Rapport BIPM-80/6

Un nouveau regard sur les probabilités a priori*

par Jörg W. Müller

Bureau International des Poids et Mesures, F-92310 Sèvres

Summary

The well-known Bayes theorem which establishes a probabilistic link between observed random events and their possible "causes", allows us for a series of measurements to dispose at any time of an updated knowledge for instance on a parameter θ which characterizes the underlying distribution law.

Previous discussion was mostly centered on the assignment of the a priori density $f(\theta)$, and in particular for the situation where nothing is known about θ; it was usually concluded that there does not exist a well defined form of $f(\theta)$ corresponding to complete ignorance. However, a paper by E.T. Jaynes actually shows that this is not the case. We describe in some detail his interesting analysis which leads for the two important cases of a Poisson and a binomial variate to an unambiguous determination of their respective prior distributions. They both differ markedly from a constant probability density which is sometimes assumed to describe ignorance.

1. Introduction

Dans toute introduction aux notions élémentaires de probabilité on rencontre nécessairement le théorème de Bayes. En effet, celui-ci n'est qu'une application particulière de la définition de probabilité conditionnelle et par conséquent il repose sur des bases que l'on ne saurait mettre en doute.

On peut considérer ce théorème comme établissant un lien entre nos connaissances, par exemple sur un paramètre connu, telles qu'elles se présentent avant et après avoir tenu compte du résultat d'une expérience qui apporte une information supplémentaire.

---

* L'essentiel de ce rapport a fait l'objet d'un exposé présenté le 6 décembre 1979 à la Section des Rayonnements Ionisants du BIPM.

Dans une écriture habituelle la formule de Bayes prend la forme

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i) \cdot P(A_i)} , \qquad (1)$$

où les $A_1, A_2, \ldots, A_n$ sont des événements, avec des probabilités (absolues) $P(A_i)$ supposées connues, qui s'excluent mutuellement et qui forment un système complet, tandis que

B est un événement pour lequel on dispose des probabilités conditionnelles $P(B \mid A_i)$.

Puisque le dénominateur dans (1), qui joue le rôle d'une constante de normalisation, est égal à $P(B) \neq 0$, il ne dépend pas de A_i. On a donc pour tout i la proportionnalité

$$P(A_i \mid B) \sim P(B \mid A_i) \cdot P(A_i) . \qquad (2)$$

Dans une interprétation courante de cette formule* on considère B comme le "résultat" des "causes" éventuelles A_i. Dans cet esprit on appelle

$P(A_i)$ les probabilités a priori et

$P(A_i \mid B)$ " " a posteriori des "causes" A_i ,

et parfois on désigne aussi $P(B \mid A_i)$ comme vraisemblance. La formule (2) exprime le fait que la probabilité a posteriori varie comme la probabilité a priori multipliée par la vraisemblance.

Les probabilités $P(A_i)$ et $P(A_i \mid B)$ représentent donc notre connaissance (ou opinion) relative aux "causes" A_i avant et après avoir effectué une expérience qui a conduit au résultat B. Le théorème de Bayes permet alors de combiner la connaissance antérieure sur A_i avec la nouvelle information provenant de l'événement B, pour en déduire une estimation renouvelée sur la probabilité des différentes "causes" A_i. Ce procédé est itératif et l'incorporation successive d'une série de résultats B permet de disposer d'une appréciation de tous les A_i qui soit constamment "à jour".

Dans le cas fréquent où les "causes" s'expriment par un seul paramètre continu θ, tandis que les événements B sont toujours discrets, le théorème de Bayes prend la forme

$$g(\theta) = \frac{P(B \mid \theta) \cdot f(\theta)}{\int P(B \mid \theta) \cdot f(\theta)\, d\theta} , \qquad (3)$$

---

* En principe, cette manière de s'exprimer n'est évidemment pas justifiée car l'existence d'une corrélation ne permet pas d'indiquer la direction d'une éventuelle liaison causale.

où $f(\theta)$ est la densité de probabilité a priori pour le paramètre θ à déterminer, tandis que $g(\theta)$ est la densité a posteriori.

