9)$$

Remarques L'espérance et la covariance présentées ici sont globales. Elles ne dépendent pas de l'état, mais de l'observation. Si l'état est θ , l'estimateur de Bayes admet évidemment aussi une espérance $E_\theta \tau_{Bayes}$ et une covariance $C_\theta \tau_{Bayes}$ locales qu'il ne faut pas confondre avec les précédentes.

L'estimateur de Bayes peut être biaisé. Nous verrons à la section 9 que ce biais traduit une information a priori.

A tout a priori correspond un estimateur de Bayes. On peut se demander si l'inverse est vrai. On aurait ainsi une matrice de covariance globale a posteriori associée à un estimateur. Pour un τ donné, il suffit de résoudre l'équation en π

$$\forall x, \int_\theta (\tau(x) - \theta) P(dx | \theta) \pi(d\theta) = 0$$

En d'autre termes, les solutions sont les probabilités contenues dans le noyau de l'opérateur intégral

$$(\tau(x) - \theta) P(dx | \theta)$$

De plus, elles doivent être positives.

2.4 Propagation des incertitudes

Si l'observateur n'est intéressé que par une combinaison $\phi(\theta)$ des composantes de l'état, l'estimation et l'incertitude relatives à cette combinaison en présence d'une observation x sont très simples à formuler. Les états se répartissent a posteriori selon $\pi(d\theta | x)$ et l'on a

$$E\phi | x = \int_\theta \phi(\theta) \pi(d\theta | x) \quad (2.10)$$

$$V\phi | x = \int_\theta (\phi(\theta) - E\phi | x)^2 \pi(d\theta | x) \quad (2.11)$$

Si ϕ est vectorielle de composantes $\phi_j(\theta)$, la généralisation est immédiate et donne le vecteur des espérances ainsi que la matrice de covariance

$$(E\phi | x)_j = \int_\theta \phi_j(\theta) \pi(d\theta | x) \quad (2.12)$$

$$(C\phi | x)_{jk} = \int_\theta (\phi_j(\theta) - (E\phi | x)_j)(\phi_k(\theta) - (E\phi | x)_k) \pi(d\theta | x) \quad (2.13)$$

En développant ϕ au premier ordre, on obtient la *loi de propagation des incertitudes* exprimant la variance de ϕ en fonction de la covariance a posteriori des θ_i . Soit

$$\hat{\theta} = E\theta | x = \tau_{Bayes}(x) \quad (2.14)$$

Supposons l'a posteriori concentré au voisinage de $\hat{\theta}$. Il vient

$$\phi(\theta) \simeq \phi(\hat{\theta}) + \sum_i \partial_i \phi(\hat{\theta})(\theta_i - \hat{\theta}_i) \quad (2.15)$$

$$E\phi | x \simeq \phi(\hat{\theta}) \quad (2.16)$$

Puis

$$V\phi | x = \int_{\Theta} (\phi(\theta) - \phi(\hat{\theta}))^2 \pi(d\theta | x) \quad (2.17)$$

$$\simeq \int_{\Theta} \left(\sum_i \partial_i \phi(\hat{\theta}) (\theta_i - \hat{\theta}_i) \right)^2 \pi(d\theta | x) \quad (2.18)$$

$$\simeq \int_{\Theta} \sum_{ik} \partial_i \phi(\hat{\theta}) \partial_k \phi(\hat{\theta}) (\theta_i - \hat{\theta}_i) (\theta_k - \hat{\theta}_k) \pi(d\theta | x) \quad (2.19)$$

$$\simeq \sum_{ik} \partial_i \phi(\hat{\theta}) \partial_k \phi(\hat{\theta}) C_{Bayes,ik}(x) \quad (2.20)$$

La matrice de covariance de Bayes étant celle de l'estimateur de l'état, on retrouve l'expression habituelle. L'approximation est d'autant meilleure que l'a posteriori est concentré autour de la vraie valeur.

2.5 Ensembles de crédibilité

L'incertitude sur les valeurs physiques peut également être présentée comme un *intervalle de confiance*. Ces intervalles sont souvent obtenus par inversion brute du modèle, donnant à la notion de confiance un caractère pseudo-probabiliste qui débouche sur des paradoxes. D'intéressantes propositions ont conduit à un protocole hybride faisant intervenir des éléments bayésiens dans un contexte fréquentiste qui sera présenté à la section 10.2.4. Ce protocole peut être avantageusement remplacé par une approche entièrement bayésienne.

L'existence d'un a posteriori permet de définir facilement des intervalles de ce type, qui sont nommés ici ensembles de *crédibilité* : un sous-ensemble $E \subset \Theta$ est dit de crédibilité γ si sa probabilité a posteriori est supérieure ou égale à γ . Cette valeur est typiquement fixée à 95% ou 99%.

$$\pi(E | x) \geq \gamma \quad (2.21)$$

Opposée à la notion vague de confiance, la crédibilité est une vraie probabilité.

2.6 Estimateurs et ensembles de crédibilité dans le cas poissonnien

Reprenons le modèle poissonnien pour une source sans bruit de fond. Ce modèle donne la probabilité d'obtenir N impulsions lorsqu'une source d'activité a est observée pendant un temps t avec un instrument de sensibilité ϵ

$$P(N | a) = \mathcal{P}(N | a\epsilon t)$$

Nous verrons à la partie suivante qu'un bon a priori sur l'activité pour ce modèle en l'absence d'information sur l'état est

$$\pi(da) \propto a^{-\frac{1}{2}} da$$

Ceci induit l'a posteriori

$$\pi(da | N) \propto e^{-a\epsilon t} a^{N-\frac{1}{2}} da$$

qui n'est autre qu'une loi gamma $\text{Ga}(da | N + \frac{1}{2}, \epsilon t)$. Les lois gamma jouent un rôle crucial lorsque apparaissent des modèles de Poisson. Il en sera plus longuement question aux sections 7.1.2 et 7.2.2.2.

L'estimateur de Bayes de l'activité et son incertitude (ici donnée par un écart-type) valent

$$\mathbf{E}a | N = \frac{N + \frac{1}{2}}{\varepsilon t} \quad (2.22)$$

$$\sqrt{\mathbf{V}a} | N = \frac{\sqrt{N + \frac{1}{2}}}{\varepsilon t} \quad (2.23)$$

Ce calcul donne une valeur strictement positive même si aucune impulsion n'a été détectée. D'autre part, on retrouve une incertitude de l'ordre de la racine du nombre d'impulsions pour les grands comptages. Les intervalles de crédibilité pour l'activité sont obtenus en utilisant les lois gamma incomplètes qui seront abordées plus loin. Les figures 2.1 à 2.3 en donnent quelques réalisations. Les intervalles choisis sont ceux de longueur minimale au sens habituel sur l'axe des états. L'usage habituel consiste à prendre des intervalles de probabilité $\gamma/2$ de chaque côté du maximum. Ce choix fonctionne bien pour des lois symétriques, mais ce n'est pas le cas ici.

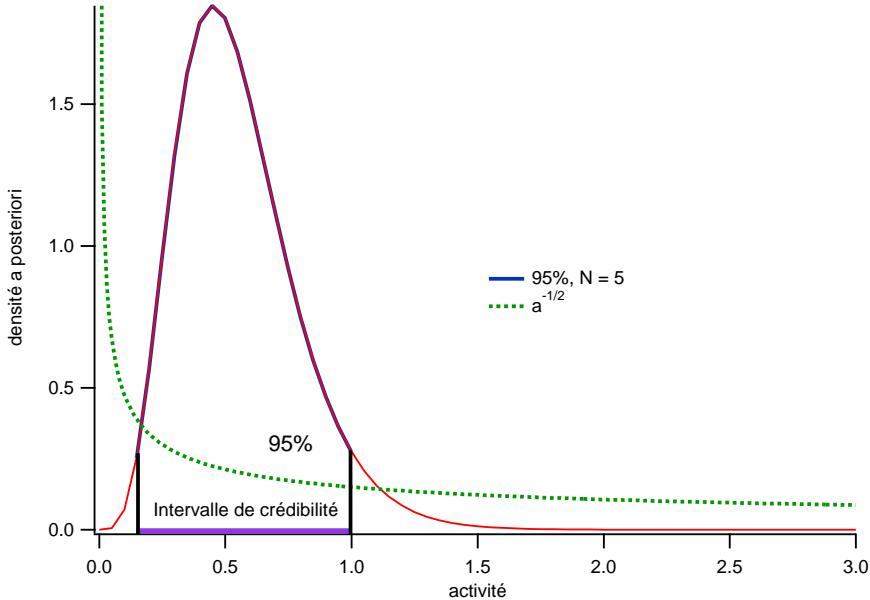


FIG. 2.1 – Densités a posteriori. L' intervalle de crédibilité à 95% est indiqué pour 5 impulsions détectées. L'efficacité du compteur est de 1 et la durée de mesure 10. La courbe en $1/\sqrt{a}$ est un cas limite qui sera discuté plus loin.

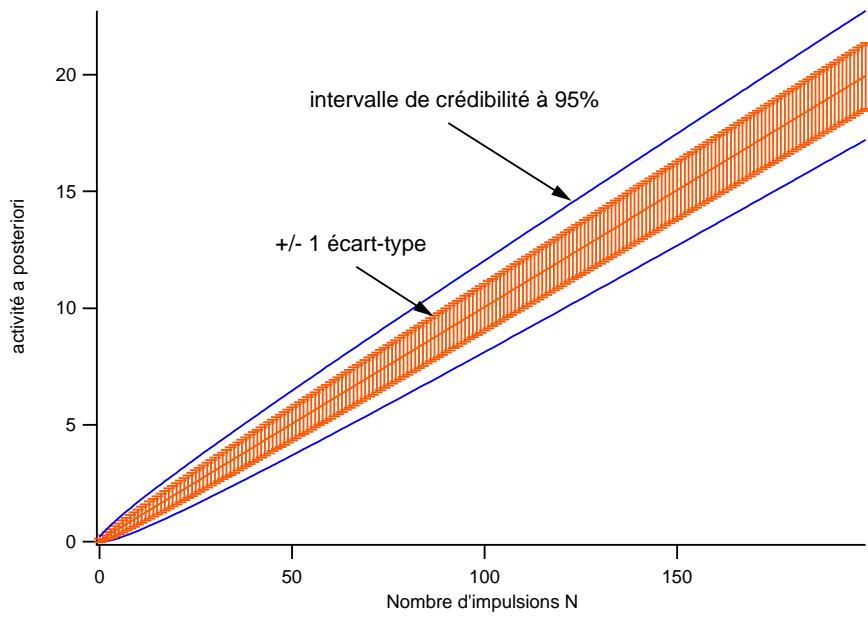


FIG. 2.2 – Variation des intervalles de crédibilité. *Dans les mêmes conditions que celles de la figure 2.1, on a représenté les intervalles de crédibilité et l'espérance a posteriori \pm un écart-type.*

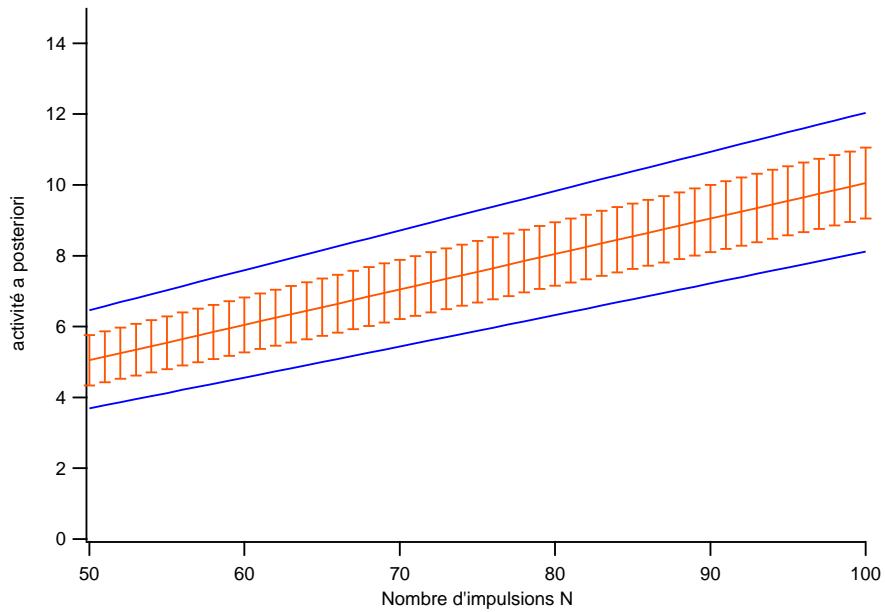


FIG. 2.3 – Variation des intervalles de crédibilité. *Agrandissement d'une plage à haut nombre de détections. On retrouve la formulation classique avec un intervalle quasiment symétrique de longueur ± 1.96 écart-type.*

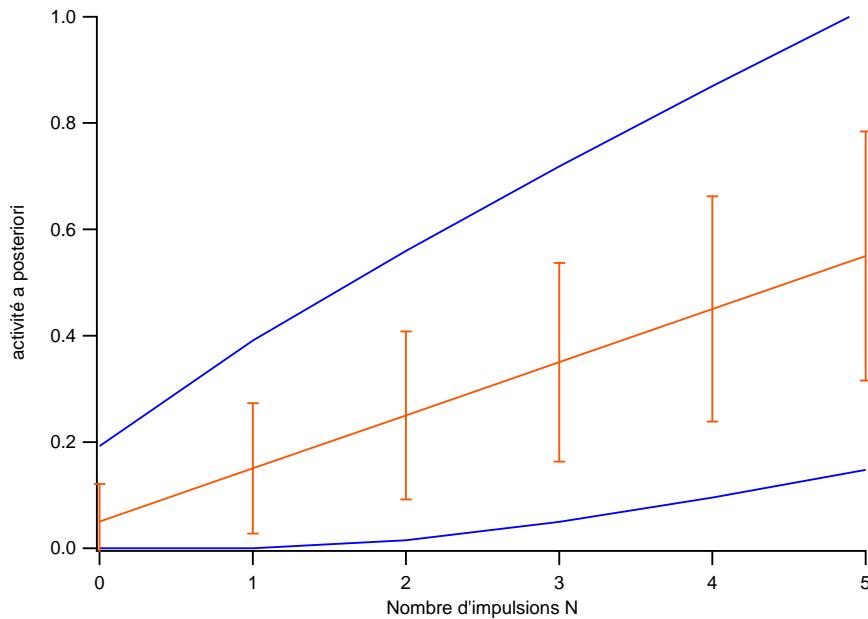


FIG. 2.4 – Variation des intervalles de crédibilité. Agrandissement de la plage à bas nombre de détections. Les intervalles ne sont plus symétriques et l'approximation classique ne se justifie plus.

Chapitre 3

Tests d'hypothèses simples

3.1 Introduction

Lorsque l'espace des décisions Θ est réduit à deux éléments, et que la décision consiste à choisir entre ces deux éléments, on parle de test d'hypothèses. Par exemple, si le problème consiste à déterminer si une source radioactive d'une activité donnée a est présente ou non lors d'une mesure, l'espace des états devient $\{0, a\}$, et la décision est l'une des affirmations : *une source d'activité a est présente* versus *il y a pas de source*.

La théorie des tests d'hypothèses s'intègre très naturellement dans le formalisme de la TBDS. Il faut remarquer que cette théorie ne fait pas intervenir d'a priori lorsque les hypothèses en opposition sont simples. Elle n'est donc pas bayesienne. Le choix entre deux hypothèses simples permet d'utiliser le modèle seul. En effet, si l'état θ_0 est en compétition avec l'état θ_1 , il suffit de comparer $P(dx | \theta_0)$ à $P(dx | \theta_1)$. La situation n'est plus du tout la même si les hypothèses sont composées. Par exemple, si pour un état critique θ_c , l'on devait opposer l'événement $\{\theta > \theta_c\}$ à $\{\theta \leq \theta_c\}$, car l'expression $P(dx | \theta > \theta_c)$ ne peut pas être obtenue depuis $P(dx | \theta)$ sans utiliser d'a priori.

3.2 Test de Neyman-Pearson

Soit l'espace des états $\Theta = \{\theta_0, \theta_1\}$. L'espace des décisions se réduit à deux éléments $\mathcal{D} = \{\mathcal{H}_0, \mathcal{H}_1\}$, où la décision \mathcal{H}_i consiste en l'affirmation que le système est dans l'état θ_i . Si le modèle est donné par une fonction de vraisemblance \mathcal{L}

$$P(dx | \theta) = \mathcal{L}(\theta | x)dx$$

une solution générale a été proposée par Neyman et Pearson [14] [15].

Ces auteurs définissent d'abord le *rapport de vraisemblance*¹ associé à une observation x

$$\mathcal{LR}(x) = \frac{\mathcal{L}(\theta_1 | x)}{\mathcal{L}(\theta_0 | x)}$$

Une stratégie de décision (au sens de la TBDS (1.5.2)) consiste alors à se donner un *niveau critique* ℓ pour ce rapport, et à associer à x la décision \mathcal{H}_1 s'il est supérieur à ℓ , ou \mathcal{H}_0 dans le cas contraire. L'ensemble des observations x conduisant à une décision \mathcal{H}_1 est dit *région critique* ou *région de réjection*

¹Plus généralement, il est possible d'utiliser ici le théorème de Radon-Nicodym et de faire le rapport des modèles $\frac{P(dx|\theta_1)}{P(dx|\theta_0)}$

(sous-entendu de l'hypothèse \mathcal{H}_0). En d'autre termes, le test de Neyman-Pearson est donné par la stratégie s

$$s(x) = \begin{cases} \mathcal{H}_0 & \text{si } \mathcal{LR}(x) \leq \ell \\ \mathcal{H}_1 & \text{si } \mathcal{LR}(x) > \ell \end{cases}$$

3.3 Probabilités d'erreur

Etant donné un modèle, pour n'importe quelle stratégie s visant à accepter une hypothèse ou l'autre, la décision peut être correcte ou non. On introduit donc les probabilités dites de premier et second type définies comme suit :

$$\alpha = P(s = \mathcal{H}_1 \mid \theta_0) \quad (3.1)$$

$$\beta = P(s = \mathcal{H}_0 \mid \theta_1) \quad (3.2)$$

3.3.1 Erreur de premier type

α est la probabilité de prendre à tort la décision que le système est dans l'état θ_1 (où d'accepter à tort l'hypothèse \mathcal{H}_1). Cette erreur est dite de *premier type* car l'hypothèse \mathcal{H}_0 est généralement associée à une absence d'effet (population neutre ou témoin, effets zéro). Il s'avère que le modèle est souvent bien connu dans ce cas. Cette probabilité est aussi appelée probabilité de *faux positif*, faux car on se trompe, positif car on admet la présence d'un effet. La grandeur $\bar{\alpha} = 1 - \alpha$ est souvent appelée *spécificité*, ou probabilité de *vrai négatif*.

3.3.2 Erreur de second type

β est la probabilité de prendre à tort la décision que le système est dans l'état θ_0 . C'est l'*erreur de second type* ou de *faux négatif*. $\bar{\beta} = 1 - \beta$ est appelée *sensibilité, puissance du test au point* θ_1 (un test est d'autant plus puissant que $\bar{\beta}$ est grand) ou encore probabilité de *vrai positif*. Cette probabilité est généralement plus difficile à obtenir que α , car l'hypothèse \mathcal{H}_0 est souvent associée à un groupe témoin qui est généralement mieux connu statistiquement que celui en test.

Le tableau 3.1 décrit les différents termes sous lesquels on trouve ces probabilités.

TAB. 3.1 – Probabilités d'erreur pour le test à deux issues

		Etat	
		θ_0	θ_1
Décision	\mathcal{H}_0	$1 - \alpha$ vrai négatif spécificité	β faux négatif second type
	\mathcal{H}_1	α faux positif premier type	$1 - \beta$ vrai positif sensibilité, puissance

Chapitre 4

Evaluation bayesienne des courbes ROC

Une manière d'évaluer un test consiste à examiner sa courbe ROC, qui sera définie ci-dessous. Cette méthode ne donne malheureusement que des comparaisons partielles. D'un point de vue global, la valeur d'un test ne prend tout son sens qu'en introduisant la notion de prévalence, qui n'est autre qu'un a priori. Le but de cette section est de présenter ces notions classiques en utilisant les concepts de la TBDS, et de déboucher sur une méthode d'optimisation des tests lorsque l'espace des observations est discret et donne des courbes ROC non différentiables, voire discontinues. Cette méthode permettra d'une part le calcul des niveaux critiques, d'autre part la comparaison des tests entre eux d'une manière globale lorsque l'on dispose d'une fonction de perte et d'un a priori.

4.1 Définition des courbes ROC

Lorsque la décision se prend avec des stratégies basées sur des niveaux critiques, la fonction de l'observation dont la valeur sera comparée au niveau critique prend le nom de *critère*, ou *fonction de test*. Soit $K(x)$ une telle fonction, et ℓ un *niveau critique*, la stratégie associée s'exprime par

$$K(x) \leq \ell \Rightarrow \mathcal{H}_0 \quad (4.1)$$

$$K(x) > \ell \Rightarrow \mathcal{H}_1 \quad (4.2)$$

Les probabilités α et β sont définies de la même manière. La littérature récente préférant se référer à $1 - \beta = \bar{\beta}$, on procédera de même ici.

$$\begin{aligned} \alpha &= P(K > \ell \mid \theta_0) \\ \bar{\beta} &= P(K > \ell \mid \theta_1) \end{aligned}$$

La paire de probabilités $(\alpha, \bar{\beta})$ décrit une courbe dans le carré unité $[0, 1]^2$ lorsque ℓ varie. Cette courbe a reçu le nom de ROC, qui est un acronyme pour *Receiver Operating Characteristic curve* [16]. Elle qualifie le système d'observation (représenté par sa fonction de test K), par le biais de la variation de ses probabilités d'erreur.

$$ROC_K : \ell \mapsto (\alpha, \bar{\beta})$$

Lorsque ℓ augmente, les probabilités α et $\bar{\beta}$ décroissent simultanément. La courbe part au point $(1, 1)$ lorsque $\ell = -\infty$, et aboutit en $(0, 0)$ lorsque $\ell = \infty$.

Les stratégies optimales minimisent les probabilités d'erreur, où, ce qui est équivalent, maximisent la puissance (la sensibilité) et la spécificité. Les courbes proches du coin supérieur gauche du carré $[0, 1]^2$

seront les meilleures. Bien que la comparaison globale de deux courbes ROC par une seule grandeur scalaire n'aie guère d'utilité pratique, cela se fait régulièrement dans la littérature. Par exemple en considérant la surface comprise sous la courbe (son intégrale), notée A_z , que l'on aimerait voisine de 1, ou bien par la probabilité α lorsque $\alpha = \beta$ (intersection avec la deuxième diagonale), que l'on aimerait aussi petite que possible.

Plus généralement, à toute stratégie peut être associé son couple de probabilités $(\alpha, \bar{\beta})$. La première diagonale représente une limite inférieure pour les stratégies admissibles. En effet, si une stratégie donne un couple de valeurs au dessous, il suffit de prendre systématiquement la décision inverse, ce qui donne un résultat statistiquement meilleur. Considérons une stratégie aléatoire s_γ ne tenant pas compte de l'observation, mais choisissant l'hypothèse \mathcal{H}_1 avec probabilité γ . Les probabilités d'erreur sont donc

$$\alpha = P(s_\gamma = \mathcal{H}_1 | \mathcal{H}_0) = \gamma \quad (4.3)$$

$$\bar{\beta} = P(s_\gamma = \mathcal{H}_1 | \mathcal{H}_1) = \gamma \quad (4.4)$$

Son point représentatif est sur la première diagonale, qui est complètement parcourue lorsque γ varie de 0 à 1.

Les tests basés sur des rapports de vraisemblance $K = \mathcal{LR}$ sont optimaux [14]. En outre, toute transformation d'une fonction de test par une fonction croissante donne un test équivalent, car il suffit de modifier les niveaux en conséquence. En particulier, on utilise souvent le logarithme.

La figure 4.1 montre trois courbes ROC associées à trois fonctions de test K_1 , K_2 et K_3 . Le test associé à K_1 est supérieur à celui de K_2 , car les probabilités d'erreur le long de la première courbe sont systématiquement inférieures à celles de la deuxième. La courbe issue de K_3 , par contre, n'est pas globalement comparable aux deux premières.

4.2 Risques et pertes pour les test d'hypothèses

Les concepts de la TBDS prennent une forme extrêmement simple dans le cas des tests d'hypothèses. L'espace des états étant réduit à deux éléments $\Theta = \{\theta_0, \theta_1\}$, une probabilité s'y exprime par deux nombres positifs dont la somme est 1. Utilisant le dirac δ_θ pour désigner une probabilité concentrée en θ , il vient

$$\pi(\theta) = p_0 \delta_{\theta_0} + p_1 \delta_{\theta_1} \quad (4.5)$$

$$p_0 + p_1 = 1 \quad (4.6)$$

Toute fonction de perte peut être définie par le tableau 4.1.

TAB. 4.1 – Fonction de perte $L(d | \theta)$ pour le test à deux issues

Décision	Etat	
	θ_0	θ_1
\mathcal{H}_0	$-V_{00}$	V_{01}
\mathcal{H}_1	V_{10}	$-V_{11}$

Les V_{ij} sont positifs et les signes sont placés de sorte à obtenir des pertes négatives en cas de bonne décision.

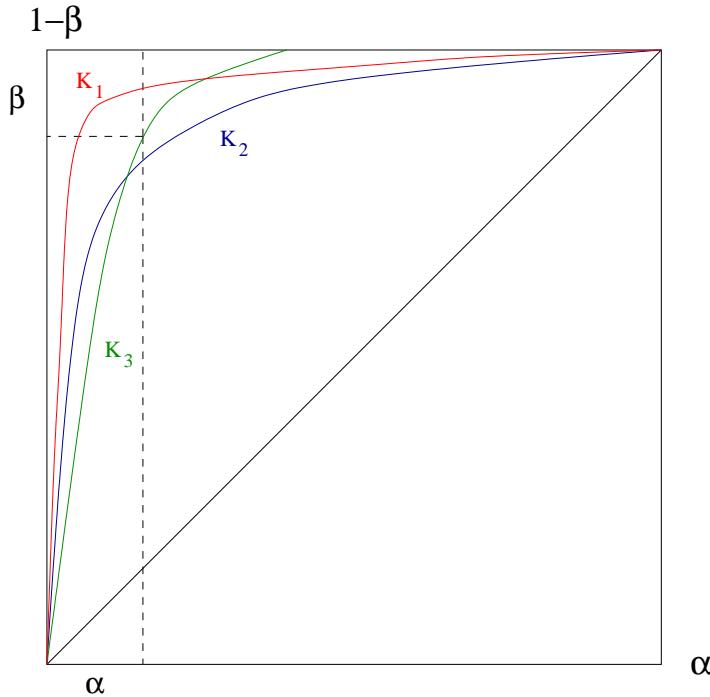


FIG. 4.1 – Exemples de courbes ROC. La courbe K_1 traduit un meilleur test que K_2 , mais ne peut être globalement comparée à K_3 .

V_{00} est le gain obtenu en prenant la bonne décision \mathcal{H}_0

V_{01} est le coût de la décision erronée \mathcal{H}_0

V_{10} est le coût de la décision erronée \mathcal{H}_1

V_{11} est le gain obtenu en prenant la bonne décision \mathcal{H}_1

4.3 Stratégies optimales

La recherche des stratégies optimales se ramène à celle des décisions de Bayes (cf sections 1.5.3 et 1.5.5) pour l'a posteriori sur les états.

Soient donc (p'_0, p'_1) les probabilités a posteriori de (θ_0, θ_1) après l'observation x et $M(x)$ la probabilité marginale de x

$$M(x) = p_0 P(x | \theta_0) + p_1 P(x | \theta_1) \quad (4.7)$$

$$p'_0 = p_0 P(x | \theta_0) / M(x) \quad (4.8)$$

$$p'_1 = p_1 P(x | \theta_1) / M(x) \quad (4.9)$$

Le risque moyen de Bayes a posteriori pour les décisions \mathcal{H}_0 et \mathcal{H}_1 devient selon (1.5.5)

$$\rho'_0 = -V_{00} p'_0 + V_{01} p'_1 \quad (4.10)$$

$$\rho'_1 = V_{10} p'_0 - V_{11} p'_1 \quad (4.11)$$

La décision à prendre est alors celle qui minimise le risque qui lui est associé. En d'autres termes, la région d'acceptance de l'hypothèse \mathcal{H}_1 est

$$\mathcal{A} = \{x \mid p'_1 < p'_0\} \quad (4.12)$$

$$= \{x \mid V_{10} p'_0 - V_{11} p'_1 < -V_{00} p'_0 + V_{01} p'_1\} \quad (4.13)$$

$$= \{x \mid (V_{00} + V_{10}) p'_0 < (V_{11} + V_{01}) p'_1\} \quad (4.14)$$

$$= \left\{ x \mid \frac{V_{00} + V_{10}}{V_{11} + V_{01}} < \frac{p_1 P(x \mid \theta_1)}{p_0 P(x \mid \theta_0)} \right\} \quad (4.15)$$

$$= \left\{ x \mid \frac{(V_{00} + V_{10}) p_0}{(V_{11} + V_{01}) p_1} < \frac{P(x \mid \theta_1)}{P(x \mid \theta_0)} \right\} \quad (4.16)$$

Définissons le niveau critique

$$\ell_C = \frac{(V_{00} + V_{10}) p_0}{(V_{11} + V_{01}) p_1} \quad (4.17)$$

qui est un terme ne dépendant que des a priori et de la fonction de perte. En notant que $\frac{P(x \mid \theta_1)}{P(x \mid \theta_0)}$ n'est autre que le rapport de vraisemblance, il vient

$$\mathcal{A} = \{x \mid \ell_C < \mathcal{LR}(x)\} \quad (4.18)$$

Le test minimisant le risque moyen est ainsi donné par une fonction de test qui n'est autre que le rapport de vraisemblance, appliqué au niveau critique ℓ_C . Le test de Neyman-Pearson ne se fait pas dans un cadre bayesien, puisqu'il ne met en oeuvre aucun a priori. Néanmoins, l'optimisation bayesienne donne exactement la même fonction de test. L'apport essentiel de cette dérivation est de relier le niveau critique lui-même à l'a priori et à la fonction de perte.

4.4 Calcul du risque

Pour une fonction de test K quelconque, il est possible de calculer le risque de Bayes associé à un niveau critique ℓ donnant des probabilités d'erreur $\alpha(\ell)$ et $\beta(\ell)$. L'application de la définition de la section 1.5.4 à ce cas particulier donne

$$\rho(\ell) = (\alpha V_{10} + (1 - \alpha)(-V_{00})) p_0 + (\beta V_{01} + (1 - \beta)(-V_{11})) p_1 \quad (4.19)$$

$$= \alpha(V_{10} + V_{00}) p_0 - \bar{\beta}(V_{01} + V_{11}) p_1 - V_{00} p_0 + V_{01} p_1 \quad (4.20)$$

Le but est de minimiser ce risque. Comme $V_{01} + V_{11} > 0$ et que la constante $-V_{00} p_0 + V_{01} p_1$ du deuxième membre de l'équation (4.20) ne joue aucun rôle dans le processus de minimisation, il suffit de considérer le risque dit *réduit*

$$\tilde{\rho}(\ell) = \frac{\rho(\ell) + V_{00} p_0 - V_{01} p_1}{(V_{01} + V_{11}) p_1} \quad (4.21)$$

qui permet d'exprimer la relation (4.20) par un produit scalaire

$$\tilde{\rho}(\ell) = (\ell_C, -1) \cdot (\alpha, \bar{\beta}) \quad (4.22)$$

où ℓ_C est bien celui défini en (4.17).

4.4.1 Cas différentiable

Si le risque $\tilde{\rho}$ est différentiable, les extrema s'obtiennent simplement par annulation de la dérivée

$$\frac{d\tilde{\rho}}{d\ell} = 0 \quad (4.23)$$

Utilisant l'équation (4.22), il vient

$$\ell_C \frac{d\alpha}{d\ell} = \frac{d\bar{\beta}}{d\ell} \quad (4.24)$$

Ou encore

$$\frac{d\bar{\beta}}{d\alpha} = \ell_C \quad (4.25)$$

Ce résultat signifie que les stratégies optimales s'obtiennent par des points de la courbe ROC dont la tangente admet une pente ℓ_C . Il faut encore s'assurer du fait que ce sont bien des minima, ce qui doit se faire de cas en cas, selon la forme de la courbe ROC, en choisissant la tangente la plus proche du coin supérieur gauche.

4.4.2 Cas non différentiable

Dans de nombreuses situations, la courbe ROC n'est pas différentiable. Dans ce cas, un raisonnement purement géométrique permet d'obtenir le risque minimal. Considérons l'axe passant par l'origine et défini par le vecteur $(\ell_C, -1)$. Le risque réduit s'obtenant par le produit scalaire de ce vecteur par le vecteur définissant la courbe ROC, le pied de la perpendiculaire abaissée du point de la courbe sur l'axe représente, avec une échelle idoine, le risque lui-même sur l'axe. Cela permet de nommer *axe des risques* le prolongement du vecteur $(\ell_C, -1)$. La figure 4.2 montre que le niveau critique optimal pour un test est donné par le(s) point(s) de sa courbe ROC dont la projection sur cet axe est minimale. La définition par projection permet non seulement le calcul sur des courbes discontinues, mais aussi la comparaison des tests entre eux.

4.5 Extension aux hypothèses composées

Tout ce qui précède se limite à des tests d'hypothèses dites simples, en ce sens que l'on oppose un état à un autre état. Ceci est possible car le modèle donne $P(dx | \theta)$ pour un état θ précis. Dans le cas d'hypothèses composées, où l'on désire opposer une disjonction d'états à une autre, c'est-à-dire tester si un état se trouve dans des sous-ensembles Θ' de Θ qui ne sont pas réduits à un seul élément, un a priori sur les états $\pi(d\theta)$ est nécessaire. En fait, il faut donner un sens à

$$P(x | \Theta')$$

pour un sous-ensemble général Θ' . Une extension peut s'effectuer de la manière suivante

$$P(dx | \Theta') = \frac{\int_{\Theta'} P(dx | \theta) \pi(d\theta)}{\int_{\Theta} P(dx | \theta) \pi(d\theta)} \quad (4.26)$$

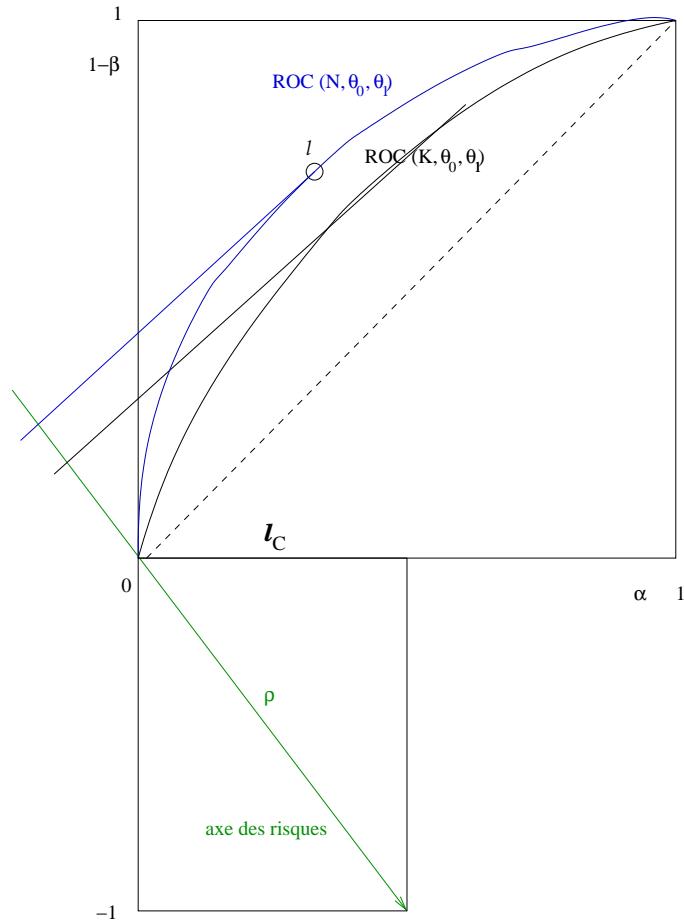


FIG. 4.2 – Courbes ROC et axe des risques. Deux courbes ROC sont représentées avec les points de décision obtenus par les tangentes perpendiculaires à l'axe des risques. La courbe dont la tangente est la plus proche du coin supérieur gauche donne une risque inférieur.

Le rapport de vraisemblance et les erreurs de premier et second types deviennent

$$\mathcal{LR} = \frac{P(dx | \Theta_1)}{P(dx | \Theta_0)} \quad (4.27)$$

$$\alpha = P(K > \ell | \Theta_0) \quad (4.28)$$

$$\bar{\beta} = P(K > \ell | \Theta_1) \quad (4.29)$$

Nous avons vu que les tests d'hypothèses simples sont construits indépendamment des a priori. Du moment qu'un a priori est nécessaire pour l'extension proposée, autant se placer carrément dans le contexte de la théorie bayesienne. Nous ne développerons donc pas cette extension hybride.

Troisième partie

Modèles, information et a priori

Ex nihilo, nihil

Lucrèce

Chapitre 5

Lois et modèles de base

L’application de la théorie de Bayes à la mesure de la radioactivité fait intervenir quelques modèles de base que nous allons passer en revue. Ces modèles sont construits sur des lois de probabilité classiques qui peuvent être discrètes (si l’observation est un comptage) ou continues (si l’observation est un temps d’attente). Ces lois de base entretiennent entre elles des rapports particuliers.

5.1 Modèle binomial

Les détecteurs utilisés pour la mesure de la radioactivité ont une certaine sensibilité. Cette *sensibilité* est définie par la probabilité ε de détecter une désintégration. ε dépend du nucléide mesuré et de la géométrie de mesure. La détermination de cette sensibilité peut se faire en comptant le nombre de détections produites par un nombre donné N de désintégrations. On suppose que les désintégrations sont indépendantes les unes des autres, et que les détections n’interfèrent pas. Ceci permet de considérer la suite des désintégrations comme autant de répétitions indépendantes au sens des probabilités. A chaque détection est associée un succès dont la probabilité est ε . Autrement dit, si pour N désintégrations, on a obtenu n détections, la loi régissant n est celle du nombre de succès en N répétitions, qui est *binomiale*.

L’état du compteur est sa sensibilité ε , qui est une nombre entre 0 et 1. L’espace des états est donc $\Theta = [0, 1]$. Le nombre de désintégrations N est fixé. L’observation est un nombre de détections, entier, ce qui implique $\mathcal{X} = \mathbb{N}$ et le modèle s’écrit

$$P(n | \varepsilon) = \mathcal{B}(n | N, \varepsilon) = \binom{N}{n} \varepsilon^n (1 - \varepsilon)^{N-n} \quad (5.1)$$

L’espérance et la variance du nombre de détections valent respectivement

$$\mathbf{E}n | \varepsilon = \varepsilon N \quad (5.2)$$

$$\mathbf{V}n | \varepsilon = \varepsilon(1 - \varepsilon)N \quad (5.3)$$

Le maximum de vraisemblance pour ce modèle est donné par $\hat{\varepsilon} = \frac{n}{N}$. On retrouve le rapport du nombre de succès au nombre d’essais qui constitue une estimation de la probabilité d’un événement par répétition de l’expérience. Quelques lois binomiales sont présentées à la figure 5.1

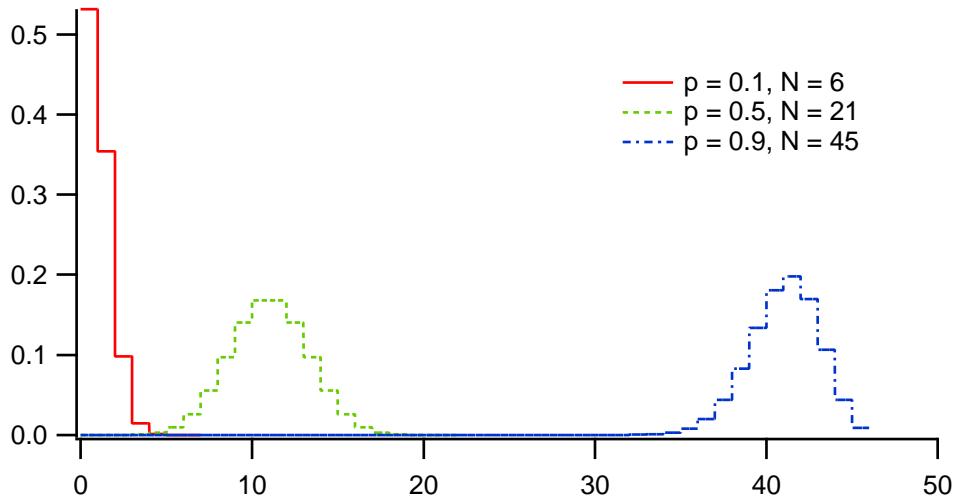


FIG. 5.1 – Exemples de lois binomiales. *Ces lois gouvernent le nombre de succès obtenus lors d'un nombre fixé de répétitions.*

5.2 Modèle poissonnien

La loi de *Poisson* est omniprésente dans le domaine de la radioactivité. Soit une source dont le nombre de noyaux est M_0 au temps $t = 0$. Le moment de la désintégration d'un atome particulier est inconnu, mais on sait que sa probabilité de désintégration par unité de temps λ est constante. La probabilité qu'un atome présent à $t = 0$ se désintègre dans un intervalle de temps dt est alors donnée par $\lambda e^{-\lambda t} dt$. La probabilité qu'un atome produise sa désintégration dans l'intervalle $[t_1, t_2]$ est

$$p_1 = \int_{t_1}^{t_2} \lambda e^{-\lambda t} dt \quad (5.4)$$

Pour M_0 noyaux indépendants, nous aurons donc une répartition binomiale pour le nombre N de désintégrations dans cet intervalle

$$P(N) = \mathcal{B}(N \mid M_0, p_1) \quad (5.5)$$

Si M_0 est assez grand et p_1 assez petit, cette loi est proche de la poissonnienne

$$P(N) \simeq \mathcal{P}(N \mid M_0 p_1) \quad (5.6)$$

Mais le théorème de la moyenne permet d'exprimer p_1 par

$$p_1 = \lambda e^{-\lambda \bar{t}} (t_2 - t_1) \quad (5.7)$$

où \bar{t} se situe entre t_1 et t_2 . Soit $\Delta t = t_2 - t_1$ et $M(\bar{t}) = M_0 e^{-\lambda \bar{t}}$, il vient

$$P(N) \simeq \mathcal{P}(N \mid \lambda M(\bar{t}) \Delta t) \quad (5.8)$$

Le terme

$$a(\bar{t}) = \lambda M(\bar{t}) \quad (5.9)$$

est dit *activité de la source* au temps \bar{t} et on peut écrire

$$P(N) = \mathcal{P}(N \mid a\Delta t) \quad (5.10)$$

Ces désintégrations sont ensuite comptées par une détecteur d'efficacité ε , ce qui signifie que chacune d'entre elles a une probabilité ε de donner une impulsion. La probabilité d'obtenir n détections s'il y a eu N désintégrations est binomiale

$$P(n \mid N) = \mathcal{B}(n \mid N, \varepsilon) \quad (5.11)$$

La probabilité d'obtenir n détections pour une durée t et une activité a s'obtient par la composition

$$P(n \mid a) = \sum_{N=0}^{\infty} P(n \mid N)P(N \mid a) \quad (5.12)$$

$$= \sum_{N=0}^{\infty} \mathcal{B}(n \mid N, \varepsilon) \mathcal{P}(N \mid a\Delta t) \quad (5.13)$$

$$= \sum_{N=n}^{\infty} \binom{N}{n} \varepsilon^n (1-\varepsilon)^{N-n} \frac{(a\Delta t)^N}{N!} e^{-at} \quad (5.14)$$

Mais (5.14) peut se réarranger de la manière suivante

$$P(n \mid a) = e^{-a\Delta t} \frac{\varepsilon^n}{n!} (a\Delta t)^n \sum_{N \geq n} \frac{(1-\varepsilon)^{N-n}}{(N-n)!} (a\Delta t)^{N-n} \quad (5.15)$$

Or la somme vaut $e^{(1-\varepsilon)a\Delta t}$, d'où

$$P(n \mid a) = \mathcal{P}(n \mid a\varepsilon\Delta t) \quad (5.16)$$

Les nombres de détections se répartissent de manière poissonnienne. Tout se passe comme si l'activité avait été multipliée par la sensibilité du détecteur. Il faut remarquer que la loi de Poisson est obtenue par le fait que l'on mesure pendant un temps donné, et non pas un certain nombre de désintégrations donné. Ceci fait que la loi, initialement binomiale, devient poissonnienne. Il n'y a pas dans ce dernier calcul d'approximation de la binomiale par la poissonnienne. La figure 5.2 montre quelques lois de Poisson.

Pour ce modèle, le nombre moyen de détections est $a\varepsilon t$, qui est aussi sa variance. Si l'on a observé n impulsions, le maximum de vraisemblance pour a est obtenu par $\hat{a} = \frac{n}{\varepsilon t}$, qui a bien la dimension d'une activité.

Rappelons que la loi de Poisson admet une variance égale à sa moyenne.