Inutile de dire que toutes ces formules sont bien connues depuis fort longtemps et que leur validité ne suscite aucun doute.

Or, aussitôt que l'on essaie de les appliquer, des problèmes surgissent concernant les probabilités a priori qui sont souvent mal connues. Leur estimation (ou, le cas échéant, celle des densités correspondantes) semble trop souvent être basée sur des raisonnements subjectifs qui, en dépit de l'information apparemment identique dont disposent différents évaluateurs, peuvent diverger de façon considérable. Ce fait a réduit de beaucoup l'intérêt pratique du théorème de Bayes. En conséquence, d'autres méthodes d'estimation dites objectives ont été développées pendant les dernières décennies (comme les "confidence intervals" ou les "fiducial intervals", respectivement par J. Neyman et R.A. Fisher), réduisant apparemment la méthode de Bayes à une curiosité intéressante, mais souvent inutilisable en pratique.

2. L'ignorance totale

Le problème de la connaissance des probabilités a priori est particulièrement ardu dans le cas où l'on commence par une situation qui correspond à l'ignorance totale. C'est à ce moment que l'on est tenté de se servir du "postulat" de Bayes pour sortir de l'impasse. Celui-ci propose que dans ce cas toutes les probabilités a priori $P(A_i)$ devraient être considérées comme étant égales. Ce postulat, aussi connu sous le nom de "principe de l'ignorance", conduit pour un paramètre continu θ à une densité de probabilité uniforme dans tout le domaine des valeurs possibles.

Or, ce postulat est fort suspect. Il a certes trouvé, depuis Laplace, des défenseurs renommés, mais d'autre part il est difficile de fermer les yeux sur le fait que son application générale mène à des contradictions logiques. Ceci semble dû au fait que la notion d'ignorance est trop vague et échappe à une définition précise. Les commentaires faits à cet égard dans les traités de statistique, même les plus réputés, sont peu satisfaisants et l'on reste sur sa faim.

C'est exactement à ce point critique que la situation, qui était bloquée et n'aboutissait qu'à des discussions aussi longues que stériles, a changé considérablement, à ce qu'il nous semble. Ce progrès est dû au physicien Edwin T. Jaynes [1] qui a publié, il y a plus de dix ans déjà, un article dans lequel le problème des distributions a priori est abordé de manière tout à fait nouvelle et probablement résolu une fois pour toutes. Ce n'est que très récemment que F.G. Perey (Oak Ridge National Laboratory, USA) a attiré notre attention sur ce travail qui semble avoir échappé à presque tout le monde. Le présent rapport n'a d'autre but que de souligner l'importance de cette publication en lui empruntant deux exemples pratiques qui nous paraissent bien illustrer son intérêt général.

L'idée de base de Jaynes consiste dans un approfondissement de la notion "ignorance". On peut presque dire (au moins rétrospectivement) que Jeffreys [2] , qui s'est penché sur ces problèmes pendant longtemps et avec beaucoup de perspicacité, était tout près de la même solution en suggérant de choisir comme distributions initiales, qui reflètent notre ignorance totale, celles qui sont invariantes sous toutes les transformations les plus plausibles des méthodes de mesure. Contrairement à l'avis de la majorité des statisticiens, il était convaincu qu'il n'y a rien d'arbitraire dans leur choix et qu'il n'existe qu'une seule distribution correcte. Il convient, cependant, de distinguer clairement entre une heureuse idée générale et son application pratique, surtout dans un domaine aussi controversé que celui-ci.

Pour faciliter, dans les exemples qui suivent, la comparaison avec le texte original de Jaynes, on ajoutera, chaque fois que cela semble utile, la numérotation de ses équations "(J ..)" à la nôtre.

3. Loi de Poisson

Pour un processus de Poisson, la probabilité qu'exactement k événements se produisent pendant une durée t est donnée par

$$P(k\,|\,\theta) = \frac{(\theta t)^k}{k\,!}\, e^{-\theta t} , \qquad (4) = (J\ 55)$$

où θ est le taux de comptage du processus que nous cherchons à estimer. Avant la première observation nous sommes complètement ignorants sur la valeur de θ, dont la dimension est s^{-1}.

Admettons deux observateurs qui utilisent différentes échelles de temps (horloges). Leurs déterminations de la durée d'un intervalle de temps sont alors liées par une relation du type $t_1 = q \cdot t_2$, et l'on a par conséquent pour les taux respectifs $\theta_2 = q \cdot \theta_1$. Pour exprimer leur ignorance, les deux observateurs utilisent des densités a priori que nous désignons par $f_1(\theta_1)$ pour l'un et par $f_2(\theta_2)$ pour l'autre. Pour que ces densités soient mutuellement cohérentes, il faut demander que $f_1(\theta_1)\, d\theta_1 = f_2(\theta_2)\, d\theta_2$, d'où l'on tire $f_1(\theta_1) = q \cdot f_2(\theta_2)$. D'autre part, le même état d'ignorance implique que $f_1(\theta) = f_2(\theta)$. La densité a priori que nous cherchons à déterminer doit donc obéir à l'équation fonctionnelle $f(\theta) = q \cdot f(q\theta)$, dont la solution est

$$f(\theta) = \frac{\text{const.}}{\theta} , \qquad (5) = (J\ 58)$$

comme on le vérifie aisément. Seule cette fonction garantit que toutes les échelles de temps sont équivalentes.

Il est vrai que (5) n'est pas normalisable, mais ce petit défaut est sans conséquence pour la détermination de la densité a posteriori puisque $f(\theta)$ apparaît dans (3) au numérateur et au dénominateur.

Appliquons maintenant ce résultat, sans doute un peu surprenant, pour en déduire la densité a posteriori correspondante. En utilisant (3) et (4) on arrive, avec

$\theta \cdot t = \mu$, à

$$g(\mu) = \frac{\frac{t}{\mu} \cdot \frac{\mu^k \cdot e^{-\mu}}{k!}}{t \int_0^\infty \frac{1}{\mu} \cdot \frac{\mu^k \cdot e^{-\mu}}{k!} \, d\mu} = \frac{\mu^{k-1} \cdot e^{-\mu}}{(k-1)!} , \tag{6}$$

puisque $\int_0^\infty \mu^k \cdot e^{-\mu} d\mu = k!$. Il s'agit donc d'une densité gamma (ou distribution d'Erlang).

Le calcul des moments de l'ordre s de μ donne

$$m_s(\mu) = \int_0^\infty \mu^s \cdot g(\mu) \, d\mu = \frac{(k+s-1)!}{(k-1)!} , \tag{7}$$

d'où l'on déduit que

$$E(\mu) = m_1(\mu) = k \quad \text{et}$$
$$V(\mu) = m_2(\mu) - m_1^2(\mu) = k . \tag{8}$$

Cette estimation de μ semble toute naturelle. Elle est beaucoup plus convaincante que l'ancien résultat obtenu en se basant sur une distribution a priori uniforme pour $\mu > 0$ et qui donnait (voir par exemple [3], ou aussi [4] pour plus de détails)

$$E^*(\mu) = V^*(\mu) = k + 1 , \tag{9}$$

résultat qui était toujours difficile à expliquer. Au lieu d'une densité a priori uniforme pour θ (ou pour μ), qui entraîne par ailleurs, les mêmes ennuis de normalisation, on a aussi proposé une densité a priori de la forme $f(\mu) = c \cdot e^{-c\mu}$ pour $\mu > 0$ [5], mais elle conduirait également au résultat (9), jugé peu satisfaisant.