5.3 Modèle gamma

Au lieu de compter des désintégrations pendant un temps donné, il est évidemment possible, à l'inverse, d'attendre qu'un nombre donné de désintégrations se soient produites. La loi régissant le temps qu'il faut attendre pour que se produise la n^e désintégration d'une source d'activité a est donnée par le modèle *gamma* [17]

$$P(dt \mid a) = \text{Ga}(dt \mid n, a) = \frac{a^n}{\Gamma(n)} t^{n-1} e^{-at} dt \quad (5.17)$$

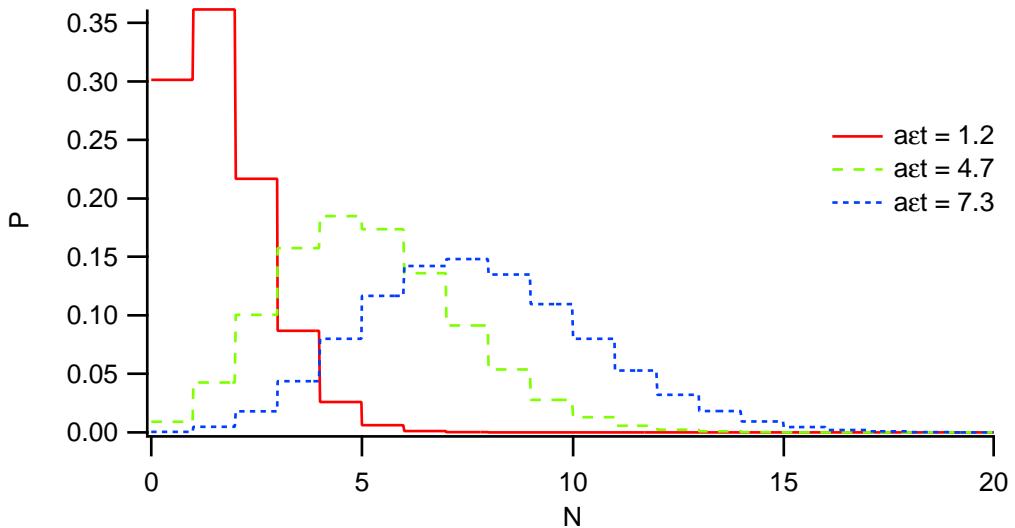


FIG. 5.2 – Exemples de lois de Poisson. *Le paramètre est ici le nombre moyen de détections attendues dans un compteur de sensibilité ϵ , pour une activité a mesurée pendant un temps t*

Son espérance et sa variance sont reliées aux paramètres a et n par

$$\begin{aligned} E &= \frac{n}{a} & V &= \frac{n}{a^2} \\ a &= \frac{E}{V} & n &= \frac{E^2}{V} \end{aligned} \quad (5.18)$$

Cette loi est définie habituellement avec les paramètres α (ici n) et β (ici $1/a$). Cette nouvelle définition permet d'une part d'éviter la confusion avec les erreurs de premier et second types qui sont définies en (3.1), d'autre part d'interpréter n et a comme un comptage et une activité. Les lois de Poisson et gamma sont reliées par la relation

$$\text{Ga}([0, t] \mid n, a) = P([n, \infty[\mid at) \quad (5.19)$$

qui exprime de deux manières la probabilité d'obtenir au moins n désintégrations en moins de t unités de temps. Il est en effet équivalent de dire que la n^e désintégration s'est produite avant t que d'affirmer que le nombre de désintégrations jusqu'à t est plus grand ou égal à n .

Quelques lois gamma sont présentées à la figure 5.3.

La figure 5.4 montre les différentes formes possibles de la loi gamma en fonction des deux paramètres n et a . Chaque point de ce plan représente une loi gamma. Sur les axes d'abscisses et d'ordonnées se trouvent des répartitions impropre, car elles ne sont pas normalisables. Pour $n \in]0, 1[$, ces lois ont une singularité intégrable en $a = 0$. Lorsque $n = 1$, on retrouve les lois exponentielles. Au dessus, la singularité disparaît, et lorsque n devient plus grand, leur forme se rapproche de la gaussienne. Si $n \rightarrow \infty$ de sorte que le rapport $\frac{n}{a}$ soit constant et égal à E , ces répartitions tendent vers le dirac en E .

Signalons en passant que les lois de χ^2 sont dérivées des lois gamma. En effet, si n est demi-entier et $a = 1$, on a

$$P(\chi^2 < x \mid v) = \text{Ga}([0, x/2] \mid v/2, 1) \quad (5.20)$$

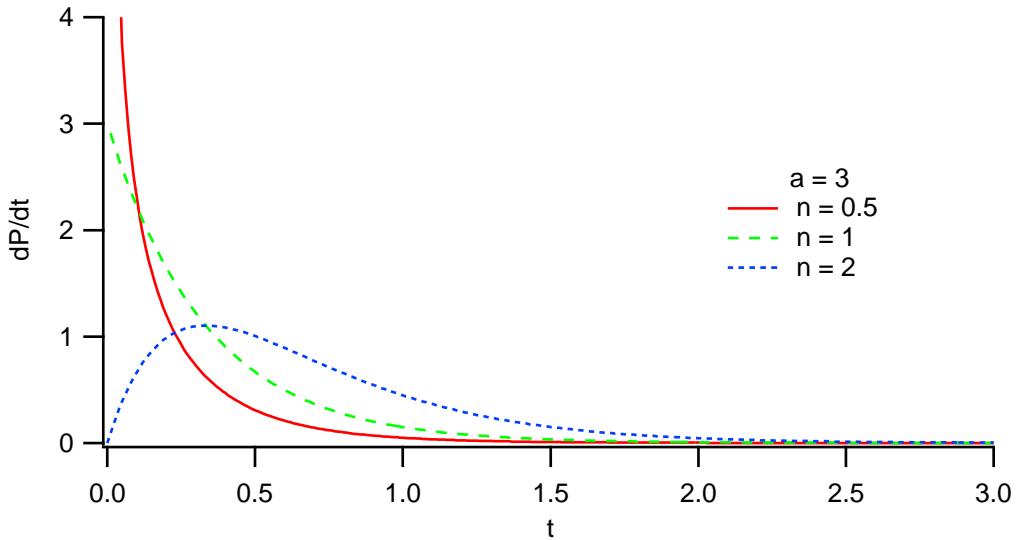


FIG. 5.3 – Exemples de lois gamma. *Les temps d'attente se répartissent selon ces lois. Elles sont densitaires sur l'intervalle $[0, \infty[$.*

5.4 Modèle négatif-binomial

Ce modèle est dual du binomial. Nommons p_f la probabilité de succès d'un évènement quelconque. Au lieu de fixer le nombre d'essais et de compter le nombre de succès, il est possible de fixer un nombre de succès à obtenir N_f , et de compter le nombre d'échecs N à subir pour y parvenir¹. On obtient ainsi la loi *négative-binomiale*

$$\mathcal{NB}(N | N_f, p_f) = \frac{\Gamma(N_f + N)}{\Gamma(N_f)N!} p_f^{N_f} (1 - p_f)^N \quad (5.21)$$

Lorsque N_f est entier on peut écrire

$$\mathcal{NB}(N | N_f, p_f) = \binom{N_f + N - 1}{N} p_f^{N_f} (1 - p_f)^N \quad (5.22)$$

Son espérance et sa variance sont reliées aux paramètres N_f et p_f par

$$\begin{aligned} E &= N_f \frac{1-p_f}{p_f} & V &= N_f \frac{1-p_f}{p_f^2} \\ N_f &= \frac{E^2}{V-E} & p_f &= \frac{E}{V} \end{aligned} \quad (5.23)$$

Si l'on suppose que l'activité d'une source est répartie selon une loi gamma de paramètres (N_f, t_f) , alors le nombre de désintégrations N de cette source durant un temps t suivra la loi négative-binomiale

$$P(N | N_f, t_f) = \mathcal{NB}(N | N_f, p_f) \quad (5.24)$$

avec $p_f = \frac{t_f}{t+t_f}$ (cf A.2).

¹Certains auteurs définissent cette loi par le nombre d'essais à faire au lieu du nombre d'échecs. Nous préférions la version avec les échecs, car elle part de zéro.

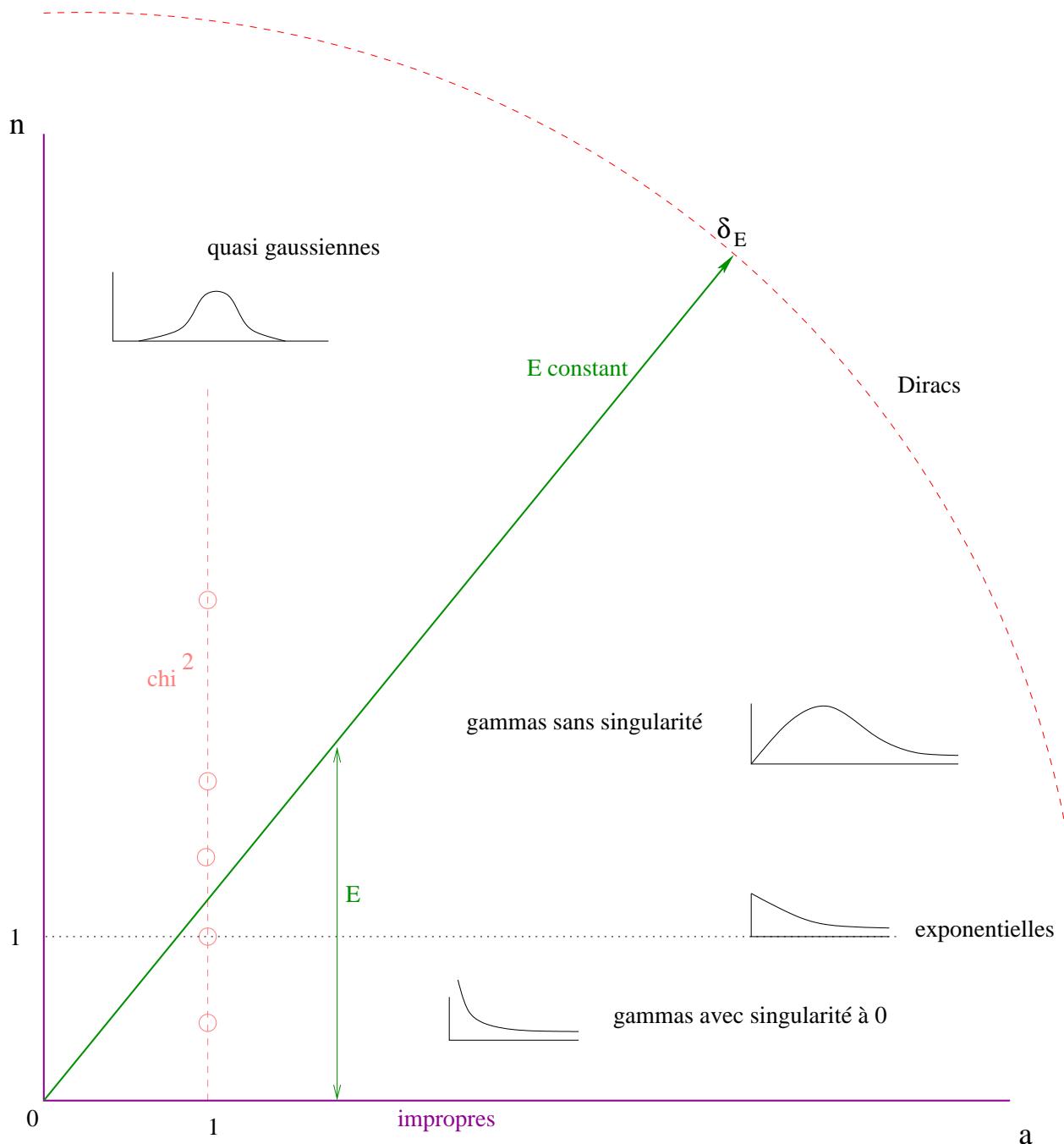


FIG. 5.4 – Plan gamma. Chaque point de ce plan représente une loi gamma dont les paramètres sont figurés sur les axes. On y remarque les différentes formes possibles et les cas limites qui sont les lois normales et les diracs.

Ainsi cette loi, dite aussi de *Pòlya*, apparaît comme une extension de la loi de Poisson et peut la remplacer lorsque l'égalité entre la moyenne et la variance d'un échantillon n'est plus admissible. La variance doit être supérieure à l'espérance et rend alors compte d'une *sur-dispersion* de l'échantillon. On trouvera quelques exemples de lois négatives-binomiales à la figure 5.5.

Remarque L'attribut *négatif* dans la dénomination de cette loi provient du fait qu'il faut utiliser le développement de Newton avec un exposant négatif pour montrer que cette loi est bien normalisée. En effet, la généralisation du binôme de Newton à des exposants quelconques s'écrit

$$(x+y)^r = \sum_{k=0}^{\infty} \binom{r}{k} x^k y^{r-k} \quad (5.25)$$

où le coefficient binomial pour un r réel est donné par

$$\binom{r}{k} = \frac{r(r-1)\dots(r-k+1)}{k!} \quad (5.26)$$

Utilisons la série de Newton avec $x = (p_f - 1)$, $y = 1$ et $r = -N_f$, il vient

$$p_f^{-N_f} = ((p_f - 1) + 1)^{-N_f} = \sum_N \frac{(-N_f)(-N_f - 1)\dots(-N_f - N + 1)}{N!} (p_f - 1)^N \quad (5.27)$$

$$= \sum_N \frac{(N_f)(N_f + 1)\dots(N_f + N - 1)}{N!} (1 - p_f)^N \quad (5.28)$$

$$= \sum_N \frac{\Gamma(N_f + N)}{\Gamma(N_f)N!} (1 - p_f)^N \quad (5.29)$$

(5.30)

Il suffit alors de multiplier les deux membres par $p_f^{N_f}$ pour obtenir la normalisation.

5.5 Modèle bêta

La loi *bêta* est définie dans l'intervalle $[0, 1]$ par la densité

$$\text{Be}(dp \mid n, m) = \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} p^{n-1} (1-p)^{m-1} dp \quad (5.31)$$

Son espérance et sa variance sont reliées aux paramètres n et m par

$$\begin{aligned} E &= \frac{n}{n+m} & V &= \frac{nm}{(n+m)^2(n+m+1)} \\ n &= E \frac{E(1-E)}{V} & m &= (1-E) \frac{E(1-E)}{V} \end{aligned} \quad (5.32)$$

L'interprétation de ce modèle apparaîtra plus loin, lors de la recherche des a priori et de la détermination de la sensibilité des détecteurs. Si une activité a est répartie selon une loi gamma de paramètres (N_f, t_f) , et que t est le temps d'attente de la N^e désintégration produite par cette activité, on peut définir une variable dérivée $p = \frac{t_f}{t_f + t}$. Cette variable suit une loi bêta de paramètres (N_f, N) (cf A.3). Cette loi apparaît aussi comme quotient de deux temps d'attente. En effet, rapport du temps d'attente des la r^e désintégration à celui de la $(r+s)^e$ est une variable aléatoire qui suit une loi bêta de paramètres (r, s) (cf par exemple [18] 4.5). Quelques unes de ces lois sont présentées à la figure 5.6.

Les lois bêta et négative-binomiale sont reliées par la relation

$$\text{Be}([0, r] \mid n, m) = \mathcal{N}\mathcal{B}([0, m-1] \mid n, r) \quad (5.33)$$

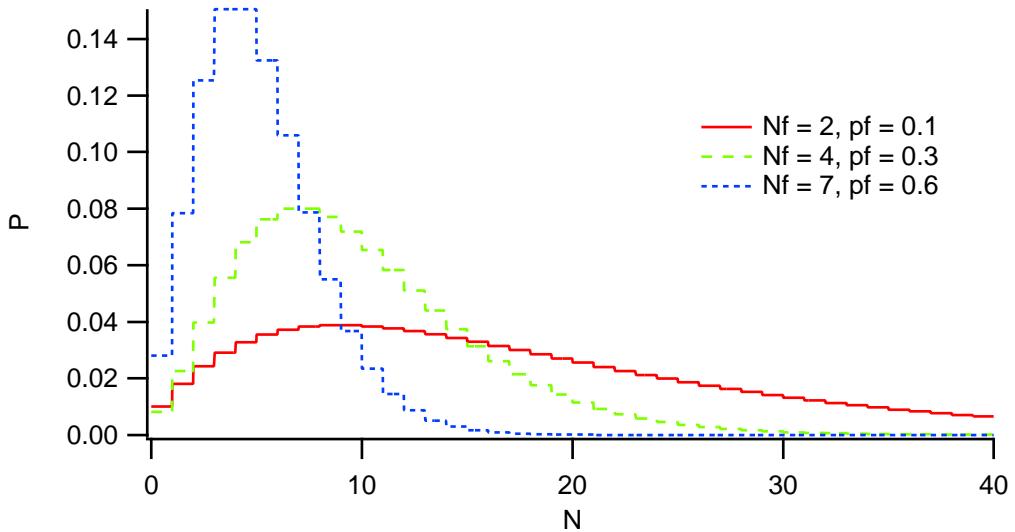


FIG. 5.5 – Exemples de lois négatives-binomiales. *Ces lois gouvernent le nombre d'échecs N à subir avant d'obtenir un nombre fixé N_f de succès.*

5.6 Modèles discrets

Lorsque les différents espaces de la TBDS sont discrets, les probabilités deviennent des vecteurs, et les modèles des matrices dites stochastiques. Pour simplifier la notation, les états et observations seront repérés par leurs indices. On définira donc

- l'espace des états $\Theta = \{1 \dots j \dots m\}$
- l'espace des observations $\mathcal{X} = \{1 \dots i \dots n\}$
- le modèle par une matrice $P_{ij} = P(i \mid j)$
- les a priori et a posteriori par des vecteurs $\pi_j = \pi(\{j\})$
- et la loi marginale sur les observations par le vecteur $M_i = \sum_j P_{ij} \pi_j$

5.6.1 Modèle discret à transition poissonnienne

Supposons qu'un observateur ait mesuré un certain nombre de détections en présence d'une source parmi m possibles, mais sans savoir laquelle. L'espace des états est celui des m activités $\Theta = \{a_1 \dots a_m\}$, et celui des observations l'ensemble des entiers naturels. Le modèle, discret, sera dit à *transition poissonnienne* car la matrice P est définie par les lois de Poisson. La probabilité que la j^e source produise n impulsions dans le compteur durant le temps t vaut

$$P_{nj} = \mathcal{P}(n \mid a_j \varepsilon t) \quad (5.34)$$

5.6.2 Modèle discret à transition gamma

Ce modèle est simplement le dual du précédent. Pour un nombre fixé n , la répartition du temps d'attente de la n^e détection en face de la j^e source est donnée par la loi gamma

$$P(dt \mid j) = \text{Ga}(dt \mid n, a_j \varepsilon) \quad (5.35)$$

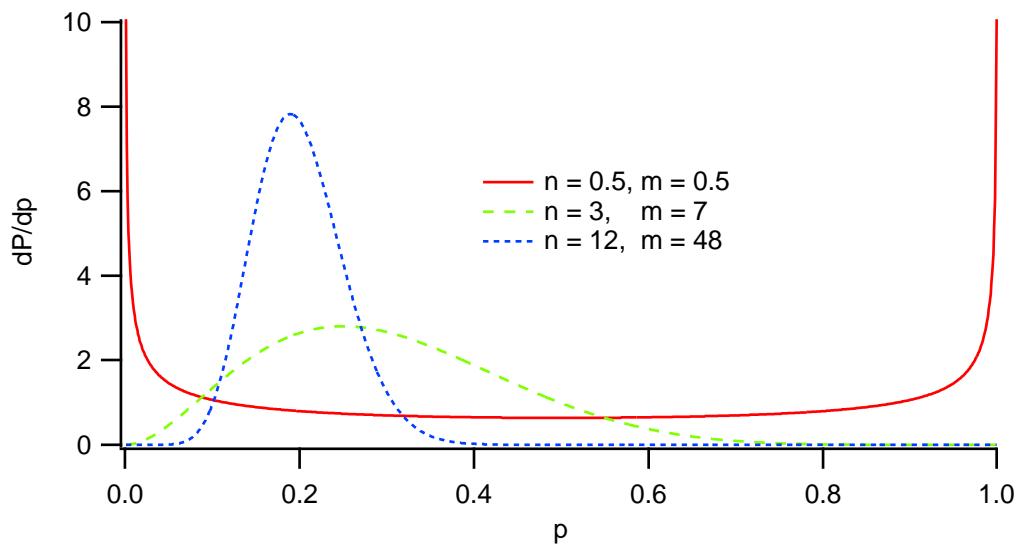


FIG. 5.6 – Exemples de lois bêta. *Ces lois sont aussi densitaires, mais sont portées par l'intervalle $[0, 1]$. La variable d'intégration est ici p .*

Chapitre 6

Information contenue dans les modèles

Dans l'approche de la mesure bayesienne, une observation transforme un a priori en a posteriori. Ainsi, l'information sur l'état apportée par l'observation peut être quantifiée par l'écart entre l'a priori et l'a posteriori. Une mesure de l'écart entre deux probabilités couramment retenue est la *divergence de Kullback-Leibler*. Un autre approche est celle de *Fisher*, qui définit une information basée uniquement sur le modèle. Le but de cette section est de décrire ces différentes notions ainsi que les relations qui les lient. Elles seront ensuite utilisées pour construire des a priori.

6.1 Information entropique

6.1.1 Divergence de Kullback-Leibler

6.1.1.1 Définition

Soient P et Q deux probabilités sur un espace mesurable Ω , de sorte que¹

$$\forall A \subset \Omega, Q(A) = 0 \Rightarrow P(A) = 0$$

la *divergence de Kullback-Leibler de P par rapport à Q* est [19]

$$D(P \parallel Q) = \int P(d\omega) \ln \frac{P}{Q}(\omega) \tag{6.1}$$

Dans le cas continu $\frac{P}{Q}(\omega)$ désigne la densité de P par rapport à Q obtenue par le théorème de *Radon-Nicodym* (cf [11], par exemple)². Dans le cas discret, c'est simplement le rapport des probabilités au point considéré.

Cette fonctionnelle est aussi appelée *divergence logarithmique*, ou *distance de Kullback*.

6.1.1.2 Propriétés

La divergence logarithmique possède les propriétés suivantes (les démonstrations peuvent se trouver par exemple dans [19] ou [20]).

¹ P est dite absolument continue par rapport à Q . On note parfois $P \ll Q$

²Cette densité, mettons f , s'écrit aussi $\frac{dP}{dQ}$ ou $\frac{P(d\omega)}{Q(d\omega)}$ et on note $P = f \cdot Q$

Positivité

$$\forall P, Q, D(P \parallel Q) \geq 0 \quad (6.2)$$

La démonstration de cette propriété fait intervenir l'inégalité de Jensen.

Invariance

La divergence est définie ici de manière intrinsèque. Si Ω est une variété différentiable, le calcul peut donc se faire dans n'importe quel système de coordonnées.

Discrimination

$$D(P \parallel Q) = 0 \implies P = Q \quad (6.3)$$

D n'est pas une distance au sens des espaces métriques car elle n'est pas symétrique et ne vérifie pas l'inégalité du triangle, mais le fait qu'elle ne s'annule que si les répartitions sont égales permet de l'utiliser pour les discriminer. Si les deux répartitions P et Q sont associées à deux hypothèses opposées \mathcal{H}_1 et \mathcal{H}_0 , la divergence logarithmique exprime l'*information discriminante* entre ces deux hypothèses. L'idée est que plus les répartitions sont proches, plus les hypothèses sous-jacentes sont difficiles à discriminer. Lorsque la divergence est utilisée pour mesurer une discrimination entre deux répartitions, il est d'usage de la symétriser en posant

$$J(P \parallel Q) = \frac{1}{2}(D(P \parallel Q) + D(Q \parallel P)) \quad (6.4)$$

Dans notre contexte, où il s'agit d'examiner l'effet d'une mesure, il n'y a pas de raison de le faire, car il y un avant et un après la mesure.

Convexité

L'application

$$(P, Q) \mapsto D(P \parallel Q)$$

est convexe, c'est-à-dire que si $(P_\lambda, Q_\lambda) = (1 - \lambda)(P_0, Q_0) + \lambda(P_1, Q_1)$, alors

$$D(P_\lambda \parallel Q_\lambda) \leq (1 - \lambda)D(P_0 \parallel Q_0) + \lambda D(P_1 \parallel Q_1) \quad (6.5)$$

Mesure de la dépendance

Soient deux variables aléatoires X et Y de répartition conjointe $P_{XY}(dx, dy)$. Nommons $P_X(dx)$ et $P_Y(dy)$ les répartitions marginales de X et Y . La divergence entre la conjointe et le produit des marginales

$$I(X, Y) = D(P_{XY}(dx, dy) \parallel P_X(dx)P_Y(dy)) \quad (6.6)$$

donne une mesure de la dépendance entre les variables aléatoires X et Y , qui est appelée *information mutuelle de X et Y* (cf par exemple [20]).

6.1.1.3 Exemples

Le tableau 6.1 donne la divergence logarithmique de quelques lois de base pour lesquelles il existe une expression analytique. Les démonstrations sont renvoyées en annexe (A.7, A.8, A.9). Ψ est la fonction digamma, dérivée du logarithme de la fonction gamma : $\Psi = (\ln \Gamma)'$.

TAB. 6.1 – Divergences logarithmiques de quelques lois de base

Type	Loi 1	Loi 2	Divergence logarithmique
Poisson	$\mathcal{P}(\alpha_1)$	$\mathcal{P}(\alpha_0)$	$\alpha_1 \ln \frac{\alpha_1}{\alpha_0} - (\alpha_1 - \alpha_0)$
gamma	$\text{Ga}(N_1, t_1)$	$\text{Ga}(N_0, t_0)$	$N_0 \ln \frac{t_1}{t_0} - N_1 \left(1 - \frac{t_0}{t_1}\right) + \ln \Gamma(N_0) - \ln \Gamma(N_1) + (N_1 - N_0) \Psi(N_1)$
bêta	$\text{Be}(\alpha_1, \beta_1)$	$\text{Be}(\alpha_0, \beta_0)$	$\ln \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} - \ln \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} + (\alpha_1 - \alpha_0)(\Psi(\alpha_1) - \Psi(\alpha_1 + \beta_1)) + (\beta_1 - \beta_0)(\Psi(\beta_1) - \Psi(\alpha_1 + \beta_1))$

6.1.2 Information moyenne apportée par la mesure

6.1.2.1 Définition

Supposons que nous avons obtenu une observation x . La divergence de l'a posteriori relative à l'a priori initial π est

$$D(x) = D(\pi(d\theta | x) || \pi(d\theta)) = \int_{\Theta} \pi(d\theta | x) \ln \frac{\pi(d\theta | x)}{\pi(d\theta)} \quad (6.7)$$

Mais on sait que pour cet a priori initial x est réparti selon

$$M(dx) = \int_{\Theta} P(dx | \theta) \pi(d\theta) \quad (6.8)$$

On définit alors l'*information moyenne apportée par la mesure* par la fonctionnelle I_H de $\pi(d\theta)$

$$I_H(\pi(d\theta)) = \int_x M(dx) D(x) \quad (6.9)$$

Soit $P(dx, d\theta) = P(dx | \theta) \pi(d\theta)$ la répartition conjointe sur $\mathcal{X} \times \Theta$. On peut remarquer que

$$I_H(\pi(d\theta)) = \int_x M(dx) \int_{\Theta} \frac{P(dx | \theta) \pi(d\theta)}{M(dx)} \ln \frac{P(dx | \theta) \pi(d\theta)}{M(dx) \pi(d\theta)} \quad (6.10)$$

$$= \int_{x, \theta} P(dx, d\theta) \ln \frac{P(dx, d\theta)}{M(x) \pi(d\theta)} \quad (6.11)$$

$$= D(P(dx, d\theta) || M(dx) \pi(d\theta)) \quad (6.12)$$

La probabilité marginale sur les observations $M(x)$ donne la répartition de la projection sur les observations. L'a priori $\pi(d\theta)$ donne la répartition de la projection sur les états. L'équation (6.12) donne donc la divergence logarithmique entre la probabilité conjointe et celle que l'on obtiendrait si les états et les observations étaient indépendants. On trouve ici l'information mutuelle entre les états et les observations (cf équ. (6.6) en prenant pour X et Y les projections sur Θ et \mathcal{X}).

En particulier, si les observations et les états sont indépendants, la conjointe s'exprime par le produit des marginales, l'information mutuelle est nulle et une mesure n'apporte rien.

6.1.2.2 Propriétés

I_H est toujours positive et est aussi définie de manière intrinsèque. De plus la convexité de la divergence implique qu'une simple transformation mathématique de l'espace des observations ne peut pas apporter d'information supplémentaire (cf section A.4).

Pour un a priori donné, l'information moyenne apportée par la mesure peut s'exprimer de trois manières

- du point de vue global, c'est l'information mutuelle entre les projections canoniques, comme nous venons de le voir

$$I_H(\pi(d\theta)) = D(P(dx, d\theta) \parallel M(dx)\pi(d\theta)) \quad (6.13)$$

- du point de vue des états, c'est la divergence moyenne entre l'a priori et l'a posteriori par définition

$$I_H(\pi(d\theta)) = \int_x M(dx)D(\pi(d\theta | x) \parallel \pi(d\theta)) \quad (6.14)$$

- du point de vue des observations, c'est la divergence moyenne entre le modèle et la marginale

$$I_H(\pi(d\theta)) = \int_\theta \pi(d\theta)D(P(dx | \theta) \parallel M(dx)) \quad (6.15)$$

Démonstration

$$\int_\theta \pi(d\theta)D(P(dx | \theta) \parallel M(dx)) = \int_\theta \pi(d\theta) \int_x P(dx | \theta) \ln \frac{P(dx | \theta)}{M(dx)} \quad (6.16)$$

$$= \int_{\theta, x} \pi(d\theta)P(dx | \theta) \ln \frac{P(dx | \theta)\pi(d\theta)}{M(dx)\pi(d\theta)} \quad (6.17)$$

$$= \int_{\theta, x} P(dx, d\theta) \ln \frac{P(dx, d\theta)}{M(dx)\pi(d\theta)} \quad (6.18)$$

$$= D(P(dx, d\theta) \parallel M(dx)\pi(d\theta)) \quad (6.19)$$

Cette information moyenne est une généralisation de l'information de Shannon. On trouvera à la section B.2 le modèle qui fait le passage de l'une à l'autre.

6.1.2.3 Exemples

Modèle poissonnien Reprenons le modèle des comptages. Sous un a priori $\text{Ga}(da | N_f, t_f)$, l'information moyenne apportée par une mesure de durée t est de la forme (cf section A.10)

$$\begin{aligned} I_H(N_f, t_f) &= N_f \ln \left(\frac{t_f + \varepsilon t}{t_f} \right) + \ln \Gamma(N_f) - N_f \frac{\varepsilon t}{t_f} \\ &+ \sum_N (N \Psi(N_f + N) - \ln \Gamma(N_f + N)) \mathcal{NB} \left(N | N_f, \frac{t_f}{t_f + \varepsilon t} \right) \end{aligned} \quad (6.20)$$

Modèle gamma Sous un a priori $\text{Ga}(dt \mid N_f, t_f)$, l'information moyenne pour le modèle du temps d'attente de la N^e détection s'écrit (cf section A.11)

$$I_H(N_f, t_f) = \ln \Gamma(N_f) - \ln \Gamma(N_f + N) - N_f \Psi(N_f) + (N_f + N) \Psi(N_f + N) - N \quad (6.21)$$

On remarque que cette expression ne dépend pas de t_f . Un changement d'unité de temps ne change en effet pas l'information. De plus elle ne dépend pas non plus de la sensibilité. En effet l'attente de la N^e détection durera plus longtemps si la sensibilité diminue, mais l'information restera la même.

6.2 Information de Fisher

L'idée maîtresse de la statistique est de synthétiser un grand nombre de résultats en quelques variables. Par exemple, lors de l'étude d'une répartition gaussienne, si l'observateur a obtenu lors de m mesures les résultats réels $(x_i)_{i=1}^m$, il peut calculer la moyenne e et la variance v de l'échantillon. Nommons \mathcal{T} la statistique $(x_i)_{i=1}^m$, et \mathcal{S} la statistique (m, e, v) . Comme il n'existe pas d'application bijective continue de \mathbb{R}^m dans \mathbb{R}^3 pour $m > 3$, le remplacement de \mathcal{T} par \mathcal{S} peut donner l'impression que l'on perd de l'information. Plus précisément, si l'observateur doit prendre une décision basée sur ses résultats, il semble évident que celles basées sur \mathcal{T} seront meilleures que celles basées sur \mathcal{S} . En fait, la prise en considération du fait que la gaussienne n'a que deux paramètres laisse entrevoir que ce n'est pas le cas. L'idée de Fisher est de définir une mesure de l'information contenue dans une statistique, de telle sorte que deux statistiques de même information donnent des décisions de même efficacité, c'est-à-dire de mêmes probabilités d'erreur.

Cette mesure, dite *information de Fisher* $I(\mathcal{T})$ est en relation étroite avec la mesure de l'incertitude et l'information de Shannon.

6.2.1 Construction

Supposons que l'ensemble des états soit réel $\Theta = \mathbb{R}^m$. La théorie de Shannon [21] donne l'information obtenue lors la réalisation de l'évènement dx par le logarithme du modèle P

$$I_S(dx, \theta) = -\ln P(dx \mid \theta) \quad (6.22)$$

Le gradient de I_S par rapport à θ exprime la variation de cette information selon l'état, ou la variabilité relative du modèle, dite *fonction de score*

$$s(x, \theta) = -\partial_\theta I_S(dx, \theta) \quad (6.23)$$

Considérées à θ fixe, les coordonnées s_i de s sont des fonctions de x . Le modèle pour l'état θ , $P(dx \mid \theta)$, donne alors une répartition qui permet de calculer l'espérance des s_i par

$$E_\theta(s_i) = \int_x P(dx \mid \theta) s_i(x, \theta) \quad (6.24)$$

et la matrice de covariance

$$I_{ij}(\theta) = \int_x P(dx \mid \theta) (s_i(x, \theta) - E_\theta(s_i))(s_j(x, \theta) - E_\theta(s_j)) = \text{Cov}_\theta(s_i, s_j) \quad (6.25)$$

Cette matrice définit l'*information de Fisher*.

Lorsque le modèle est donné par une fonction de vraisemblance, la fonction de score en est la dérivée logarithmique

$$s(x, \theta) = \frac{\partial_{\theta} \mathcal{L}(\theta | x)}{\mathcal{L}(\theta | x)} \quad (6.26)$$

Son espérance mathématique est nulle car $\int dx \mathcal{L}(\theta | x) = 1$, en effet

$$\int dx \mathcal{L}(\theta | x) s(x, \theta) = \int dx \mathcal{L}(\theta | x) \frac{\partial_{\theta} \mathcal{L}(\theta | x)}{\mathcal{L}(\theta | x)} \quad (6.27)$$

$$= \int dx \partial_{\theta} \mathcal{L}(\theta | x) \quad (6.28)$$

$$= \partial_{\theta} \int dx \mathcal{L}(\theta | x) = 0 \quad (6.29)$$

Remplaçant $P(dx | \theta)$ par son expression densitaire $dx \mathcal{L}(\theta | x)$, l'information de Fisher prend la forme

$$I_{ij}(\theta) = \int P(dx | \theta) s_i(x, \theta) s_j(x, \theta) \quad (6.30)$$

$$= \int dx \mathcal{L}(\theta | x) \frac{\partial \ln \mathcal{L}(\theta | x)}{\partial \theta_i} \frac{\partial \ln \mathcal{L}(\theta | x)}{\partial \theta_j} \quad (6.31)$$

On montre (cf par exemple [22], p373) que cette information peut aussi s'écrire

$$I_{ij}(\theta) = - \int dx \mathcal{L}(\theta | x) \frac{\partial^2 \ln \mathcal{L}(\theta | x)}{\partial \theta_i \partial \theta_j} \quad (6.32)$$

ou encore, sous la forme utilisée par Rodriguez ([23], section 3)

$$I_{ij}(\theta) = \int dx \frac{\partial \sqrt{\mathcal{L}}}{\partial \theta_i} \frac{\partial \sqrt{\mathcal{L}}}{\partial \theta_j} \quad (6.33)$$

Cette information a les propriétés suivantes

- *Positivité* : la matrice est définie positive puisque c'est une matrice de covariance
- *Invariance* : cette matrice est un tenseur deux fois covariant qui définit une métrique sur l'espace des états dite *informationnelle*. L'espace des états devient ainsi une variété riemannienne.

6.2.2 Lien avec la divergence de Kullback

L'information de Fisher donne la variation locale de la divergence de Kullback. On a

$$D(P(dx | \theta) || P(dx | \theta + d\theta)) = \frac{1}{2} \sum_{i,j} I_{ij}(\theta) d\theta_i d\theta_j \quad (6.34)$$

Démonstration Par développement au premier ordre du rapport des vraisemblances et au deuxième ordre pour le logarithme, on a

$$D(P(dx | \theta) || P(dx | \theta + d\theta)) = - \int dx \mathcal{L}(x | \theta) \ln \frac{\mathcal{L}(x | \theta + d\theta)}{\mathcal{L}(x | \theta)} \quad (6.35)$$

$$\simeq - \int dx \mathcal{L} \ln \left(1 + \sum \frac{\partial \ln \mathcal{L}}{\partial \theta_i} d\theta_i \right) \quad (6.36)$$

$$\simeq - \int dx \sum_i \frac{\partial \mathcal{L}}{\partial \theta_i} - \frac{1}{2} \int dx \mathcal{L} \sum_i \frac{\partial \ln \mathcal{L}}{\partial \theta_i} \sum_j \frac{\partial \ln \mathcal{L}}{\partial \theta_j} \quad (6.37)$$

$$\simeq \frac{1}{2} \sum_{i,j} I_{ij}(\theta) d\theta_i d\theta_j \quad (6.38)$$

vu que le premier terme du second membre est nul car \mathcal{L} est normalisée. Un calcul du même type donne

$$D(P(dx | \theta + d\theta) || P(dx | \theta)) \simeq D(P(dx | \theta) || P(dx | \theta + d\theta))$$

6.2.3 Inégalité de Cramér-Rao

Le but de cette section est de mettre en évidence la relation entre l'information de Fisher et l'incertitude définie par la variance. Afin de simplifier l'exposé, nous nous limiterons à des états de dimension un. La matrice de Fisher se réduit à une seule fonction $I(\theta)$.

I joue un rôle important dans la mesure de l'optimalité des estimateurs, un estimateur étant d'autant meilleur que sa variance est petite. Dans ce contexte l'espace des décisions \mathcal{D} s'identifie à celui des états Θ . Soit l'estimateur

$$\tau : \mathcal{X} \rightarrow \Theta$$

Si $P(dx | \theta)$ est le modèle, notons k son espérance lorsque l'état est θ :

$$k(\theta) = \int_x \tau(x) P(dx | \theta) \quad (6.39)$$

$$= E\tau | \theta \quad (6.40)$$

τ est dit *non biaisé* si $k(\theta) = \theta$.

L'inégalité de *Cramér-Rao*³ (cf par exemple [22]) donne alors une borne inférieure pour la variance de l'estimateur faisant intervenir la dérivée de k

$$V\tau | \theta \geq \frac{k'(\theta)^2}{I(\theta)} \quad (6.41)$$

τ est dit *efficient* si cette borne est atteinte. En particulier, aucun estimateur non biaisé ne peut avoir un écart-type inférieur à $\frac{1}{\sqrt{I}}$.

Réécrivons cette inégalité *à la Heisenberg* pour mettre en évidence le produit de deux incertitudes (cf par exemple [25] p37)

$$V\tau | \theta \cdot I(\theta) \geq k'(\theta)^2 \quad (6.42)$$

Précision : $V\tau | \theta$ est un terme d'incertitude propre à l'estimateur, qui exprime sa *précision*.

³Qui est un cas particulier de l'inégalité de Schwarz.

Exactitude : k'^2 est en relation avec le biais $b = (k - \theta)^2$, qui mesure l'*exactitude* de l'estimateur.

Sensibilité : $I(\theta)$ est la variance de la fonction de score. Comme cette dernière mesure la variabilité relative du modèle, l'information de Fisher est une mesure de la sensibilité (i.e. de la qualité) du système de mesure au voisinage de l'état considéré.

Un modèle très variable donnera la possibilité d'estimateurs précis. Ici, nous entendons par exactitude une convergence de la moyenne de l'estimateur vers la vraie valeur sous-jacente. L'exactitude maximale est obtenue par $k'(\theta) = 1$. La fonction de score et l'estimateur se comportent comme des variables conjuguées. A biais constant, une grande variance de la fonction de score donnera une petite variance pour les estimateurs et réciproquement.

L'inégalité de Cramér-Rao servira à justifier le choix de certains a priori.

6.2.4 Illustrations

6.2.4.1 Modèle normal unidimensionnel

Considérons le cas de la mesure d'une variable physique numérique, par un système donnant une réponse normale dont la moyenne et la variance sont des fonctions de cette variable.

$$P(dx | \theta) = \mathcal{N}(dx | \mu(\theta), v(\theta)) \quad (6.43)$$

Un rapide calcul (cf section A.13) donne l'information

$$I = \frac{\mu'^2}{v} + \frac{1}{2} \left(\frac{v'}{v} \right)^2 \quad (6.44)$$

Cette information est d'autant plus grande que

- v est petit : un instrument précis donne une grande information
- μ' ou v' sont grands : un instrument dont la réponse moyenne varie vite avec l'état est sensible

Il est intéressant de remarquer que si μ ne varie pas, c'est-à-dire que la réponse moyenne de l'instrument ne dépend pas de l'état, mais que sa variance varie, l'information n'est pas nulle. En effet, la dispersion des mesures donne alors à elle seule une indication sur l'état.

Exemple En dosimétrie individuelle par thermoluminescence, la mesure de la dose D peut se modéliser à l'aide d'une gaussienne dont la moyenne est D et l'écart-type proportionnel à D . On a ici

$$\mu(D) = D \quad v(D) = a^2 D^2 \quad I = \frac{2 + 1/a^2}{D^2}$$

Soit x la réponse du dosimètre, que nous prendrons comme estimateur de la dose $\hat{D}(x) = x$. L'espérance de cet estimateur lorsque la dose vaut D est

$$E\hat{D} | D = \frac{1}{\sqrt{2\pi a^2 D^2}} \int_x x e^{-\frac{(x-D)^2}{2a^2 D^2}} dx = D \quad (6.45)$$

montrant qu'il n'est pas biaisé. Sa variance est évidemment

$$V\hat{D} | D = \frac{1}{\sqrt{2\pi a^2 D^2}} \int_x (x - D)^2 e^{-\frac{(x-D)^2}{2a^2 D^2}} dx = a^2 D^2 \quad (6.46)$$

L'inégalité de Cramér-Rao (6.42) s'écrit

$$a^2 D^2 \cdot \frac{2+1/a^2}{D^2} = 1 + a^2 > 1 \quad (6.47)$$

prouvant qu'il n'est pas efficient car $k'(\theta)^2 = 1$ n'est pas atteint.

6.2.4.2 Modèle poissonnien

Dans ce cas, le log de la fonction de vraisemblance est

$$\ln \mathcal{L}(a | N) = \ln \left(\frac{(a\epsilon t)^N}{N!} e^{-a\epsilon t} \right) = N \ln(a\epsilon t) - \ln N! - a\epsilon t \quad (6.48)$$

donc

$$\frac{\partial^2 \mathcal{L}}{\partial a^2} = -\frac{N}{a^2} \quad (6.49)$$

et l'information de Fisher est

$$I(a) = - \sum_N -\frac{N}{a^2} \mathcal{P}(N | a\epsilon t) = \frac{\epsilon t}{a} \quad (6.50)$$

Soit l'estimateur de l'activité donné par le maximum de vraisemblance

$$\hat{a}(N) = \frac{N}{\epsilon t} \quad (6.51)$$

On a

$$E\hat{a} | a = \sum_N \hat{a}(N) \mathcal{P}(N | a\epsilon t) = \frac{1}{\epsilon t} a\epsilon t = a \quad (6.52)$$

$$(6.53)$$

L'estimateur est non biaisé et $k'(a)^2 = 1$. Mais

$$V\hat{a} | a = \frac{1}{(\epsilon t)^2} a\epsilon t = \frac{a}{\epsilon t} \quad (6.54)$$

donc $V\hat{a} | a \cdot I(a) = 1$ et l'estimateur est efficient.

L'estimateur au maximum de vraisemblance est un bon choix car l'absence de biais implique une convergence vers la vraie valeur de l'activité et sa variance est minimale.

Soit $\bar{N}(a)$ le nombre moyen d'impulsions lorsque l'activité vaut a . On a $\bar{N}(a) = \frac{a}{\epsilon t} = V\hat{a} | a$. On retrouve ainsi le traditionnel écart-type en $\sqrt{\bar{N}}$ des comptages poissonniens.

6.2.4.3 Modèle gamma

Ici, on a

$$\ln \mathcal{L}(a \mid t) = N \ln(a\epsilon) - \ln \Gamma(N) + (N-1) \ln t - a\epsilon t \quad (6.55)$$

$$\frac{\partial \mathcal{L}}{\partial a} = \frac{N}{a} - \epsilon t \quad (6.56)$$

$$\frac{\partial^2 \mathcal{L}}{\partial a^2} = -\frac{N}{a^2} \quad (6.57)$$

(6.58)

ce qui donne, vu que cette dernière expression ne dépend pas de t

$$I(a) = \frac{N}{a^2} \quad (6.59)$$

L'estimateur du maximum de vraisemblance est

$$\hat{a} = \frac{N}{\epsilon t} \quad (6.60)$$

Un rapide calcul montre que

$$E\hat{a} \mid a = \frac{N}{N-1}a \quad (6.61)$$

$$k'(a)^2 = \left(\frac{N}{N-1} \right)^2 \quad (6.62)$$

$$V\hat{a} \mid a = \frac{N^2 a^2}{(N-1)^2(N-2)} \quad (6.63)$$

$$V\hat{a} \mid a \cdot I(a) = \frac{N^2 a^2}{(N-1)^2(N-2)} \cdot \frac{N}{a^2} = \frac{N}{N-2} k'(a)^2 > k'(a)^2 \quad (6.64)$$

Cet estimateur est biaisé et non efficient. De plus, pour $N = 1$, expérience correspondant à l'attente de la première détection, son espérance diverge. Il est néanmoins sans biais et efficient asymptotiquement, car

$$\lim_{N \rightarrow \infty} E\hat{a} \mid a = a \quad (6.65)$$

$$\lim_{N \rightarrow \infty} V\hat{a} \mid a \cdot I(a) = \lim_{N \rightarrow \infty} k'(a)^2 = 1 \quad (6.66)$$

Chapitre 7

Choix des a priori

Les a priori sont divisés en deux grandes classes. L'une est celle des non informatifs, qui sont ceux de départ, lorsque l'observateur n'a initialement rien à sa disposition, si ce n'est son instrument de mesure. L'autre est celle des informatifs, qui permettent d'inclure une connaissance sur l'état préexistante à la mesure. Avant de les aborder, nous allons décrire un type d'a priori qui est très utile pour calculer facilement des a posteriori.

7.1 Conjugaison entre l'a priori et l'a posteriori

7.1.1 Définition

Reprenons le théorème de Bayes donnant l'a posteriori associé à un modèle $P(dx | \theta)$ et un a priori $\pi(d\theta)$

$$\pi(d\theta | x) = \frac{P(dx | \theta)\pi(d\theta)}{M(dx)} \quad (7.1)$$

$$M(dx) = \int_{\Theta} P(dx | \theta)\pi(d\theta) \quad (7.2)$$

Habituellement, les intégrales portent sur la variable x du modèle et sont bien connues. Par contre, les intégrales faisant intervenir l'a posteriori portent sur l'état, autrement dit sur les paramètres des répartitions. Par exemple, dans le cas gaussien, si la variance ν du modèle est fixée et que seule la moyenne μ est à déterminer, le problème est simple, puisque μ se trouve au même niveau que x , et n'intervient que comme paramètre de position. Par contre, si la variance elle-même n'est pas connue, les intégrales portent sur les variables qui ont des positions nettement moins favorables. En effet, la probabilité conjointe issue des a priori $\pi(d\mu)$ et $\pi(d\nu)$ devient

$$P(dx, d\mu, d\nu) = \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{(x-\mu)^2}{2\nu}} \pi(d\mu)\pi(d\nu)dx$$

Une manière d'appréhender la situation est de se restreindre à des classes d'a priori ayant un comportement plus simple lors de la transformation par le modèle et permettant d'exprimer analytiquement la distribution des états a posteriori.

Soit \mathcal{F} une famille de probabilités sur Θ . on dit que cette famille est *conjuguée* au modèle $P(dx | \theta)$ si tout a priori dans \mathcal{F} donne un a posteriori qui est aussi dans \mathcal{F} . Une telle famille est *stable par posteriorisation*.

Examinons quelques cas particuliers.

7.1.2 Lois gamma et modèle de Poisson

Reprendons le modèle poissonnien associé à une mesure de source radioactive de durée t avec une sensibilité ϵ .

$$P(N | a) = \mathcal{P}(N | a\epsilon t) = \frac{(a\epsilon t)^N}{N!} e^{-a\epsilon t} \quad (7.3)$$

La famille de lois *gamma*

$$\left\{ \text{Ga}(da | N_f, t_f) = \frac{t_f^{N_f}}{\Gamma(N_f)} a^{N_f-1} e^{-at_f} da \mid N_f > 0, t_f > 0 \right\} \quad (7.4)$$

est conjuguée à ce modèle. En effet, soit un a priori $\pi(da) = \text{Ga}(da | N_f, t_f)$, l'a posteriori s'exprime

$$\pi(da | N) \propto (a\epsilon t)^N e^{-a\epsilon t} a^{N_f-1} e^{-at_f} da \quad (7.5)$$

$$\propto a^{N_f+N-1} e^{-a(t_f+\epsilon t)} da \quad (7.6)$$

qui n'est autre qu'une loi gamma

$$\pi(da | N) = \text{Ga}(da | N_f + N, t_f + \epsilon t) \quad (7.7)$$

Remarquons tout d'abord que les rôles du temps et de l'activité sont échangés par rapport à l'introduction des lois gamma pour les temps d'attente (cf section 7.1.3). La signification des paramètres de la loi gamma devient alors celle d'un comptage et d'un temps de mesure. En effet, si l'on traduit l'a priori $\text{Ga}(da | N_f, t_f)$ par *la source a déjà produit N_f désintégrations pendant t_f* , l'ajout de l'information *obtenir N impulsions pendant t* résulte bien en *avoir obtenu la somme des nombres de désintégrations $N_f + N$ pendant le temps total $t_f + \epsilon t$* si l'on admet que mesurer pendant t avec une sensibilité ϵ revient au même que mesurer pendant ϵt avec une sensibilité 1. ϵt apparaît comme un *temps équivalent*. C'est celui qu'on devrait prendre pour obtenir le même résultat avec un compteur d'efficacité 1. La moyenne et la variance de l'a posteriori sont

$$Ea | N = \frac{N_f + N}{t_f + \epsilon t} \quad (7.8)$$

$$Va | N = \frac{N_f + N}{(t_f + \epsilon t)^2} \quad (7.9)$$

On retrouve ainsi un taux pour l'espérance ainsi qu'une déviation-standard croissant avec la racine carrée du nombre de détections.

Pour un a priori gamma, la répartition marginale des comptages est la loi *négative-binomiale*. Soit

$$p_f = \frac{t_f}{t_f + \epsilon t} \quad (7.10)$$

alors

$$M(N) = \mathcal{NB}(N | N_f, p_f) = \frac{\Gamma(N_f + N)}{\Gamma(N_f)\Gamma(N+1)} p_f^{N_f} (1-p_f)^N \quad (7.11)$$

En effet, nous avons vu que pour une activité a observée par un compteur de sensibilité ϵ , la répartition du nombre d'impulsions est poissonnienne de paramètre $a\epsilon$. Il suffit d'appliquer le résultat de la section A.2.

Remarque Le fait que l'activité soit donnée par une répartition introduit une sur-dispersion par rapport à la loi de Poisson, car la variance de la négative-binomiale est toujours supérieure à sa moyenne (cf 5.23).

7.1.3 Lois gamma et modèle gamma

Lors d'une expérience consistant à mesurer l'activité en notant le temps d'attente de la N^e détection dans le compteur, l'espace des observations devient réel positif. Le modèle pour un compteur d'efficacité ε est donné par

$$P(dt | a) = \text{Ga}(dt | N, a\varepsilon) \quad (7.12)$$

On voit facilement que si l'a priori sur l'activité est donné par une loi gamma

$$\pi(da) = \text{Ga}(da | N_f, t_f) \quad (7.13)$$

l'a posteriori prend la forme

$$\pi(da | t) = \text{Ga}(da | N_f + N, t_f + \varepsilon t) \quad (7.14)$$

qui est exactement le même que celui des comptages. Les lois gamma sont donc aussi conjuguées au modèle des temps d'attente, et les paramètres (N_f, t_f) admettent la même interprétation physique d'un comptage et d'un temps de mesure.

Si l'on prend pour variable $p = \frac{t_f}{t_f + \varepsilon t}$ au lieu de t , la marginale prend la forme d'une distribution bêta (cf A.3)

$$M(dp) = \text{Be}(dp | N_f, N) = \frac{\Gamma(N_f + N)}{\Gamma(N_f)\Gamma(N)} p^{N_f-1} (1-p)^{N-1} dp \quad (7.15)$$

7.1.4 Lois bêta et modèles binomiaux

Pour le modèle binomial

$$P(n | \varepsilon) = \mathcal{B}(n | N, \varepsilon) \quad (7.16)$$

on voit facilement que la famille de lois bêta

$$\left\{ \text{Be}(d\varepsilon | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \varepsilon^{\alpha-1} (1-\varepsilon)^{\beta-1} d\varepsilon \mid \alpha > 0, \beta > 0 \right\} \quad (7.17)$$

est conjuguée. Posons $\alpha = n_f$ et $\beta = N_f - n_f$. Soit $\text{Be}(dp | n_f, N_f - n_f)$ un a priori sur la sensibilité, que l'on peut interpréter par *avoir obtenu n_f détections lors de N_f désintégrations*. Cet a priori est compatible avec l'idée que $\varepsilon \sim \frac{n_f}{N_f}$. L'a posteriori consécutif à n détections lors d'une nouvelle série de N désintégrations s'écrit

$$\pi(d\varepsilon | n) = \text{Be}(d\varepsilon | n_f + n, (N_f + N) - (n_f + n)) \quad (7.18)$$

Ce qui est bien ce que l'on attend avec l'interprétation *obtenir $n_f + n$ détections en $N_f + N$ désintégrations*.

Cette famille est aussi conjuguée au modèle négatif-binomial. Le tableau 7.1 résume les différents modèles présentés ci-dessus. Les trois dernières lignes de sa dernière colonne utilisent les symboles définis ci-dessous :

- $p = \frac{t_f}{t_f + \varepsilon t}$
- Bb est la loi binomiale-bêta (cf [10] p117)
- Nbb est la loi négative-binomiale-bêta (cf [10] p118)

TAB. 7.1 – Lois et modèles de base concernant la détermination de l’activité des sources et de la sensibilité des compteurs

a priori	modèle	a posteriori	marginale
$\text{Ga}(da N_f, t_f)$	$\frac{\mathcal{P}(N a\epsilon t)}{\text{Ga}(dt N, a\epsilon)}$	$\text{Ga}(da N_f + N, t_f + \epsilon t)$	$\mathcal{N}\mathcal{B}(N N_f, p)$ $\text{Be}(dp N_f, N)$
$\text{Be}(d\epsilon n_0, N_0 - n_0)$	$\frac{\mathcal{B}(n N, \epsilon)}{\mathcal{N}\mathcal{B}(n N, \epsilon)}$	$\text{Be}(d\epsilon n_f + n, N_f + N - (n_f + n))$	$\text{Bb}(n n_f, N_f - n_f, N)$ $\text{Nbb}(n n_f, N_f - n_f, N)$

7.2 A priori sans information préalable

7.2.1 A priori constants

Une manière d’envisager le manque d’information sur un ensemble d’états est de le traduire par une notion d’invariance. Suivant Laplace, on donnera une probabilité égale à tous les états, c’est-à-dire invariante par translation. Ceci est illustré par son fameux problème :

Sachant que le soleil s’est levé tous les matins depuis la création, quelle est la probabilité qu’il se lève demain ?

L’état du soleil est caractérisé par sa probabilité de se lever $p \in [0, 1]$. Le modèle donne la probabilité d’obtenir n levers lors de N journées, qui est la binomiale

$$P(n | p) = \binom{N}{n} p^n (1-p)^{N-n} \quad (7.19)$$

A partir d’un a priori sur l’état $\pi(dp)$, le théorème de Bayes donne l’a posteriori suivant

$$\pi(dp | n) = \frac{\binom{N}{n} p^n (1-p)^{N-n} \pi(dp)}{M(n)} \quad (7.20)$$

$$M(n) = \int_p \binom{N}{n} p^n (1-p)^{N-n} \pi(dp) \quad (7.21)$$

Evaluons les a priori constants obtenus selon divers choix possibles de la variable décrivant l’état.

1) A priori sur la probabilité Laplace donne aux éléments dp une probabilité a priori uniforme sur $[0, 1]$, qui est un *invariant de lieu* car tous les points ont le même poids

$$\pi_1(dp) = dp$$

et obtient

$$\pi_1(dp | n) = C_1 \cdot p^n (1-p)^{N-n} dp \quad (7.22)$$

$$= \text{Be}(dp | n+1, N-n+1) \quad (7.23)$$

C’est une loi bêta qui admet l’espérance suivante

$$E_1 p = \frac{n+1}{N+2} \quad (7.24)$$

2) A priori sur les chances de succès Comme p n'est qu'une des multiples caractérisations de l'état, on aurait très bien pu prendre ses *chances* (*anglais odds*) de succès au lieu de sa probabilité d'occurrence,

$$o = \frac{p}{1-p}$$

et y appliquer une probabilité uniforme a priori

$$\pi_2(dp) = do = d \frac{p}{1-p} = \frac{dp}{(1-p)^2}$$

on aurait alors

$$\pi_2(dp | n) = C_2 \cdot p^n (1-p)^{N-n-2} dp \quad (7.25)$$

$$= \text{Be}(dp | n+1, N-n-1) \quad (7.26)$$

dont la moyenne est

$$E_2 p = \frac{n+1}{N} \quad (7.27)$$

Signalons d'emblée que cet a priori va poser un problème lorsque $n = N$. En effet, l'a posteriori n'est pas normalisable dans ce cas et donne une valeur aberrante pour la moyenne.

3) A priori sur l'information relative à la probabilité On pourrait aussi associer à p son contenu informationnel $\ln \frac{1}{p}$ et obtenir l'*invariant d'échelle*

$$\pi(dp) = d \ln p = \frac{dp}{p}$$

on aurait alors

$$\pi_3(dp | n) = C_3 \cdot p^{n-1} (1-p)^{N-n} dp \quad (7.28)$$

$$= \text{Be}(dp | n, N-n+1) \quad (7.29)$$

avec la moyenne

$$E_3 p = \frac{n}{N+1} \quad (7.30)$$

4) A priori de l'analyse de la variance Enfin, on pourrait écrire l'état par $\arcsin \sqrt{p}$, induit par une transformation de stabilisation de la variance, ce qui donnerait

$$\pi_4(dp) = \frac{4}{\pi} \frac{dp}{\sqrt{p} \sqrt{1-p}}$$

$$\pi_4(dp | n) = C_4 \cdot p^{n-\frac{1}{2}} (1-p)^{N-n-\frac{1}{2}} dp \quad (7.31)$$

$$= \text{Be}(dp | n + \frac{1}{2}, N-n+\frac{1}{2}) \quad (7.32)$$

avec la moyenne

$$E_{4p} = \frac{n + \frac{1}{2}}{N + 1} \quad (7.33)$$

On peut remarquer que les moyennes a posteriori obtenues par ces différents a priori sont égales à $O(\frac{1}{N})$ près, c'est-à-dire asymptotiquement les mêmes. Cette limite est celle de l'analyse fréquentiste. Si N n'est pas grand, ces différences deviennent importantes. En particulier, si l'évènement s'est produit chaque fois (ce qui est le cas du problème de Laplace), le deuxième modèle donne un a posteriori non normalisable. La figure 7.1 montre les différentes estimations de cette probabilité

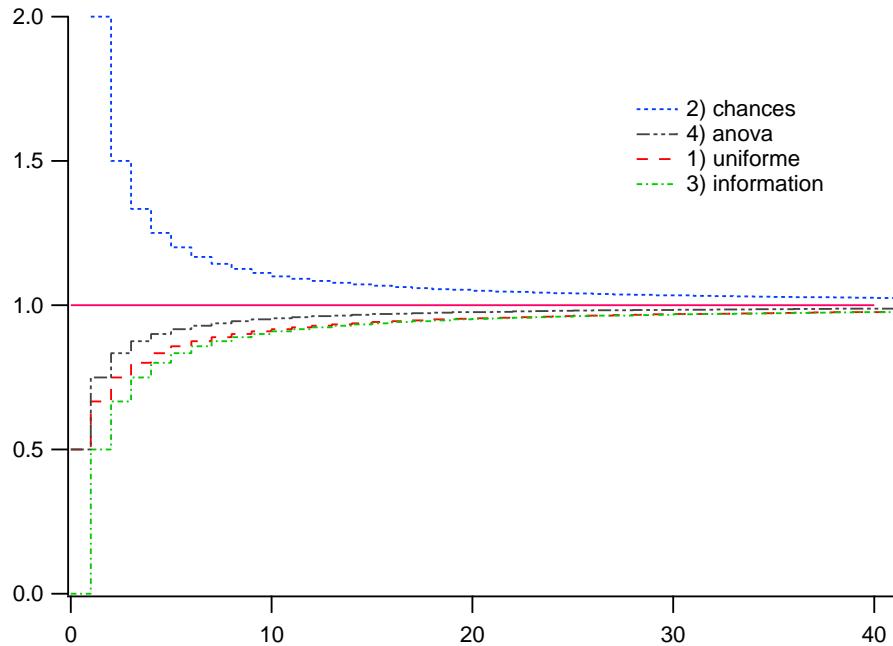


FIG. 7.1 – Estimation de la probabilité de lever du soleil. *Lors de N répétitions, l'évènement s'est produit chaque fois. Les courbes donnent l'estimation selon quatre types d'a priori. L'a priori 2 basé sur les chances de succès donne des valeurs aberrantes car l'a posteriori n'est pas normalisable.*

Ces différences ont poussé certains auteurs à rechercher des a priori universels, comme nous allons le voir plus bas.

7.2.2 A priori de Jeffreys

7.2.2.1 Définition

Un des résultats les plus importants de la discussion générale sur les a priori en l'absence de données préalables, a priori dits *non informatifs*, est qu'il est illusoire d'en trouver un qui convienne à tous les modèles. Néanmoins, pour les espaces d'états qui possèdent une structure de variété différentiable, il est possible d'en trouver qui, s'il dépendent du modèle, ne sont pas tributaires de la paramétrisation.

L'a priori de *Jeffreys*, avec les liens qu'il entretient avec la théorie de l'information et la résolution des estimateurs, s'est largement imposé.

La matrice d'information de Fisher est un tenseur métrique sur l'espace Θ , qui devient ainsi une variété de Riemann. Cette métrique induit une mesure au sens de la théorie de l'intégration [26]. On obtient ainsi l'a priori de *Jeffreys*, qui est défini de manière intrinsèque sur la variété

$$\pi_J(d\theta) = \sqrt{|I(\theta)|} d\theta \quad (7.34)$$

L'information de Fisher décrivant localement celle de Shannon, cet a priori est en relation étroite avec la notion générale d'information

Il faut remarquer que souvent cette densité n'est pas normalisable et donne des a priori dits *impropres*. C'est le cas par exemple pour les a priori uniforme da ou invariant d'échelle da/a sur la droite réelle positive. L'application du théorème de Bayes à de tels a priori donne la plupart du temps des a posteriori normalisables qui sont parfaitement utilisables.

Réécrivons l'inégalité de Cramér-Rao (6.41) dans le cas des estimateurs non biaisés :

$$\frac{1}{V\tau|\theta|} \leq I(\theta) \quad (7.35)$$

L'inverse de la variance peut constituer une mesure de la *résolution* d'un estimateur. Cette résolution est gage de sa qualité. L'inégalité de Cramér-Rao montre que localement, la résolution maximale ne peut excéder l'information de Fisher. En choisissant un a priori de Jeffreys, l'observateur donne plus de poids aux régions de l'espace des états qui sont susceptibles de fournir de bons estimateurs.

L'utilisation de la métrique de Fisher pour décrire la topologie de l'espace des états consiste à déplier cet espace selon l'information contenue en ses points. L'accentuation des régions où le modèle donne une grande sensibilité traduit le fait qu'une théorie et une expérimentation physique sont restreintes à une partie de la réalité.

7.2.2.2 Illustrations

Modèle poissonnien L'information de Fisher pour le modèles des comptages est donnée par l'équation (cf 6.50)

$$I(a) = \frac{\varepsilon t}{a} \quad (7.36)$$

Ce qui donne un a priori de Jeffreys

$$\pi_J(da) \propto a^{-\frac{1}{2}} da \quad (7.37)$$

Cet a priori est impropre, à mi-chemin entre l'a priori de lieu $a da$ et l'a priori d'échelle $a^{-1} da$, et peut être considéré comme limite d'une loi gamma

$$\pi_J(da) = \text{Ga}\left(da \mid \frac{1}{2}, 0\right) \quad (7.38)$$

L'a posteriori issu d'un comptage N est donc

$$\pi_J(da \mid N) = \text{Ga}\left(da \mid N + \frac{1}{2}, \varepsilon t\right) \quad (7.39)$$

dont les moments sont

$$E a | N = \frac{N + \frac{1}{2}}{\varepsilon t} \quad (7.40)$$

$$V a | N = \frac{N + \frac{1}{2}}{(\varepsilon t)^2} \quad (7.41)$$

L'estimateur bayesien (moyenne de l'a posteriori) est proche du $\frac{N}{\varepsilon t}$ induit par le maximum de vraisemblance. On retrouve aussi un écart-type en \sqrt{N} . Il est intéressant de remarquer qu'en l'absence de détection ($N = 0$), l'estimation bayesienne de l'activité n'est pas nulle.

Modèle gamma Selon la dérivation (6.59), l'information de Fisher s'écrit

$$I(a) = \frac{N}{a^2} \quad (7.42)$$

Ce qui donne un a priori de Jeffreys

$$\pi_J(da) \propto a^{-1} da \quad (7.43)$$

qui n'est autre que l'invariant d'échelle et peut aussi être considéré comme limite d'une loi gamma

$$\pi_J(da) = \text{Ga}(da | 0, 0) \quad (7.44)$$

L'a posteriori induit par un temps d'attente t suit alors aussi une loi gamma

$$\pi_J(da | t) = \text{Ga}(da | N, \varepsilon t) \quad (7.45)$$

dont les moments sont

$$E a | N = \frac{N}{\varepsilon t} \quad (7.46)$$

$$V a | N = \frac{N}{(\varepsilon t)^2} \quad (7.47)$$

Dans ce modèle, $N = 0$ n'est pas admis, et t ne s'annule presque sûrement pas. L'estimateur de la moyenne de l'a posteriori garde un sens pour $N = 1$, contrairement à l'approche du maximum de vraisemblance.

7.2.3 A priori maximisant l'information

Un autre approche consiste à faire en sorte que la mesure ait le maximum d'influence sur l'a posteriori. En termes informationnels, cela revient à trouver un a priori qui maximise l'information moyenne apportée par la mesure. Selon la définition 6.9, on cherche

$$\pi = \arg \max_{\pi'} I_H(\pi') \quad (7.48)$$

Cet a priori éloigne le plus possible la conjointe $P(dx, d\theta)$ du produit $P(dx)\pi(d\theta)$ (cf (6.13)). Il maximise la divergence avec son a posteriori (cf (6.15)), il est sensible à la mesure. Un observateur ayant ce type d'a priori est susceptible de changer facilement d'avis. Selon la troisième formulation (6.14), une telle répartition éloigne le plus possible la réponse à un état particulier de la répartition globale de l'état,

ce qui correspond à ne pas privilégier un état à un autre. A l'opposé, un a priori qui décorrélait les états des observations annulerait l'influence de la mesure et serait impossible à modifier. C'est le cas par exemple des répartitions déterministes δ_θ .

La recherche de ce type d'a priori est problématique, car la maximisation n'est souvent pas possible, ou pire, donne des solutions *dégénérées* attribuant des probabilités nulles à certaines régions de l'espace des états. Comme le support de l'a posteriori est toujours contenu dans celui de l'a priori, ces états sont définitivement exclus de l'expérience, car aucune observation ne pourra leur donner par la suite une probabilité non nulle. L'exemple du cas fini est abordé en annexe à la section B.4. Une proposition a été faite par Berger [9] et Bernardo [10] pour sortir de cette impasse. Au lieu de maximiser l'information pour une mesure, on cherche à la maximiser pour une suite de mesures indépendantes. Lorsque le nombre de mesures tend vers l'infini, les a priori obtenus ne sont plus dégénérés. Ils sont dits *de référence* (cf B.3.2).

Dans le contexte de la mesure de la radioactivité, il est possible de se restreindre à la famille des lois gamma, et quelques calculs assez complexes suggèrent que les a priori de Jeffreys, invariant de lieu et invariants d'échelle sont des solutions [27].

Une autre classe d'a priori, dite *MDIP* (Pour Maximum Data Information Prior) a été proposée par Zellner [28]. Cette approche théorique donne l'a priori de Jeffreys pour le modèle de Poisson, et un curieux a priori limite $\text{Ga}(2, 0)$ pour le modèle gamma. Le lecteur en trouvera une définition à la section B.3.1.

7.3 A priori avec information préalable sur les états

Le concept d'information basé sur les définitions de Shannon et Fisher permet d'intégrer les résultats des mesures préalables dans la recherche d'un a priori optimal. Ce type d'a priori est dit *entropique basé sur des données préalables*. D'autre part, l'itération du théorème de Bayes fait aussi évoluer la répartition de l'état en tenant compte d'un ensemble de mesures. Ces deux méthodes sont en fait complémentaires. Afin de délimiter les domaines d'application de ces deux méthodes, il convient de faire la différence entre les états stationnaires ou non.

7.3.1 Stationnarité de l'état

Lors de la répétition d'une mesure, l'état sera dit *stationnaire* s'il ne varie pas d'une mesure à l'autre. Le modèle est le produit des modèles de base

$$P(x \mid \theta) = \prod_{i=1}^r P(x_i \mid \theta) \quad (7.49)$$

Et la marginale sur les observations devient si $x = (x_i)_{i=1}^r$

$$P_{\text{stat}}(x \mid \pi) = \int \pi(d\theta) \prod_{i=1}^r P(x_i \mid \theta) \quad (7.50)$$

La méthode à appliquer pour obtenir un a priori est décrite ci-dessous. C'est l'itération du théorème de Bayes depuis un a priori initial non informatif.

Par contre, si l'état a varié au cours des mesures, mais que son occurrence est régie par une probabilité π , la marginale sur l'ensemble des mesures préalables prend la forme

$$P_{non\ stat}(x | \pi) = \prod_i P(x_i | \pi) = \prod_i \int P(x_i | \theta) \pi(d\theta) \quad (7.51)$$

Dans ce cas, l'itération du théorème de Bayes ne s'applique évidemment pas telle quelle, mais il est possible de l'utiliser en considérant comme état non plus les éléments de Θ , mais les probabilités sur ce dernier.

7.3.2 Cas stationnaire, itération du théorème de Bayes

L'estimation des variables d'état d'un système physique perturbé par un bruit se fait souvent par la répétition de la même expérience, seules les composantes de bruit étant censées avoir changé, et d'une manière indépendante au sens statistique. La théorie bayesienne de la décision statistique permet l'intégration de mesures répétées, proposant ainsi un pont vers la notion fréquentiste de la probabilité.

Considérons un système pour lequel nous avons une répartition a priori des états $\pi_0(d\theta)$. L'obtention d'une observation x_1 sur ce système nous conduit par application du théorème de Bayes à l'a posteriori $\pi_0(d\theta | x_1)$. Si nous supposons que l'état ne change pas, et que nous effectuons une nouvelle mesure qui donne comme résultat l'observation x_2 , il est maintenant possible de trouver un nouvel a posteriori tenant compte de cette observation, mais en partant de l'a posteriori issu de x_1 . En d'autres termes, l'a posteriori issu d'une mesure sert d'a priori à la suivante.

Répétant le procédé, nous aboutissons à une suite de répartitions sur Θ . A la suite de résultats

$$x = (x_1 \dots x_r)$$

correspond la suite de répartitions

$$\pi_0(d\theta) \quad (7.52)$$

$$\pi_1(d\theta) = \pi_0(d\theta | x_1) = \frac{P(x_1 | \theta) \pi_0(d\theta)}{\int_{\theta'} P(x_1 | \theta') \pi_0(d\theta')} \quad (7.53)$$

...

$$\pi_r(d\theta) = \pi_{r-1}(d\theta | x_r) \quad (7.54)$$

Désignons la répartition obtenue au bout des r mesures par

$$\pi^{iter}(d\theta) = \pi_r(d\theta)$$

On peut alors se poser la question de ce qui se passe si l'on fait tendre r vers l'infini. L'intuition dit qu'à la limite, la répartition devrait être le dirac δ_{θ} , où θ est l'état effectif dans lequel se trouve le système. On aurait alors le pendant d'une loi des grands nombres. En outre, il se pourrait bien que l'a priori originel π_0 ne joue pratiquement plus aucun rôle à la limite, accroissant la robustesse de la méthode bayesienne. Ce problème général dépassant le propos de cette étude, nous allons nous limiter à montrer l'équivalence de l'itération du théorème de Bayes avec une approche dans l'espace produit induit par une suite de mesures. La limite de l'itération sera calculée dans le cas des lois normales, poissonniennes et gamma.

7.3.2.1 Produit des espaces d'observation

Considérons la suite d'observations (x_1, x_2, \dots, x_r) comme un seul point x dans l'espace produit \mathcal{X}^r . Les mesures se faisant indépendamment au sens probabiliste, et l'état du système ne changeant pas, le modèle de mesure sur \mathcal{X}^r est donné par

$$P^r(dx | \theta) = \prod_{i=1}^r P(dx_i | \theta) \quad (7.55)$$

Partons à nouveau de l'a priori π_0 du paragraphe précédent, l'application du théorème de Bayes à ce nouveau modèle donne

$$\pi^{prod}(d\theta) = \pi_0(d\theta | x) = \frac{P^r(x | \theta) \pi_0(d\theta)}{\int_{\theta'} P^r(x | \theta') \pi_0(d\theta')} = \frac{\pi_0(d\theta) \prod_i P(x_i | \theta)}{\int_{\theta'} \pi_0(d\theta') \prod_i P(x_i | \theta')} \quad (7.56)$$

7.3.2.2 Equivalence de l'itération et du produit

Développons l'approche itérative

$$\begin{aligned} \pi_1(d\theta') &= \frac{\pi_0(d\theta') P(x_1 | \theta')}{\int_{\theta} \pi_0(d\theta) P(x_1 | \theta)} \\ \pi_2(d\theta'') &= \frac{\pi_1(d\theta'') P(x_2 | \theta'')}{\int_{\theta_1} \pi_1(d\theta') P(x_2 | \theta')} \\ &= \frac{\frac{\pi_0(d\theta'') P(x_1 | \theta'')}{\int_{\theta} \pi_0(d\theta) P(x_1 | \theta)} P(x_2 | \theta'')}{\int_{\theta'} \frac{\pi_0(d\theta') P(x_1 | \theta')}{\int_{\theta} \pi_0(d\theta) P(x_1 | \theta)} P(x_2 | \theta')} \\ &= \frac{\pi_0(d\theta'') P(x_1 | \theta'') P(x_2 | \theta'')}{\int_{\theta} \pi_0(d\theta') P(x_1 | \theta') P(x_2 | \theta')} \\ &\dots \\ \pi_r(d\theta) &= \frac{\pi_0(d\theta) P(x_1 | \theta) \dots P(x_r | \theta)}{\int_{\theta} \pi_0(d\theta') P(x_1 | \theta') \dots P(x_r | \theta')} \end{aligned}$$

ce qui est exactement l'expression de π^{prod} . Il revient donc au même de considérer d'un coup les résultats d'une mesure multiple que de les traiter séquentiellement.

7.3.2.3 Modèles poissonnien ou gamma

Les suites des répartitions pour ces modèles sont particulièrement simples. En effet, supposons un observateur partant d'un a priori gamma $\text{Ga}(da | N_f, t_f)$ et accumulant des comptages N_i durant des temps t_i (ou attendant des temps t_i pour des comptages N_i), devant une source constante d'activité inconnue a_0 . Après la n^e mesure, la répartition de l'état est

$$\pi_n(da) = \text{Ga}(da | N_f + \sum_i N_i, t_f + \varepsilon \sum_i t_i) \quad (7.57)$$

Si la valeur de l'état stationnaire est a_0 , les N_i sont produits de manière indépendante par la loi de Poisson $\mathcal{P}(N | a_0 \varepsilon t)$. Lorsque $n \rightarrow \infty$, on peut appliquer la loi des grands nombres et

$$\frac{N_f + \sum_i N_i}{\varepsilon \sum_i t_i} \rightarrow a_0 \quad (7.58)$$

Les propriétés de lois gamma donnent alors à la limite

$$\lim_{n \rightarrow \infty} \pi_n(da) = \delta_{a_0} \quad (7.59)$$

7.3.2.4 Modèle gaussien

Plaçons-nous dans le cas où l'ensemble des états et celui des observations sont tout deux \mathbb{R} , et supposons le modèle gaussien avec une variance fixe

$$P(dx | \theta) = \mathcal{N}(dx | \theta, v)$$

Partons d'un a priori également gaussien

$$\pi(d\theta) = \mathcal{N}(d\theta | \theta_0, v_0)$$

et examinons l'évolution des répartitions a posteriori. La propriété de composition des lois normales donne un a posteriori consécutif à l'observation x_1 qui est aussi gaussien. L'inverse de sa variance est la somme des inverses des variances du modèle et de l'a priori, et sa moyenne la pondération selon les inverses des variances de leurs moyennes. On a

$$\theta_1 = \frac{\frac{\theta_0}{v_0} + \frac{x_1}{v}}{\frac{1}{v_0} + \frac{1}{v}} \quad (7.60)$$

$$\frac{1}{v_1} = \frac{1}{v_0} + \frac{1}{v} \quad (7.61)$$

L'application de ce lemme à la suite de mesures $(x_1 \dots x_r)$ donne la suite de variances et de moyennes suivantes pour les a posteriori

$$\theta_r = \frac{\frac{\sum_i x_i}{v} + \frac{\theta_0}{v_0}}{\frac{r}{v} + \frac{1}{v_0}} \quad (7.62)$$

$$\frac{1}{v_r} = \frac{r}{v} + \frac{1}{v_0} \quad (7.63)$$

L'équation (7.63) montre que la variance tends vers 0 avec $\frac{1}{n}$, ce qui est conforme à l'intuition. Réécrivons l'équation (7.62) sous la forme :

$$\theta_r = \frac{\frac{1}{r} \sum_i x_i + \frac{v\theta_0}{rv_0}}{1 + \frac{v}{rv_0}} \quad (7.64)$$

Si l'état effectif du système est θ_{eff} , le terme $\frac{1}{r} \sum_i x_i$ tends vers θ_{eff} par la loi des grands nombres et les autres termes tendent vers zéro. Donc

$$\lim_{r \rightarrow \infty} \theta_r = \theta_{eff} \quad (7.65)$$

De plus, la moyenne et la variance de la répartition a priori originelle ne jouent pas de rôle à la limite. Il faut néanmoins remarquer que la convergence sera d'autant plus lente que sa variance est petite et que sa moyenne est éloignée de l'état effectif, traduisant le fait que les mauvais a priori sont d'autant plus dangereux qu'ils sont pointus.

7.3.2.5 Accumulation de l'information

En d'autres termes, il revient au même d'envisager une suite de mesures indépendantes sur un système stationnaire d'une manière séquentielle que de considérer l'ensemble des résultats comme un tout, à condition d'accumuler l'information le long de la séquence. Cette accumulation se fait en partant à chaque étape d'un a priori issu de toutes les mesures précédentes et en intégrant chaque mesure via le théorème de Bayes. Il est intéressant de remarquer que l'ordre des résultats n'a pas d'importance et que pour une large classe d'a priori originels, la forme de leur répartition n'influe pas sur l'a posteriori limite qui finit par être concentré en l'état effectif du système, à condition que l'a priori original ne soit pas dégénéré.

7.3.3 Cas non stationnaire, Bayes appliqué à des états probabilistes

7.3.3.1 La probabilité comme état

Si les observations ont été obtenues en présence d'états différents, il est licite d'imaginer que ces derniers se sont présentés aux observateurs selon un processus aléatoire produit par une probabilité inconnue sur Θ . Cette probabilité sera dite *effective* et notée π_{eff} . Au lieu d'un état particulier, c'est une répartition sur les états qu'il faut trouver. Notons par $\mathcal{M}_1(\Theta)$ l'espace des probabilités sur Θ . La probabilité π_{eff} est un élément de cet ensemble, qui sera un nouvel espace d'états auquel nous allons appliquer le formalisme de la TBDS.

Dans le cas discret, une probabilité sur Θ est un vecteur $\pi = (\pi_1 \dots \pi_m)$ dont le nombre de composantes est égal au nombre d'éléments de Θ . Ces composantes sont soumises aux contraintes suivantes

$$0 \leq \pi_i \leq 1 \quad (7.66)$$

$$\sum_i \pi_i = 1 \quad (7.67)$$

Ces contraintes définissent un sous-espace de \mathbb{R}^m de dimension $m - 1$ qui se nomme un *simplexe* de dimension $m - 1$. Un point du simplexe peut être repéré par ses $m - 1$ premières coordonnées. Une répartition sur les probabilités sur Θ pourra être donnée par une densité, fonction de ces $m - 1$ coordonnées

$$\mu(d\pi) = f_\mu(\pi_1 \dots \pi_{m-1}) d\pi_1 \dots d\pi_{m-1} \quad (7.68)$$

Ici, nous utiliserons μ pour désigner une probabilité sur les probabilités, afin de bien la démarquer des probabilités originales π sur Θ .

7.3.3.2 Obtention d'une probabilité a posteriori

Le problème est d'obtenir une probabilité sur $\mathcal{M}_1(\Theta)$. Il faut tout d'abord définir un nouveau modèle. Partons du modèle de base $P(dx | \theta)$ sur Θ . Il est possible de le généraliser d'une manière naturelle à $\mathcal{M}_1(\Theta)$ en posant

$$\tilde{P}(dx | \pi) = \int_{\Theta} \pi(d\theta) P(dx | \theta) \quad (7.69)$$

Ceci traduit bien le fait que x est obtenu à partir du modèle original, mais avec un état qui est choisi au hasard selon $\pi(d\theta)$.

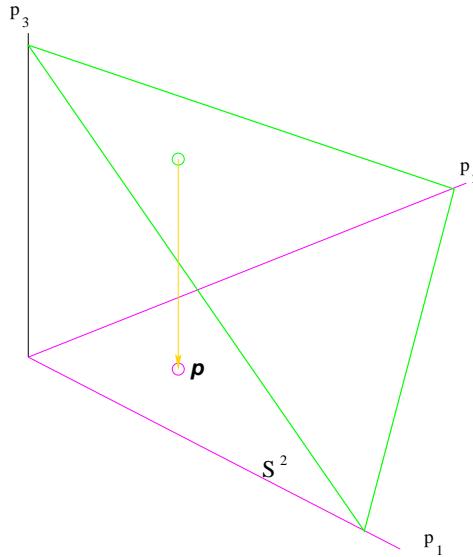


FIG. 7.2 – Simplexe probabiliste de dimension 2. *Le simplexe de dimension $m - 1$ est défini par l'ensemble des points de \mathbb{R}^m dont les coordonnées sont positives et dont la somme vaut 1. C'est donc l'ensemble des probabilités sur une ensemble fini à m éléments. Lorsque l'une ou plusieurs de ces coordonnées sont nulles, la probabilité se trouve sur le bord du simplexe et dite dégénérée.*

Supposons maintenant que nous ayons obtenu une suite d'observations $x = (x_1 \dots x_r)$ produites par ce nouveau modèle. Le théorème de Bayes s'écrit simplement

$$\mu(d\pi | x) = \frac{\mu(d\pi) \prod_k \tilde{P}(dx_k | \pi)}{\tilde{M}(dx)} \quad (7.70)$$

$$\tilde{M}(dx) = \int_{\pi} \mu(d\pi) \prod_k \tilde{P}(dx_k | \pi) \quad (7.71)$$

Si l'on dispose d'une probabilité $\mu(d\pi)$ a priori sur $\mathcal{M}_1(\Theta)$, la suite d'observations permet d'obtenir une probabilité a posteriori $\mu(d\pi | x)$ sur cet espace.

D'une manière générale, à une probabilité quelconque $v(d\pi)$ sur $\mathcal{M}_1(\Theta)$, il est possible d'associer une probabilité $\check{v}(d\theta)$ sur Θ par intégration. Soit A un sous-ensemble de Θ , définissons \check{v} par

$$\check{v}(A) = \int_{\pi} \pi(A) v(d\pi) \quad (7.72)$$

Cette fonction de A est une probabilité sur Θ qui est la moyenne de v , formellement

$$\check{v} = \int_{\pi} \pi v(d\pi) \quad (7.73)$$

Si ϕ est une fonction quelconque sur Θ , son intégrale se calcule par

$$\int_{\theta} \phi(\theta) \check{v}(d\theta) = \int_{\pi} v(d\pi) \int_{\theta} \phi(\theta) \pi(d\theta) \quad (7.74)$$

en considérant l'intégrale intérieure comme fonction de π .

Appliquant ce procédé à l'a posteriori $\mu(d\pi | x)$, celui-ci donnera une probabilité $\check{\mu}(d\theta | x)$ sur Θ qui résout le problème. Un observateur ayant à sa disposition r observations préalables, mais non forcément issues du même état, peut donc utiliser cet a posteriori comme a priori pour sa propre mesure. Il lui faut évidemment encore disposer d'un $\mu(d\pi)$ convenable. Dans le cas discret, il existe un a priori de Jeffreys. Nous nous limiterons à ce cas dans les lignes suivantes.

7.3.3.3 Particularisation au cas discret

Selon les notations de la section 5.6, nous avons :

- les états $\Theta = \{\theta_1 \dots \theta_m\}$
- les observations possibles $\mathcal{X} = \{x_1 \dots x_n\}$
- le modèle décrit par une matrice stochastique $P_{ij} = P(x_i | \theta_j)$

Posons $\pi_m = 1 - \sum_{j=1}^{m-1} \pi_j$ et $d\pi = d\pi_1 \dots d\pi_{m-1}$. Le modèle généralisé prend la forme

$$\tilde{P}(x_i | \pi) = \sum_{j=1}^m P_{ij} \pi_j \quad (7.75)$$

$$= \sum_{j=1}^{m-1} P_{ij} \pi_j + P_{im} \left(1 - \sum_{j=1}^{m-1} \pi_j \right), \quad i = 1 \dots n \quad (7.76)$$

Calculons la matrice d'information de Fisher

$$\frac{\partial \ln \tilde{P}(x_i | \pi)}{\partial \pi_{j_0}} = \frac{P_{ij_0} - P_{im}}{\sum_{j=1}^{m-1} P_{ij} \pi_j + P_{im} \left(1 - \sum_{j=1}^{m-1} \pi_j \right)} \quad (7.77)$$

$$\frac{\partial^2 \ln \tilde{P}(x_i | \pi)}{\partial \pi_{j_0} \partial \pi_{j_1}} = - \frac{(P_{ij_0} - P_{im})(P_{ij_1} - P_{im})}{\left(\sum_{j=1}^{m-1} P_{ij} \pi_j + P_{im} \left(1 - \sum_{j=1}^{m-1} \pi_j \right) \right)^2} \quad (7.78)$$

$$I_{j_0 j_1}(\pi) = - \sum_{i=1}^n \tilde{P}(x_i | \pi) \frac{\partial^2 \ln \tilde{P}(x_i | \pi)}{\partial \pi_{j_0} \partial \pi_{j_1}} \quad (7.79)$$

$$= \sum_{i=1}^n \frac{(P_{ij_0} - P_{im})(P_{ij_1} - P_{im})}{\sum_{j=1}^{m-1} P_{ij} \pi_j + P_{im} \left(1 - \sum_{j=1}^{m-1} \pi_j \right)} \quad (7.80)$$

La densité de l'a priori de Jeffreys est la racine du déterminant de cette $(m-1) \times (m-1)$ -matrice indexée ici par j_0 et j_1

$$g_J(\pi) = \left| \sum_{i=1}^n \frac{(P_{ij_0} - P_{im})(P_{ij_1} - P_{im})}{\sum_{j=1}^{m-1} P_{ij} \pi_j} \right|^{\frac{1}{2}} \quad (7.81)$$

Après r observations $x = (x_{i_1} \dots x_{i_r})$, il vient

$$g_J(\pi | x) \propto g(\pi) \prod_k \sum_{j=1}^m P_{i_k j} \pi_j \quad (7.82)$$

Résultat L'a priori tenant compte des données préalables $x = (x_{i_1} \dots x_{i_r})$ est

$$(\check{u}_J \mid x)_{j_0} \propto \int d\pi \pi_{j_0} g_J(\pi) \prod_{k=1}^r \sum_{j=1}^m P_{i_k j} \pi_j \quad (7.83)$$

$$g_J(\pi) = \left| \sum_{i=1}^n \frac{(P_{ij_0} - P_{i j_0})(P_{ij_1} - P_{i m})}{\sum_{j=1}^m P_{ij} \pi_j} \right|^{\frac{1}{2}} \quad (7.84)$$

Cet a priori calculé comme a posteriori d'une suite de mesures en présence d'un état non stationnaire sera désigné par *a priori de Jeffreys avec données préalables*. Cet a priori peut se calculer par Monte Carlo.

Cas particulier Si l'espace des états est formé de deux éléments $\{\theta_1, \theta_2\}$ et que ces états se manifestent de manière déterministe par deux observations possibles $\theta_1 \rightarrow x_1$ et $\theta_2 \rightarrow x_2$. Le modèle de base s'exprime par la matrice identité à deux dimensions et l'on obtient

$$g_J(\pi_1) = \sqrt{\frac{1}{\pi_1} + \frac{1}{1 - \pi_1}} \quad (7.85)$$

$$\int_0^1 g_J(\pi_1) d\pi_1 = \pi \quad (7.86)$$

$$\int_0^1 \pi_1 g_J(\pi_1) d\pi_1 = \frac{\pi}{2} \quad (7.87)$$

$$\pi_{J1} = \frac{1}{2} \quad (7.88)$$

Ce qui est compatible avec la symétrie en π_1 et $1 - \pi_1$.

7.3.3.4 Application au modèle poissonnien

Considérons l'expérience suivante : un observateur est en face d'une source radioactive qui fait partie d'un ensemble de m sources d'activités connues, mais il ignore laquelle. On lui communique une suite de r comptages, mais sans lui indiquer en présence de quelles sources ils ont eu lieu. La méthode développée ci-dessus lui permet d'obtenir un a priori pour sa propre mesure qui utilise cette information.

L'espace des états se ramène à l'ensemble des activités possibles

$$\Theta = \{a_1 \dots a_m\} \quad (7.89)$$

Les r comptages seront désignés par

$$N = (N_1 \dots N_r) \quad (7.90)$$

Le problème consiste à déterminer la probabilité qu'a cet observateur d'être en présence d'une source donnée. Le modèle de base donnant la répartition du nombre d'impulsions N en présence de la source j est poissonnien :

$$P_{Nj} = \mathcal{P}(N \mid a_j \varepsilon t) \quad (7.91)$$

ce qui donne le modèle généralisé

$$P_{N\pi} = \sum_{j=1}^m \frac{(a_j \varepsilon t)^N}{N!} e^{-a_j \varepsilon t} \pi_j \quad (7.92)$$

La méthode développée ci-dessus implique

$$g_J(\pi) = \left| \sum_{N=0}^{\infty} \frac{(P_{Nj_0} - P_{Nm})(P_{Nj_1} - P_{Nm})}{\sum_{j=1}^m P_{Nj} \pi_j} \right|^{\frac{1}{2}} \quad (7.93)$$

Si v_N représente le nombre de fois que N apparaît dans la suite des r observations, et que $m_N = \sum_{j=1}^m P_{Nj} \pi_j$ désigne la marginale associée à π , l'observateur devra partir de l'a priori sur les sources donné par

$$g_J(\pi | N) = g_J(\pi) \prod_{N=0}^{\infty} m_N^{v_N} \quad (7.94)$$

$$\pi_J(j | N) = C \cdot \int \pi_j g(\pi | N) d\pi \quad (7.95)$$

Une simulation basée sur les valeurs du tableau 7.2 a été effectuée. La figure 7.3 montre la convergence de la probabilité de Jeffreys avec données préalables vers la probabilité effective pour une suite de plus en plus longue d'observations préalables. Le temps de mesure est 2 et la sensibilité du détecteur 0.6

TAB. 7.2 – Paramètres de la simulation de l'a priori de Jeffreys avec données préalables

source	activité	probabilité effective
1	2	0.1
2	8	0.2
3	6	0.3
4	1	0.4

S'il n'y a que deux sources possibles, ce qui se produit par exemple si l'observateur doit décider de la présence ou de l'absence d'une source radioactive dans un rayonnement ambiant, les paramètres des lois de Poisson (7.91) deviennent

$$\theta_0 = bt \quad (7.96)$$

$$\theta_1 = (a\epsilon + b)t \quad (7.97)$$

La procédure ci-dessus s'applique directement. Soit v_N le nombre d'occurrences du comptage N dans les observations préalables et π_0 la probabilité a priori à attribuer à la situation sans source, il vient

$$m_N = \pi_0 \mathcal{P}(N | \theta_0) + (1 - \pi_0) \mathcal{P}(N | \theta_1) \quad (7.98)$$

$$\Delta_N = \mathcal{P}(N | \theta_0) - \mathcal{P}(N | \theta_1) \quad (7.99)$$

$$g_J(\pi_0) = \sqrt{\sum_N \frac{\Delta_N^2}{m_N}} \quad (7.100)$$

$$h(\pi_0) = \prod_N m_N^{v_N} \quad (7.101)$$

$$\pi_0 | v = \frac{\int \pi_0 h(\pi_0) g_J(\pi_0) d\pi_0}{\int h(\pi_0) g_J(\pi_0) d\pi_0} \quad (7.102)$$

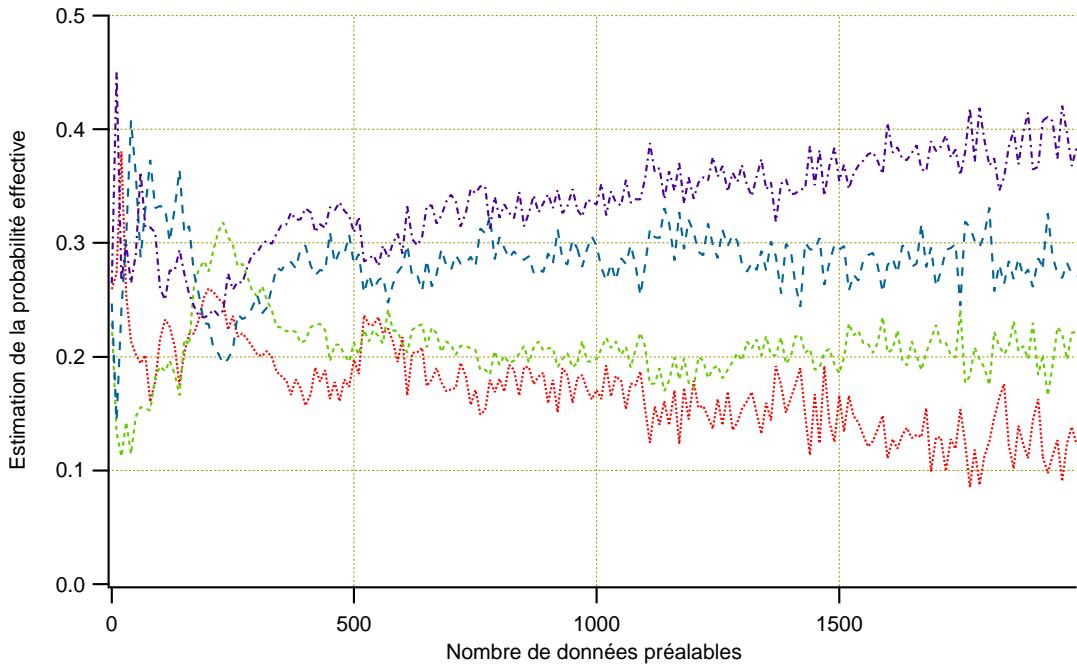


FIG. 7.3 – Evolution de l'a priori de Jeffreys avec données préalables pour un modèle poissonnien avec une activité non stationnaire. *L'estimation des composantes de la probabilité est calculée au fil de l'accumulation des données préalables, depuis un a priori non informatif sur les probabilités elles-mêmes. On remarque la convergence vers les valeurs effectives qui sont ici 0.1, 0.2, 0.3 et 0.4*

En l'absence de données préalables, g_J s'exprime par

$$g_J(\pi_0) = \sqrt{\sum_{N=0}^{\infty} \frac{\theta_0^N}{N!} e^{-\theta_0} \frac{\left(1 - \left(\frac{\theta_1}{\theta_0}\right)^N e^{-\theta_1+\theta_0}\right)^2}{\pi_0 + (1 - \pi_0) \left(\frac{\theta_1}{\theta_0}\right)^N e^{-\theta_1+\theta_0}}} \quad (7.103)$$

Comme le nombre d'états est deux, il suffit de donner la probabilité a priori à associer à la situation sans source θ_0 . Un intégration numérique de g_J a donné les valeurs présentées à la figure 7.4. Cet a priori est proche de $\frac{1}{2}$ si les états sont proches. La situation étant invariante par échange de θ_0 et θ_1 , seules les valeurs pour $\theta_1 > \theta_0$ sont représentées. Elles situent dans l'intervalle $[\frac{1}{2}, \frac{2}{3}]$.

Espaces continus Lorsque les espaces ne sont plus discrets, le problème est nettement plus complexe. Une piste est indiquée à la section B.5 de l'annexe.

7.4 Conclusion

Pour le calcul des activités avec des modèles poissonnien, les a priori non informatifs de choix sont des lois gamma. L'intégration des données préalables peut souvent s'effectuer en se limitant à ces lois. Dans les applications du chapitre suivant, nous nous bornerons donc aux a priori de Jeffreys, invariant de lieu et invariants d'échelles, en considérant les deux derniers comme limites de lois gamma.

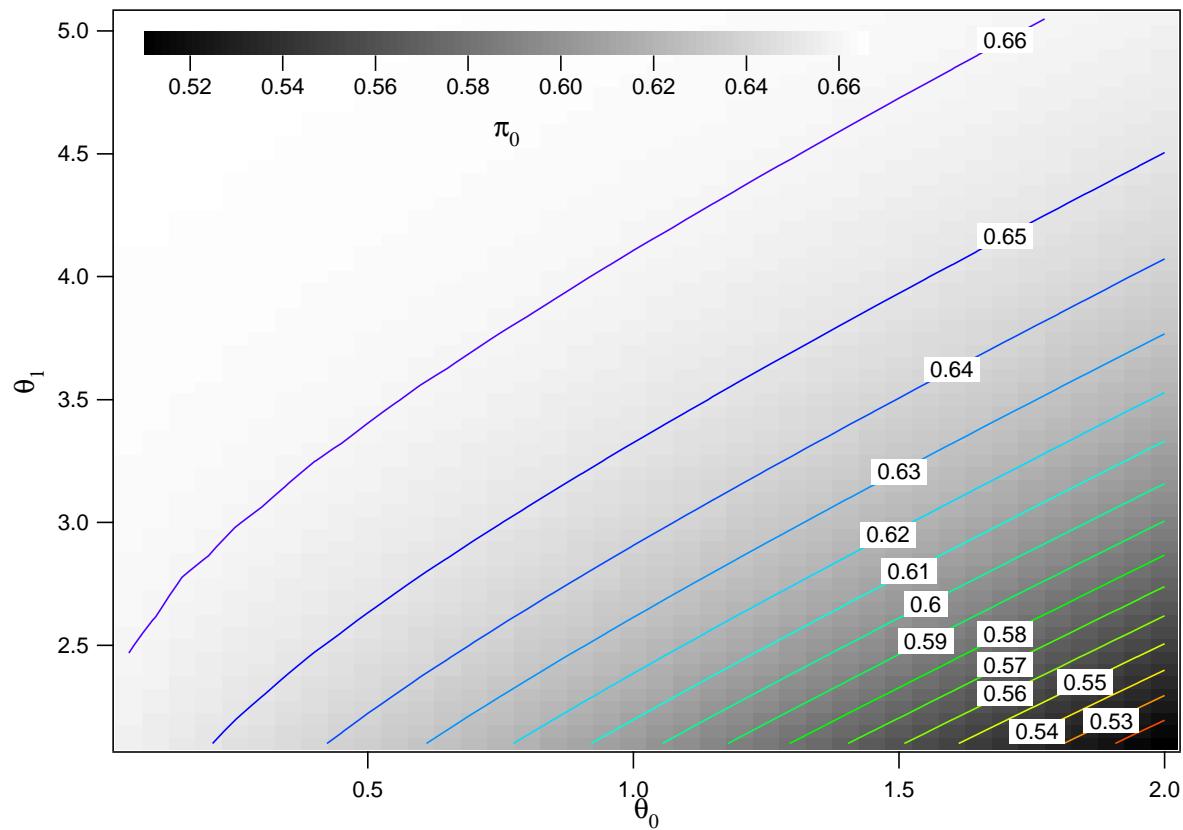


FIG. 7.4 – A priori pour un modèle à deux états. *Deux environnements produisant des nombres d'impulsions moyens θ_0 et θ_1 sont possibles. La probabilité a priori π_0 à associer à la situation θ_0 est donnée en fonction de θ_0 et θ_1 .*

Quatrième partie

Application de la TBDS à la mesure de la radioactivité

- *Bien, et ça sert à quoi ?*
- *Jean, tu ne réponds pas à cette question.*

Extrait d'un souper de thèse sur les algèbres de Lie

Chapitre 8

Introduction

Le but de cette partie est l’application de la théorie générale de la décision statistique à des problèmes apparaissant en physique des radiations. Ces applications seront de trois types. Le premier concerne des développements particuliers dont les résultats sont théoriques. Le deuxième contient des simulations qui traduisent des cas rencontrés en mesure de la radioactivité, mais qui n’ont pas fait l’objet d’une expérience précise en laboratoire. Le dernier type englobe le traitement de résultats d’expériences effectivement réalisées.

Tout d’abord, les conséquences de l’introduction d’un a priori informatif seront explicitées en termes de biais et de précision dans le cas d’un modèle poissonnien simple.

Dans le domaine des mesures monocanales à activité constante, les limites de détection seront interprétées à l’aide de la TBDS. Ces limites, le niveau critique L_C et la limite de détection L_D associée à l’activité minimale détectable¹ sont mieux comprises en différenciant correctement les états des observations. Ceci nous conduira assez naturellement à une discussion autour des intervalles de confiance, et des problèmes apparaissant lors de mesures de courte durée. Enfin, les incertitudes sur l’efficacité du compteur et la durée de mesure seront intégrées au modèle.

La spectrométrie γ permet une mesure multidimensionnelle des nucléides à schéma complexe. Après la construction du modèle, celui-ci sera appliqué à titre d’exemple au calcul des activités de l’europium-152, du thorium-230 et de l’uranium-235 présents dans des échantillons de l’environnement, depuis les comptages obtenus dans diverses zones du même spectre.

Nous aborderons ensuite les mesures avec activité décroissante. La durée de vie d’un atome radioactif ou la durée d’un état métastable lors d’une désintégration sont régis par des lois exponentielles. L’estimation du paramètre régissant ces temps sera estimé selon la TBDS, à partir d’une mesure du temps par chronométrage.

La diminution du nombre de noyaux par désintégration conduit à une décroissance de l’activité. L’estimation de ce nombre implique un modèle plus complexe, où intervient une constante radioactive. Le calcul de ces deux variables sera effectué à partir des temps d’apparition des détections dans un intervalle durée fixe.

¹MDA : Minimum Detectable Amount, est souvent utilisé

Chapitre 9

Information, biais et précision dans un modèle poissonnien simple

Le but de cette section est la comparaison de l'estimateur de Bayes (cf 2.3), donné par la moyenne de l'a posteriori, et de l'estimateur de Fisher, donné par le maximum de vraisemblance (cf 6.51). A cet effet nous nous sommes placé dans le cas simple d'une source sans bruit de fond, mesurée par un détecteur d'efficacité 1 durant un temps t . Le modèle est donné par la loi de Poisson

$$P(N | a) = \mathcal{P}(N | at) \quad (9.1)$$

L'information de Fisher pour ce modèle est donnée par (cf 6.50)

$$I(a) = \frac{t}{a} \quad (9.2)$$

Soit $\tau_F(N) = \frac{N}{t}$ l'estimateur au maximum de vraisemblance. La valeur de son espérance lorsque l'état est a vaut $k_F(a) = a$, ce qui montre qu'il est non biaisé. De plus, sa variance est alors donnée par $V\tau_F | a = \frac{a}{t}$. Le produit de la variance et de l'information de Fisher valant 1, cet estimateur est donc efficient au sens de Cramér-Rao.

Partons d'un a priori gamma sur l'activité $\pi(da) = \text{Ga}(da | N_f, t_f)$. Cet a priori correspond à une expérience qui aurait donné N_f impulsions en un temps t_f . L'estimateur bayésien associé est donné par l'espérance de l'a posteriori, qui, par la propriété de conjugaison des lois gamma (cf 7.7) donne

$$\tau_B(N) = \frac{N_f + N}{t_f + t} \quad (9.3)$$

Lorsque l'état est a , les moments de cet estimateur valent

$$E\tau_B | a = \sum_N \frac{N_f + N}{t_f + t} \mathcal{P}(N | at) = \frac{N_f + at}{t_f + t} = a \frac{1}{1 + \frac{t_f}{t}} + \frac{N_f}{t_f + t} \quad (9.4)$$

$$V\tau_B | a = \frac{at}{(t + t_f)^2} = \frac{a}{t} \frac{1}{\left(1 + \frac{t_f}{t}\right)^2} \quad (9.5)$$

Cet estimateur est donc biaisé et

$$k'_B(a) = \frac{1}{1 + \frac{t_f}{t}} \quad (9.6)$$

Mais le produit de sa variance et de l'information de Fisher est égale à $k'_B(a)^2$:

$$\frac{a}{t} \frac{1}{\left(1 + \frac{t_f}{t}\right)^2} \cdot \frac{t}{a} = \left(\frac{1}{1 + \frac{t_f}{t}}\right)^2 \quad (9.7)$$

donc τ_B est efficient au sens de Cramér-Rao.

On constate que la variance de l'estimateur bayesien est meilleure que celle de l'estimateur fréquentiste

$$V\tau_B = V\tau_F \frac{1}{\left(1 + \frac{t_f}{t}\right)^2} \quad (9.8)$$

Lorsque l'a priori devient plus pointu (par exemple si t_f augmente), la précision de l'estimateur augmente.

Remarques Un priori peut être donné par une mesure effectuée dans un premier laboratoire. Une autre mesure par un deuxième laboratoire peut alors utiliser cette information. Si la donnée primaire est sous la forme d'une valeur et d'une incertitude exprimée par un écart-type, un bon a priori pourra être une loi gamma de mêmes moyenne et écart-type.

Le biais de l'estimateur de Bayes est (cf 2.2)

$$b(a) = \left(\frac{N_f + N}{t_f + t} - a \right)^2 = \left(\frac{N_f - at_f}{t_f + t} \right)^2 \quad (9.9)$$

Si l'a priori a été obtenu en effectuant une mesure précédente très longue en présence d'une activité effective a_{eff} , la loi gamma résultante est concentrée autour de cette valeur et l'on peut écrire $a_{eff} \simeq \frac{N_f}{t_f}$. Supposons encore que le temps t de la mesure actuelle soit petit devant t_f , il vient

$$b(a) \simeq \left(\frac{a_{eff}t_f - at_f}{t_f + t} \right)^2 = \left(\frac{a_{eff} - a}{1 + \frac{t}{t_f}} \right)^2 \simeq (a_{eff} - a)^2 \quad (9.10)$$

L'apparition d'un biais pour l'estimateur bayesien est à première vue gênante. En effet, le biais est généralement à éviter, car un résultat précis mais inexact n'a que peu d'utilité. Les méthodes fréquentistes répètent la même mesure et nécessitent un estimateur non biaisé, puisque seul celui-ci converge en moyenne vers la vraie valeur. Les méthodes bayesiennes peuvent aussi répéter la même mesure, mais doivent chaque fois considérer un a priori différent, qui n'est autre que l'a posteriori issu de la mesure précédente. Le biais traduit ici le fait que l'on se rapproche de l'état effectif en prenant en compte l'information obtenue lors des mesures antérieures. Le résultat (9.10) montre que le biais est nul pour $a = a_{eff}$, et qu'il croît lorsque a s'éloigne de a_{eff} . En effet, si l'a priori est digne de confiance, celui-ci doit biaiser l'estimateur pour les activités a éloignées de a_{eff} , qui sont peu probables.

Chapitre 10

Mesures monocanales à activité constante

10.1 Limites de détection

10.1.1 Processus

Le cadre classique [29] [30] [31] [32] pour la détection de sources consiste en trois étapes :

- mesurer le nombre d’impulsions du bruit de fond durant t_0 ,
- mesurer le nombre d’impulsions N_1 durant t provenant de bruit de fond et de l’éventuelle source,
- évaluer le comptage net $N = N_1 - N_0 \frac{t}{t_0}$ et décider qu’il y a effectivement une source si ce compte net est supérieur à un certain niveau.

C’est une application de la théorie des tests d’hypothèses avec stratégies à niveau critique développée à la section 3. Ce cadre s’inscrit donc très naturellement dans la TBDS.

10.1.2 Théorie

Supposons que la source recherchée est d’activité a et que le système de comptage détecte les désintégrations avec une sensibilité ϵ , qui dépend de la géométrie source-détecteur et des caractéristiques de ce dernier. Si b est le taux moyen de bruit de fond exprimé dans l’espace des états, les distributions de N_0 et N_1 sont données par les lois de Poisson :

$$P(N_0 | b) = \mathcal{P}(N_0 | bt_0) \quad (10.1)$$

$$P(N_1 | a, b) = \mathcal{P}(N_1 | (a\epsilon + b)t) \quad (10.2)$$

Sous l’hypothèse que la source, le bruit et les deux mesures sont indépendants, la moyenne et la variance du comptage net $N = N_1 - N_0$ sont

$$EN = a\epsilon t + bt - bt_0 \frac{t}{t_0} \quad (10.3)$$

$$= a\epsilon t \quad (10.4)$$

$$VN = (a\epsilon + b)t + bt_0 \left(\frac{t}{t_0} \right)^2 \quad (10.5)$$

Lorsque le nombre d’impulsions est assez grand, il est possible d’approximer les lois de Poisson par des normales. Il vient

$$P(N | a, b) \simeq \mathcal{N} \left(N | a\epsilon t, a\epsilon t + bt \left(1 + \frac{t}{t_0} \right) \right) \quad (10.6)$$

Lorsque la mesure du bruit de fond seul est assez longue, b peut être remplacé par son estimation $\frac{N_0}{t_0}$, et le problème est réduit à l'estimation de a :

$$P(N | a) \simeq \mathcal{N}\left(N | a\epsilon t, a\epsilon t + N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)\right) \quad (10.7)$$

10.1.3 Définitions des limites L_C et L_D

La première est une *limite de décision* donnée par le niveau de comptage L_C au dessus duquel l'observateur ne peut plus admettre qu'il n'y a que du bruit de fond. Cette limite est déterminée d'après la donnée d'une probabilité α qu'un bruit de fond soit considéré à tort comme contenant une source. On reconnaît ici l'erreur de premier type (cf 3.1). Comme la répartition du comptage est connue via l'équ.(10.6), le niveau critique L_C associé à α est donné par

$$P(N > L_C | a = 0) = \alpha \quad (10.8)$$

où $| a = 0$ signifie *étant donné qu'il n'y a pas de source*

Ensuite, il est aisément de calculer la probabilité β de commettre l'erreur de second type, consistant à décider qu'il n'y a pas de source, alors qu'il en a une d'activité $a > 0$:

$$P(N \leq L_C | a) = \beta \quad (10.9)$$

Généralement, si a augmente, β diminue. La donnée d'une probabilité de non détection β et d'une limite de décision L_C conduit par la relation (10.9) à une activité minimale a_D au-dessous de laquelle le système de mesure ne peut plus être considéré comme fiable. La *limite de détection* L_D est définie par rapport à cette activité :

$$L_D = a_D \epsilon t \quad (10.10)$$

Ainsi, la signification de la limite L_D est la suivante : si une source radioactive génère un comptage moyen au-dessous de L_D , la stratégie définie par le niveau critique L_C manque la source avec une probabilité plus grande que β . L_D est une caractéristique du processus de mesure. Ces limites sont illustrées à la figure 10.1.

Nous voyons immédiatement la différence conceptuelle entre les deux limites. L_C est lié au domaine des comptages, c'est-à-dire des observations, alors que L_D est en fait dérivée de a_D , qui se trouve dans le domaine des activités, autrement dit dans l'espace des états. La confusion entre les espaces des états et des observations ouvre la porte à des erreurs d'interprétation. Nous verrons plus bas que L_C est plus petit que L_D . Une question souvent posée est alors la suivante : *Le système de mesure admet une limite de détection L_D , alors, comment se fait-il que l'on puisse décider de la présence d'une source sur une limite L_C qui est plus basse, en particulier si le comptage net se trouve entre les deux limites ?*

10.1.4 Valeurs des limites

Si Φ désigne la cumulative de la loi normale réduite

$$k_\alpha = \Phi^{-1}(1 - \alpha) \quad (10.11)$$

$$k_\beta = \Phi^{-1}(1 - \beta) \quad (10.12)$$

la relation (10.7) conduit aux expressions suivantes :

$$L_C = k_\alpha \sqrt{N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0} \right)} \quad (10.13)$$

$$L_D = L_C + \frac{k_\beta^2}{2} \pm \frac{k_\beta}{2} \sqrt{4L_C + k_\beta^2 + 4 \left(\frac{L_C}{k_\alpha} \right)^2} \quad (10.14)$$

Le calcul est détaillé à la section A.19.

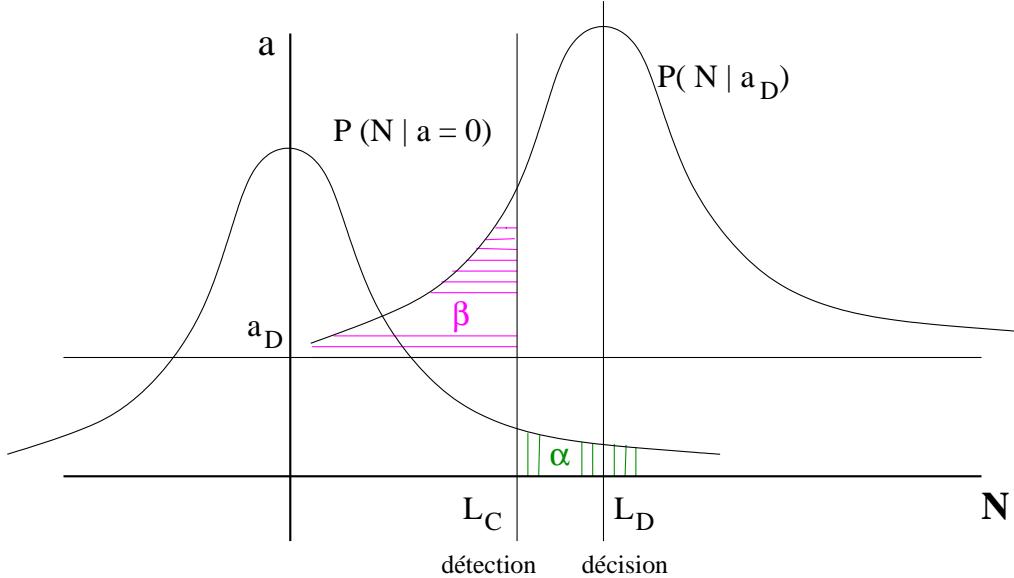


FIG. 10.1 – Limites de détection. La répartition des comptages pour une source d'activité minimale détectable $P(N | a_D)$ ainsi que celle sans source $P(N | a = 0)$ ont été rabattues sur le plan produit des observations (comptages) par les états (activités). Les surfaces des zones hachurées correspondent aux probabilités d'erreur de premier et second types

La relation (10.14) donne deux solutions pour L_D . Généralement, le signe moins peut être éliminé afin d'obtenir un L_D positif. Dans certains cas rarissimes, les valeurs avec les deux signes sont positives et il faut décider en faisant intervenir d'autres contraintes.

Si le signe positif est retenu, L_C est plus bas que L_D . Un paradoxe issu de la confusion des espaces des états et des observations apparaît lorsque le comptage net se trouve entre les deux limites. L'observateur doit conclure à la présence d'une source alors que l'observation est au dessous de la limite de détection ! Cette apparente contradiction ne peut se comprendre que dans le contexte de la TBDS.

On notera encore que le développement effectué ci-dessus repose sur deux hypothèses : d'une part que l'on puisse approximer les lois poissonniennes par des normales, ce qui n'est pas vérifié aux faibles comptages, d'autre part que le temps de mesure du bruit de fond soit suffisamment long pour pouvoir le remplacer par son estimation.

10.2 Intervalles de confiance

10.2.1 Définition

Repartons de la relation (10.7) décrivant le modèle du comptage pour une activité a . Habituellement, le résultat d'une telle mesure s'exprime par la valeur moyenne et une incertitude-type multipliée par un facteur d'élargissement k :

$$N = a\epsilon t \pm k \sqrt{a\epsilon t + N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)} \quad (10.15)$$

En pratique, cette relation est alors résolue en isolant le premier terme $a\epsilon t$ et en remplaçant le second terme $a\epsilon t$ se trouvant à l'intérieur de la racine par l'observation correspondante N :

$$a\epsilon t \simeq N \pm k \sqrt{N + N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)} \quad (10.16)$$

Cette relation se simplifie en exprimant les différentes grandeurs par leurs taux mesurés :

- net : $R = N/t$
- brut : $R_1 = N_1/t$
- zéro : $R_0 = N_0/t_0$

L'estimation de a prend alors la forme couramment rencontrée dans la littérature :

$$a = \frac{1}{\epsilon} \left(R \pm k \sqrt{\frac{R_1}{t} + \frac{R_0}{t_0}} \right) \quad (10.17)$$

qui donne une valeur pour l'activité avec une incertitude élargie. Si nous prenons pour k la valeur $k_{1-\gamma}$, nous obtenons un intervalle de confiance γ pour l'activité présumée a . En pratique, γ est souvent fixé à 95% ce qui donne $k \simeq 2$. Cette substitution d'une variable d'observation par une variable d'état conduit, comme nous le verrons plus bas, à une notion de confiance qui n'est *pas* une probabilité.

10.2.2 Problèmes

Lorsque l'activité de la source est basse, l'expression (10.17) peut donner des valeurs négatives pour a . La solution traditionnelle consiste à poser $a = 0$ sans autre forme de procès ...

La situation est mieux comprise avec le diagramme bidimensionnel de la figure (10.2). Soit maintenant N le comptage brut obtenu en présence d'une source d'activité a et d'un bruit de taux b connu. Pour chaque activité a , nous avons reporté sur la parallèle aux abscisses les limites sur les comptages pour une confiance γ . Lorsque a varie, cet intervalle dit N -intervalle génère un domaine nommé *bande de confiance*. Le N -intervalle est bien défini par le modèle, et donne une probabilité γ d'y mesurer un comptage. Cette bande de confiance peut alors être tranchée par une parallèle aux ordonnées, à la valeur observée de comptage N . Cet a -intervalle est nommé intervalle de confiance pour l'activité. Mais *rien ne prouve que l'activité réelle ait effectivement une probabilité γ de s'y trouver !*

L'observateur peut très bien obtenir un comptage pour lequel le a -intervalle est vide. Cette possibilité a conduit certains auteurs à modifier la définition du N -intervalle, comme nous allons l'expliquer brièvement. Nous proposerons ensuite une solution bayesienne.

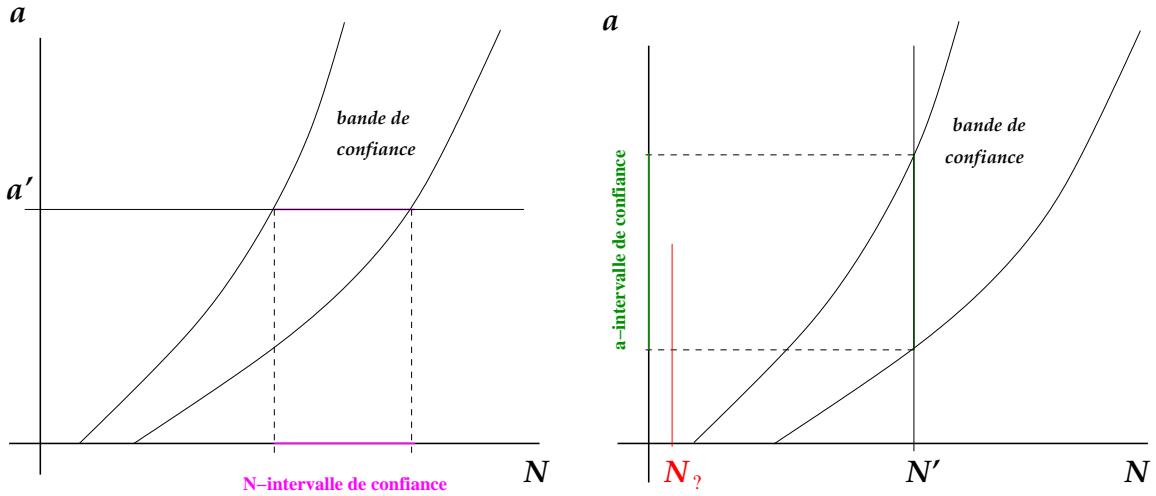


FIG. 10.2 – Bande de confiance. *Les intervalles de confiance sont de type N ou a selon que $a = a'$ ou $N = N'$ est fixé. L'obtention d'un comptage $N_?$ trop bas peut impliquer un intervalle vide pour a .*

10.2.3 Intervalles de Feldman-Cousins

Dans leur publication [33], Gary J. Feldman and Robert D. Cousins donnent une définition des intervalles de confiance dans le cas particulier du modèle poissonnien :

$$P(N | a) = \mathcal{P}(N | (a\epsilon + b)t) \quad (10.18)$$

La première constatation est que les intervalles pour une confiance donnée ne sont pas uniques. Avec une approximation gaussienne, ils sont d'habitude choisis centrés. Si la distribution n'est pas symétrique, on prend souvent l'intervalle de longueur euclidienne minimale ayant la confiance donnée. Mais le problème des a -intervalles vides demeure.

Feldman et Cousins donnent la stratégie suivante pour construire un intervalle résolvant le problème. Soit a_0 l'activité pour laquelle un N -intervalle de confiance γ est désiré. La loi de Poisson étant discrète, nous pouvons le construire par accrétion. Pour chaque N , considérons la fonction de vraisemblance $a \rightarrow P(N | a)$ et son maximum $a_{best}(N)$. Puis définissons le rapport

$$R(N) = \frac{P(N | a_0)}{P(N | a_{best}(N))}$$

et classons les N par R décroissants. L'intervalle est ensuite construit en prenant les N l'un après l'autre et additionnant leurs probabilités jusqu'à l'obtention de la confiance γ .

Cette idée a immédiatement été acceptée par le *Particle Data Group* [34], et rapidement, une autre proposition a été faite par C. Giunti [35], tirant le processus dans le domaine bayesien.

10.2.4 Intervalles de Giunti

C. Giunti utilise le théorème de Bayes pour obtenir une répartition a posteriori pour a étant donné N , sous un a priori uniforme $\pi(da) = da$. Nous avons

$$\pi(da | N) = \frac{\mathcal{P}(N | (a\epsilon + b)t)da}{M(N)} \quad (10.19)$$

$$M(N) = \int_0^\infty \mathcal{P}(N | (a\epsilon + b)t)da \quad (10.20)$$

Un rapide calcul donne

$$\pi(da | N) = ((a\epsilon + b)t)^N e^{-a\epsilon t} \left(N! \sum_{k=0}^N \frac{(bt)^k}{k!} \right)^{-1} da \quad (10.21)$$

Il définit ensuite a_{ref} comme moyenne de a a posteriori

$$a_{ref} = \int_0^\infty a \pi(da | N) \quad (10.22)$$

$$= \frac{\sum_{k=0}^N \frac{(bt)^k (N-k+1)}{k!}}{\sum_{k=0}^N \frac{(bt)^k}{k!}} \quad (10.23)$$

puis construit son intervalle de la même manière que Feldman-Cousins, en remplaçant a_{best} par a_{ref} .

10.2.5 Solution bayesienne

L'approche de Giunti incite à utiliser directement la TBDS. En effet, les intervalles de crédibilité bayesiens (cf 2.21) donnent automatiquement des bandes de confiance dont les tranches à a constant ne sont pas vides. Au lieu de l'a priori uniforme, prenons celui de Jeffreys :

$$\pi_J(da) = (a\epsilon t + b)^{-\frac{1}{2}} da \quad (10.24)$$

qui induit l'a posteriori¹ dont les formes différentielle et cumulative sont les suivantes :

$$\pi_J(da | N) = \frac{((a\epsilon + b)t)^{N-\frac{1}{2}} e^{-(a\epsilon + b)t} da}{\Gamma(N + \frac{1}{2}, bt)} \quad (10.25)$$

$$\pi_J([0, a] | N) = 1 - \frac{\Gamma(N + \frac{1}{2}, (a\epsilon + b)t)}{\Gamma(N + \frac{1}{2}, bt)} \quad (10.26)$$

Les intervalles de crédibilité γ satisfont à

$$\pi_J([a_1, a_2] | N) = \gamma \quad (10.27)$$

On choisira généralement ceux dont la longueur $a_2 - a_1$ est minimale. La figure 10.3 montre ces divers intervalles.

¹ $\Gamma(N + \frac{1}{2}, b) = \int_b^\infty x^{n-\frac{1}{2}} e^{-x} dx$ est une loi gamma incomplète ([36] 6.5.3)

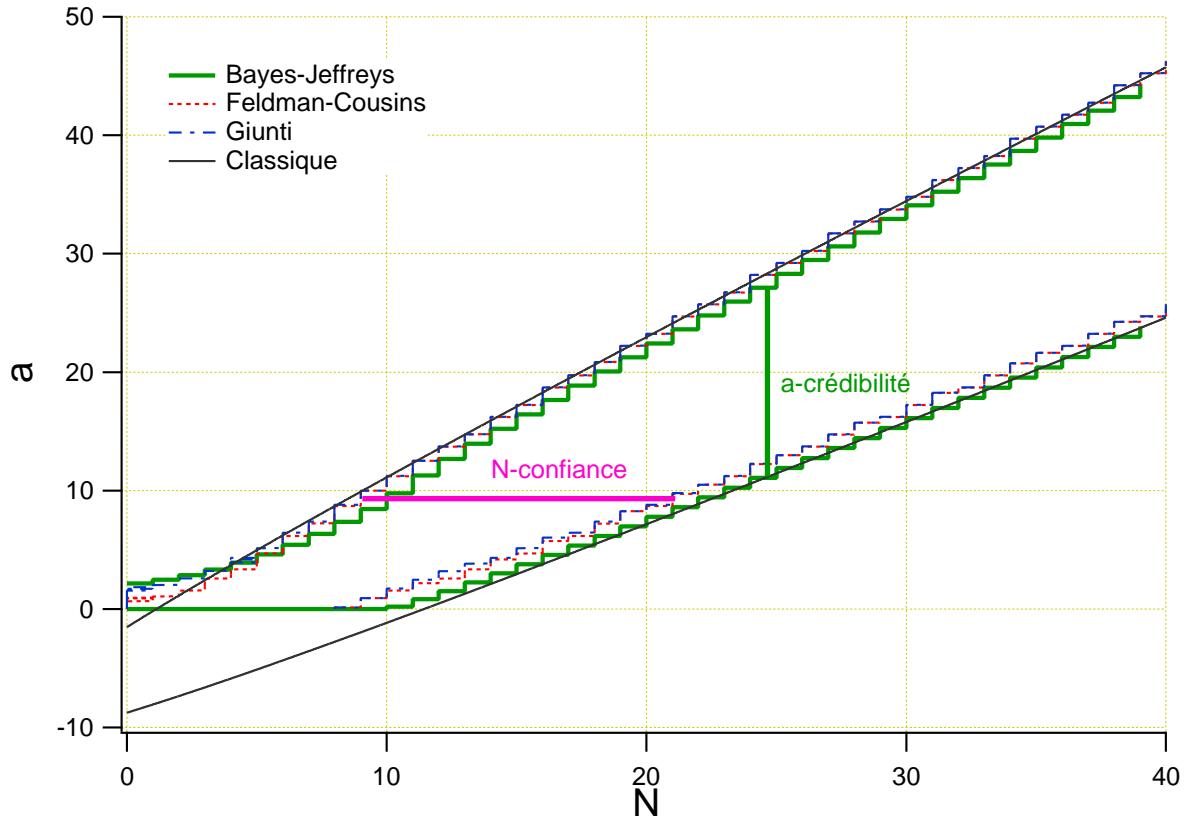


FIG. 10.3 – Intervalles de confiance à 90% de Feldman-Cousins, Giunti et Bayes-Jeffreys. Le taux de bruit est $b = 5.3$, le temps de mesure $t = 1$ et l'efficacité $\epsilon = 1$. Les a -intervalles non classiques ne sont jamais vides, même pour de petites valeurs de N . De plus, ils n'ont pas partie négative.

10.3 Introduction de l'incertitude sur la durée de mesure et l'efficacité dans un modèle poissonnien

Le schéma général pour la prise en compte d'incertitudes sur certains paramètres consiste à les introduire comme composantes de l'état. Ces nouvelles composantes prennent le nom de *paramètres de nuisance*. Les répartitions associées à ces nouvelles composantes seront choisies de sorte que leurs moyenne et écart-type correspondent aux données de mesure. Le modèle final sera obtenu ensuite par intégration.

10.3.1 Construction du modèle

Partant du modèle de Poisson

$$P(N | a) = \mathcal{P}(N | a\epsilon t) \quad (10.28)$$

nous allons supposer que la durée de mesure t et la sensibilité ϵ sont données d'une manière classique par leurs valeurs et des incertitudes exprimées par un écart-type

$$t = \bar{t} \pm \sigma_t \quad (10.29)$$

$$\epsilon = \bar{\epsilon} \pm \sigma_\epsilon \quad (10.30)$$

L'idée consiste à ajouter t et ε comme composantes à l'état. Le nouvel espace devient donc l'ensemble des triplets (a, t, ε) . La relation (10.29) est traduite par une répartition pour t dont l'espérance sera \bar{t} et la variance σ_t^2 . Comme t est une variable essentiellement positive, un choix naturel est une loi gamma :

$$\pi(dt) = \text{Ga}(dt | N_t, a_t) \quad (10.31)$$

$$N_t = \frac{\bar{t}^2}{\sigma_t^2} \quad (10.32)$$

$$a_t = \frac{\bar{t}}{\sigma_t^2} \quad (10.33)$$

dont les paramètres N_t et a_t donnent l'espérance et l'écart-type indiqués (cf 5.18).

Quant à l'efficacité ε , c'est une probabilité de détection comprise entre 0 et 1. Ici, le choix naturel est une loi bêta (cf 5.32) :

$$\pi(d\varepsilon) = \text{Be}(d\varepsilon | n_\varepsilon, \bar{n}_\varepsilon) \quad (10.34)$$

Si l'on pose $n = n_\varepsilon + \bar{n}_\varepsilon$ alors l'espérance et l'écart-type donnés en (10.30) sont obtenus par le choix des paramètres suivants :

$$n = \frac{\bar{\varepsilon}(1 - \bar{\varepsilon})}{\sigma_\varepsilon^2} - 1 \quad (10.35)$$

$$n_\varepsilon = \bar{\varepsilon}n \quad (10.36)$$

$$\bar{n}_\varepsilon = (1 - \bar{\varepsilon})n \quad (10.37)$$

Supposons encore les trois composantes indépendantes, le modèle généralisé prend la forme :

$$P(N | a, t, \varepsilon) = \mathcal{P}(N | a\varepsilon t) \pi(dt) \pi(d\varepsilon) = \mathcal{P}(N | a\varepsilon t) \text{Ga}(dt | N_t, a_t) \text{Be}(d\varepsilon | n_\varepsilon, \bar{n}_\varepsilon) \quad (10.38)$$

Comme la composante intéressante est a , il suffit d'intégrer sur t et ε :²

$$P(N | a) = \int_{\varepsilon, t} \mathcal{P}(N | a\varepsilon t) \text{Ga}(dt | N_t, a_t) \text{Be}(d\varepsilon | n_\varepsilon, \bar{n}_\varepsilon) \quad (10.39)$$

10.3.2 Etude du modèle généralisé, moments a posteriori

Le modèle obtenu peut se simplifier en utilisant le fait que la succession d'une loi de Poisson et d'une loi gamma donne une loi négative-binomiale (cf A.2). On obtient

$$P(N | a) = \int_\varepsilon \mathcal{NB}\left(N | N_t, \frac{a_t}{a_t + a\varepsilon}\right) \text{Be}(d\varepsilon | n_\varepsilon, \bar{n}_\varepsilon) \quad (10.40)$$

Supposons maintenant que l'a priori sur l'activité soit $\text{Ga}(da | N_a, t_a)$. L'obtention des moments de l'activité a posteriori met en scène les intégrales

$$M_m = \int_{a, \varepsilon} a^m \text{Ga}(da | N_a, t_a) \mathcal{NB}\left(N | N_t, \frac{a_t}{a_t + a\varepsilon}\right) \text{Be}(d\varepsilon | n_\varepsilon, \bar{n}_\varepsilon) \quad (10.41)$$

²On dit que l'on marginalise t et ε .

Introduisons la *fonction de Kummer de deuxième espèce*³ U ([36], p505, 13.2.5). Un calcul simple montre que

$$\begin{aligned} \int da a^m \text{Ga}(da | N_a, t_a) \mathcal{N}(\mathcal{B} \left(N | N_t, \frac{a_t}{a_t + a\epsilon} \right)) = \\ \frac{\Gamma(N_t + N)}{\Gamma(N_t)N!} \frac{\Gamma(m + N_a + N)}{\Gamma(N_a)} \left(\frac{a_t t_a}{\epsilon} \right)^{N_a} \left(\frac{a_t}{\epsilon} \right)^m \\ \cdot U \left(m + N_a + N, m + N_a - N_t + 1, \frac{a_t t_a}{\epsilon} \right) \end{aligned} \quad (10.42)$$

Les moments (non normalisés) peuvent se calculer en une seule quadrature

$$\begin{aligned} M_m = \frac{\Gamma(N_t + N)}{\Gamma(N_t)N!} \frac{\Gamma(m + N_a + N)}{\Gamma(N_a)} \int_{\epsilon} \text{Be}(d\epsilon | n_{\epsilon}, \bar{n}_{\epsilon}) \\ \cdot \left(\frac{a_t t_a}{\epsilon} \right)^{N_a} \left(\frac{a_t}{\epsilon} \right)^m \cdot U \left(m + N_a + N, m + N_a - N_t + 1, \frac{a_t t_a}{\epsilon} \right) \end{aligned} \quad (10.43)$$

En particulier, l'espérance et la variance de l'activité a posteriori s'expriment par

$$\mathbf{E}a | N = M_1 / M_0 \quad (10.44)$$

$$\mathbf{V}a | N = M_2 / M_0 - (M_1 / M_0)^2 \quad (10.45)$$

On voit que dans cette application existe une formulation analytique, ce qui est très favorables au calcul.

10.3.3 Comparaison avec le calcul classique

Le calcul classique s'effectue par division et propagation des erreurs

$$a = \frac{N}{\epsilon t} \quad (10.46)$$

$$\frac{\sigma_a}{a} = \sqrt{\frac{N}{N^2} + \left(\frac{\sigma_t}{t} \right)^2 + \left(\frac{\sigma_{\epsilon}}{\epsilon} \right)^2} \quad (10.47)$$

Une comparaison des deux estimations de l'activité a été faite dans le cas particulier suivant :

Paramètre	Valeur ± 1 incertitude-type
Sensibilité ϵ	0.7 ± 0.1
Durée de mesure t	5 ± 0.5

Une simulation partant d'une activité effective $a_{eff} = 4$ a donné 1000 comptages selon le modèle généralisé. Dans chaque cas, cette activité a été estimée de manière classique, puis bayesienne en partant de deux-écart types gamma, le premier dit large correspondant à une activité à priori 4.55 ± 4.55 , l'autre dit pointu correspondant à 4.17 ± 0.83 .⁴

Pour chaque estimation a de l'activité et de son incertitude s_a , il est possible d'obtenir un écart à la valeur effective relatif et normalisé défini par $\frac{a - a_{eff}}{s_a}$ qui devrait suivre approximativement une loi normale réduite dans le cas classique. La figure 10.4 montre les histogrammes de ces écarts. On remarque le resserrement de l'estimation bayesienne autour de la valeur effective pour l'a priori pointu, ainsi que la similitude avec l'estimation classique pour l'a priori large.

³Dite aussi *fonction de Tricomi*, de *Gordon* ou encore *confluente hypergéométrique*

⁴L'a priori large correspond grossièrement à une attente de la première détection de 4.55, alors que le pointu serait issu d'un temps d'attente de 21 pour la cinquième détection.

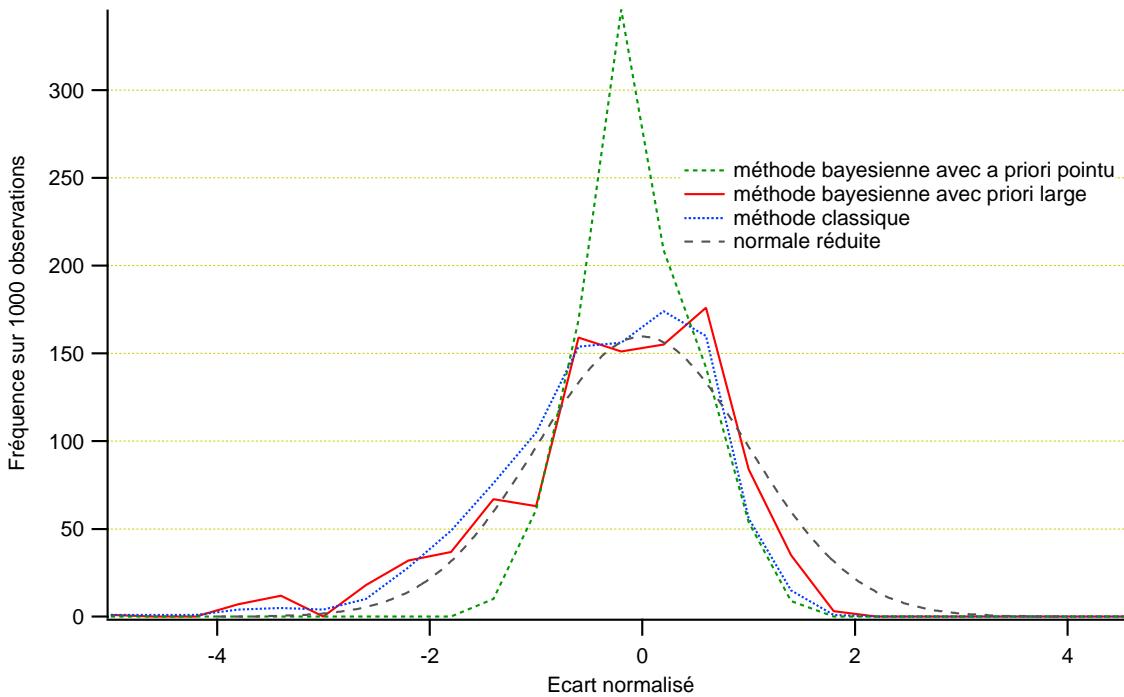


FIG. 10.4 – Fréquences des écarts à l’activité effective normalisés pour deux estimateurs bayésiens et un estimateur classique. *L’estimation classique est équivalente à une bayésienne avec a priori large. Un a priori plus pointu resserre les écarts autour de la valeur effective*

Remarque Le calcul est effectué par un code C utilisant la librairie GSL [37]. Cet exemple est à la limite de précision de la machine (double). Pour des activités et temps plus longs, il faudrait notamment accroître la précision. Ceci serait possible en utilisant des librairies de calcul en précision arbitraire ([38] par exemple). Ceci suppose aussi une réécriture des algorithmes d’intégration numérique.

Chapitre 11

Mesures multicanales à activité constante

La méthode de mesure d'une source radioactive en présence d'un rayonnement ambiant décrite en 10.1.1 peut être généralisée à des observations contenant simultanément plusieurs comptages. En spectro-métrie γ , il est possible d'évaluer les nombres d'impulsions récoltées dans des plages d'énergie disjointes du spectre. Ces plages sont généralement situées aux voisinages des pics produits par les nucléides présents dans la source. Le but de cette section est d'utiliser la TBDS pour tenir compte de cette information, qui devient vectorielle.

11.1 Etablissement du modèle

11.1.1 Comptages produits par la source

Soient donc d plages d'énergie disjointes $[E_k, E_{k+1}]_{k=1,d}$. Un détecteur admet dans chacune de ces plages une sensibilité ε_k qui correspond à la probabilité d'obtenir une détection dans la plage $[E_k, E_{k+1}]$ pour d'une désintégration dans la source. Remarquons que cette sensibilité est définie pour un nuclide donné, et non pas, comme c'est le cas habituellement, pour un photon d'une certaine énergie.

Supposons que la source ait produit N_E désintégrations indépendantes lors d'une mesure de durée t . Ajoutons encore une plage virtuelle 0 qui accumule les désintégrations non détectées dans une des plages réelles¹. La sensibilité associée à cette plage virtuelle est évidemment

$$\varepsilon_0 = 1 - \sum_{k>0} \varepsilon_k \quad (11.1)$$

Comme chaque désintégration produit une incrémentation dans une et une seule plage, la répartition des comptages $N^A = (N_0^A, N_1^A, \dots, N_d^A)$ est multinomiale :

$$P(N^A | N_E) = \frac{N_E!}{N_0^A! N_1^A! \dots N_d^A!} \varepsilon_0^{N_0^A} \varepsilon_1^{N_1^A} \dots \varepsilon_d^{N_d^A} \quad (11.2)$$

Mais si l'activité de la source est a , N_E est réparti selon la poissonnienne $\mathcal{P}(N_E | at)$ et

$$P(N^A | a) = \sum_{N_E} \mathcal{P}(N_E | at) \frac{N_E!}{N_0^A! N_1^A! \dots N_d^A!} \varepsilon_0^{N_0^A} \varepsilon_1^{N_1^A} \dots \varepsilon_d^{N_d^A} \quad (11.3)$$

¹Cette plage n'est évidemment pas mesurable, mais simplifie le calcul.

qui peut se réécrire en développant les poissonniennes

$$P(N^A | a) = \sum_{N_0^A} \mathcal{P}(N_0^A | a\epsilon_0) \prod_{k>0} \mathcal{P}(N_k^A | a\epsilon_k) \quad (11.4)$$

$$= \prod_{k>0} \mathcal{P}(N_k^A | a\epsilon_k) \quad (11.5)$$

prouvant que les N_k^A sont indépendants, chacun suivant une répartition de Poisson de paramètre $a\epsilon_k$.

Il est intéressant de remarquer que les canaux sont rendus indépendants par le fait que l'on mesure durant un temps fixé.

11.1.2 Comptages produits par le bruit de fond

Ici, nous ferons l'hypothèse que le rayonnement ambiant génère dans chaque plage des comptages indépendants et poissonniens. A chaque canal k est donc associé un taux de bruit b_k qui est mesuré durant un temps t_k . Supposons que les comptages pour la mesure du bruit soient $M = (M_1 \dots M_d)$, désignant par b le vecteur des taux de bruit $(b_1 \dots b_d)$, il vient pour la mesure sans source

$$P(M | b) = \prod_{k>0} \mathcal{P}(M_k | b_k t_k) \quad (11.6)$$

Soient enfin $N^B = (N_1^B \dots N_d^B)$ les comptages produits par le bruit lors de la mesure avec source. Le temps de mesure est ici évidemment le même pour toutes les plages et l'on obtient :

$$P(N^B | b) = \prod_{k>0} \mathcal{P}(N_k^B | b_k t_k) \quad (11.7)$$

11.1.3 Modèle global

Le bruit et la source sont supposés indépendants. Soit $N = N^A + N^B$. Tenant compte de l'additivité des lois de Poisson, il vient

$$P(N) = \prod_{k>0} \mathcal{P}(N_k | (a\epsilon_k + b_k)t) \quad (11.8)$$

L'hypothèse d'indépendance des deux mesures donne alors

– l'espace des états

$$\Theta = \mathbb{R}_+ \times \mathbb{R}_+^d \ni (a, b) \quad (11.9)$$

– l'espace des observations

$$\mathcal{X} = \mathbb{N}^d \times \mathbb{N}^d \ni (N, M) \quad (11.10)$$

– le modèle

$$P(N, M | a, b) = \prod_{k>0} [\mathcal{P}(N_k | (a\epsilon_k + b_k)t) \mathcal{P}(M_k | b_k t_k)] \quad (11.11)$$

11.1.4 A priori et a posteriori de Jeffreys

Un calcul simple montre que seules la première ligne, la première colonne et la diagonale de la matrice d'information de Fisher sont non nulles

$$I(a, b) = \begin{pmatrix} \sum_k \frac{\varepsilon_k^2 t}{a\varepsilon_k + b_k} & \frac{\varepsilon_1 t}{a\varepsilon_1 + b_1} & \dots & \frac{\varepsilon_k t}{a\varepsilon_k + b_k} & \dots \\ \frac{\varepsilon_1 t}{a\varepsilon_1 + b_1} & \frac{t}{a\varepsilon_1 + b_1} + \frac{t_1}{b_1} & 0 & \dots & \dots \\ \frac{\varepsilon_k t}{a\varepsilon_k + b_k} & 0 & \dots & 0 & \frac{t}{a\varepsilon_k + b_k} + \frac{t_k}{b_k} \\ \dots & \dots & \dots & \dots & 0 & \dots \end{pmatrix} \quad (11.12)$$

ou encore, en appliquant les traditionnelles recettes pour le calcul des déterminants :

$$|I(a, b)| = \begin{vmatrix} 0 & -\frac{\varepsilon_1 t_1}{b_1} & \dots & -\frac{\varepsilon_k t_k}{b_k} & \dots \\ \frac{\varepsilon_1 t}{a\varepsilon_1 + b_1} & \frac{t}{a\varepsilon_1 + b_1} + \frac{t_1}{b_1} & 0 & \dots & \dots \\ \frac{\varepsilon_k t}{a\varepsilon_k + b_k} & 0 & \dots & 0 & \frac{t}{a\varepsilon_k + b_k} + \frac{t_k}{b_k} \\ \dots & \dots & \dots & \dots & 0 & \dots \end{vmatrix} \quad (11.13)$$

11.1.4.1 Cas particulier de la mesure dans une seule plage

Lorsque $d = 1$, le déterminant (11.13) est de dimension 2 (compte tenu de la mesure du bruit seul) et prend une valeur particulièrement simple :

$$|I_1(a, b_1)| = \frac{\varepsilon_1^2 t_1 t}{(a\varepsilon_1 + b_1)b_1} \quad (11.14)$$

Sous la priori de Jeffreys qui est la racine de cette dernière expression, le théorème de Bayes donne l'a posteriori

$$\pi_J(da, db_1 | N_1, M_1) \propto (a\varepsilon_1 + b_1)^{N_1 - \frac{1}{2}} b_1^{M_1 - \frac{1}{2}} e^{-a\varepsilon_1 t - b_1(t_1 + t)} da db_1 \quad (11.15)$$

Introduisons les moments d'ordre n pour la source et m pour le bruit (non normés) et supprimons les indices pour simplifier la notation sauf pour la durée de mesure du bruit (t_1)

$$Mo_{nm}(N, M, \varepsilon, t, t_1) = \int a^n b^m \pi_J(da, db | N, M) \quad (11.16)$$

Les moments normés a posteriori s'obtiennent par

$$Ea^n b^m = \frac{Mo_{nm}(N, M, \varepsilon, t, t_1)}{Mo_{00}(N, M, \varepsilon, t, t_1)} \quad (11.17)$$

et permettent de calculer les moyennes, variances et covariances. Il se trouve que ces moments ont la forme analytique suivante

$$\begin{aligned}
 Mo_{nm}(N, M, \varepsilon, t, t_1) = & \frac{1}{\Gamma(\frac{1}{2} - N)} \left[\right. \\
 & \Gamma(\frac{1}{2} + m + M) \Gamma(\frac{1}{2} - N) \Gamma(\frac{1}{2} + n + N) \\
 & {}_2F_1(\frac{1}{2} + m + M, \frac{1}{2} - N, \frac{1}{2} - n - N, \frac{t}{t + t_1}) \\
 & \varepsilon^{-\frac{1}{2} + N} (\varepsilon t)^{-\frac{1}{2} - n - N} (t + t_1)^{-\frac{1}{2} - m - M} \\
 & + \Gamma(1 + n) \Gamma(-\frac{1}{2} - n - N) \Gamma(1 + m + M + n + N) \\
 & {}_2F_1(1 + m + M + n + N, 1 + n, \frac{3}{2} + n + N, \frac{t}{t + t_1}) \\
 & \left. \varepsilon^{-1 - n} (t + t_1)^{-1 - m - M - n - N} \right] \quad (11.18)
 \end{aligned}$$

Cette élégante formule (sortie par le logiciel Mathematica via la transformation de Laplace) fait intervenir la fonction hypergéométrique ${}_2F_1$ [36]. Une simulation Monte Carlo a permis de vérifier cette expression dans quelques cas simples.

11.1.4.2 Mesure d'activité réalisée à l'aide de plusieurs plages

Dans le cas multidimensionnel, le résultat obtenu à la section A.12 donne, en remplaçant η_k par $\frac{1}{\varepsilon_k}$, α_k par $\frac{\varepsilon_k t}{a\varepsilon_k + b_k}$ et β_k par $\frac{\varepsilon_k t_k}{b_k}$:

$$|I(a, b)| = \sum_k \frac{\varepsilon_k^2}{\left(\frac{b_k}{t_k} + \frac{a\varepsilon_k + b_k}{t}\right)} \cdot \prod_k \frac{t_k t}{(a\varepsilon_k + b_k) b_k} \left(\frac{b_k}{t_k} + \frac{a\varepsilon_k + b_k}{t}\right) \quad (11.19)$$

L'a posteriori de Jeffreys devient

$$\begin{aligned}
 \pi_J(da, db \mid N, M) \propto & da db \sqrt{\sum_k \frac{\varepsilon_k^2}{\frac{a\varepsilon_k + b_k}{t} + \frac{b_k}{t_k}}} \cdot \\
 & \cdot \prod_k \sqrt{\frac{a\varepsilon_k + b_k}{t} + \frac{b_k}{t_k}} (a\varepsilon_k + b_k)^{N_k - \frac{1}{2}} e^{-(a\varepsilon_k + b_k)t} b_k^{M_k - \frac{1}{2}} e^{-b_k t_k} \quad (11.20)
 \end{aligned}$$

Les diverses projections sont corrélées car les racines empêchent une décomposition en produit.

Les moments de cet a posteriori peuvent s'obtenir par *Monte Carlo*. Soit

$$u_k = \frac{a\varepsilon_k + b_k}{t} + \frac{b_k}{t_k}$$

Si

$$\begin{aligned}
 Mo_{nm} = & \frac{\Gamma(n+1)}{(t \sum_k \varepsilon_k)^{n+1}} \prod_k \frac{\Gamma(m_k + M_k + \frac{1}{2})}{(t_k + t)^{m_k + M_k + \frac{1}{2}}} \cdot \\
 & \cdot \int \sqrt{\sum_k \frac{\varepsilon_k^2}{u_k}} \cdot \prod_k \sqrt{u_k} (a\varepsilon_k + b_k)^{N_k - \frac{1}{2}} \cdot \\
 & \cdot \text{Ga} \left(da \left| n+1, t \sum_k \varepsilon_k \right. \right) \prod_k \text{Ga} (db_k \mid m_k + M_k + \frac{1}{2}, t_k + t) \quad (11.21)
 \end{aligned}$$

les moments normalisés s'obtiennent par

$$Ea^n b^m = \frac{Mo_{nm}}{Mo_{00}} \quad (11.22)$$

11.2 Application à la mesure d'une source d'europium-152

L'europium-152 est un nucléide à schéma complexe produisant des spectres γ avec de nombreux pics. La méthode utilisant plusieurs plages autour de ses pics dominants a fait l'objet d'une simulation. Le modèle utilisé est donné par la relation (11.11).

11.2.1 Bases de la simulation

Cette simulation est effectuée avec le code Monte Carlo *GEANT 3.21* développé par le *CERN* [39]. L'instrument de mesure simulé est un détecteur au germanium de haute pureté, et les émissions de l'europium-152 lors d'une désintégration sont modélisées par le logiciel *ad-hoc SCH2FOR* [40].

Les calculs impliqués par la TBDS sont généralement assez complexes. La librairie *l_detdec* [41], basée sur la librairie scientifique *GSL* [37] a été développée spécialement dans ce sens. En particulier, la conjonction de plusieurs pics utilise un algorithme optimalisé basé sur la *transformée de Fourier rapide*.

Les comptages seront simulés selon le modèle (11.11) et utilisés pour estimer l'activité de la source. Les courbes ROC relatives à diverses combinaisons de plages seront ensuite comparées.

11.2.2 Détermination des plages et des sensibilités

Les émissions choisies sont les quatre photons de plus grande intensité émis par ce nucléide au-dessus de 100keV. Une simulation de 3.5 millions de désintégrations a été effectuée pour obtenir les sensibilités (ε_k) au voisinage des énergies choisies. Les plages envisagées sont des intervalles de $\pm 3\sigma$ autour de l'énergie des photons émis. La sensibilité dans une telle plage est définie par la probabilité qu'une désintégration y soit détectée. Précisons que toutes les émissions du schéma peuvent être comptées, et non pas seulement les photons effectuant un effet photoélectrique dans le cristal. σ est basé sur la résolution connue du détecteur, qui est exprimée en largeurs à mi-hauteur (FWHM). Les sensibilités obtenues sont données dans le tableau 11.1.

11.2.3 A priori sur les taux de bruit de fond

Les taux de bruit de fond dans ces plages sont issus d'un spectre expérimental (cf tableau 11.2). Les répartitions a priori pour ces taux ont été choisies uniforme sur un intervalle couvrant 20 fois le taux

TAB. 11.1 – Emissions photoniques dominantes de l'europium-152

Energie [keV]	Intensité [% par désint.]	FWHM [keV]	Sensibilité ε_k [%]
121.78	28.4	1.170	4.463
344.28	26.4	1.361	5.233
1408.02	20.8	1.844	0.407
964.13	14.4	1.679	0.413

expérimental. En d'autres termes, si b_{exp} est le taux donné expérimentalement, on aura

$$\pi(db) = \frac{\mathbf{1}_{[0, 20b_{exp}](b)}}{20b_{exp}} db \quad (11.23)$$

TAB. 11.2 – Taux de bruit de fond étendus aux plages

Energie [keV]	Taux [cps]	Largeur de plage [keV]
121.78	.01065	3.227
344.28	.00797	3.723
1408.02	.00145	4.964
964.13	.00229	4.468

11.2.4 A priori sur l'activité

Pour une activité effective a_{eff} de la source, la pratique de ce genre de mesure indique qu'un a priori uniforme couvrant un intervalle allant de l'activité nulle à 10 fois l'activité effective est un bon point de départ. Nous prendrons donc

$$\pi(da) = \frac{\mathbf{1}_{[0, 10a_{eff}](a)}}{10a_{eff}} da \quad (11.24)$$

11.2.5 Calcul des a posteriori

Supposant une activité effective de $1Bq$, nous avons simulé une observation par échantillonnages poissonniens basés sur les taux, l'activité et les sensibilités. Pour des durées de mesure de 500s, les comptages obtenus sont donnés dans le tableau 11.3.

TAB. 11.3 – Comptages simulés pour le modèle (11.11)

Energie [keV]	121.78	344.28	1408.02	964.13
sans source, M	6	1	0	5
avec source, N	23	29	4	3

Les espérances et écarts-types a posteriori obtenus par application du théorème de Bayes au modèle 11.11 sont mis en relation avec leur valeur effective dans le tableau 11.4. Ces valeurs sont en bon accord.

TAB. 11.4 – Estimations a posteriori

A posteriori	Moyenne	Ecart-type	Effectif
bruit 121.78	0.014	0.005	0.011
bruit 344.28	0.004	0.003	0.008
bruit 1408.02	0.002	0.002	0.001
bruit 964.13	0.012	0.005	0.002
activité	0.917	0.152	1.000

11.2.6 Effet de la prise en compte de plusieurs plages sur les courbes ROC

Afin de mettre en évidence l'effet cumulatif des plages, six courbes ROC ont été calculées. Les quatre premières sont relatives aux plages isolées, la cinquième correspond aux quatre plages traitées simultanément, et la sixième est obtenue par fusion des quatre plages, en sommant les taux de bruit de fond et les comptages.

Le tableau 11.5 et la figure 11.1 montrent que la meilleure méthode est bien celle considérant les quatre plages simultanément. La fusion des plages donne un résultat meilleur que chacune d'entre elles. La table indique les surfaces A_z sous les courbes ROC, les risques de Bayes pour une fonction de perte valant 1 pour une erreur et 0 pour un succès (ce qui donne une valeur de 0.5 pour une stratégie aléatoire donnant l'une ou l'autre des hypothèses de manière équiprobable). Les variables (α, β) sont les probabilités d'erreur de type 1 et 2, si l'on applique la stratégie optimale pour cette fonction de perte et une prévalence 50% en faveur de la présence d'une telle source. Notons que cette prévalence n'est nécessaire que pour calculer le risque.

TAB. 11.5 – Qualifications des courbes ROC

Pics	A_z	Risque	α	β
121.782	0.9516	0.1236	0.1294	0.1179
344.282	0.9797	0.0759	0.0760	0.0759
1408.02	0.6817	0.3713	0.3764	0.3663
964.131	0.6531	0.3932	0.4000	0.3864
simultanés	0.9967	0.0274	0.0281	0.0267
fusionnés	0.9920	0.0443	0.0464	0.0421

11.2.7 Conclusion

L'utilisation de plusieurs plages simultanément pour la mesure de l'activité de nucléides à schéma complexe donne de meilleurs résultats qu'en ne considérant qu'une seule plage, même formée de leur fusion. Les activités estimées sont mieux résolues, les risques moins grands et les probabilités d'erreur plus

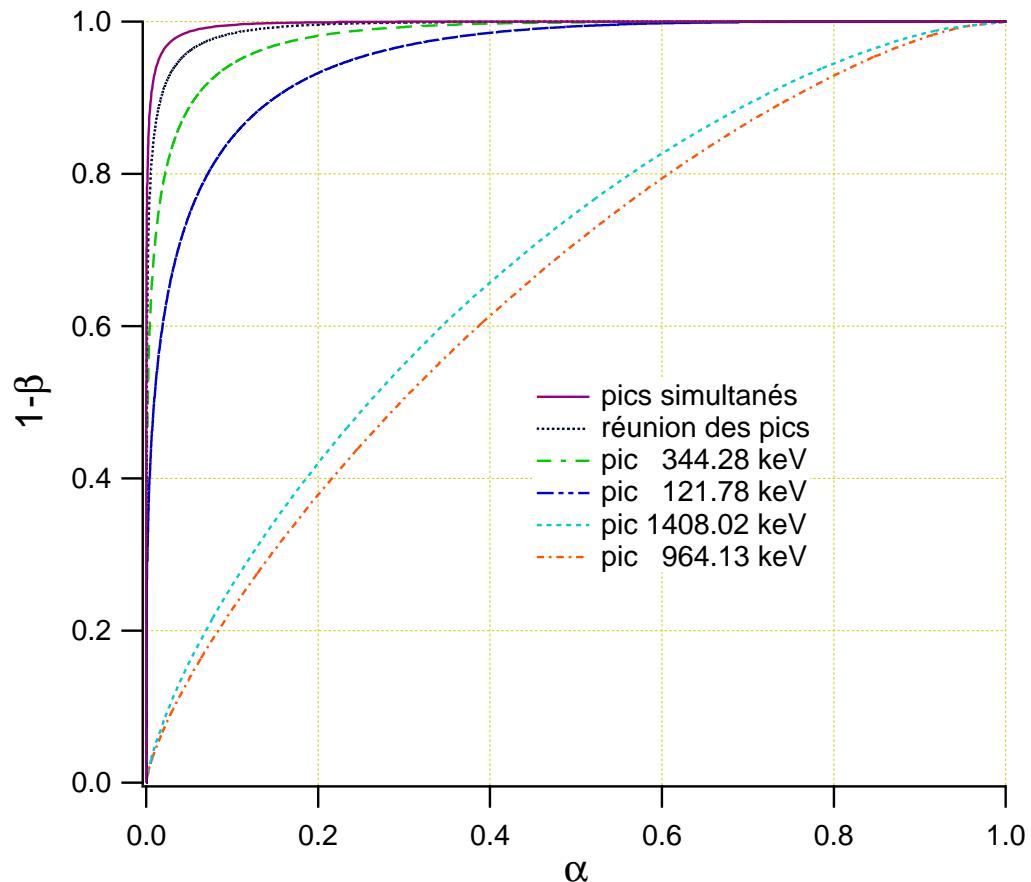


FIG. 11.1 – Courbes ROC relatives à la détection de l'europium-152

faibles. Néanmoins, le calcul se passe difficilement d'une machine et de quelques puissants algorithmes numériques.

11.3 Mesure d'uranifères dans l'environnement

La détection d'uranifères doit parfois se faire sur des activités très basses. Certains de ces nucléides ont un schéma de désintégration complexe qui produit un spectre où plusieurs pics apparaissent. C'est par exemple le cas de l'uranium-235 qui sera traité à la section suivante en appliquant le modèle 11.11. D'autre part, la détermination de l'activité d'un nucléide peut se faire par celle de ses filles. Si ces dernières sont en équilibre avec la mère, la connaissance de leurs activités peut donner un a priori pour l'estimation de celle de la mère. Cette approche sera illustrée par un exemple relatif au thorium-230.

11.3.1 Détermination de l'activité de l'uranium-235

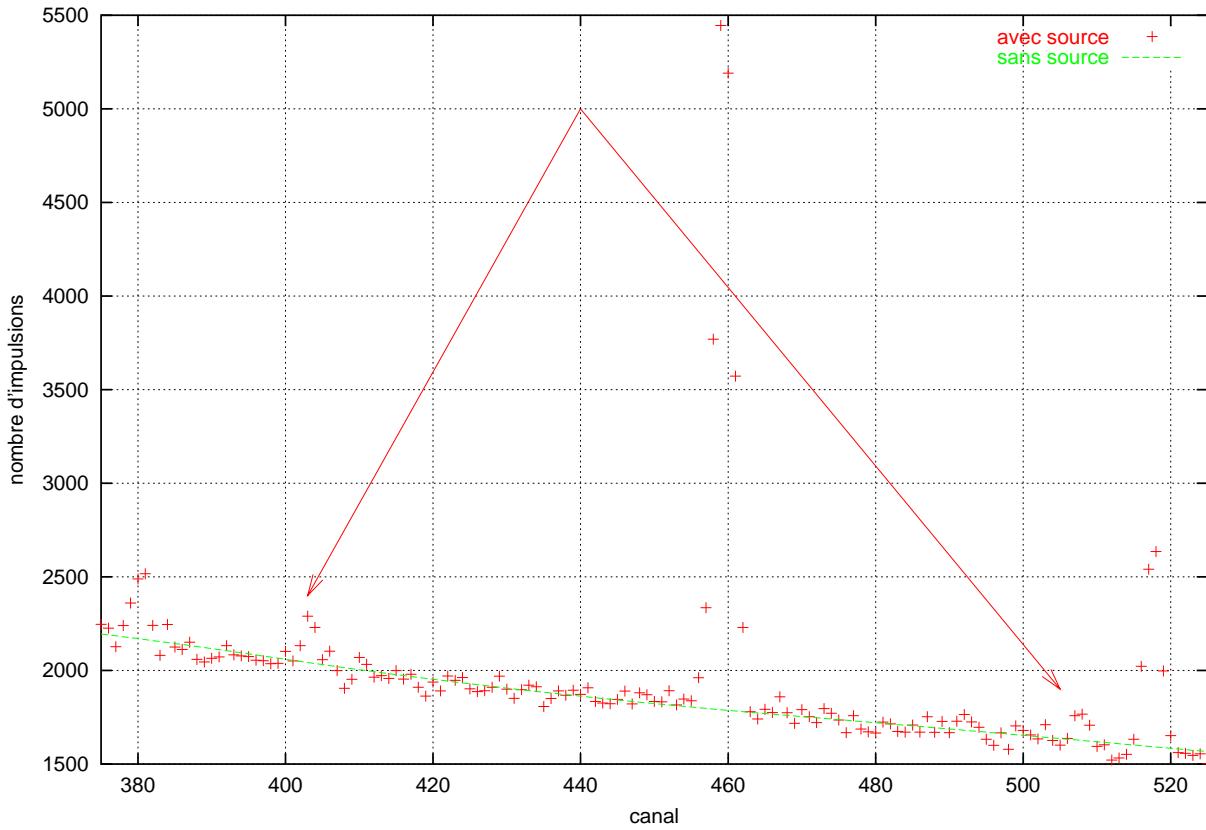


FIG. 11.2 – Spectre brut pour la détermination de l'activité d'un échantillon d'uranium-235. *Les flèches indiquent les deux pics utilisés à 163keV et 205keV*

Dans cette situation réelle, les deux pics d'intérêt sont ceux de 163keV et 205keV qui sont indiqués par des flèches à figure 11.2. Comme ces énergies sont dans la partie basse du spectre, une évaluation du bruit de fond peut être faite sans passer par une mesure autonome. En effet, les voisinages immédiats de pics donnent une bonne image de la situation sans source représentée par la droite *sans source*. Les données de cette estimation sont présentées au tableau 11.6. Les sensibilités sont données par la courbe de réponse du détecteur avec une précision suffisante pour être utilisée telle quelle. Les a priori utilisés

proviennent de la connaissance expérimentale de ce genre de mesure, il n'y a donc pas lieu d'utiliser ici l'a priori non informatif de Jeffreys. Les largeurs de plages sont calculées automatiquement par un programme ad-hoc d'analyse spectrale.

TAB. 11.6 – Données pour une estimation de l'activité de l'uranium-235

Paramètre	Unité	Pic 163keV	Pic 205keV
Durée de mesure	s	300'000	
Sensibilité du compteur	%	0.252	0.219
A priori sur l'activité	Bq	uniforme de 0 à 5	
A priori sur le bruit de fond	imp/s	uniforme entre 0.005 et 1.0	
Comptage sans source	imp	10547.30	5185.1
Comptage avec source	imp	11174.98	5539.8

La méthode, appliquée successivement aux pics individuels puis simultanés, a donné les résultats du tableau 11.7. On remarque une amélioration sensible de l'incertitude lors de l'utilisation simultanée, la méthode bayesienne permettant la prise en compte de plus d'information sur l'état. Il n'y a pas ici de validation du calcul car l'activité de cet échantillon n'est pas connue.

TAB. 11.7 – Activité a posteriori

Pics [keV]	Activité [Bq]	Ecart-type	
		absolu [Bq]	relatif [%]
163	0.829	0.195	24
205	0.549	0.157	29
simultanés	0.659	0.123	19

L'analyse des courbes ROC pour les erreurs de premier et second types $\alpha = \beta = 5\%$ a donné les activités détectables minimales du tableau 11.8.

TAB. 11.8 – Activités détectables minimales

pic [keV]	Classique	$a_D[Bq]$	
		Bayes	
		b inconnu	b connu
163	0.636	0.780	0.445
205	0.514	0.899	0.367
simultanés	-	0.548	0.281

Pour chaque pic pris séparément, les limites données par le calcul classique sont plus basses que celles obtenues par celui de Bayes. Cela s'explique par le fait que les formules classiques utilisées ici considèrent que le bruit est connu, sur-estimant la finesse du modèle de mesure. La colonne *b connu*

du tableau 11.8 donne les activités minimales que l'on obtiendrait avec le calcul bayesien dans ces conditions. La prise en considération des deux pics simultanément par la méthode bayesienne conduit alors à une activité minimale détectable significativement plus basse.

11.3.2 Calcul de l'activité du thorium-230 utilisant un a priori issu du radium-226

Le thorium-230, membre de la chaîne de l'uranium-238 décroît vers le radium-226, dont l'activité peut être déterminée avec nettement plus de précision. Ces deux nucléides se trouvent souvent dans la nature non loin de l'équilibre séculaire, ce qui implique que leurs activités sont pratiquement égales. La mesure du radium donne une répartition pour son activité qui, en vertu de l'égalité des activités, peut être reprise comme a priori pour la mesure du thorium. Afin de mesurer l'effet de cette information, avons comparé le résultat obtenu de cette manière avec celui que donnerait un a priori empirique plus large sur l'activité du thorium. Un même spectre a servi à l'estimation des deux nucléides et du bruit de fond.

Les données de mesure pour ce cas sont données au tableau 11.9.

TAB. 11.9 – Données de mesure pour le thorium-230

Paramètre	Unité	Valeur
Durée de mesure	s	89799.0
Sensibilité	%	0.02425
A priori large	Bq	uniforme sur [0,100]
A priori basé sur le radium	Bq	gaussien 4.8 ± 1.0
Comptage sans source	imp	856.0
Comptage avec source	imp	883.0

Le tableau 11.10 résume les résultats du calcul. Les erreurs de premier et second types sont toujours fixés à $\alpha = \beta = 5\%$. N_{class} désigne le comptage net pour l'estimation classique et L_C et sa limite de décision. Les rapports de vraisemblance sont exprimés par leur logarithme. $\ln \ell_{10}$ donne le rapport de vraisemblance de la situation avec source à la situation sans source qui est comparé au niveau critique $\ln \ell_C$. Les activités minimale détectables sont désignées par a_D .

TAB. 11.10 – Estimation de l'activité du thorium

Analyse bayesienne		
a priori activité [Bq]	gaussien 3.62 ± 0.82	uniforme 1.72 ± 1.09
$\ln \ell_{10}$	-2.24	-2.26
$\ln \ell_C$.05	-.08
a_D [Bq]	4.52	4.53

Analyse classique	
N_{class}	27.0
L_C	68.1
L_D	138.9
a_D [Bq]	6.38

On remarque que le comptage classique N_{class} et les rapports de vraisemblance $\ln \ell_{10}$ sont tous au-dessous des niveaux critiques et conduisent à une décision d'absence de source. Par contre, l'a priori gaussien conduit à une détermination de l'activité à 23% d'incertitude relative. La figure 11.3 montre les répartitions a posteriori de l'activité du thorium. Si la présence de la source est discutable pour l'a priori large, l'allure de la courbe issue de l'a priori du radium ne laisse aucun doute en ce qui concerne la présence de la source.

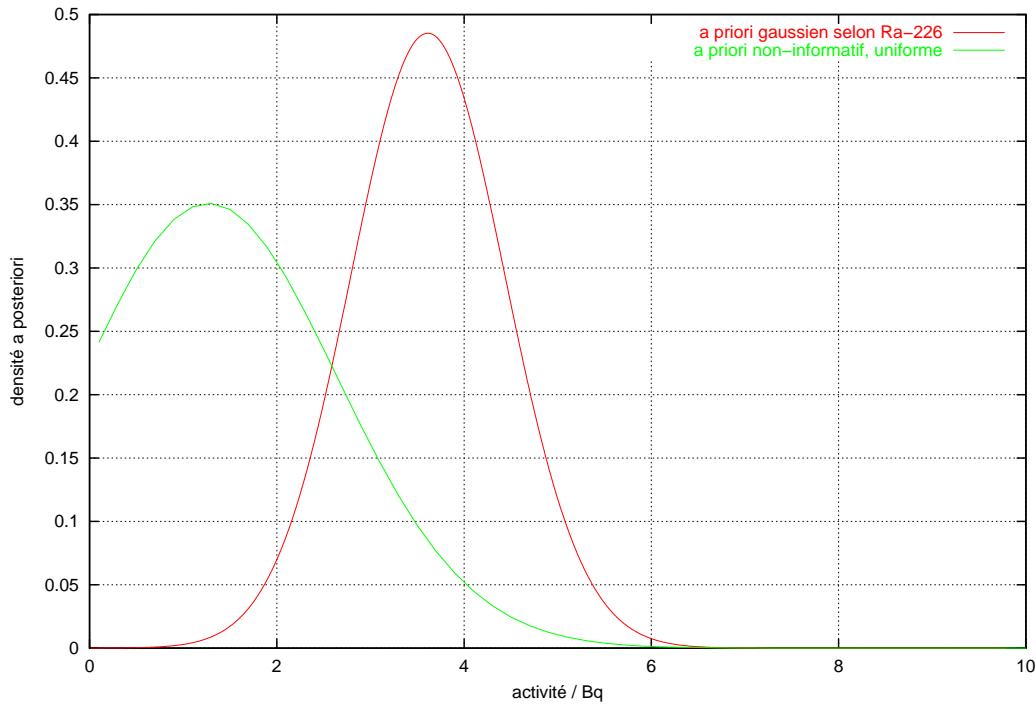


FIG. 11.3 – Comparaison des activités a posteriori du thorium-230. *L'a priori large ne permet pas de conclure à la présence d'une activité, alors que la source est bien visible si l'on utilise un a priori informatif issu de sa fille radium-226*

Chapitre 12

Temps d'attente et modèles à activité décroissante

12.1 Estimation d'un temps d'attente à répartition exponentielle

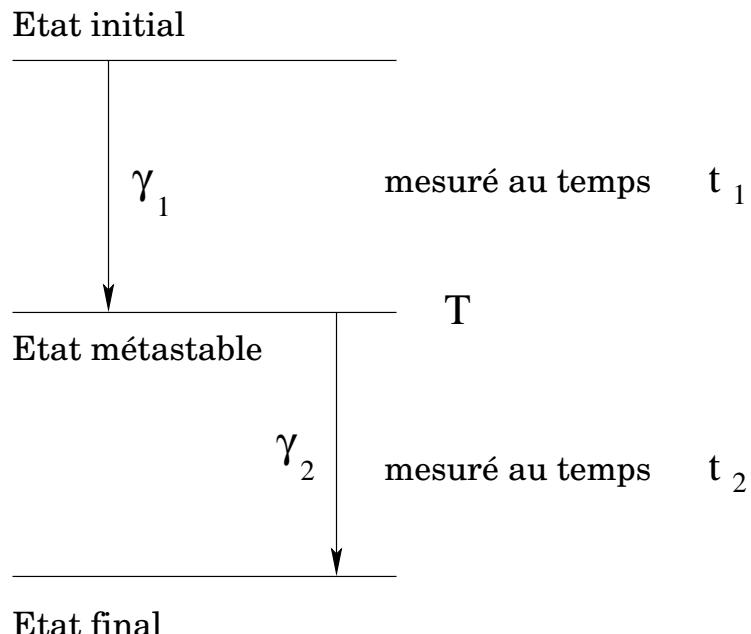


FIG. 12.1 – Nucléide dont le schéma de désintégration comporte deux photons émis en cascade passant par un état métastable de période T

Un des apports de la théorie bayesienne concerne le traitement des mesures ayant une grande incertitude. L'approche classique conduit souvent dans ce cas à des estimations incohérentes. Imaginons par exemple une mesure de la durée d'un état métastable, pour un nucléide de schéma simple connu figuré en 12.1. Les émissions photoniques γ_1 et γ_2 se succèdent, mais un certain délai s les sépare. Ce délai est une variable aléatoire régie par une loi exponentielle dont le paramètre est le temps moyen de séjour θ

dans l'état métastable. Ce temps est généralement donné en terme de période par $T = \theta \ln 2$. On a

$$P(ds | \theta) = e^{-\frac{s}{\theta}} \frac{ds}{\theta} \quad (12.1)$$

Supposons que la détermination de θ s'effectue avec deux chronomètres indépendants dont les incertitudes standard sont σ_1 et σ_2 . Si les temps de détection des photons sont t_1 et t_2 , un calcul habituel donne l'estimation de θ par

$$d = (t_2 - t_1) \pm \sqrt{\sigma_1^2 + \sigma_2^2} \quad (12.2)$$

On remarque que l'aléatoire de la loi exponentielle n'est pas pris en compte dans l'incertitude. Par ailleurs, lorsque θ est petit, la différence des temps observés peut devenir négative, violant le principe de causalité. Il est possible de sortir de cette impasse en appliquant le calcul bayesien décrit ci-dessous.

12.1.1 Délai non aléatoire

Commençons par traiter le cas d'un délai non aléatoire θ .

t_1 est le résultat de la mesure d'un temps effectif τ_1 et admet une répartition normale

$$P(dt_1 | \tau_1) = \mathcal{N}(dt_1 | \tau_1, \sigma_1^2) \quad (12.3)$$

Similairement t_2 est réparti selon $\mathcal{N}(dt_2 | \tau_1 + \theta, \sigma_2^2)$ d'où

$$P(dt_2 | \tau_1, \theta) = \mathcal{N}(dt_2 | \tau_1 + \theta, \sigma_2^2) \quad (12.4)$$

Les chronomètres étant supposés indépendants l'un de l'autre, la répartition du couple (t_1, t_2) est

$$P(dt_1, dt_2 | \tau_1, \theta) = \mathcal{N}(dt_1 | \tau_1, \sigma_1^2) \mathcal{N}(dt_2 | \tau_1 + \theta, \sigma_2^2) \quad (12.5)$$

Le changement de variable $(t_1, t_2) \rightarrow (t_1, \Delta t = t_2 - t_1)$ implique

$$P(dt_1, d\Delta t | \tau_1, \theta) = \frac{dt_1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(t_1-\tau_1)^2}{2\sigma_1^2}} \frac{d\Delta t}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(\Delta t + t_1 - \tau_1 - \theta)^2}{2\sigma_2^2}} \quad (12.6)$$

Avec une répartition à priori $\pi(d\tau_1, d\theta)$, le théorème de Bayes donne l'a posteriori

$$\pi(d\tau_1, d\theta | t_1, \Delta t) \propto e^{-\frac{(t_1-\tau_1)^2}{2\sigma_1^2}} e^{-\frac{(\Delta t + t_1 - \tau_1 - \theta)^2}{2\sigma_2^2}} \pi(d\tau_1, d\theta) \quad (12.7)$$

L'homogénéité du temps conduit à exprimer cet a priori avec un invariant de lieu sur τ_1

$$\pi(d\tau_1, d\theta) = \pi(d\theta) d\tau_1 \quad (12.8)$$

Par intégration sur τ_1 , la dépendance en t_1 disparaît ; avec l'incertitude combinée $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$, on voit aisément que l'a posteriori sur θ devient

$$\pi(d\theta | \Delta t) \propto \mathcal{N}(d\theta | \Delta t, \sigma^2) \pi(d\theta) \quad (12.9)$$

On retrouve l'expression (12.2) à $\pi(d\theta)$ près.

Le choix d'un a priori uniforme sur θ , mais de support positif, puisque l'on sait que les photons γ_1 et γ_2 se succèdent, permet d'obtenir la répartition de θ

$$\begin{aligned}\pi_1(d\theta) = \pi(d\theta | \Delta t) &= \frac{e^{-\frac{(\theta-\Delta t)^2}{2\sigma^2}} d\theta}{M_0(\Delta t)}, \theta > 0 \\ M_0(\Delta t) &= \int_{\theta>0} e^{-\frac{(\theta-\Delta t)^2}{2\sigma^2}} d\theta\end{aligned}$$

Quelques manipulations donnent alors la moyenne et la variance de cette distribution au moyen de la gaussienne cumulée Φ

$$\begin{aligned}E_1\theta &= \Delta t + \frac{\sigma e^{-\frac{\Delta t^2}{2\sigma^2}}}{\sqrt{2\pi}\Phi\left(\frac{\Delta t}{\sigma}\right)} \\ V_1\theta &= \sigma^2 - (E_1\theta - \Delta t)E_1\theta\end{aligned}$$

L'évaluation de l'incertitude est un peu plus basse que celle du calcul classique σ . La figure 12.2 montre comment cette estimation varie en fonction de l'intervalle de temps mesuré.

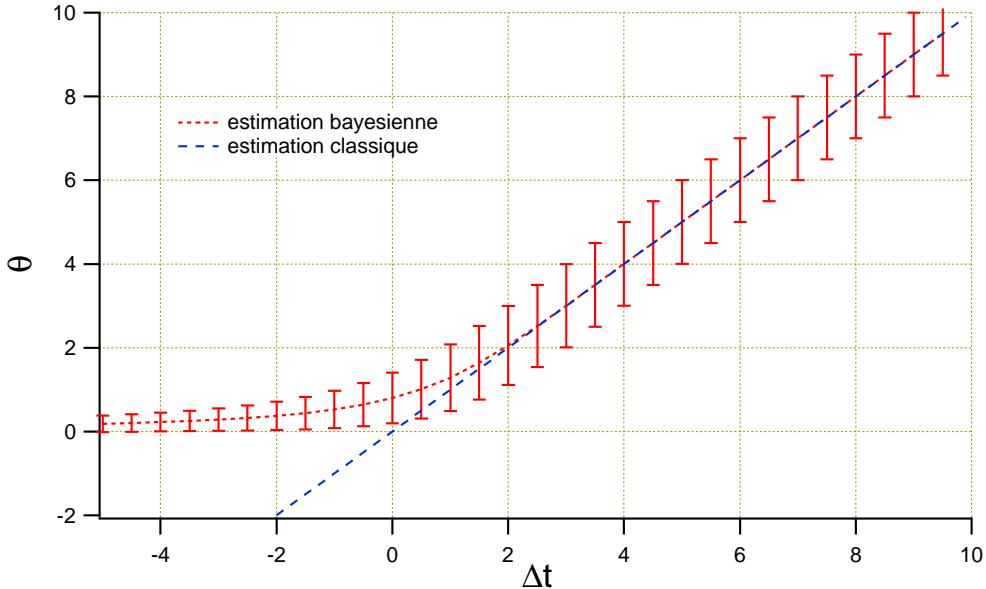


FIG. 12.2 – Estimation bayésienne du délai. L'incertitude-type sur la mesure du temps est 1, L'estimation du délai θ en fonction de la différence des temps Δt ne donne jamais de valeur négative, même lorsque Δt est négatif. L'incertitude sur l'estimation est de un écart-type.

12.1.2 Délai aléatoire

Abordons maintenant le cas d'un délai s réparti exponentiellement. La répartition du temps t_2 s'exprime par

$$P(dt_2 | \tau_1, \theta) = \int_s \mathcal{N}(dt_2 | \tau_1 + s, \sigma_2^2) e^{-\frac{s}{\theta}} \frac{ds}{\theta} \quad (12.10)$$

Ceci implique la répartition

$$P(dt_1, dt_2 | \tau_1, \theta) = \mathcal{N}(dt_1 | \tau_1, \sigma_1^2) \int_{s=0}^{\infty} \mathcal{N}(dt_2 | \tau_1 + s, \sigma_2^2) e^{-\frac{s}{\theta}} \frac{ds}{\theta} \quad (12.11)$$

Un rapide calcul donne

$$P(dt_1, dt_2 | \tau_1, \theta) = \frac{dt_1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(t_1-\tau_1)^2}{2\sigma_1^2}} \frac{dt_2}{\theta} e^{-\frac{1}{\theta}(t_2-\tau_1-\frac{\sigma_2^2}{2\theta})} \Phi\left(\frac{1}{\sigma_2}\left(t_2-\tau_1-\frac{\sigma_2^2}{\theta}\right)\right) \quad (12.12)$$

Si le premier chronomètre est parfait¹, cela correspond à une mesure directe de l'intervalle et on obtient avec $\sigma = \sigma_2$:

$$P(d\Delta t | \theta) = \frac{d\Delta t}{\theta} e^{-\frac{1}{\theta}(\Delta t - \frac{\sigma^2}{2\theta})} \Phi\left(\frac{1}{\sigma}\left(\Delta t - \frac{\sigma^2}{\theta}\right)\right) \quad (12.13)$$

Plaçons-nous dans le cas $\sigma_1 = \sigma_2 = 1$. Un choix respectant l'homogénéité temporelle pour l'a priori sur τ_1 impose l'invariant de lieu

$$\pi(d\tau_1) \propto d\tau_1 \quad (12.14)$$

Le changement de variable $(t_1, t_2) \rightarrow (t_1, \Delta t)$ avec $\Delta t = t_2 - t_1$, suivi de la marginalisation sur τ_1 conduit à

$$P(dt_1, d\Delta t | \theta) = dt_1 \frac{d\Delta t}{\theta} \int_{\tau_1} e^{-\frac{(t_1-\tau_1)^2}{2}} e^{-\frac{1}{\theta}(\Delta t + t_1 - \tau_1 - \frac{1}{2\theta})} \Phi(\Delta t + t_1 - \tau_1 - \frac{1}{\theta}) d\tau_1 \quad (12.15)$$

L'intégrale ne dépend pas de t_1 , ce qui permet d'obtenir

$$\begin{aligned} P(d\Delta t | \theta) &= \frac{d\Delta t}{\theta} \int_{\varepsilon} e^{-\frac{\varepsilon^2}{2}} e^{-\frac{1}{\theta}(\Delta t - \varepsilon - \frac{1}{2\theta})} \Phi(\Delta t - \varepsilon - \frac{1}{\theta}) d\varepsilon \\ &= \frac{d\Delta t}{\theta} e^{-\frac{1}{\theta}(\Delta t - \frac{1}{\theta})} \Phi\left(\frac{\Delta t - \frac{2}{\theta}}{\sqrt{2}}\right) \end{aligned} \quad (12.16)$$

On retrouve la même expression que pour le cas où le premier chronomètre est parfait, avec $\sigma = \sqrt{2}$, ce qui est cohérent.

Choisissons pour a priori sur θ l'invariant de lieu $\pi(d\theta) = d\theta$. L'a posteriori étant donné l'observation Δt s'exprime

$$\pi(d\theta | \Delta t) \propto \frac{d\theta}{\theta} e^{-\frac{1}{\theta}(\Delta t - \frac{1}{\theta})} \Phi\left(\frac{\Delta t - \frac{2}{\theta}}{\sqrt{2}}\right) \quad (12.17)$$

Cette densité converge vers un nombre strictement positif lorsque $\theta \rightarrow 0$, et se comporte en $\frac{1}{\theta}$ lorsque $\theta \rightarrow \infty$ (cf figure 12.3). Le choix de l'a priori invariant de lieu sur θ impose une petite discussion.

D'une manière générale, les a priori non normalisables sont admis (voire conseillés), mais l'a posteriori, lui, doit être une véritable probabilité. Or le comportement en $\frac{1}{\theta}$ de (12.16) au voisinage de $\theta = \infty$ interdit sa normalisation. Selon l'approche bayesienne, *il n'est donc pas possible d'estimer le paramètre*

¹La première gaussienne devient un dirac

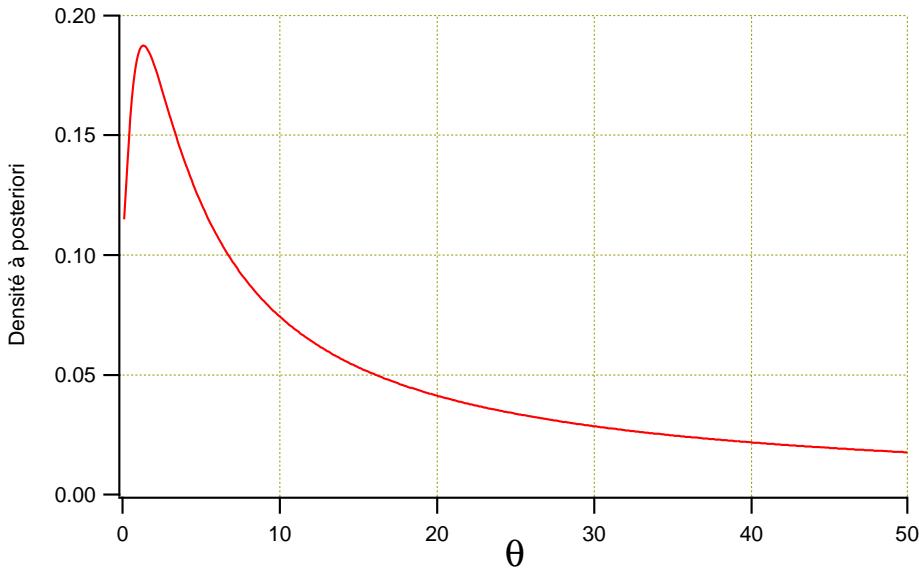


FIG. 12.3 – Densité a posteriori pour θ . *L'a priori est un invariant de lieu et la mesure a donné $\Delta t = 2$. La courbe admet une asymptote en $1/\theta$ lorsque $\theta \rightarrow \infty$*

θ . Ceci se retrouve d'ailleurs dans l'impossibilité d'estimer un temps d'attente poissonnien en une seule mesure en partant d'a priori non informatifs. Ceci reviendrait à estimer une activité avec une seule détection, ce qui n'est pas une estimation statistique très convaincante !

Supposons maintenant que cette mesure soit effectuée r fois. Si les résultats sont $\Delta t = (\Delta t_i)_{i=1}^r$, il vient

$$\pi(d\theta | \Delta t) \propto \frac{d\theta}{\theta^r} e^{-\frac{r}{\theta}(\frac{1}{r} \sum \Delta t_i - \frac{1}{\theta})} \prod_i \Phi\left(\frac{\Delta t_i - \frac{2}{\theta}}{\sqrt{2}}\right)$$

On remarque que

- pour $r = 1$, l'échec de la normalisation permet tout au plus l'utilisation du maximum de vraisemblance
- pour $r = 2$, la normalisation est assurée, mais il n'y a ni moyenne, ni variance ; néanmoins, des intervalles de confiance sont définissables
- pour $r = 3$, la moyenne existe, donnant un estimateur, mais l'absence de variance exclut une mesure de l'incertitude
- ce n'est qu'à partir de $r = 4$ que la situation s'arrange ; il faut $r + 2$ mesures pour obtenir un moment d'ordre r

Lorsque r augmente, cette probabilité converge vers une loi de plus en plus normale et pointue, établissant un lien avec la statistique classique. La figure 12.4 montre les a posteriori résultants d'une suite de simulations avec un nombre de répétitions croissant qui confirme cette convergence.

12.1.3 Résumé

Cet exemple illustre la problématique bayésienne. La conception du problème de mesure est plus complète (l'aléatoire du délai peut être introduit facilement) et l'observateur est à même d'estimer l'état

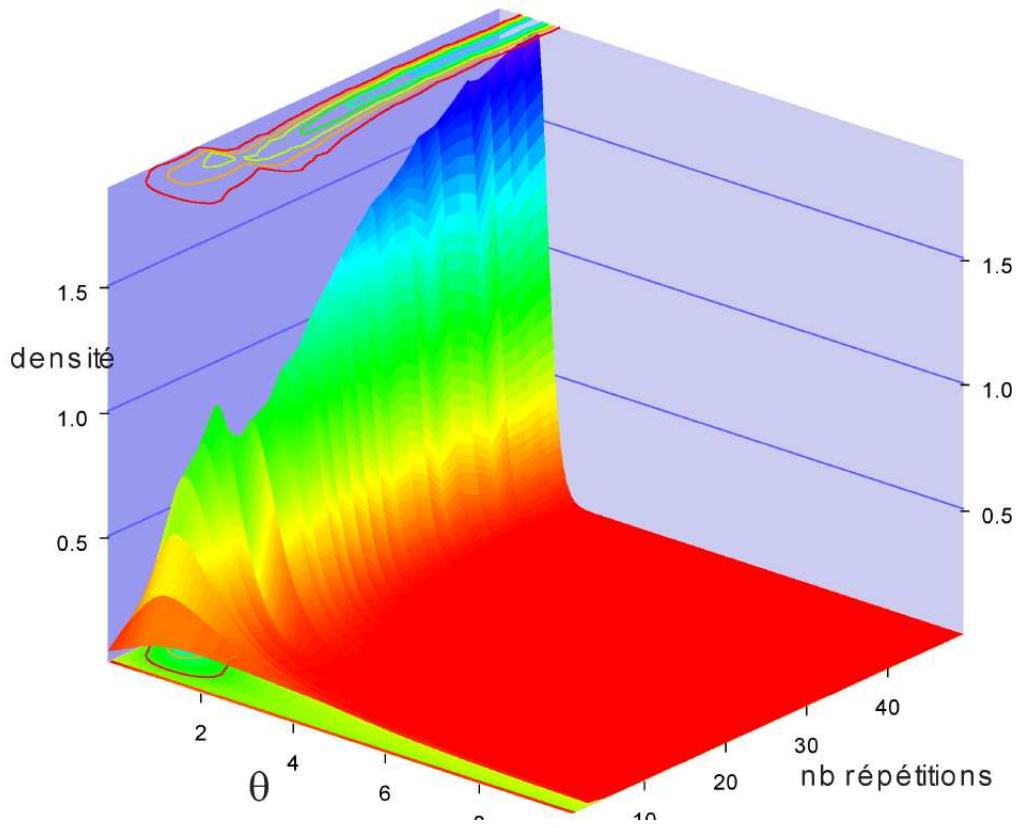


FIG. 12.4 – Evolution de l'a posteriori sur le paramètre du délai. 50 mesures avec un nombre croissant de répétitions sont simulées avec un paramètre de délai effectif θ_{eff} fixé à 1 unité de temps.

par une répartition qui reste dans un domaine physiquement acceptable (ici un délai positif). Un estimateur de l'état est donné par l'espérance mathématique de cette répartition, avec une incertitude reliée à sa variance, si ces deux grandeurs existent.

Par ailleurs, il est possible d'introduire dans l'a priori une information préalable. Ici l'a priori choisi est non informatif, mais si l'on a la certitude que le délai ne peut se situer en dehors d'un certain intervalle, il est parfaitement licite de restreindre l'a priori à cet intervalle. Ceci élimine les problèmes de convergence des intégrales et donne une répartition a posteriori avec des moments d'ordre quelconque.

En revanche, le calcul est devenu nettement plus complexe que dans le cas classique. La convergence des intégrales doit être examinée de près, et il faut tenir compte de la singularité en $\theta = 0$. Le choix de l'a priori peut s'avérer délicat, mais la répétition de l'expérience arrange en général assez rapidement les choses.

12.2 Modèles avec activité décroissante

Jusqu'ici, nous avons appliqué la TBDS à des mesures de sources dont l'activité pouvait être considérée comme constante. Pour les sources à décroissance rapide, ce modèle n'est évidemment plus valable, et la loi de Poisson doit être abandonnée. La première partie de cette section sera consacrée à la détermination de l'activité d'une source par l'observation d'un nombre d'impulsions dans un intervalle de temps

donné, en supposant la constante de désintégration λ connue. La seconde partie montrera une utilisation des temps de détection pour déterminer cette constante.

12.2.1 Détermination de l'activité par les comptages avec constante de désintégration connue

12.2.1.1 Modèle

Soient N noyaux caractérisés par une constante de désintégration λ au temps² $t=0$. Un observateur observe les désintégrations de ces noyaux avec un détecteur de sensibilité ϵ et observe n impulsions dans un intervalle de temps $[d, f]$ postérieur à $t = 0$. La question est de trouver N .

La probabilité qu'un noyau effectue une désintégration dans l'intervalle $[d, f]$ est

$$p_1 = \int_d^f \lambda e^{-\lambda t} dt = e^{-\lambda d} - e^{-\lambda f} \quad (12.18)$$

donc celle d'avoir \tilde{n} désintégrations dans cet intervalle pour N noyaux initiaux est

$$P(\tilde{n} | N) = \mathcal{B}(\tilde{n} | N, p_1) \quad (12.19)$$

Mais chacune de ces désintégrations n'a qu'une probabilité ϵ d'être vue par le détecteur. Donc, pour \tilde{n} désintégrations, nous obtiendrons n détections avec la probabilité

$$P(n | \tilde{n}) = \mathcal{B}(n | \tilde{n}, \epsilon) \quad (12.20)$$

Le modèle donnant le nombre de détections en partant du nombre de noyaux est alors

$$P(n | N) = \sum_{\tilde{n}} P(n | \tilde{n}) P(\tilde{n} | N) \quad (12.21)$$

$$= \sum_{\tilde{n}} \mathcal{B}(n | \tilde{n}, \epsilon) \mathcal{B}(\tilde{n} | N, p_1) \quad (12.22)$$

En utilisant la propriété de semi-goupe (cf A.1), il vient

$$P(n | N) = \mathcal{B}(n | N, \epsilon p_1) \quad (12.23)$$

12.2.1.2 Inférence bayesienne du nombre de noyaux

Soit maintenant un a priori $\pi_0(N)$ sur le nombre de noyaux au temps $t = 0$. Le théorème de Bayes permet d'obtenir un a posteriori tenant compte de l'observation de n impulsions dans le détecteur

$$\pi_1(N | n) = \frac{P(n | N) \pi_0(N)}{M(n)} \quad (12.24)$$

Où

$$M(n) = \sum_N P(n | N) \pi_0(N) \quad (12.25)$$

²Toute la présente problématique ne dépendant pas du choix de l'origine des temps, nous prendrons $t = 0$ pour simplifier les expressions.

est la répartition marginale des comptes de détections sous l'a priori $\pi_0(N)$. Remplaçant les probabilités par leurs expressions, il vient

$$\pi_1(N | n) \propto \mathcal{B}(n | N, \varepsilon p_1) \pi_0(N) \quad (12.26)$$

$$\propto \binom{N}{n} (\varepsilon p_1)^n (1 - \varepsilon p_1)^{N-n} \pi_0(N) \quad (12.27)$$

On retrouve ici la forme de la loi négative binomiale (5.21) qui donne

$$\pi_1(N | n) \propto \mathcal{NB}(N - n | n + 1, \varepsilon p_1) \pi_0(N) \quad (12.28)$$

Partant d'un a priori uniforme, l'a posteriori devient simplement négatif-binomial translaté

$$\pi_1(N | n) = \mathcal{NB}(N - n | n + 1, \varepsilon p_1) \quad (12.29)$$

$$(12.30)$$

L'espérance et la variance de N valent donc :

$$EN = n + \frac{(n+1)(1 - \varepsilon p_1)}{\varepsilon p_1} \quad (12.31)$$

$$VN = \frac{(n+1)(1 - \varepsilon p_1)}{(\varepsilon p_1)^2} \quad (12.32)$$

Si $\eta = \varepsilon p_1$ désigne la probabilité de détection d'un noyau dans la plage $[d, f[$, on peut écrire

$$EN = \frac{n+1}{\eta} - 1 \quad (12.33)$$

$$\sigma_N = \frac{1}{\eta} \sqrt{(n+1)(1 - \eta)} \quad (12.34)$$

Remarques La translation de la loi négative binomiale implique que N ne peut être inférieur à n , le nombre de détections ne pouvant évidemment pas dépasser celui des noyaux. En première approximation³, on a $N\eta = n \pm \sqrt{n}$. Connaissant N , l'activité à $t = 0$ s'obtient par $N\lambda$.

12.2.1.3 Application

Un estimation de l'activité initiale du fluor-18 par un scintillateur au iodure de sodium a été effectuée par cette méthode, basée sur des résultats expérimentaux. Environ 6000 mesures ont été faites. Toute les 10 secondes, le nombre de détections a servi à une estimation du nombre de noyaux initiaux. Les résultats obtenus correspondent exactement à ceux obtenus par la méthode classique. La figure 12.5 montre l'évolution de l'estimation de l'activité initiale déduite de ce nombre en fonction de la date de la mesure. L'incertitude augmente avec le temps car le nombre de noyaux restants diminue exponentiellement.

12.2.2 Détermination de la constante de désintégration radioactive par les temps de détection

12.2.2.1 Introduction

La *constante radioactive* λ est la probabilité qu'un noyau se désintègre par unité de temps. Elle est souvent exprimée par la période $T_{\frac{1}{2}}$ représentant le temps moyen nécessaire à la disparition de la moitié

³Si l'on avait pris un a priori invariant d'échelle, ce serait même exact

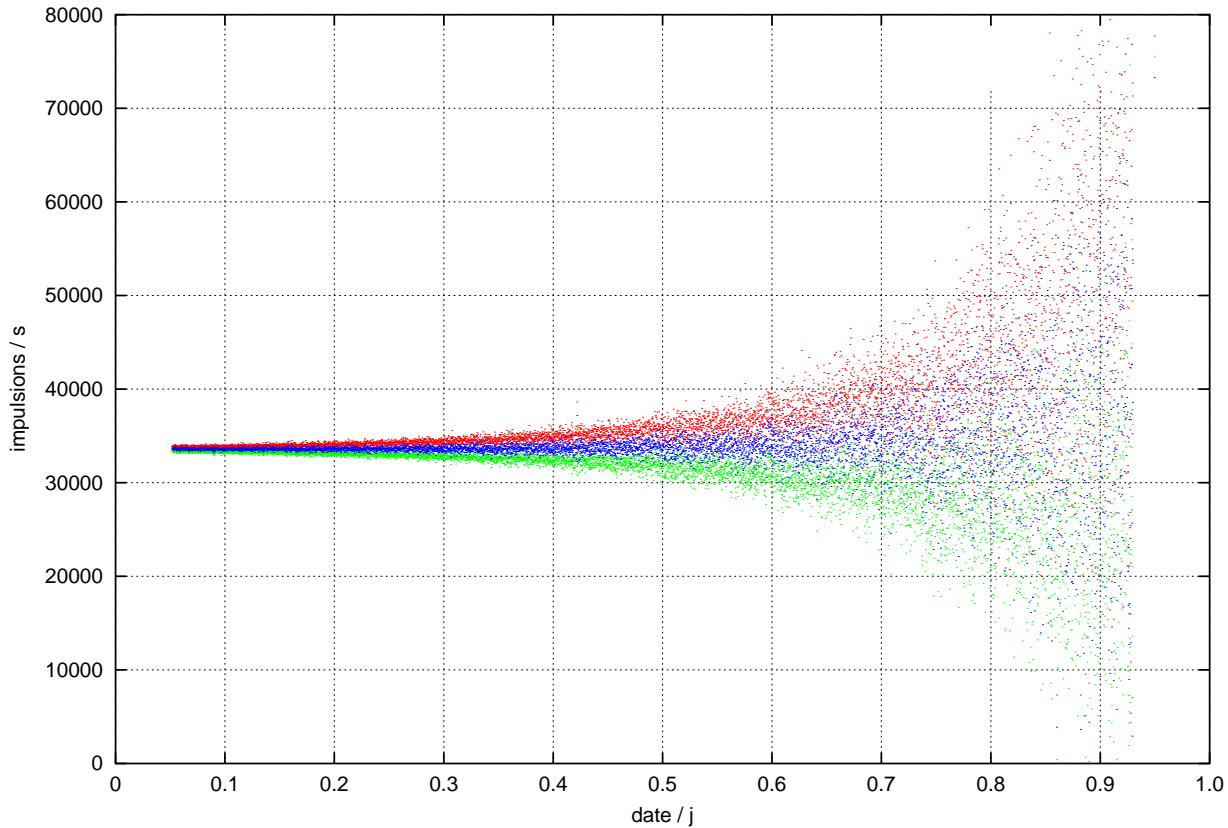


FIG. 12.5 – Estimation bayesienne de l’activité initiale d’une source de fluor-18. Pour 6000 plages successives de 10s, l’estimation du nombre de noyaux est obtenue depuis le nombre de détections enregistrées dans un scintillateur au iodure de sodium.

d’un ensemble de noyaux. Cette formulation s’apparente à celle des délais de métastabilité rencontrés en 12.1.

Une méthode habituelle pour déterminer cette constante consiste à utiliser un détecteur permettant de compter les désintégrations par l’intermédiaire des particules qui sont émises par le nucléide dans des plages successives de temps. Les nombres d’impulsions décroissant selon une loi exponentielle, une régression linéaire sur les logarithmes des comptages donne alors une estimation de la constante.

Une autre manière de faire consiste à déterminer les temps d’apparition des détections et à appliquer une modèle bayesien. Dans un premier temps, nous ferons le calcul pour une source dans un environnement non bruité ; ensuite, nous introduirons un bruit aléatoire poissonnien de taux constant.

12.2.2.2 Calcul bayesien sans bruit de fond

Modèle La source est constituée de N noyaux au temps $t = 0$ et l’intervalle d’observation noté $[d, f]$. La sensibilité du détecteur est ε . L’intervalle d’observation ainsi que la sensibilité sont considérés comme connus, ce qui limite l’ensemble des états (ou des paramètres à estimer) Θ aux couples (N, λ) . Les observations sont des suites de temps de détections dans l’intervalle $[d, f] : (t_i)_{i=1}^n$ pour n variable. L’espace

des observations \mathcal{X} est donc

$$\mathcal{X} = \bigcup_{n=0}^N [d, f]^n$$

D'une manière générale, on montre (cf section A.14) que le modèle de mesure est donné par

$$P(t_1, \dots, t_n | N, \lambda) = \mathcal{B}(n | N, \varepsilon p_1(\lambda)) \prod_{i=1}^n \frac{\lambda e^{-\lambda t_i}}{p_1(\lambda)} dt_i \quad (12.35)$$

Où \mathcal{B} est la loi binomiale et

$$p_1(\lambda) = e^{-\lambda d} - e^{-\lambda f} \quad (12.36)$$

Dans notre cas, le nombre de détections n est connu⁴ et le modèle se simplifie pour donner la fonction de vraisemblance

$$\mathcal{L}(\lambda | t_1, \dots, t_n) = \prod_{i=1}^n \frac{\lambda e^{-\lambda t_i}}{p_1(\lambda)} \quad (12.37)$$

$$= \left(\frac{\lambda e^{-\lambda \bar{t}}}{p_1(\lambda)} \right)^n \quad (12.38)$$

où \bar{t} est la moyenne des observations t_i .

Il est intéressant de remarquer la réduction d'information, qui permet de ne tenir compte que de la moyenne et du nombre des temps d'observation, en évacuant le nombre de noyaux initiaux ainsi que la sensibilité du détecteur. Ces deux derniers paramètres jouent en fait un rôle implicite en agissant sur le nombre n de temps observés, qui lui, influence la précision des estimations.

A priori Le paramètre à déterminer λ étant un évaluateur positif, nous utiliserons une loi gamma comme a priori :

$$\pi_0(\lambda) = \text{Ga}(\lambda | N_f, t_f) \quad (12.39)$$

A posteriori L'application du théorème de Bayes donne immédiatement

$$\pi_1(\lambda) \propto \left(\frac{\lambda e^{-\lambda \bar{t}}}{p_1(\lambda)} \right)^n \frac{t_f^{N_f}}{\Gamma(N_f)} \lambda^{N_f-1} e^{-\lambda t_f} d\lambda \quad (12.40)$$

Ce type d'intégrale est difficile à évaluer, car n est souvent grand (quelques centaines). La section A.15 décrit le calcul des moments de l'a posteriori en utilisant l'approximation de Laplace. Une limite de cette dérivation est la condition $N_f > 1$, qui exclut le choix d'a priori trop larges.

Application La figure 12.6 représente l'estimation de λ en fonction de sa valeur effective pour une simulation dont les paramètres sont définis à la table 12.1. L'a priori choisi est une loi gamma d'espérance 1 et d'écart-type relatif 9.5% qui est une répartition assez pointue. Si la valeur effective est éloignée de cette espérance, la méthode ne peut plus rendre compte des observations. C'est ce qui se produit ici pour des λ supérieurs à 4 ou inférieurs à $\frac{1}{2}$.

⁴Lorsque le temps f est dépassé, il ne peut plus y avoir de détection supplémentaire.

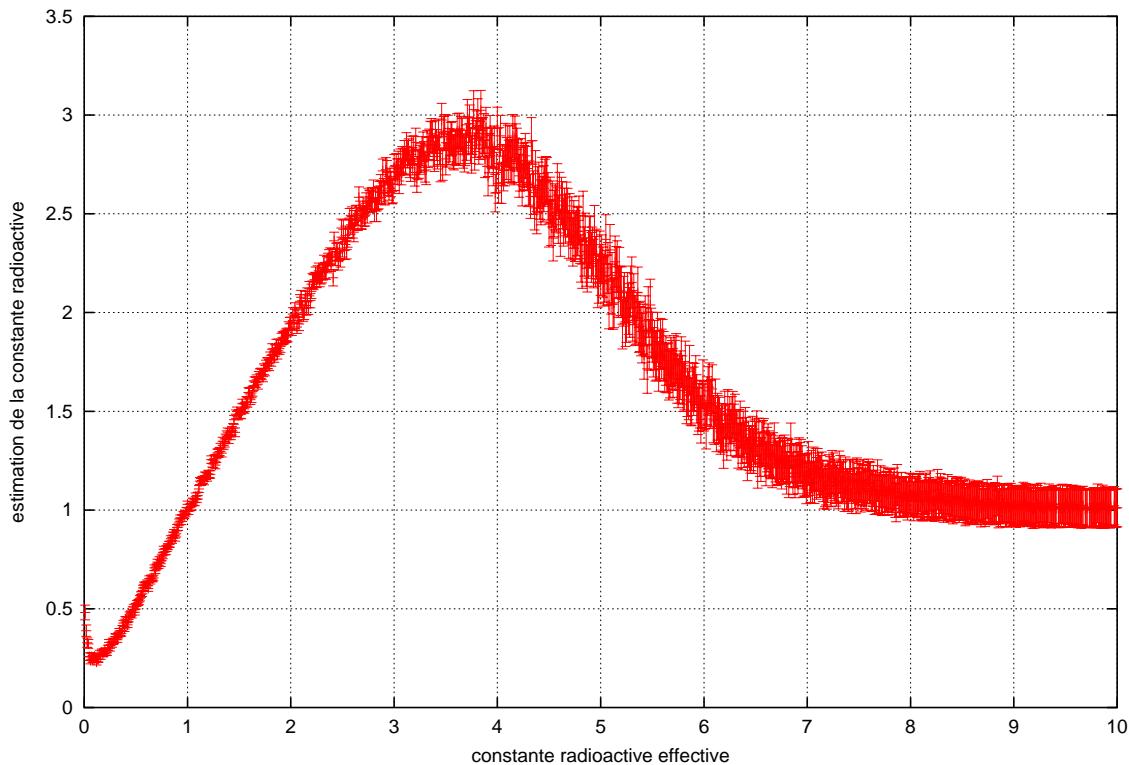


FIG. 12.6 – Estimation de λ par approximation de Laplace. L'estimation de λ est donnée en fonction de sa valeur effective, avec une incertitude bayésienne. Cette estimation est bonne pour des valeurs effectives de 0.5 à 2.5. Au delà et en deçà, l'a priori devient progressivement dominant.

12.2.2.3 Détermination en présence de bruit de fond

Méthode Plaçons maintenant la source dans un bruit poissonnien de taux constant b . Rien ne permet de distinguer les impulsions du bruit de ceux de la source. Il est néanmoins possible d'obtenir une estimation de la constante par suppression aléatoire de temps de détection selon la loi du bruit de fond. Supposons que nous ayons observé n temps de détection. Si le nombre de d'impulsions du bruit de fond est n_b , nous pouvons supprimer aléatoirement selon une loi uniforme n_b temps de détection et estimer λ sans bruit de fond avec les $n - n_b$ temps restants. Par itération de cette procédure en choisissant aléatoirement n_b selon la loi de Poisson $\mathcal{P}(n_b | b(f - d))$, il est alors possible d'effectuer une moyenne pondérée des espérances et variances des λ a posteriori. Pour l'estimation de la variance a posteriori, nous avons ajouté une composante dite externe qui est simplement celle rendant compte du processus de suppression lui-même. La procédure est décrite à la section A.16.

Application Cette méthode a été validée en évaluant par simulation la période de l'yttrium-90 en se basant sur des données réelles. Partant d'un échantillon provenant de l'environnement, λ a été estimé par régression linéaire sur 46 intervalles de mesure d'une durée de 4h. La valeur obtenue pour cette période sert originellement à estimer la pureté de l'échantillon. Cette analyse donne 4836 atomes initiaux (avec une sensibilité unité). Les temps de détection ont ensuite été simulés en partant de ce nombre initial

TAB. 12.1 – Paramètres de simulation pour une estimation de λ sans bruit de fond

Variable	Symbole	Valeur
début de la mesure	d	1
fin de la mesure	f	3
nombre de noyaux initiaux	N	100'000
sensibilité	ϵ	0.4
Paramètres de l'a priori gamma		
forme	N_f	110
échelle	t_f	110
espérance	E_0	1
déviation standard	$\sqrt{V_0}$	0.095

d'atomes et d'un bruit de fond réaliste $b = 19.76 \text{ h}^{-1}$.

L'analyse bayesienne a ensuite été conduite depuis un a priori gamma dont la moyenne est la constante effective de l'yttrium-90 ($\lambda = 0.01083 \text{ h}^{-1}$) et un écart-type relatif de 90%.

L'a posteriori sur l'activité calculé par la méthode décrite ci-dessus est représenté par sa moyenne et son incertitude-type à différents temps de mesure à la figure 12.7. Cette incertitude est grande au départ car le temps de mesure est court, et augmente aussi lorsque le temps devient trop long, car le bruit de fond prend une place de plus en plus importante. L'optimal se trouve ici aux environs de 300 heures. Les nombres d'impulsions cumulées provenant de la source et du bruit sont reportés à la figure 12.8. Celles provoquées par la source se stabilisent car le nombre de noyaux est fini. Par contre, le bruit de fond cumulé augmente constamment. L'oscillation est due au fait qu'il s'agit pour chaque temps d'une simulation différente.

Les écarts entre l'estimation et la valeur effective sont exprimées en *p-values*. Ces p-values sont obtenues de la manière suivante. Tout d'abord, les écarts normalisés entre les estimations et la valeur effective sont calculés en divisant leur écart par l'incertitude-type estimée. Ces écarts normalisés devraient se comporter à peu près selon une loi normale réduite. En transformant ces écarts normalisés par l'inverse de la gaussienne cumulée, on obtient les p-values, qui devraient se répartir approximativement de manière uniforme entre 0 et 1. Ces grandeurs sont présentées dans notre cas sous forme d'un histogramme à la figure 12.9, qui montre que leur répartition est sensiblement uniforme.

12.2.2.4 Conclusion

L'utilisation de l'approximation de Laplace donne de bons résultats pour la détermination des constantes de désintégration par inférence bayesienne. Le traitement statistique du bruit de fond par suppression aléatoire ramène le problème au modèle sans bruit.

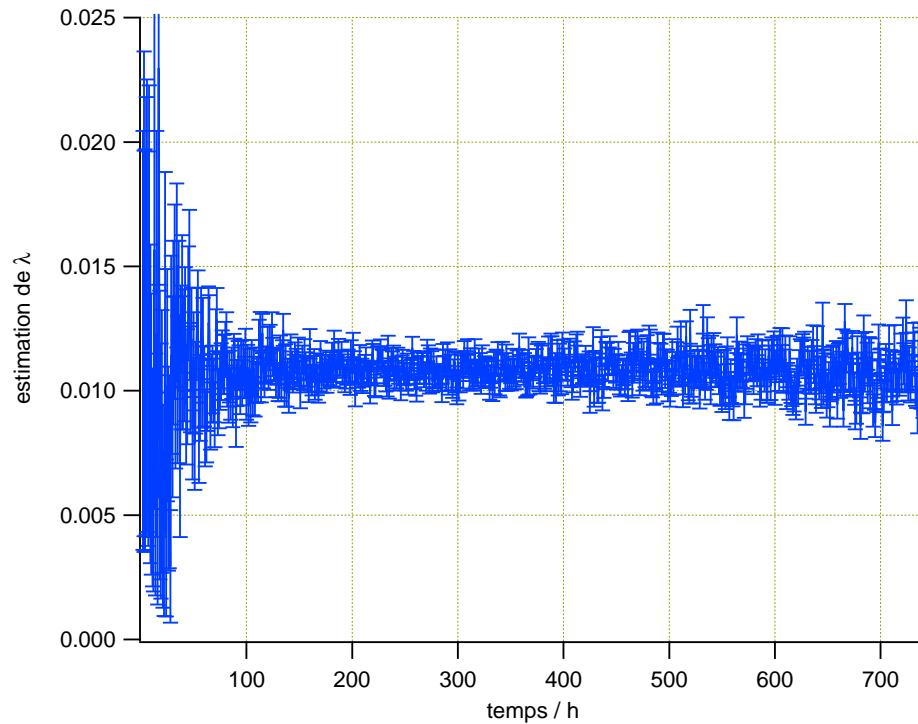


FIG. 12.7 – Estimation du λ de l'yttrium-90 en fonction de la durée de la mesure.

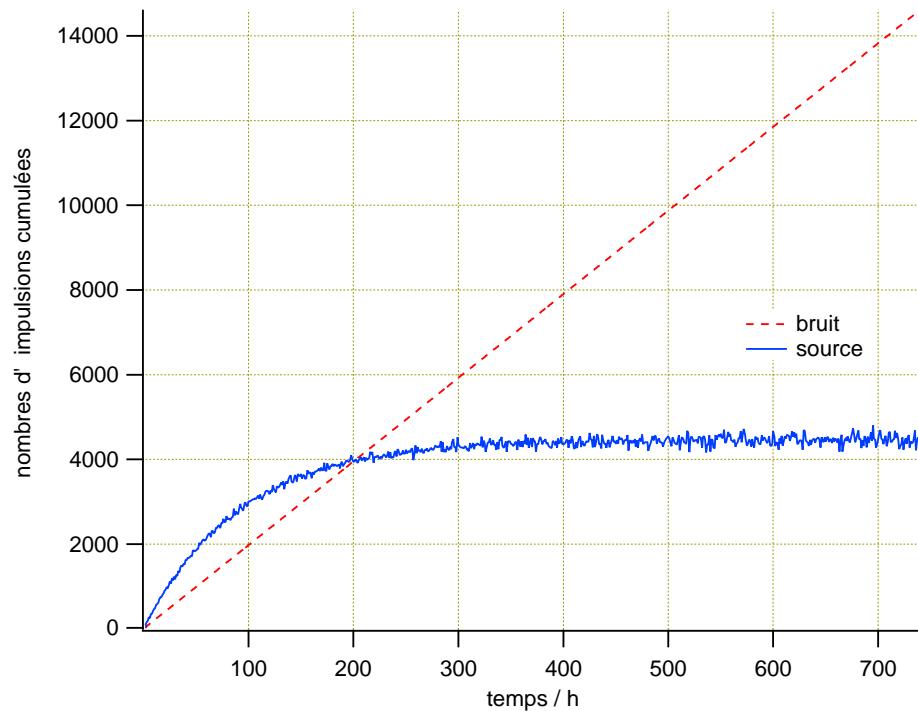


FIG. 12.8 – Nombres d'impulsions cumulées pour l'estimation du λ de l'yttrium-90.

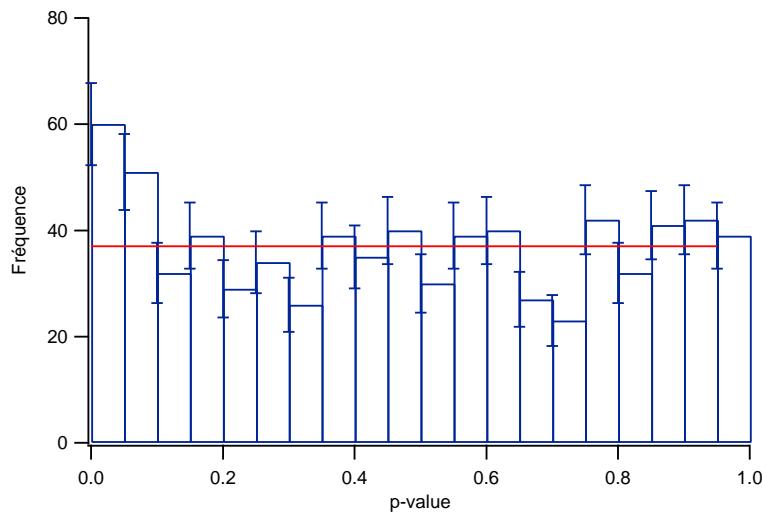


FIG. 12.9 – Validation statistique de l'estimation de λ de l'yttrium-90. *Les p-values se répartissent de manière à peu près uniforme, ce qui est signe que la méthode est bonne. Les barres d'erreur représentent la racine carrée de la fréquence.*

Cinquième partie

Conclusions et perspectives

Conclusions

La théorie bayesienne de la décision statistique a été utilisée ici pour l'analyse de résultats issus d'expériences traitant d'événements rares. En particulier, lors de la recherche de radionucléides présents sous formes de traces, les compteurs donnent de petits nombres d'impulsions et les évaluations de la statistique classique ne sont plus fiables. D'autre part, les approximations gaussiennes ne sont plus justifiées, et donnent par exemple des valeurs aberrantes pour l'estimation de l'activité.

Ces considérations ont conduit à utiliser un formalisme qui différencie clairement l'état du système physique à évaluer des observations obtenues par la mesure. Ce formalisme est celui de la théorie de la décision statistique, dont l'espace de base est le produit cartésien de l'espace des observations par celui des états. La relation statistique régissant le résultat de mesure en présence d'un état donné du système physique est le modèle, qui par application du théorème de Bayes, peut être inversé et donner une répartition des états compte tenu de l'observation obtenue par l'expérimentateur.

Cette inversion nécessite un a priori sur l'espace des états qui peut être de deux types. Le premier est dit non informatif. Dans ce cas, nous avons montré, à l'aide d'une interprétation de l'inégalité de Cramér-Rao, que les répartitions de Jeffreys sont d'une grande utilité pour les problèmes physiques, où le modèle et l'instrument de mesure sont bien définis. Ces a priori, construits sur les concepts d'information de Shannon-Kullback et de Fisher, jettent un pont entre l'analyse fréquentiste et celle de Bayes. Le deuxième type d'a priori est dit informatif et recouvre les cas où l'expérimentateur possède des indications sur les états possibles du système physique. Ces indications peuvent être des résultats de mesures antérieures, ou des considérations limitant le domaine des états possibles.

Plusieurs applications, pratiques, théoriques et simulées, ont mis en évidence que cette méthode conduit à de meilleures estimations et à des décisions de plus bas risque. En particulier, nous avons pu obtenir des limites de détection plus basses pour des radionucléides à schéma complexe, ainsi que des valeurs plus précises de leur activité par la prise en compte de leurs produits de filiation.

Apports spécifiques de ce travail

L'utilisation d'une théorie cohérente pour l'estimation des activités en présence de faibles taux de comptage a permis d'une part l'élimination de certaines estimations incohérentes, d'autre part une diminution de l'incertitude par intégration de données préalables,

D'un point de vue théorique, les notions de subjectivité et d'objectivité de la probabilité ont été réinterprétées dans le cadre de l'analyse d'une mesure. Nous avons vu que la répartition des observations donnée par le modèle peut être qualifiée d'objective, car la répétition de l'expérience est possible et donne ainsi une probabilité fréquentiste, alors que la répartition des états obtenue par le théorème de Bayes est subjective, en ce sens qu'elle décrit une incertitude qui est le résultat de l'application d'un a priori propre à l'observateur. L'avantage principal de cette estimation est de donner une répartition complète, globale, à partir de laquelle il est possible de retrouver l'expression traditionnelle de l'incertitude par la variance, ou la covariance dans le cas de mesures multidimensionnelles.

L'intégration des données préalables à une mesure a abouti à l'importante distinction entre état stationnaire ou non. Lorsque des observations sont collectées lors d'une suite de mesures pour laquelle l'état du système physique ne varie pas, ce dernier est qualifié de stationnaire, et il suffit d'itérer le théorème de Bayes. Si cela n'est pas le cas, nous avons proposé une extension de son application où les états sont devenus les répartitions elles-mêmes sur les états.

Le modèle de base de ce travail est la loi de Poisson. La particularisation de la théorie à ce cas donne des lois gamma qui doivent se substituer aux gaussiennes. En effet, le domaine de l'activité est la droite réelle positive et l'utilisation des lois normales à basse activité donne fréquemment une probabilité non négligeable à des valeurs négatives qui n'ont aucun sens physique. Les lois gamma sont impliquées directement par le modèle de Poisson par conjugaison et sont aussi les a priori de Jeffreys associés.

Les calculs complexes impliqués par l'approche bayesienne ont été rendus possibles par l'écriture d'une librairie ad-hoc car ils nécessitent une grande puissance informatique.

Perspectives techniques

L'application du théorème de Bayes aux suites d'observations produites par un état non stationnaire n'a été développée que dans le cas discret. Le cas d'espaces continus suppose un développement théorique plus important. Néanmoins, si les probabilités sur l'espace des états peuvent s'exprimer dans des bases dénombrables, le procédé du cas discret devrait pouvoir s'étendre sans trop de difficulté. En particulier, si l'espace des états est la droite réelle positive, les combinaisons linéaires de lois gamma recouvrent l'essentiel des cas pouvant apparaître en physique. Cette décomposition peut être ramenée à une décomposition en polynômes de Laguerre. D'un manière plus générale, il existe une relation entre les polynômes orthonormaux et les variables aléatoires qui pourrait ouvrir quelques pistes, par exemple Hermite-Gauss, Laguerre-Gamma ou Tchebitchev-bêta [43][44] [45].

Le calcul des a posteriori fait souvent intervenir des quadratures qui ne peuvent se calculer que numériquement. La précision exigée peut alors dépasser celle des types float standard (généralement 16 décimales). Il existe des librairies dites *en précision infinie* [38] qui permettent des évaluations plus fiables, mais au prix d'un ralentissement considérable de la vitesse de calcul. La méthode de Monte Carlo est aussi utilisable, mais le temps de calcul devient vite rédhibitoire, car l'écart-standard statistique ne diminue qu'avec la racine carrée de ce temps. Les cas où une solution analytique existe sont évidemment nettement plus favorables, tant du point de vue du calcul que de celui de la compréhension du phénomène. Dans le contexte de la mesure de l'activité, les modèles de base sont dérivés des lois poissonniennes, gamma, binomiales ou bêta, dont les interrelations n'ont probablement pas été assez exploitées.

C'est en fait le calcul de ces intégrales qui reste l'une des principales difficultés de cette approche. Son champ d'application a été restreint ici aux faibles activités, où des solutions par analyse numérique existent. A l'opposé on trouve les mesures à haute activité, pour lesquelles les approximations asymptotiques de la théorie classique sont suffisantes. Entre ces deux situations, tout reste à faire.

Perspectives théoriques

L'introduction de la théorie de l'information apporte un nouvel éclairage sur la réalité physique. Le fait que l'information de Fisher soit d'une part l'expression locale de celle de Shannon-Kullback, et d'autre part la base de l'a priori de Jeffreys, permet d'envisager des variétés riemanniennes informationnelles où l'on devrait pouvoir réinterpréter toutes les notions de la géométrie différentielle (dérivée covariante, courbures, etc). Cette approche n'est pas sans rappeler celle de la relativité générale. Les développements dans cette voie sont en cours depuis un certain temps déjà [24] [42].

Le modèle de la section B.2.2 reliant l'entropie de Shannon à celle de Kullback est un exemple où l'état ne peut être résolu. Il y a une maille minimale de l'espace des états au-dessous de laquelle l'observation ne peut rien apporter. Cette situation se retrouve lors de l'utilisation de géométries non

commutatives [46], qui introduisent un flou de base. La loi des grands nombres, qui assure la convergence d'un estimateur fréquentiste non biaisé vers la vraie valeur de l'état, se traduit par une évolution des a posteriori bayésiens vers des formes de plus en plus pointues, mais il n'est pas démontré que la limite est une distribution de Dirac.

Sixième partie

Remerciements

Institut de Radiophysique Appliquée

La plupart des applications proviennent du cadre de la radiométrie, où j'ai profité des compétences de nombreux collègues, en particulier Thierry Buchillier, Gilles Triscone, Marta Garcia-Talavera, Jean-Jacques Gostely et le regretté Marc Décombaz.

Loterie Romande

Le calcul des chances est trop important pour être laissé au hasard. C'est pourquoi je salue la Loterie Romande, qui m'a permis d'acquérir une expérience essentielle et concrète dans ce domaine.

Aurelio Bay

Le professeur Aurelio Bay, rapporteur de l'Unil, qui a suivi tout le développement de ce travail, m'a fait bénéficier de son implication physique à haut niveau d'énergie.

Marcel Py

Quelques temps après avoir commencé un groupe de travail sur les limites de détection avec Jean-François Valley et François Bochud, j'ai reçu un téléphone de Marcel Py :

Je suis en train d'organiser un cours sur le bruit dans le cadre de l'AVCP. Comme je me suis souvenu que tu faisais de la musique, j'ai pensé que tu serais intéressé à y participer.

Marcel avait probablement été marqué par quelques expériences saugrenues de musique contemporaine. C'est après ce cours qui a eu lieu à Vercorin en 2000, que j'ai eu l'idée de transformer ce travail en thèse.

Max-Olivier Hongler

Dans ce même séminaire j'ai croisé Max-Olivier Hongler, qui n'a pas économisé son temps comme expert externe. Quelques discussions musclées sur les processus stochastiques m'ont donné envie de faire une deuxième thèse ...

François Bochud et Jean-François Valley

Les mathématiques sont un art concis et précis. Sans l'apport du Prof. Jean-François Valley, directeur de thèse et celui du Dr. François Bochud, ce travail aurait eu à peu près l'allure présentée à la page suivante. Je doute que le rectorat de l'Université de Lausanne admette des thèses de moins d'une page, c'est pourquoi le nombre considérable de leurs remarques m'a été salutaire.

TBDS

Définition Un *espace de Bayes* est la donnée d'un quintuplet

$$\square = (\Theta, \mathcal{X}, P, \mathcal{D}, \pi)$$

où Θ , \mathcal{X} et \mathcal{D} sont des espaces mesurables, P une probabilité de transition de Θ vers \mathcal{X} et π une mesure positive sur Θ .

Théorème Soit $M = \int P(\cdot | \theta) \pi(d\theta)$. Il existe une unique probabilité de transition π_1 de \mathcal{X} vers Θ qui vérifie

$$\pi_1(d\theta | x) M(dx) = P(dx | \theta) \pi(d\theta)$$

La démonstration est triviale⁵.

Définition Soient $\mu << \nu$ deux probabilités sur un espace mesurable Ω , la *divergence de Kullback* est donnée par

$$D(\mu || \nu) = \int d\mu \ln \frac{\mu}{\nu}$$

La démonstration de l'existence de cette fonction est laissée au lecteur⁶.

Définition Si Θ est une variété différentiable, l'*information de Fisher* I est le tenseur donnant la variation au deuxième ordre de la probabilité de transition P au sens de la divergence de Kullback.

Définition La *mesure de Jeffreys* π_J associée à l'espace de Bayes \square est donnée par la densité volumétrique induite par l'information de Fisher.

Exemple Les espaces de Bayes peuvent être mis à profit pour le traitement des mesures de la radioactivité.

⁵Indication : utiliser le théorème de Radon-Nicodym.

⁶Voir note précédente.

Septième partie

Annexe

Annexe A

Technique

A.1 Semi-groupe binomial

La loi binomiale possède la propriété de composition suivante

$$\sum_m \mathcal{B}(n \mid m, p) \mathcal{B}(m \mid l, q) = \mathcal{B}(n \mid l, pq) \quad (\text{A.1})$$

En termes matriciels, désignons par $B(q)_{nm}$ la probabilité $\mathcal{B}(n \mid m, p)$. Avec le produit habituel des matrices, il vient

$$B(p)B(q) = B(pq) \quad (\text{A.2})$$

qui traduit une propriété dite de semi-groupe. Pour des raisons qui apparaîtront plus tard, nous définissons le *semi-groupe binomial* par

$$\mathcal{B}(\phi)_{nm} = \mathcal{B}(n \mid m, e^\phi) \quad (\text{A.3})$$

Ans, nous aurons

$$\mathcal{B}(\phi)\mathcal{B}(\psi) = \mathcal{B}(\phi + \psi) \quad (\text{A.4})$$

Démonstration La transformée de Fourier de la loi binomiale est

$$\mathcal{F}(\mathcal{B}(\mid N, p))(\xi) = = \left(1 + p(e^{i\xi} - 1)\right)^N \quad (\text{A.5})$$

Calculons la transformée de Fourier du produit matriciel

$$\sum_n e^{i\xi n} \sum_m \mathcal{B}(n \mid m, p) \mathcal{B}(m \mid l, q) = \quad (\text{A.6})$$

$$= \sum_m \sum_n e^{i\xi n} \mathcal{B}(n \mid m, p) \mathcal{B}(m \mid l, q) = \quad (\text{A.7})$$

$$= \sum_m \left(1 + p(e^{i\xi} - 1) \right)^m \mathcal{B}(m \mid l, q) = \quad (\text{A.8})$$

$$= \sum_m \left(1 + p(e^{i\xi} - 1) \right)^m \binom{l}{m} q^m (1-q)^{l-m} = \quad (\text{A.9})$$

$$= \sum_m \binom{l}{m} \left((1 + p(e^{i\xi} - 1)) q \right)^m (1-q)^{l-m} = \quad (\text{A.10})$$

$$= \left((1 + p(e^{i\xi} - 1)) q + (1-q) \right)^l = \quad (\text{A.11})$$

$$= \left(1 + pq(e^{i\xi} - 1) \right)^l \quad (\text{A.12})$$

qui est bien la transformée de Fourier de $\mathcal{B}(\mid l, pq)$.

D'un manière générale, la propriété de semi-groupe (A.4) conduit à une expression en exponentielle de matrice. Ici, définissons la matrice suivante

$$\mathbf{R} = \begin{pmatrix} 0 & -1 & & & \\ & 1 & -2 & & \\ & & 2 & -3 & \\ & & & \ddots & \end{pmatrix} \quad (\text{A.13})$$

On peut montrer que

$$\mathcal{B}(\phi) = \exp(-\phi \mathbf{R}) \quad (\text{A.14})$$

A.2 Composition d'une loi gamma et d'une loi de Poisson

Soit une activité a répartie selon une loi gamma $\text{Ga}(da \mid N_f, t_f)$, alors le nombre de désintégrations produites par cette activité durant un temps t suit la loi négative-binomiale

$$\mathcal{NB}(N \mid N_f, p_f) = \frac{\Gamma(N_f + N)}{\Gamma(N_f) \Gamma(N + 1)} p_f^{N_f} (1 - p_f)^N \quad (\text{A.15})$$

où

$$p_f = \frac{t_f}{t_f + t} \quad (\text{A.16})$$

Démonstration

$$P(N \mid N_f, t_f) = \int_a \mathcal{P}(N \mid a) \text{Ga}(da \mid N_f, t_f) \quad (\text{A.17})$$

$$= \int_a \frac{a^N}{N!} e^{-at} \frac{t_f^{N_f}}{\Gamma(N_f)} a^{N_f-1} e^{-at_f} da \quad (\text{A.18})$$

$$= \frac{t^N t_f^{N_f}}{N! \Gamma(N_f)} \int_a a^{N+N_f-1} e^{-a(t+t_f)} da \quad (\text{A.19})$$

$$= \frac{t^N t_f^{N_f}}{N! \Gamma(N_f)} \frac{\Gamma(N_f + N)}{(t+t_f)^{N_f+N}} \quad (\text{A.20})$$

$$= \frac{\Gamma(N_f + N)}{\Gamma(N_f) \Gamma(N+1)} \left(\frac{t}{t_f+t} \right)^N \left(\frac{t_f}{t_f+t} \right)^{N_f} \quad (\text{A.21})$$

A.3 Composition de deux lois gamma

Soit un activité a répartie selon une loi gamma de paramètres (N_f, t_f) et t le temps d'attente de la N^e désintégration, alors la variable $p = \frac{t_f}{t_f+t}$ suit la loi bêta

$$P(dp \mid N_f, t_f) = \text{Be}(dp \mid N_f, N) \quad (\text{A.22})$$

Démonstration

$$P(dt \mid N_f, t_f) = \int_a \text{Ga}(dt \mid N, a) \text{Ga}(da \mid N_f, t_f) \quad (\text{A.23})$$

$$= dt \int_a \frac{a^N}{\Gamma(N)} t^{N-1} e^{-at} \frac{t_f^{N_f}}{\Gamma(N_f)} a^{N_f-1} e^{-at_f} da \quad (\text{A.24})$$

$$= dt t^{N-1} \frac{t_f^{N_f}}{\Gamma(N) \Gamma(N_f)} \int_a a^{N+N_f-1} e^{-a(t_f+t)} da \quad (\text{A.25})$$

$$= dt t^{N-1} \frac{t_f^{N_f}}{\Gamma(N) \Gamma(N_f)} \frac{\Gamma(N_f + N)}{(t_f+t)^{N_f+N}} \quad (\text{A.26})$$

$$= \frac{\Gamma(N_f + N)}{\Gamma(N_f) \Gamma(N)} \left(\frac{t_f}{t_f+t} \right)^{N_f-1} \left(\frac{t}{t_f+t} \right)^{N-1} \frac{t_f dt}{(t_f+t)^2} \quad (\text{A.27})$$

et le résultat s'obtient par le changement de variable $p = \frac{t_f}{t_f+t}$ qui implique $dp = -\frac{t_f dt}{(t_f+t)^2}$.

A.4 Perte d'information par changement de variable

Montrons que l'information moyenne apportée par une mesure ne peut pas augmenter lors d'une transformation de l'espace des observations. Soit un modèle $P : \Theta \rightsquigarrow \mathcal{X}$. Une fonction $f : \mathcal{X} \rightarrow \mathcal{X}'$ induit un modèle $P' : \Theta \rightsquigarrow \mathcal{X}'$ par $P'(dx' \mid \theta) = P(f^{-1}(dx) \mid \theta)$. Alors pour tout a priori $\pi(d\theta)$, on a

$$I'_H(\pi(d\theta)) \leq I_H(\pi(d\theta)) \quad (\text{A.28})$$

Lemme 1

$$M'(dx') \stackrel{\text{def}}{=} \int_{\Theta} P'(dx' | \theta) \pi(d\theta) = \int_{\Theta} P(f^{-1}(dx') | \theta) \pi(d\theta) = M(f^{-1}(dx')) \quad (\text{A.29})$$

Lemme 2

$$\pi(d\theta | x') \stackrel{\text{def}}{=} \frac{P'(dx' | \theta) \pi(d\theta)}{M'(dx')} \quad (\text{A.30})$$

$$= \frac{\int_{x \in f^{-1}(dx')} P(dx | \theta) \pi(d\theta)}{M'(dx')} \quad (\text{A.31})$$

$$= \frac{\int_{x \in f^{-1}(dx')} \pi(d\theta | x) M(dx)}{M(f^{-1}(dx'))} \quad (\text{A.32})$$

$$= \int_x \pi(d\theta | x) M(dx | f^{-1}(dx')) \quad (\text{A.33})$$

Lemme 3

$$M(dx) = \int_{x'} M(dx | f^{-1}(x')) M(f^{-1}(dx')) = \int_{x'} M(dx | f^{-1}(x')) M'(dx') \quad (\text{A.34})$$

Démonstration de l'inégalité La divergence logarithmique est convexe (cf 6.5), on a donc

$$D\left(\int_x \pi(d\theta | x) M(dx | f^{-1}(x')) || \pi(d\theta)\right) \leq \int_x M(dx | f^{-1}(x')) D(\pi(d\theta | x) || \pi(d\theta)) \quad (\text{A.35})$$

Intégrons les deux termes de l'inégalité par rapport $M'(dx')$. Il vient

$$\int M'(dx') D(\pi(d\theta | x') || \pi(d\theta)) \leq \int_{x'} M(dx') \int_x M(dx | f^{-1}(x')) D(\pi(d\theta | x) || \pi(d\theta)) \quad (\text{A.36})$$

et le lemme 3 donne le résultat par croisement des intégrales.

A.5 Divergence logarithmique et géodésiques

Selon Dabak [47], soient p_0 et p_1 deux densités de probabilité sur Ω . On définit la *densité exponentielle croisée* par

$$p_t = \frac{p_0^{1-t} p_1^t}{N_t} \quad (\text{A.37})$$

$$N_t = \int_{\Omega} p_0(\omega)^{1-t} p_1(\omega)^t d\omega \quad (\text{A.38})$$

Cette famille de probabilités dépendant de t donne une information de Fisher

$$I(t) = \int \left(\frac{d \ln p_t(\omega)}{dt} \right)^2 p_t(\omega) d\omega \quad (\text{A.39})$$

Alors, on a les relations entre la divergence de Kullback et l'information de Fisher

$$\frac{\partial^2 D(p_t || p_u)}{\partial t^2} = I(t) + (t - u) I'(t) \quad (\text{A.40})$$

$$D(p_1 || p_0) = \int_0^1 u I(u) du \quad (\text{A.41})$$

A.6 Divergence logarithmique de deux lois multinormales

Soient

$$\mathcal{N}_i = \mathcal{N}(\mu_i, \Sigma_i), \quad i = 0, 1 \quad (\text{A.42})$$

deux lois multinormales de dimension n de moyennes μ_i et de matrices de covariance Σ_i . Alors

$$2D(\mathcal{N}_1 \parallel \mathcal{N}_0) = \ln \frac{|\Sigma_0|}{|\Sigma_1|} - n + \text{Tr}(\Sigma_0^{-1} \Sigma_1) + {}^t(\mu_1 - \mu_0) \Sigma_0^{-1} (\mu_1 - \mu_0) \quad (\text{A.43})$$

Démonstration

$$D(\mathcal{N}_1 \parallel \mathcal{N}_0) = \int \frac{dx}{(2\pi)^{n/2} |\Sigma_1|^{1/2}} e^{-\frac{1}{2} {}^t(x-\mu_1) \Sigma_1^{-1} (x-\mu_1)} \ln \frac{\frac{1}{(2\pi)^{n/2} |\Sigma_1|^{1/2}} e^{-\frac{1}{2} {}^t(x-\mu_1) \Sigma_1^{-1} (x-\mu_1)}}{\frac{1}{(2\pi)^{n/2} |\Sigma_0|^{1/2}} e^{-\frac{1}{2} {}^t(x-\mu_0) \Sigma_0^{-1} (x-\mu_0)}} \quad (\text{A.44})$$

$$\begin{aligned} &= \frac{1}{2} \ln \frac{|\Sigma_0|}{|\Sigma_1|} + \\ &- \frac{1}{2} \int \frac{dx}{(2\pi)^{n/2} |\Sigma_1|^{1/2}} e^{-\frac{1}{2} {}^t(x-\mu_1) \Sigma_1^{-1} (x-\mu_1)} {}^t(x-\mu_1) \Sigma_1^{-1} (x-\mu_1) + \\ &+ \frac{1}{2} \int \frac{dx}{(2\pi)^{n/2} |\Sigma_1|^{1/2}} e^{-\frac{1}{2} {}^t(x-\mu_1) \Sigma_1^{-1} (x-\mu_1)} {}^t(x-\mu_0) \Sigma_0^{-1} (x-\mu_0) \end{aligned} \quad (\text{A.45})$$

Effectuons le changement de variables

$$y = \sqrt{\Sigma_1^{-1}} x \quad (\text{A.46})$$

La deuxième partie du second membre de (A.45) s'écrit¹.

$$M_2 = -\frac{1}{2} \int \frac{dy}{(2\pi)^{n/2}} {}^t y y e^{-\frac{1}{2} {}^t y y} = -\frac{n}{2} \quad (\text{A.47})$$

La troisième devient

$$M_3 = +\frac{1}{2} \int \frac{dx}{(2\pi)^{n/2}} e^{-\frac{1}{2} {}^t y y} (\sqrt{\Sigma_1} y + \mu_1 - \mu_0) \Sigma_0^{-1} (\sqrt{\Sigma_1} y + \mu_1 - \mu_0) \quad (\text{A.48})$$

Développons le produit interne

$$\begin{aligned} {}^t (\sqrt{\Sigma_1} y + \mu_1 - \mu_0) \Sigma_0^{-1} (\sqrt{\Sigma_1} y + \mu_1 - \mu_0) &= \\ &{}^t y \sqrt{\Sigma_1} \Sigma_0^{-1} \sqrt{\Sigma_1} y + \\ &+ 2 {}^t y \sqrt{\Sigma_1} \Sigma_0^{-1} (\mu_1 - \mu_0) + \\ &+ {}^t (\mu_1 - \mu_0) \Sigma_0^{-1} (\mu_1 - \mu_0) \end{aligned} \quad (\text{A.49})$$

On vérifie facilement que le premier terme intégré donne $\text{Tr}(\sqrt{\Sigma_1} \Sigma_0^{-1} \sqrt{\Sigma_1})$, qui devient $\text{Tr}(\Sigma_0^{-1} \Sigma_1)$ par invariance de la trace par rotation interne. L'intégrale du deuxième est nulle, et le troisième est constant.

¹Les Σ et $\sqrt{\Sigma}$ sont symétriques

A.7 Divergence logarithmique de deux lois de Poisson

Soient

$$\mathcal{P}_i = \mathcal{P}(\alpha_i), \quad i = 0, 1 \quad (\text{A.50})$$

deux lois de Poisson. Alors

$$D(\mathcal{P}_1 \parallel \mathcal{P}_0) = \alpha_1 \ln \frac{\alpha_1}{\alpha_0} - (\alpha_1 - \alpha_0) \quad (\text{A.51})$$

Démonstration

$$D(\mathcal{P}_1 \parallel \mathcal{P}_0) = \sum_n \frac{\alpha_1^n}{n!} e^{-\alpha_1} \ln \frac{\frac{\alpha_1^n}{n!} e^{-\alpha_1}}{\frac{\alpha_0^n}{n!} e^{-\alpha_0}} \quad (\text{A.52})$$

$$= \sum_n \frac{\alpha_1^n}{n!} e^{-\alpha_1} n \ln \frac{\alpha_1}{\alpha_0} + \sum_n \frac{\alpha_1^n}{n!} e^{-\alpha_1} (-\alpha_1 + \alpha_0) \quad (\text{A.53})$$

$$= \alpha_1 \ln \frac{\alpha_1}{\alpha_0} - (\alpha_1 - \alpha_0) \quad (\text{A.54})$$

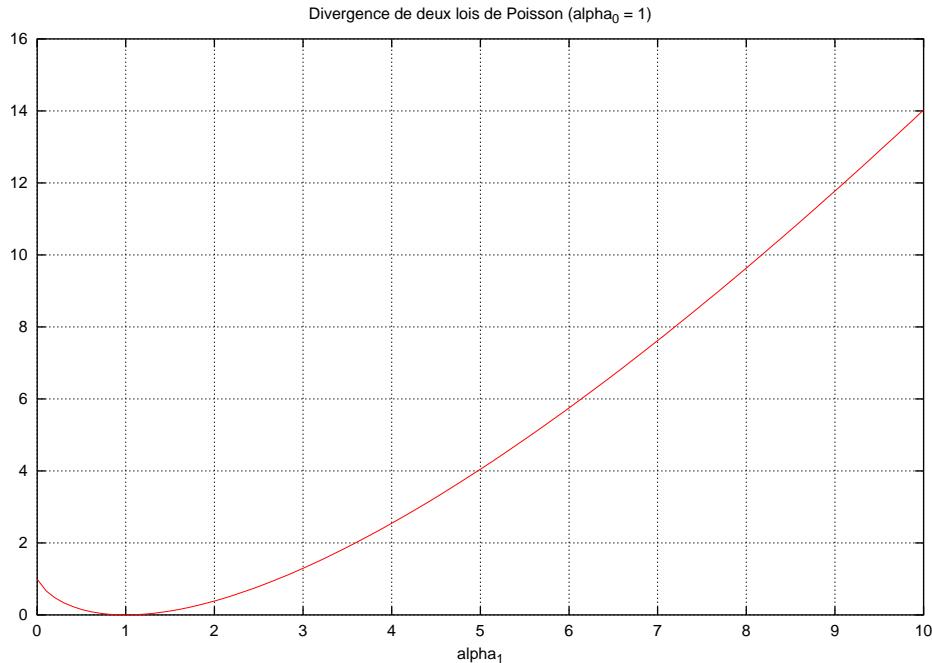


FIG. A.1 – Divergence de deux lois de Poisson, $\alpha_0 = 1$

A.8 Divergence logarithmique de deux lois gamma

Soient

$$\text{Ga}_i = \text{Ga}(N_i, t_i), \quad i = 0, 1 \quad (\text{A.55})$$

deux lois gamma. Alors

$$D(\text{Ga}_1 \parallel \text{Ga}_0) = N_0 \ln \frac{t_1}{t_0} - N_1 \left(1 - \frac{t_0}{t_1} \right) + \ln \Gamma(N_0) - \ln \Gamma(N_1) + (N_1 - N_0) \Psi(N_1) \quad (\text{A.56})$$

Lemme Si $\Psi = (\ln \Gamma)'$ est la fonction *digamma*, on a

$$\int \text{Ga}(db \mid N, t) \ln b = \Psi(N) - \ln t \quad (\text{A.57})$$

Par définition et dérivation par rapport à N

$$\Gamma(N) = \int dx x^{N-1} e^{-x} \quad (\text{A.58})$$

$$\Psi(N) = \int dx \ln x^{N-1} e^{-x} \quad (\text{A.59})$$

Donc, avec le changement de variable $x = bt$

$$\int \text{Ga}(db \mid N, t) \ln b = \int db \frac{t^N}{\Gamma(N)} b^{N-1} e^{-bt} \ln b \quad (\text{A.60})$$

$$= \int dx \frac{1}{\Gamma(N)} x^{N-1} (\ln x - \ln t) \quad (\text{A.61})$$

$$= \Psi(N) - \ln t \quad (\text{A.62})$$

Qed

Passons au calcul de la divergence des lois gamma

$$D(\text{Ga}_1 \parallel \text{Ga}_0) = \int db \frac{t_1^{N_1}}{\Gamma(N_1)} b^{N_1-1} e^{-bt_1} \ln \frac{\frac{t_1^{N_1}}{\Gamma(N_1)} b^{N_1-1} e^{-bt_1}}{\frac{t_0^{N_0}}{\Gamma(N_0)} b^{N_0-1} e^{-bt_0}} \quad (\text{A.63})$$

$$= \int db \frac{t_1^{N_1}}{\Gamma(N_1)} b^{N_1-1} e^{-bt_1} (N_1 \ln t_1 - N_0 \ln t_0 - \ln \Gamma(N_1) + \ln \Gamma(N_0) + (N_1 - N_0) \ln b - b(t_1 - t_0)) \quad (\text{A.64})$$

$$= N_0 \ln \frac{t_1}{t_0} - N_1 \left(1 - \frac{t_0}{t_1} \right) + \ln \Gamma(N_0) - \ln \Gamma(N_1) + (N_1 - N_0) \Psi(N_1) \quad (\text{A.65})$$

Structurellement

$$D(\text{Ga}_1 \parallel \text{Ga}_0) = N_0 D \left(\mathcal{P} \left(\frac{N_1}{N_0} \right) \parallel \mathcal{P} \left(\frac{N_1/t_1}{N_0/t_0} \right) \right) + \ln \Gamma(N_0) - (\ln \Gamma(N_1) + (N_0 - N_1) \Psi(N_1)) \quad (\text{A.66})$$

Remarque Si $N_s = N_0(1-s) + N_1s$ et $t_s = t_0(1-s) + t_1s$, alors la *densité exponentielle croisée* de Ga_0 et Ga_1 est Ga_s . On peut appliquer la relation de la section (A.5).

A.9 Divergence logarithmique de deux lois bêta

Soient

$$\text{Be}_i = \text{Be}(\alpha_i, \beta_i), \quad i = 0, 1 \quad (\text{A.67})$$

deux lois bêta. Alors

$$\begin{aligned} D(\text{Be}_1 \parallel \text{Be}_0) &= \ln \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} - \ln \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \\ &+ (\alpha_1 - \alpha_0)\Psi(\alpha_1) + (\beta_1 - \beta_0)\Psi(\beta_1) \\ &- ((\alpha_1 + \beta_1) - (\alpha_0 + \beta_0))\Psi(\alpha_1 + \beta_1) \end{aligned} \quad (\text{A.68})$$

Lemme

$$\int \text{Be}(dp \mid \alpha, \beta) \ln p = \Psi(\alpha) - \Psi(\alpha + \beta) \quad (\text{A.69})$$

$$\int \text{Be}(dp \mid \alpha, \beta) \ln(1 - p) = \Psi(\beta) - \Psi(\alpha + \beta) \quad (\text{A.70})$$

Par définition et dérivation par rapport à α

$$\int p^{\alpha-1}(1-p)^{\beta-1}dp = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad (\text{A.71})$$

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int \ln p p^{\alpha-1}(1-p)^{\beta-1}dp = \Psi(\alpha) - \Psi(\alpha + \beta) \quad (\text{A.72})$$

La seconde relation s'obtient par $\alpha \leftrightarrow \beta$, et $p \leftrightarrow 1 - p$

Passons au calcul de la divergence des lois bêta

$$D(\text{Be}_1 \parallel \text{Be}_0) = \int dp \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} p^{\alpha_1-1}(1-p)^{\beta_1-1} \ln \frac{\frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} p^{\alpha_1-1}(1-p)^{\beta_1-1}}{\frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} p^{\alpha_0-1}(1-p)^{\beta_0-1}} \quad (\text{A.73})$$

$$\begin{aligned} &= \ln \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \frac{\Gamma(\alpha_0)\Gamma(\beta_0)}{\Gamma(\alpha_0 + \beta_0)} \\ &+ \int \text{Be}(dp \mid \alpha_1, \beta_1) ((\alpha_1 - \alpha_0) \ln p + (\beta_1 - \beta_0) \ln(1 - p)) \end{aligned} \quad (\text{A.74})$$

$$\begin{aligned} &= \ln \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \frac{\Gamma(\alpha_0)\Gamma(\beta_0)}{\Gamma(\alpha_0 + \beta_0)} \\ &+ (\alpha_1 - \alpha_0)(\Psi(\alpha_1) - \Psi(\alpha_1 + \beta_1)) \\ &+ (\beta_1 - \beta_0)(\Psi(\beta_1) - \Psi(\alpha_1 + \beta_1)) \end{aligned} \quad (\text{A.75})$$

$$\begin{aligned} &= \ln \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \frac{\Gamma(\alpha_0)\Gamma(\beta_0)}{\Gamma(\alpha_0 + \beta_0)} \\ &+ (\alpha_1 - \alpha_0)\Psi(\alpha_1) + (\beta_1 - \beta_0)\Psi(\beta_1) \\ &- ((\alpha_1 + \beta_1) - (\alpha_0 + \beta_0))\Psi(\alpha_1 + \beta_1) \end{aligned} \quad (\text{A.76})$$

Remarque Si $\alpha_s = \alpha_0(1 - s) + \alpha_1s$ et $\beta_s = \beta_0(1 - s) + \beta_1s$, alors la *densité exponentielle croisée* de Be_0 et Be_1 est Be_s . On peut appliquer la relation de la section (A.5).

A.10 Information moyenne apportée par une mesure poissonnienne d'a priori gamma

Soit le modèle poissonnien associé à un comptage de durée t avec une sensibilité ε

$$P(N | a) = \mathcal{P}(N | a\varepsilon t) \quad (\text{A.77})$$

sous un a priori gamma

$$\pi(da) = \text{Ga}(da | N_f, t_f) \quad (\text{A.78})$$

l'information moyenne apportée par une mesure est donnée par

$$\begin{aligned} I_H(N_f, t_f) &= N_f \ln \left(\frac{t_f + \varepsilon t}{t_f} \right) + \ln \Gamma(N_f) - N_f \frac{\varepsilon t}{t_f} \\ &+ \sum_N (N \psi(N_f + N) - \ln \Gamma(N_f + N)) \mathcal{NB} \left(N | N_f, \frac{t_f}{t_f + \varepsilon t} \right) \end{aligned} \quad (\text{A.79})$$

où $\psi = (\ln \Gamma)'$ est la fonction digamma et \mathcal{NB} la répartition négative-binomiale.

Démonstration

D'un manière générale,

$$I_H(\pi) = \int_x P(dx | \pi) D(\pi(\cdot | x) || \pi) \quad (\text{A.80})$$

Calculons tout d'abord la répartition marginale des comptages

$$P(N | \text{Ga}(N_f, t_f)) = \int_a \mathcal{P}(N | a\varepsilon t) \text{Ga}(da | N_f, t_f) \quad (\text{A.81})$$

$$= \int_a \frac{(a\varepsilon t)^N}{N!} e^{-a\varepsilon t} \frac{t_f^{N_f}}{\Gamma(N_f)} a^{N_f-1} e^{-at_f} da \quad (\text{A.82})$$

$$= \frac{(\varepsilon t)^N t_f^{N_f}}{N! \Gamma(N_f)} \int_a a^N e^{-a(\varepsilon t + t_f)} da \quad (\text{A.83})$$

$$= \frac{(\varepsilon t)^N t_f^{N_f}}{N! \Gamma(N_f)} \frac{\Gamma(N_f + N)}{(\varepsilon t + t_f)^{N_f + N}} \quad (\text{A.84})$$

$$= \frac{\Gamma(N_f + N)}{\Gamma(N_f) \Gamma(N + 1)} \left(\frac{\varepsilon t}{\varepsilon t + t_f} \right)^N \left(\frac{t_f}{\varepsilon t + t_f} \right)^{N_f} \quad (\text{A.85})$$

$$= \mathcal{NB} \left(N | N_f, \frac{t_f}{t_f + \varepsilon t} \right) \quad (\text{A.86})$$

La divergence de deux lois gamma est données par l'équation (A.65)

$$\begin{aligned} D(\text{Ga}(N_f + N, t_f + \varepsilon t) || \text{Ga}(N_f, t_f)) &= N_f \ln \frac{t_f + \varepsilon t}{t_f} - (N_f + N) \left(1 - \frac{t_f}{t_f + \varepsilon t} \right) \\ &+ \ln \Gamma(N_f) - \ln \Gamma(N_f + N) + N \Psi(N_f + N) \end{aligned} \quad (\text{A.87})$$

Mais l'espérance de la négative-binomiale donne

$$\sum_N N \mathcal{NB} \left(N | N_f, \frac{t_f}{t_f + \varepsilon t} \right) = N_f \frac{\varepsilon t}{t_f} \quad (\text{A.88})$$

d'où le résultat.

A.11 Information moyenne apportée par une mesure de temps d'attente d'a priori gamma

Soit le modèle gamma associé à une attente de la N^e détection avec une sensibilité ε

$$P(dt | a) = \text{Ga}(dt | N, a\varepsilon) \quad (\text{A.89})$$

sous un a priori gamma

$$\pi(da) = \text{Ga}(da | N_f, t_f) \quad (\text{A.90})$$

l'information moyenne apportée par une mesure est donnée par

$$I_H(N_f, t_f) = \ln \Gamma(N_f) - \ln \Gamma(N_f + N) - N_f \psi(N_f) + (N_f + N) \psi(N_f + N) - N \quad (\text{A.91})$$

Démonstration

La marginale sur les observations est

$$P(dt) = \frac{\Gamma(N_f + N)}{\Gamma(N) \Gamma(N_f)} \frac{t_f^{N_f} t^{N-1}}{(t_f + t)^{N_f + N}} dt \quad (\text{A.92})$$

$$= \frac{\Gamma(N_f + N)}{\Gamma(N) \Gamma(N_f)} \left(\frac{t_f}{t_f + t} \right)^{N_f - 1} \left(\frac{t}{t_f + t} \right)^{N-1} \frac{t_f}{(t_f + t)^2} dt \quad (\text{A.93})$$

Posant $\tau = \frac{t}{t_f + t}$ on a

$$P(d\tau) = \text{Be}(d\tau | N, N_f) \quad (\text{A.94})$$

Reprendons l'équation (A.65) en remplaçant par τ

$$\begin{aligned} D(\text{Ga}(N_f + N, t_f + t) || \text{Ga}(N_f, t_f)) &= -N_f \ln(1 - \tau) + \ln \Gamma(N_f) - \ln \Gamma(N_f + N) \\ &\quad + N \psi(N_f + N) - \tau(N_f + N) \end{aligned} \quad (\text{A.95})$$

Or

$$\int \ln(1 - \tau) \text{Be}(d\tau | N, N_f) = \psi(N_f) - \psi(N_f + N) \quad (\text{A.96})$$

L'information moyenne vaut donc

$$\begin{aligned} I_H(N_f, t_f) &= -N_f \int_0^1 \ln(1 - \tau) \text{Be}(d\tau | N, N_f) \\ &\quad + \ln \Gamma(N_f) - \ln \Gamma(N_f + N) + N \psi(N_f + N) - N \end{aligned} \quad (\text{A.97})$$

$$\begin{aligned} &= -N_f (\psi(N_f) - \psi(N_f + N)) \\ &\quad + \ln \Gamma(N_f) - \ln \Gamma(N_f + N) + N \psi(N_f + N) - N \end{aligned} \quad (\text{A.98})$$

$$= \ln \Gamma(N_f) - N_f \psi(N_f) - \ln \Gamma(N_f + N) + (N_f + N) \psi(N_f + N) - N \quad (\text{A.99})$$

Qed.

A.12 Matrice de Fisher pour le problème à deux voies

Soit la matrice

$$I = \begin{pmatrix} 0 & \beta_1 & \beta_2 & \dots & \beta_d \\ \alpha_1 & \eta_1(\alpha_1 - \beta_1) & 0 & \dots & 0 \\ \alpha_1 & 0 & \eta_1(\alpha_1 - \beta_1) & \dots & 0 \\ \dots & & & & \\ \alpha_d & 0 & 0 & \dots & 0 \end{pmatrix} \quad (\text{A.100})$$

Alors

$$|I| = (d-1)! \sum_k \frac{1}{\eta_k \left(\frac{1}{\alpha_k} - \frac{1}{\beta_k} \right)} \cdot \prod_k \eta_k \alpha_k \beta_k \left(\frac{1}{\alpha_k} - \frac{1}{\beta_k} \right) \quad (\text{A.101})$$

Démonstration Le calcul a été effectué par le logiciel Maple *currie_ifish.mws*

A.13 Calcul de l'information de Fisher pour le modèle normal à un seul paramètre

Pour un espace d'observations réel monodimensionnel, le modèle normal est caractérisé par un état à deux dimensions qui sont sa moyenne et sa variance

$$\theta = (\mu, \nu) \quad (\text{A.102})$$

$$P(dx | \mu, \nu) = \frac{1}{\sqrt{2\pi\nu}} e^{\frac{(x-\mu)^2}{2\nu}} dx \quad (\text{A.103})$$

L'information de Fisher pour ce modèle est donnée par

$$\partial_\mu \ln P = \frac{x-\mu}{\nu} \quad (\text{A.104})$$

$$\partial_\nu \ln P = -\frac{1}{2\nu} + \frac{(x-\mu)^2}{2\nu^2} \quad (\text{A.105})$$

$$\partial_\mu^2 \ln P = -\frac{1}{\nu} \quad (\text{A.106})$$

$$\partial_\nu^2 \ln P = \frac{1}{2\nu^2} - \frac{(x-\mu)^2}{\nu^3} \quad (\text{A.107})$$

$$\partial_{\mu\nu}^2 \ln P = -\frac{x-\mu^2}{\nu^2} \quad (\text{A.108})$$

$$(\text{A.109})$$

Intégrons les dérivées secondes sur x , il vient

$$I = \begin{pmatrix} \frac{1}{\nu} & 0 \\ 0 & \frac{1}{2\nu^2} \end{pmatrix} \quad (\text{A.110})$$

La métrique de Fisher (ou son expression locale en termes de divergence logarithmique) prend la forme

$$2D(P(dx | \mu + d\mu, \nu + d\nu) || P(dx | \mu, \nu)) = \frac{d\mu^2}{\nu} + \frac{d\nu^2}{2\nu^2} \quad (\text{A.111})$$

Si θ dépend en fait d'un état physique sous-jacent θ_0 par l'intermédiaire de deux fonctions $\mu(\theta_0)$ et $\nu(\theta_0)$, le modèle à considérer est

$$P_0(dx | \theta_0) = P(dx | \mu(\theta_0), \nu(\theta_0)) \quad (\text{A.112})$$

La métrique de Fisher étant intrinsèque (c'est-à-dire invariante par changement de coordonnées), on obtient immédiatement sa valeur pour le modèle P_0

$$I_0(\theta_0) = \frac{\mu'(\theta_0)^2}{\nu(\theta_0)} + \frac{\nu'(\theta_0)^2}{2\nu(\theta_0)^2} \quad (\text{A.113})$$

et il vient

$$I_0 = \frac{\mu'^2}{\nu} + \frac{1}{2} \left(\frac{\nu'}{\nu} \right)^2 \quad (\text{A.114})$$

A.14 Détermination du modèle pour l'estimation de la constante radioactive

La probabilité qu'un atome existant au temps $t = 0$ effectue sa désintégration dans l'intervalle dt de \mathbb{R}_+ est donnée par $\lambda e^{-\lambda t} dt$. La probabilité p_1 qu'il a de se désintégrer dans l'intervalle de mesure s'obtient par intégration de cette densité

$$p_1 = \int_d^f \lambda e^{-\lambda t} dt = e^{-\lambda d} - e^{-\lambda f} \quad (\text{A.115})$$

Pour N atomes initiaux, la probabilité d'obtenir \tilde{n} désintégrations pendant la mesure est alors donné par la binomiale

$$P(\tilde{n} | N) = \mathcal{B}(\tilde{n} | N, p_1) = \binom{N}{\tilde{n}} p_1^{\tilde{n}} (1 - p_1)^{N - \tilde{n}} \quad (\text{A.116})$$

De plus, chacune de ces désintégrations est détectée avec une probabilité ε , donc pour \tilde{n} désintégrations, la probabilité d'obtenir n détections est aussi une binomiale

$$P(n | \tilde{n}) = \mathcal{B}(n | \tilde{n}, \varepsilon) \quad (\text{A.117})$$

Utilisant une propriété bien connue des lois binomiales, nous avons

$$P(n | N) = \sum_{\tilde{n}} P(n | \tilde{n}) P(\tilde{n} | N) = \sum_{\tilde{n}} \mathcal{B}(n | \tilde{n}, \varepsilon) \mathcal{B}(\tilde{n} | N, p_1) = \mathcal{B}(n | N, \varepsilon p_1) \quad (\text{A.118})$$

Mais ces détections doivent avoir lieu au bon moment. Il vient

$$P(t_1, \dots, t_n | N, \lambda) = \mathcal{B}(n | N, \varepsilon p_1(\lambda)) \prod_{i=1}^n \frac{\lambda e^{-\lambda t_i}}{p_1(\lambda)} dt_i \quad (\text{A.119})$$

A.15 Calcul de l'a posteriori pour la constante radioactive à l'aide de l'approximation de Laplace

Pour le calcul du m^e moment a posteriori de λ , nous utiliserons l'approximation dite de *Laplace* présentée par Bernardo [10] p340-1. Avec ses notations, nous avons

$$p(t | \lambda) = \left(\frac{\lambda e^{-\lambda \bar{t}}}{p_1} \right)^n \quad (\text{A.120})$$

$$p(\lambda) = \frac{t_f^{N_f}}{\Gamma(N_f)} \lambda^{N_f-1} e^{-\lambda t_f} \quad (\text{A.121})$$

$$g(\lambda) = \lambda^m \quad (\text{A.122})$$

La fonction h faisant office de logarithme de fonction de vraisemblance s'exprime

$$h(\lambda) = -\frac{1}{n} [n(\ln \lambda - \lambda \bar{t} - \ln p_1) + (N_f - 1 + m) \ln \lambda - \lambda t_f + N_f \ln t_f - \ln \Gamma(N_f)] \quad (\text{A.123})$$

Et nous pouvons écrire

$$u = \frac{N_f - 1 + m}{n} + 1 \quad (\text{A.124})$$

$$v = \bar{t} + \frac{t_f}{n} \quad (\text{A.125})$$

$$h = -u \ln \lambda + \lambda v + \ln p_1 - \frac{N_f \ln t_f - \ln \Gamma(N_f)}{n} \quad (\text{A.126})$$

$$h' = -\frac{u}{\lambda} + v + (\ln p_1)' \quad (\text{A.127})$$

$$h'' = \frac{u}{\lambda^2} + (\ln p_1)'' \quad (\text{A.128})$$

Désignons par \hat{h} la fonction h pour $m = 0$ et h^* celle avec un m donné, nous pouvons obtenir $\hat{\lambda}$ par résolution de $\hat{h}' = 0$ et λ^* par $h^{*'} = 0$. L'approximation de Laplace donne ensuite

$$\hat{\sigma} = \hat{h}''(\hat{\lambda})^{-\frac{1}{2}} \quad (\text{A.129})$$

$$\sigma^* = h^{*''}(\lambda^*)^{-\frac{1}{2}} \quad (\text{A.130})$$

$$E_1(\lambda^m) \simeq \frac{\sigma^*}{\hat{\sigma}} \exp \left[-n \left(h^*(\lambda^*) - \hat{h}(\hat{\lambda}) \right) \right] \quad (\text{A.131})$$

Le calcul est simplifié par

$$\mu = \lambda(f - d) \quad (\text{A.132})$$

$$\tilde{v} = \frac{f - v}{f - d} \quad (\text{A.133})$$

$$\frac{h'}{f - d} = \frac{1}{1 - e^{-\mu}} - \frac{u}{\mu} - \tilde{v} \quad (\text{A.134})$$

$$\frac{h''}{f - d} = \frac{d}{d\mu} \frac{h'}{f - d} \frac{d\mu}{d\lambda} \quad (\text{A.135})$$

$$= \left(\frac{-e^{-\mu}}{(1 - e^{-\mu})^2} + \frac{u}{\mu^2} \right) (f - d) \quad (\text{A.136})$$

Et le zéro de l'équation (A.134) s'obtient facilement par itération

$$\frac{u}{\mu} = \frac{1}{1 - e^{-\mu}} - \tilde{v} \quad (\text{A.137})$$

Retenant les définitions de u et \tilde{v} , définissons une fonction générique

$$H_u(\mu) = \frac{1}{1 - e^{-\mu}} - \frac{u}{\mu} \quad (\text{A.138})$$

il ne reste plus qu'à discuter de la résolubilité de

$$\tilde{v} = H_u(\mu) \quad (\text{A.139})$$

$$u = \frac{N_f - 1 + m}{n} + 1 \quad (\text{A.140})$$

$$\tilde{v} = \frac{f - \left(\bar{t} + \frac{t_f}{n}\right)}{f - d} \quad (\text{A.141})$$

La solution obtenue par maximum de vraisemblance peut s'obtenir par limite de la loi gamma avec $N_f = 1$ and $t_f = 0$. On peut remarquer que $n \rightarrow \infty$ aboutit au même résultat, ce qui permet de retrouver la convergence des solutions bayesiennes vers les solutions fréquentistes.

Si $N_f > 1$, nous avons $u > 1$ et $H(0) = -\infty$ and $H(\infty) = 1$. Mais $d < \bar{t} < f$ implique que le domaine de variation de \tilde{v} est $]-\frac{t_f}{n(f-d)}, 1 - \frac{t_f}{n(f-d)}[$, donnant toujours une solution. Avec un a priori gamma plus étendu, il est possible que $u < 1$, et dans ce cas le sup h ne peut s'obtenir par l'annulation des dérivées que pour des \bar{t} proches de d^2 .

A.16 Elimination du bruit de fond par itération pour la détermination de la constante radioactive

Soit n le nombre de détections dues à source et au bruit de fond. On effectue K fois³ la procédure suivante

- Tirer au hasard n_b selon la poissonienne $\mathcal{P}(n_b | b(f - d)) = q$ et retenir la probabilité q
- Tirer au hasard successivement n_b points uniformément dans l'intervalle $[d, f]$ et éliminer le temps de détection le plus proche
- Déterminer par l'inférence bayesienne sans bruit de fond l'espérance et la variance a posteriori de λ avec les $n - n_b$ temps restants

On obtient ainsi une suite d'estimateurs λ_k et de variances v_k , ainsi qu'une suite de poids

$$w_k = \frac{q_k}{\sum_k q_k}$$

L'estimateur final de λ est défini par la moyenne pondérée

$$\hat{\lambda} = \sum_k w_k \lambda_k$$

²L'hypothèse impérative de décroissance des sources implique que \bar{t} se trouve en moyenne dans la première moitié de l'intervalle de mesure. Néanmoins, des valeurs supérieures peuvent apparaître, particulièrement si les comptes sont bas ou si la source est quasi constante.

³Il semble que $K = 100$ donne déjà de bons résultats.

Pour estimer l'incertitude sur cet estimateur, il faut encore tenir compte de l'aléatoire introduit par la méthode de suppression. Nous définirons donc une variance interne

$$V_{int} = \sum_k w_k v_k$$

et une variance externe

$$V_{ext} = \sum_k w_k \lambda_k^2 - \left(\sum_k w_k \lambda_k \right)^2$$

L'incertitude finale sera ensuite estimée par la somme des variances externe et interne.

A.17 Modèle des comptages

– Modèle : Poisson

$$P(N | a) = \mathcal{P}(N | at)$$

– Conjuguées : gamma

$$(N, t) \cdot \text{Ga}(N_f, t_f) = \text{Ga}(N_f + N, t_f + t)$$

– A priori de Jeffreys

$$\pi_J(da) = \text{Ga}(da | \frac{1}{2}, 0) = \frac{da}{\sqrt{a}}$$

– A posteriori de Jeffrey

$$\pi_J(da | N) = \text{Ga}(da | N + \frac{1}{2}, t)$$

– Marginale : Négative-binomiale

$$P(N) = \mathcal{NB} \left(N | N_f, \frac{t_f}{t_f + t} \right) = \frac{\Gamma(N_f + N) t_f^{N_f} t^N}{\Gamma(N + 1) \Gamma(N_f) (t_f + t)^{N_f + N}}$$

– Marginale de Jeffrey

$$P_J(N) = \frac{\Gamma(N + \frac{1}{2})}{\Gamma(N + 1)}$$

A.18 Modèle des temps d'attente

– Modèle : gamma

$$P(dt | N, a) = \text{Ga}(dt | N, a)$$

– Conjuguées : gamma

$$(N, t) \cdot \text{Ga}(N_f, t_f) = \text{Ga}(N_f + N, t_f + t)$$

– A priori de Jeffreys

$$\pi_J(da) = \text{Ga}(da | 0, 0) = \frac{da}{a}$$

– A posteriori de Jeffreys

$$\pi_J(da | t) = \text{Ga}(da | N, t)$$

– Marginale

$$P(dt) = \frac{\Gamma(N_f + N) t_f^{N_f} t^{N-1} dt}{\Gamma(N_f) \Gamma(N) (t_f + t)^{N_f + N}}$$

– Marginale Beta pour $\tau = \frac{t}{t_f + t}$

$$P(d\tau) = \text{Be}(d\tau | N, N_f) = \frac{\Gamma(N_f + N)}{\Gamma(N_f) \Gamma(N)} \tau^{N-1} (1 - \tau)^{N_f - 1} d\tau$$

– Marginale de Jeffreys

$$P_J(dt) = \text{Ga}(dt | 0, 0) = \frac{dt}{t} \quad (\text{A.142})$$

A.19 Calcul des limites L_C et L_D

Reprendons l'approximation (10.7) et appliquons la définition (10.8) de L_C . Posons $a = 0$, il vient

$$\alpha = \int_{L_C}^{\infty} \mathcal{N}\left(0, N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)\right) \quad (\text{A.143})$$

$$= \frac{1}{\sqrt{2\pi N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)}} \int_{L_C}^{\infty} dx \exp\left(-\frac{x^2}{2N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)}\right) \quad (\text{A.144})$$

Effectuons le changement de variable $y = \frac{x}{\sqrt{N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)}}$

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{\frac{\sqrt{N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)}}{\sqrt{L_C}}}^{\infty} dy e^{y^2/2} \quad (\text{A.145})$$

$$= 1 - \Phi\left(\frac{L_C}{\sqrt{N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)}}\right) \quad (\text{A.146})$$

Ainsi

$$L_C = \Phi(1 - \alpha) \sqrt{N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)} \quad (\text{A.147})$$

$$= k_{\alpha} \sqrt{N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)} \quad (\text{A.148})$$

Pour la limite L_D , replaçant dans (10.7) $a\varepsilon_{Dt}$ par L_D et $N_0 \frac{t}{t_0} \left(1 + \frac{t}{t_0}\right)$ par $\frac{L_C^2}{k_{\alpha}^2}$, nous devons résoudre

$$\int_{-\infty}^{L_C} \mathcal{N}\left(L_D, L_D + \frac{L_C^2}{k_{\alpha}^2}\right) = \beta \quad (\text{A.149})$$

Effectuons un nouveau changement de variable

$$y = \frac{x - L_D}{\sqrt{L_D + \frac{L_C^2}{k_\alpha^2}}} \quad (\text{A.150})$$

il vient

$$\Phi\left(\frac{L_C - L_D}{\sqrt{L_D + \frac{L_C^2}{k_\alpha^2}}}\right) = \beta \quad (\text{A.151})$$

Avec $k_\beta = -\Phi^{-1}(\beta)$ cela revient à résoudre

$$\frac{L_C - L_D}{\sqrt{L_D + \frac{L_C^2}{k_\alpha^2}}} = -k_\beta \quad (\text{A.152})$$

mais

$$(L_C - L_D)^2 = k_\beta^2 \left(L_D + \frac{L_C^2}{k_\alpha^2} \right) \quad (\text{A.153})$$

$$L_D^2 - (k_\beta^2 + 2L_C)L_D + L_C^2 - k_\beta^2 \frac{L_C^2}{k_\alpha^2} = 0 \quad (\text{A.154})$$

Le discriminant de cette équation est

$$\Delta = k_\beta^2 \left(4L_C + k_\beta^2 + 4 \left(\frac{L_C}{k_\alpha} \right)^2 \right) \quad (\text{A.155})$$

Ainsi

$$L_D = L_C + \frac{k_\beta^2}{2} \pm \frac{k_\beta}{2} \sqrt{4L_C + k_\beta^2 + 4 \left(\frac{L_C}{k_\alpha} \right)^2} \quad (\text{A.156})$$

Annexe B

Notes

B.1 Modèle des clés

B.1.1 Un traitement plus complet

Il faut en fait considérer que l'on est en face d'un système physique à deux composantes. La première est le lieu de recherche, la seconde le lieu de présence des clés. Les états possibles sont les couples LL , LO , OL et OO . L'a priori sur ces états est connu en ce qui concerne la composante de recherche. Si le professeur choisit le réverbère cet a priori est concentré en L , sinon, il est concentré en O . C'est sur la composante de présence que l'on trouve les a priori du paragraphe précédent. Pour une recherche de type L , l'a priori sur les couples est donné par

		Présence	
		L	O
Recherche	L	$\pi(L)$	$\pi(O)$
	O	0	0

alors que pour le type O , on a

		Présence	
		L	O
Recherche	L	0	0
	O	$\pi(L)$	$\pi(O)$

Remarquons en passant que si le professeur choisit aléatoirement la lumière avec une probabilité μ_L et l'obscurité avec une probabilité $\mu_O = 1 - \mu_L$, l'a priori prend la forme générale

		Présence	
		L	O
Recherche	L	$\mu_L \pi(L)$	$\mu_L \pi(O)$
	O	$\mu_O \pi(L)$	$\mu_O \pi(O)$

Le résultat de la recherche ne peut prendre que deux valeurs : trouvé T ou non trouvé $\neg T$. Le modèle, qui associe à chaque état une répartition sur ces résultats possibles, s'exprime par l'intermédiaire des

efficacités de recherche

$$\begin{aligned} P(T | LL) &= S_L \\ P(T | LO) &= 0 \\ P(T | OL) &= 0 \\ P(T | OO) &= S_O \end{aligned}$$

Dans le cas d'une recherche de type L , la probabilité effective de trouver est la marginale

$$\begin{aligned} M_L(T) &= P(T | LL)\pi_L(LL) + P(T | LO)\pi_L(LO) + P(T | OL)\pi_L(OL) + P(T | OO)\pi_L(OO) \\ &= S_L\pi(L) \end{aligned}$$

et dans le cas O

$$\begin{aligned} M_O(T) &= P(T | LL)\pi_O(LL) + P(T | LO)\pi_O(LO) + P(T | OL)\pi_O(OL) + P(T | OO)\pi_O(OO) \\ &= S_O\pi(O) \end{aligned}$$

Après l'échec de la recherche de type L , l'a posteriori bayesien s'écrit

$$\pi_L(LL | \neg T) = \frac{P(\neg T | LL)\pi_L(LL)}{M_L(\neg T)} = \frac{(1 - S_L)\pi(L)}{1 - S_L\pi(L)}$$

qui est bien la valeur obtenue précédemment.

B.1.2 Une extension introduisant la durée de la recherche

Au lieu de définir l'efficacité de recherche comme probabilité de trouver dans un temps donné, il est possible d'introduire des probabilités s_L et s_O de trouver par unité de temps en imaginant un système de recherche aléatoire. La probabilité de trouver en t unités de temps est $S_L(t) = 1 - e^{-s_L t}$ sous le réverbère et $S_O(t) = 1 - e^{-s_O t}$ dans l'obscurité. Les changements de lieu de recherche se font aux temps où la différence des marginales change de signe.

B.2 Divergence de Kullback-Leibler et entropie de Shannon

B.2.1 Entropie

L'entropie de Shannon est associée à l'incertitude sur les issues d'une expérience. Soit une expérience admettant un nombre fini N d'issues possibles. Si ces issues admettent les probabilités d'occurrence $p = (p_i)_{i=1}^n$, on définit l'entropie de l'expérience par

$$H(p) = - \sum_i p_i \ln p_i \tag{B.1}$$

On trouvera par exemple dans [48] une discussion extensive de ce concept. Remarquons simplement que H s'annule si la probabilité est concentrée en un point (et seulement dans ce cas), et qu'elle prend son maximum lorsque les issues sont équiprobables.

Le lien avec la mesure physique pose un problème car le résultat d'un expérience peut généralement prendre un nombre infini de valeurs, ce qui empêche l'utilisation directe de la définition ci-dessus. Ce phénomène est analysé en détail par Brillouin dans [49]. La solution consiste à discréteriser l'espace de

mesure. Soit par exemple l'expérience consistant à déterminer une grandeur ℓ que l'on sait être dans un intervalle $[\ell_1, \ell_2]$ de longueur L . Après une mesure, il résulte que cette grandeur est dans un sous intervalle $[d_1, d_2]$ de $[\ell_1, \ell_2]$ ayant une longueur Δ . Si l'on découpe l'intervalle de départ en un grand nombre de mailles de longueur ε , et que fait une hypothèse d'équiprobabilité des issues, il y a au départ $\frac{L}{\varepsilon}$ issues équiprobables, et l'entropie initiale de l'expérience est

$$H_0 = - \sum \frac{\varepsilon}{L} \ln \frac{\varepsilon}{L} = \ln \frac{L}{\varepsilon} \quad (\text{B.2})$$

Après la mesure, il ne reste plus que $\frac{\Delta}{\varepsilon}$ issues équiprobables et l'entropie de l'expérience est alors

$$H_1 = - \sum \frac{\varepsilon}{\Delta} \ln \frac{\varepsilon}{\Delta} = \ln \frac{\Delta}{\varepsilon} \quad (\text{B.3})$$

La différence d'entropie vaut

$$\Delta H = \ln \frac{L}{\Delta} \quad (\text{B.4})$$

On remarque que ε a disparu, ce qui permet de le faire tendre vers zéro sans problème. La différence d'entropie est positive, l'incertitude a diminué, on a acquis une information sur le système.

Si dans ce cas l'entropie initiale n'a pas de sens pour $\varepsilon \rightarrow 0$, la différence d'entropie en garde un, ce qui permet de définir une information dans le cas d'espaces continus.

B.2.2 Bayes, Shannon et Kullback

Considérons le modèle simple suivant. L'ensemble des états Θ est muni d'une partition finie $\mathcal{A} : \Theta = \sum_{i=1}^n A_i$ (cf figure B.1). Admettons que la réponse du système est simplement l'indice du sous-ensemble contenant son état actuel. Le modèle est donc

$$P(i \mid \theta) = 1_{A_i}(\theta) = \delta_\theta(A_i) \quad (\text{B.5})$$

c'est-à-dire que le système répond i avec la probabilité 1 si $\theta \in A_i$ et 0 sinon. C'est un modèle déterministe car la réponse du système est univoque, même si elle ne permet pas de déterminer exactement l'état. Faire une mesure sur un tel système correspond à poser la question structurée par la partition \mathcal{A} : *Dans quel sous ensemble se trouve l'état ?*. Si au départ l'ensemble des états est muni d'une probabilité (a priori) π , la réponse à une telle question est incertaine, la probabilité de la réponse i valant $\pi(A_i)$. Il est donc possible d'associer à la partition \mathcal{A} et à la probabilité initiale π une entropie

$$H(\mathcal{A}) = - \sum_i \pi(A_i) \ln \pi(A_i) \quad (\text{B.6})$$

C'est l'entropie de la question, qui décrit l'incertitude relative à ses réponses possibles.

Posons la question (i.e. faisons la mesure) et supposons que le résultat soit l'indice i_0 . Comme nous connaissons le modèle et l'a priori, le théorème de Bayes permet le calcul de l'a posteriori

$$\pi(d\theta \mid i_0) = \frac{P(i_0 \mid \theta) \pi(d\theta)}{P(i_0)} \quad (\text{B.7})$$

$$P(i_0) = \int_\Theta P(i_0 \mid \theta) \pi(d\theta) = \pi(A_{i_0}) \quad (\text{B.8})$$

$$\begin{aligned} \pi(d\theta \mid i_0) &= \begin{cases} \frac{\pi(d\theta)}{\pi(A_{i_0})} & \text{si } \theta \in A_{i_0} \\ 0 & \text{sinon} \end{cases} \\ &= \pi(d\theta \mid A_{i_0}) \end{aligned} \quad (\text{B.9})$$

$$(\text{B.10})$$

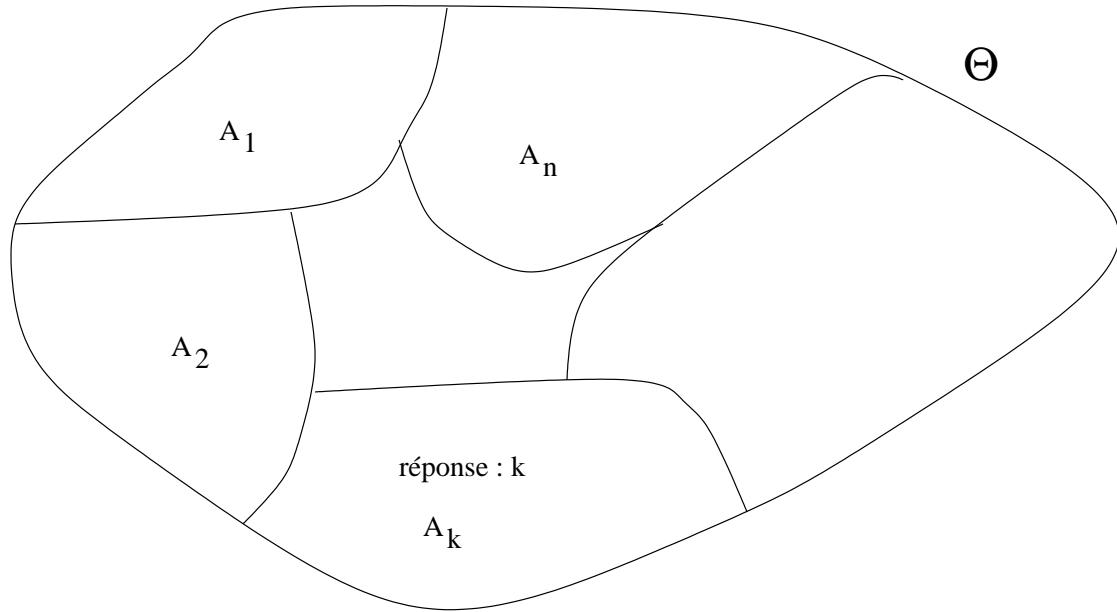


FIG. B.1 – Partition définissant une question

L'a posteriori de Bayes est la probabilité conditionnelle étant donné le sous-ensemble associé à la réponse¹.

Calculons encore l'information moyenne apportée par la mesure

$$I_H = \sum_i P(i) \int_{\theta} \pi(d\theta | i) \ln \frac{\pi(d\theta | i)}{\pi(d\theta)} \quad (\text{B.11})$$

$$= \sum_i \pi(A_i) \pi(d\theta | A_i) \ln \frac{\pi(d\theta | A_i)}{\pi(d\theta)} \quad (\text{B.12})$$

$$= \sum_i \int_{A_i} \pi(d\theta) \ln \frac{1}{\pi(A_i)} \quad (\text{B.13})$$

$$= - \sum_i \pi(A_i) \ln \pi(A_i) \quad (\text{B.14})$$

$$= H(\mathcal{A}) \quad (\text{B.15})$$

Résultat L'information moyenne apportée par la mesure est exactement l'entropie de Shannon de la question associée à la partition.

¹Le fait que $\pi(d\theta | i) = \pi(d\theta | A_i)$ peut sembler trivial, mais il faut mettre en évidence que la barre | est entendue comme a posteriori à gauche, et comme conditionnement probabiliste habituel à droite. L'égalité n'est évidemment pas un hasard, mais ne peut se passer d'une démonstration.

B.3 A priori

B.3.1 A priori MDIP

Une utilisation de la théorie de l'information pour obtenir des a priori non informatifs a été proposée par Zellner [28] en 1971. Avec un modèle densitaire, cet a priori (Maximum Data Information Prior) est défini par

$$\pi(\theta) = e^{\int dx P(x|\theta) \ln P(x|\theta)} \quad (\text{B.16})$$

ou ce qui revient au même, en utilisant la divergence de Kullback

$$\pi(d\theta) = e^{D(P_\theta || P_0)} d\theta \quad (\text{B.17})$$

avec P_0 judicieusement choisie. Cette approche a été généralisé par Rodriguez [24]. Dans le contexte de la mesure de la radioactivité, les a priori obtenus sont assez proches de ceux de Jeffreys, c'est pourquoi nous nous bornerons à présenter les cas poissonniers et gamma

B.3.1.1 Le cas poissonnien

Posons $\alpha = a\epsilon t$, l'a priori MDIP s'exprime ici par l'entropie de la loi de Poisson

$$\pi(\alpha) = e^{-H(\mathcal{P}(\alpha))} \quad (\text{B.18})$$

Cette entropie n'admet pas de forme analytique simple, mais converge vers celle de la loi normale $\frac{1}{2} \ln(2\pi e\alpha)$ lorsque α croît. On retrouve donc asymptotiquement une forme en $\alpha^{-\frac{1}{2}}$ qui est celle de Jeffreys (cf figure B.2),

B.3.1.2 Le cas gamma

Ici, il faut calculer e^D avec

$$D = \int dt \frac{a^N}{\Gamma(N)} t^{N-1} e^{-at} \ln \left(\frac{a^N}{\Gamma(N)} t^{N-1} e^{-at} \right) \quad (\text{B.19})$$

$$= \int \text{Ga}(dt | N, a) (N \ln a - \ln \Gamma(N) + (N-1) \ln t - at) \quad (\text{B.20})$$

$$= N \ln a - \ln \Gamma(N) + (N-1)(\psi(N) - \ln a) - a \frac{N}{a} \quad (\text{B.21})$$

en tenant compte du lemme A.8. Il vient donc

$$\pi(da) \propto a da = \text{Ga}(da | 2, 0) \quad (\text{B.22})$$

ce qui est assez curieux, celui de Jeffreys étant ici $a^{-1}da$

B.3.2 Maximisation asymptotique de l'information

Supposons que l'état soit décrit par un couple $\theta = (\mu, v)$, où μ est la partie déterminante pour la prise des décisions, et v un paramètre dit *de nuisance*. Si l'on dispose d'un a priori $\pi(dv)$ sur la nuisance et que les deux composantes de l'état sont indépendantes, les a priori sur l'état s'expriment

$$\pi(d\theta) = \pi(d\mu)\pi(dv)$$

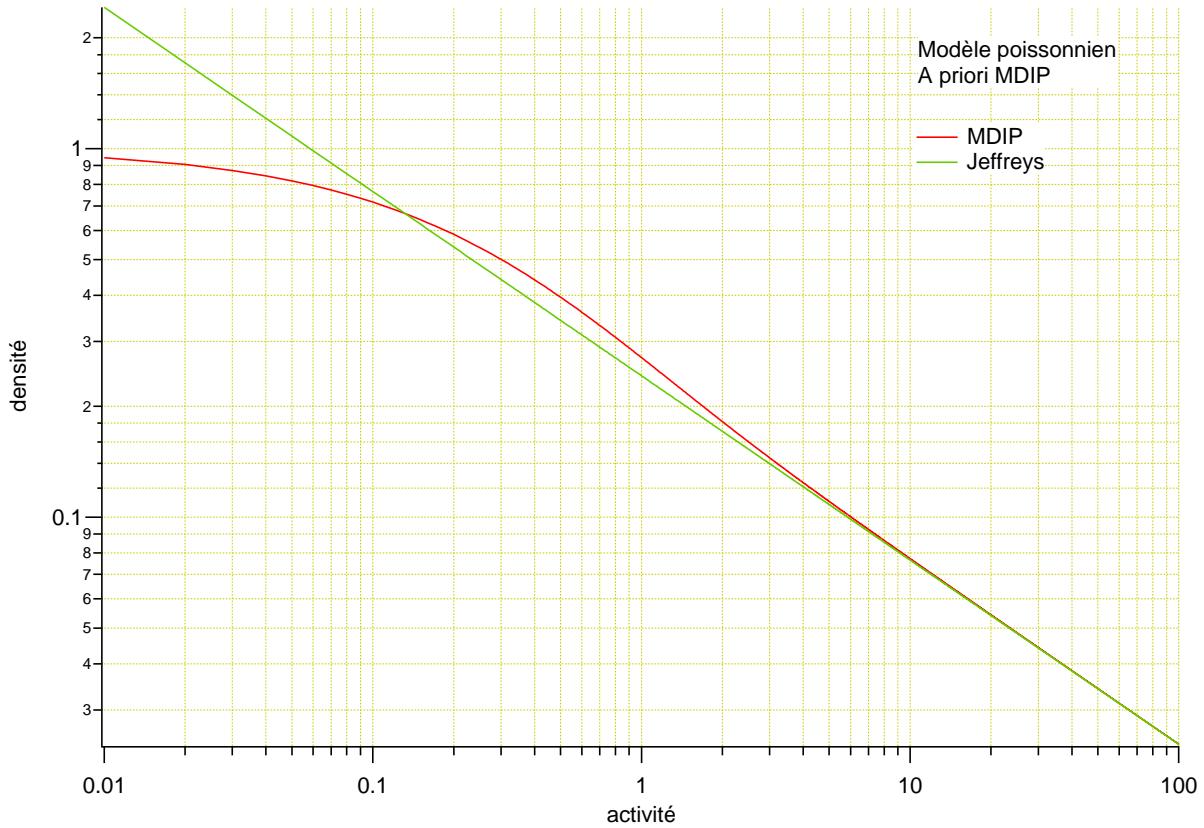


FIG. B.2 – A priori MDIP pour le modèle poissonnien

Effectuons n répétitions indépendantes de la même expérience, l'état étant stationnaire, le modèle devient

$$P^n(dx | \mu, v) = \prod_{i=1}^n P(dx_i | \mu, v) \quad (\text{B.23})$$

Marginalisons cette répartition en tenant compte de notre connaissance de la nuisance²

$$P^n(dx | \mu) = \int_v \left(\prod_{i=1}^n P(dx_i | \mu, v) \right) \pi(dv) \quad (\text{B.24})$$

Pour une répartition $\pi(d\mu)$, l'information moyenne apportée par ce type de mesure est

$$I_H^n(\pi(d\mu)) = \int_{x, \theta} P^n(dx, d\mu) \ln \frac{P^n(dx, d\mu)}{P^n(dx) \pi(d\mu)} \quad (\text{B.25})$$

Lorsque $n \rightarrow \infty$, $I_H^n(\pi(d\mu))$ représente toute l'information que l'on peut obtenir sur l'état si l'on prend $\pi(d\mu)$ comme a priori. L'idée (cf Bernardo [10]) consiste à prendre comme a priori une répartition maximisant cette limite. Ceci peut se faire de deux manières

² $P^n(dx | \mu) = \prod_i (P(dx_i | \mu, v) \pi(dv))$ introduit la subjectivité de l'a priori sur v dans l'objectivité du modèle. Une telle expression serait vraie si le bruit lui-même était objectivement aléatoire !

– Premièrement

$$I_H^\infty(\pi(d\mu)) = \lim_n I_H^n(\pi(d\mu))$$

puis

$$\pi_B(d\mu) = \arg \max_{\pi} I_H^\infty(\pi(d\mu))$$

– Deuxièmement

$$\pi^n(d\mu) = \arg \max_{\pi} I_n^\infty(\pi(d\mu))$$

puis

$$\pi'_B(d\mu) = \lim_n \pi^n(d\mu)$$

L'a priori *de référence* est alors cette limite, si elle existe. Cet a priori est aussi associé au modèle.

B.3.2.1 Application au cas fini

Dans le cas d'espaces finis, l'a priori de Bernardo est uniforme [10].

B.3.2.2 Application au cas poissonnien

Supposons que nous avons obtenu, en face du même état en r mesures indépendantes du durée t les comptages $N = (N_1 \dots N_r)$. Si l'activité est a , la probabilité d'une telle suite est

$$P(N \mid a) = \prod_{i=1}^r \mathcal{P}(N_i \mid a\epsilon t) \quad (\text{B.26})$$

donc sous l'a priori $\text{Ga}(N_f, t_f)$

$$P(N \mid N_f, t_f) = \int_a \text{Ga}(da \mid N_f, t_f) \prod_{i=1}^r \mathcal{P}(N_i \mid a\epsilon t) \quad (\text{B.27})$$

qui peut s'écrire, avec $|N| = \sum_i N_i$

$$P(N \mid N_f, t_f) = \frac{|N|!}{\prod_i N_i!} \mathcal{N}(\mathcal{B}(|N| \mid N_f, p_f^r)) \quad (\text{B.28})$$

avec $p_f^r = \frac{t_f}{t_f + rt}$. Mais l'a posteriori vaut alors $\text{Ga}(N_f + |N|, t_f + rt)$ et l'information moyenne apportée par cet ensemble de r mesures vaut

$$\begin{aligned} I_H^r(N_f, t_f \mid t) &= \sum_N D(\text{Ga}(N_f + |N|, t_f + rt) \mid\mid \text{Ga}(N_f, t_f)) \\ &\quad \cdot \frac{|N|!}{\prod_i N_i!} \mathcal{N}(\mathcal{B}(|N| \mid N_f, p_f^r)) \end{aligned} \quad (\text{B.29})$$

$$= \sum_N D(\text{Ga}(N_f + N, t_f + rt) \mid\mid \text{Ga}(N_f, t_f)) \mathcal{N}(\mathcal{B}(N \mid N_f, p_f^r)) \quad (\text{B.30})$$

$$= I_H^1(N_f, t_f \mid rt) \quad (\text{B.31})$$

qui est l'information moyenne pour une seule mesure de durée rt .

B.3.2.3 Application au modèle gamma

Selon [27], l'a priori de Jeffreys $\frac{da}{a}$ pour ce modèle est aussi de référence.

B.4 A priori maximisant l'information apportée par la mesure

B.4.1 Le paradoxe du cas fini

Plaçons nous dans le cas fini où

$$\Theta = \{1, \dots, j, \dots, m\} \quad (\text{B.32})$$

$$\mathcal{X} = \{1, \dots, i, \dots, n\} \quad (\text{B.33})$$

$$P(i | j) = P_{ij} \quad (\text{B.34})$$

Un a priori est ici un vecteur à m dimensions dans le $m - 1$ simplexe

$$\pi(j) = \pi_j, \pi_j \geq 0, \sum_j \pi_j = 1 \quad (\text{B.35})$$

La marginale s'exprime par un vecteur à n dimensions dans le $n - 1$ simplexe

$$M_i = \sum_j P_{ij} \pi_j \quad (\text{B.36})$$

Un rapide calcul donne l'information moyenne apportée par une mesure

$$I_H(\pi) = - \sum_i M_i \ln M_i + \sum_j \pi_j \sum_i P_{ij} \ln P_{ij} \quad (\text{B.37})$$

$$= H(M) - \sum_j \pi_j H(P_{.j}) \quad (\text{B.38})$$

On peut chercher à maximiser I_H par la méthode des multiplicateurs de Lagrange

$$\Phi = I_H + \lambda \sum_j \pi_j \quad (\text{B.39})$$

et l'on tombe sur

$$\sum_i P_{ij_0} \ln M_i = \lambda + \sum_i P_{ij_0} \ln P_{ij_0}, \quad j_0 = 1 \dots m \quad (\text{B.40})$$

On est donc en présence de m équations à n inconnues (les $\ln M_i$). Si $m > n$ la solution est compromise. Par contre si $m < n$, c'est l'obtention des p_j depuis les M_j qui est compromise, puisque il y a de nouveau trop d'équations. Cette maximisation ne donnerait alors de solution que pour $m = n$. Ceci montre que dans le cas général, la maximisation de l'information moyenne ne peut pas s'obtenir par l'annulation des dérivées. Mais le simplexe accueillant les π_j est compact, donc un tel maximum existe. Donc la solution se trouve sur un bord du simplexe, certains éléments du vecteur des probabilités π sont nuls, et l'a priori obtenu est dégénéré.

B.4.2 Application au modèle poissonnien

Dans ce cas, cherchons à maximiser l'information en nous limitant à des a priori de type gamma. Pour un a priori $\text{Ga}(N_f, t_f)$, on peut montrer que l'information moyenne apportée par une mesure est

$$\begin{aligned} I_H &= E \cdot \psi(N_f) - N_f \left(\ln p_f + \frac{1 - p_f}{p_f} \right) \\ &\quad - \sum_{i=0}^{\infty} \left(1 - \sum_{N=0}^i \mathcal{N}(\mathcal{B}(N | N_f, p_f)) \right) \ln(N_f + i) \\ &\quad + \sum_{i=0}^{\infty} \left(E - \sum_{N=0}^i N \cdot \mathcal{N}(\mathcal{B}(N | N_f, p_f)) \right) \frac{1}{N_f + i} \end{aligned} \quad (\text{B.41})$$

où

$$E = E\mathcal{N}\mathcal{B}(N_f, p_f) = N_f \frac{1 - p_f}{p_f}$$

et

$$p_f = \frac{t_f}{t_f + \epsilon t}$$

L'unité de temps disparaît. Lorsque $p_f \rightarrow 0$, cette information diverge, suggérant des a priori du type $\text{Ga}(N_f, 0)$.

B.4.3 Application au modèle gamma

Pour le modèle des temps d'attente, le modèle gamma donne lieu à une information moyenne (cf 6.21)

$$I_H(N_f, t_f) = \ln \Gamma(N_f) - \ln \Gamma(N_f + N) - N_f \psi(N_f) + (N_f + N) \psi(N_f + N) - N \quad (\text{B.42})$$

Asymptotiquement, on a (cf [36] 6.1.41, 6.3.18)

$$\ln \Gamma(z) = \left(z - \frac{1}{2}\right) \ln z - z - \frac{1}{2} \ln 2\pi + \frac{1}{12z} + O\left(\frac{1}{z^2}\right) \quad (\text{B.43})$$

$$\psi(z) = \ln z - \frac{1}{2z} - \frac{1}{12z^2} + O\left(\frac{1}{z^3}\right) \quad (\text{B.44})$$

Donc

$$I_H(N_f, t_f) = \frac{1}{2} \ln \left(1 + \frac{N}{N_f}\right) + O\left(\frac{1}{N_f^2}\right) \xrightarrow{N_f \rightarrow \infty} 0 \quad (\text{B.45})$$

On a

$$\ln \Gamma(z+1) = \ln z + \ln \Gamma(z) \quad (\text{B.46})$$

$$\psi(z+1) = \psi(z) + \frac{1}{z} \quad (\text{B.47})$$

$$\psi(1) = \gamma \quad (\text{B.48})$$

donc

$$\begin{aligned} \lim_{N_f \rightarrow 0} I_H(N_f, t_f) &= \lim_{N_f \rightarrow 0} \left(\ln \Gamma(N_f + 1) - \ln N_f - N_f \left(\psi(N_f + 1) - \frac{1}{N_f} \right) \right) \\ &\quad - \ln \Gamma(N) + N \psi(N) - N \end{aligned} \quad (\text{B.49})$$

$$= \infty \quad (\text{B.50})$$

ou encore

$$\lim_{N_f \rightarrow 0} \frac{I_H(N_f, t_f)}{-\ln N_f} = 1 \quad (\text{B.51})$$

Les a priori maximisant l'information sont donc les $\text{Ga}(0, t_f)$.

B.5 A priori de Jeffreys avec données préalables, extension continue Poisson-gamma

Plaçons nous dans le contexte d'une mesure de source sans bruit et supposons que la probabilité effective ayant donné des observations préalables ($N_1 \dots N_r$) soit de type $\text{Ga}(N_f, t_f)$. L'espace des états se restreint à $\mathbb{N}^* \times \mathbb{R}_+^*$ et le modèle non stationnaire s'écrit

$$P(N | N_f, t_f) = \mathcal{N}\mathcal{B}(N | N_f, \frac{t_f}{t_f + \varepsilon t}) \quad (\text{B.52})$$

L'a priori de Jeffreys étant invariant, utilisons la variable $p_f = \frac{t_f}{t_f + \varepsilon t}$

$$P(N | N_f, p_f) = \frac{\Gamma(N_f + N)}{\Gamma(N_f)N!} p_f^{N_f} (1 - p_f)^N \quad (\text{B.53})$$

$$\ln P = \ln \Gamma(N_f + N) - \ln \Gamma(N_f) - \ln N! + N_f \ln p_f + N \ln(1 - p_f) \quad (\text{B.54})$$

$$\frac{\partial \ln P}{\partial N_f} = \psi(N_f + N) - \psi(N_f) + \ln p_f \quad (\text{B.55})$$

$$\frac{\partial \ln P}{\partial p_f} = \frac{N_f}{p_f} - \frac{N}{1 - p_f} \quad (\text{B.56})$$

$$\frac{\partial^2 \ln P}{\partial N_f^2} = \psi'(N_f + N) - \psi'(N_f) \quad (\text{B.57})$$

$$\frac{\partial^2 \ln P}{\partial N_f \partial p_f} = \frac{1}{p_f} \quad (\text{B.58})$$

$$\frac{\partial^2 \ln P}{\partial p_f^2} = -\frac{N_f}{p_f^2} - \frac{N}{(1 - p_f)^2} \quad (\text{B.59})$$

Les éléments de la matrice de Fisher sont

$$I_{11} = \psi'(N_f) - \sum_N \psi'(N_f + N) \mathcal{N}\mathcal{B}(N | N_f, p_f) \quad (\text{B.60})$$

$$I_{12} = -\frac{1}{p_f} \quad (\text{B.61})$$

$$I_{22} = \frac{N_f}{p_f^2} + \frac{N_f \frac{1-p_f}{p_f}}{(1 - p_f)^2} \quad (\text{B.62})$$

$$= \frac{N_f}{p_f} \left(\frac{1}{p_f} + \frac{1}{1 - p_f} \right) \quad (\text{B.63})$$

Et son déterminant s'écrit

$$\begin{aligned} |I|(N_f, p_f) &= \frac{N_f}{p_f} \left(\frac{1}{p_f} + \frac{1}{1 - p_f} \right) \left(\psi'(N_f) - \sum_N \psi'(N_f + N) \mathcal{N}\mathcal{B}(N | N_f, p_f) \right) \\ &\quad - \frac{1}{p_f^2} \end{aligned} \quad (\text{B.64})$$

L'a priori de Jeffreys avec données préalables est alors

$$\pi_{JD}(dN_f, dp_f) \propto \left(\prod_k \frac{\Gamma(N_f + N_k)}{\Gamma(N_f)N_k!} \right) p_f^{rN_f} (1 - p_f)^{\sum_k N_k} \sqrt{|I|(N_f, p_f)} dN_f dp_f \quad (\text{B.65})$$

et l'a priori sur l'activité à utiliser est $\text{Ga}(N_g, t_g)$ avec

$$N_g = \int N_f \pi_{JD}(dN_f, dp_f) \quad (\text{B.66})$$

$$\frac{t_g}{t_g + \varepsilon t} = \int p_f \pi_{JD}(dN_f, dp_f) \quad (\text{B.67})$$

Remarque Si le modèle de base est celui du temps d'attente de la N^e détection, et que la suite des observations t_i est obtenue en face d'un état non stationnaire de répartition $\text{Ga}(N_f, t_f)$ (dont les paramètres sont évidemment inconnus), nous avons vu que la répartition des t_i devrait suivre une loi $\text{Be}(dp | N_f, N)$, où $p = \frac{t_f}{t_f + \varepsilon t}$. L'ennui est ici que les t_i ne permettent pas d'obtenir des p_i car t_f est inconnu.

Bibliographie

- [1] Thomas Bayes, Essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society of London*, 1764
- [2] E.T. Jaynes, *Papers on Probability, Statistics, and Statistical Physics*. A reprint collection. R.D. Rozenkrantz (Ed.). Reidel, Dordrecht.
- [3] Andrey Kolmogorov, *Foundations of the theory of probability*, Chelsea, 1950
- [4] Association des collaborateurs de Nicolas Bourbaki, *Eléments de Mathématiques*, Dunod, Paris, 1960-
- [5] Jérôme Segal, *Théorie de l'information : sciences, techniques et société de la seconde guerre mondiale à l'aube du XXIe siècle*, Thèse de Doctorat, Université Lumière Lyon (livre en préparation aux éditions Syllepse, Paris, 2003)
- [6] K. Weise, W. Woeger, *A Bayesian Theory of Measurement Uncertainty*, *Mens. Sci. Technol.*, 7 (1991) 1-11
- [7] G. D'Agostini, Bayesian reasoning versus conventional statistics in high energy physics (in *Maximum Entropy and Bayesian Methods*. Garching, Germany 1998, p168), ed. von der Linden, Dose, Fisher and Preuss, Kluwer Academic Publishers.
- [8] International Vocabulary of Basic and General Terms in Metrology, Second Edition, International Organization for Standardisation, 1993, ISBN 92-67-01075-1
- [9] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer (1985)
- [10] J. M. Bernardo, A. F. M. Smith, *Bayesian Theory*, Wiley (2000)
- [11] J. Neveu, *Bases mathématiques du calcul des probabilités*, Masson (1970)
- [12] Lotfi A. Zadeh, *Fuzzy sets*, *Inf. Control* 8, 338-353, 1965
- [13] Lotfi A. Zadeh, *Probability measures of fuzzy events*, *Jour. Math. Analysis and Appl.* 23, 421- 427, 1968
- [14] J. Neyman, E. S. Pearson, *Phil. Trans. Roy. Soc. London*, A231, p 332, 1932
- [15] B. L. van der Waerden, *Statistique mathématique*, p 259, Dunod 1967
- [16] J. A. Swets, *Signal detection theory and ROC analysis in psychology and diagnostics*, Lawrence Erlbaum Associates, 1996
- [17] M. Girault, *Processus Aléatoires*, Dunod, Paris 65, p17
- [18] D. Foata, A. Fuchs, *Processus stochastiques*, Dunod, Paris 2002
- [19] S. Kullback, *Information theory and statistic*. Dover Publications, Mineola, New York 1968
- [20] Thomas M. Cover, Joy A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, Wiley 1991

- [21] C.E. Shannon, A Mathematical Theory of Communication, Board of Trustees of the University of Illinois, 1949
- [22] Robert V. Hogg, Allen T. Craig, Introduction to Mathematical Statistics, Prentice Hall, 1995
- [23] C.C. Rodriguez, Entropic priors, State of New York University at Albany, 4 oct 1991
- [24] C.C. Rodriguez, Are we cruising a hypothesis space ?, arXiv :physics/9809009
- [25] B. Roy Frieden, Physics from Fisher Information, Cambridge University Press, 1998
- [26] Léon Brillouin, Les tenseurs en Mécanique et en Elasticité, Masson, 1960
- [27] Ruoyong Yang and James O. Berger, A Catalog of Non Informative Priors, Purdue University, August 1996
- [28] A. Zellner, "Past and recent results on maximal data information priors," tech. rep., H.G.B. Alexander Research Foundation. Graduate School of Business, University of Chicago, 1995.
- [29] L. A. Currie, Detection in analytical chemistry, American Chemical Society, Washington DC (1998)
- [30] L. A. Currie, International recommendations offered on analytical and quantification concepts and nomenclature, *Analytica Chimica Acta* Vol 391 No 2, may 31, 1999
- [31] L. A. Currie, Nomenclature in evaluation of analytical methods including detection and quantification capabilities, *ibid*
- [32] L. A. Currie, Detection and quantification limits : origins and historical overview, *ibid*
- [33] G. J. Feldman and R. D. Cousins, A unified approach to the classical statistical analysis of small signals, *Phys. Rev. D* 57, 3873, (1998)
- [34] Particle Data Group, <http://pdg.lbl.gov/>
- [35] C. Giunti, A new ordering principle for the classical statistical analysis of Poisson processes with background, *Phys. Rev. D* 59, 053001 (1999)
- [36] M. Abramowitz, I. Stegun, Handbook of Mathematical Functions 6.3, Dover Publications, inc. New-York, 1972
- [37] GNU scientific library, Mark Galassi, Cygnus solutions and Los Alamos National Laboratory, Jim Davis & al.
- [38] Landon Curt Noll, <http://www.isthe.com/chongo/tech/comp/calc/>, Calc - C-style arbitrary precision calculator,
- [39] http://wwwinfo.cern.ch/asdoc/geant_html3/geantall.html
- [40] J.-P. Laedermann, M. Decombaz, Simulation of nuclear decay, *Applied Radiation and Isotopes* 52 (2000) 419-425
- [41] J.-P. Laedermann, l_detdec, a C-software library for bayesian multichannel analysis in gamma spectrometry, *IRA*, 2001-2
- [42] H. Snoussi, A. Mohammad-Djafari, Information geometry and prior selection, Laboratoire des Signaux et Systèmes, Gif-sur-Yvette, 2002
- [43] Dongbin Xiu and George Em Karniadakis, Modeling Uncertainty in Steady State Diffusion Problems via Generalized Polynomial Chaos, Brown University Providence, July 25 2002
- [44] Dongbin Xiu and George Em Karniadakis, The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations, *SIAM J. Sci. Comput.* Vol. 24, No. 2, pp-619-644, 2002

- [45] Nobuhiro Asai, Izumi Kubo, Hui-Hsiung Kuo, Multiplicative Renormalization and Generating Functions, Taiwanese Journal of Mathematics, Vol. 7, No. 1, pp. 89-101, March 2003
- [46] A. Connes, Noncommutative geometry, Academic Press, 1994
- [47] Anand G. Dabak, Don H. Johnson, Relations between Kullback-Leibler distance and Fisher information, Texs Instruments and Rice University, Texas, 2000
- [48] A.M. Yaglom et I.M. Yaglom, Probabilité et Information, Dunod 1969
- [49] Léon Brillouin, Science and Information Theory, Academic Press, New York 1956

Index

- états, 33
- a posteriori, 36
- a priori, 36
- a priori dégénéré, 89
- a priori de référence, 89
- activité, 63
- anonyme, 19
- axe des risques, 55
- bêta, 67
- bande de confiance, 112
- biais, 42
- binomiale, 61
- conjointe, 35
- conjuguée, 81
- constante radioactive, 139
- courbe ROC, 51
- Cramér-Rao, 77
- critère, 51
- D'Agostini, 29
- décisions, 33
- décroissance, 137
- divergence de Kullback-Leibler, 71
- divergence logarithmique, 71
- ensemble de crédibilité, 44
- estimateur de Bayes, 42
- eupropium-152, 123
- faux négatif, 50
- faux positif, 50
- Feldman-Cousins, 113
- fonction de perte, 38
- fonction de score, 75
- fonction de test, 51
- gamma, 64
- Giunti, 114
- GSL, 123
- information de Fisher, 75
- information discriminante, 72
- information moyenne, 73
- information mutuelle, 72
- intervalle de confiance, 44
- itération, 90
- Jeffreys, 87
- Jeffreys avec données préalables, 96
- Kullback, 176
- limite de décision, 110
- limite de détection, 110
- Loi des Grands Nombres, 41
- Lucrèce, 59
- métrique informationnelle, 76
- marginale, 36
- MDIP, 89
- modèle, 35
- Monte Carlo, 123
- négative-binomiale, 65
- Neyman-Pearson, 50
- niveau critique, 51
- non informatifs, 87
- observations, 33
- p-value, 142
- période, 139
- Pôlya, 67
- Paysan, 19
- Poisson, 62
- possibilité, 41
- précision, 42
- premier type, 50
- propagation des incertitudes, 43

puissance, 50
région de réjection, 50
radium-226, 129
risque bayesen moyen, 39
risque de Bayes, 40
risque fréquentiste, 39
second type, 50
sensibilité, 50, 61
Shannon, 176
simplexe, 93
spécificité, 50
stationnaire, 89
stratégie, 33
Théorème Central Limite, 41
thorium-230, 129
transition poissonnienne, 69
uranium-235, 127
uranium-238, 129
vrai négatif, 50
vrai positif, 50
vraisemblance, 37