Il est intéressant, dans notre exemple, de voir comment la formule de Bayes s'applique dans le cas d'une situation répétitive où il importe de tenir compte de nouveaux résultats au fur et à mesure que ceux-ci sont disponibles. Puisque à chaque étape de renouvellement l'ancienne densité a posteriori devient la nouvelle densité a priori, une écriture un peu plus générale est pratique. Définissons des densités gammas par

$$G_r(\mu) = \frac{1}{(\gamma - 1)!} \cdot \mu^{\gamma - 1} \cdot e^{-\mu} , \tag{10}$$

où les résultats des r premières mesures, c'est-à-dire $k_1, k_2, \ldots, k_r$, entrent sous la forme

$$\gamma = K_r = k_o + \sum_{j=1}^{r} k_j . \tag{11}$$

Le paramètre k_o décrit, sous forme d'un résultat réel ou fictif, nos connaissances sur μ avant de commencer les expériences. Si après $r-1$ mesures nous pouvons identifier la densité a priori à une densité gamma telle que

$$f(\mu) = G_{r-1}(\mu) , \tag{12}$$

on peut alors montrer que l'on reste toujours à l'intérieur de la famille gamma et que

$$g(\mu) = G_r(\mu) . \tag{13}$$

La démonstration est donnée dans l'Annexe A.

D'après (8) on a de façon générale pour le paramètre μ, qui est maintenant décrit par (13), la relation

$$E(\mu) = V(\mu) = K_r . \tag{14}$$

Tandis que pour une densité initiale rectangulaire (ou uniforme), ainsi que pour le choix $f(\mu) = G_o(\mu) = e^{-\mu}$, on trouve $k_o = 1$, la densité obtenue par Jaynes $f(\mu) \sim 1/\mu$ correspond à $k_o = 0$. Son emploi comme représentation de notre ignorance initiale conduit seul à l'estimation

$$E(\mu) = V(\mu) = \sum_{j=1}^{r} k_j , \qquad \text{pour } r \geqslant 1 , \tag{15}$$

qui correspond à ce que l'on obtient par d'autres méthodes d'estimation ainsi qu'au résultat suggéré par le "bon sens".

4. Loi binomiale

Ce deuxième exemple, que nous empruntons également au travail de Jaynes, est un peu plus compliqué, mais fort intéressant. Il nous amène à un résultat assez surprenant, au moins à première vue, mais qui se révélera comme possédant des propriétés tout à fait appropriées.

Pour des épreuves de Bernoulli il est bien connu que la probabilité d'observer k "succès" pour n "tentatives" ou "essais" s'exprime par la loi binomiale

$$P(k \mid n, \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k} , \qquad (16) = (J\ 59)$$

où θ est la probabilité pour un "succès". De nouveau on se propose de déterminer la distribution a priori $f(\theta)$ qui corresponde à l'ignorance complète que nous avons à l'égard de θ avant d'avoir effectué une expérience appropriée.

Suivons brièvement le raisonnement fort intéressant de Jaynes* [1]. On suppose que $f(\theta)$ décrit la répartition des opinions personnelles d'un grand nombre d'individus. Est-il possible alors - bien que chaque personne ait une certaine opinion (peut-être un préjugé) - que cette population dans son ensemble ignore tout sur θ? Dans l'affirmative, il se posera le problème de déterminer la forme de la distribution $f(\theta)$ qui décrit cette confusion totale des esprits sur la valeur de θ.

Supposons maintenant qu'avant d'entreprendre une expérience une nouvelle information (appelée B) soit donnée simultanément à tout membre de la population considérée. En conséquence, chacun changera son opinion en incorporant cette nouvelle pièce d'information à sa manière. Cette adaptation sera décrite par la formule de Bayes. Avec la notation utilisée dans (1) et en désignant par A l'opinion (personnelle) originelle (c'est-à-dire avant la connaissance de B), nous pouvons dire que la probabilité a priori pour un succès est

$$P(A) = \theta_1 \quad \text{avant d'avoir pris connaissance de B,}$$

mais (17a)

$$P(A \mid B) = \theta_2 \quad \text{après avoir appris B.}$$

Les événements A et $\overline{A}$ ("non A") sont exhaustifs, d'où

$$P(A) + P(\overline{A}) = 1 \, . \tag{17b}$$

D'après la formule de Bayes (1) on a donc pour tout membre de la population la relation

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B \mid A) \cdot P(A) + P(B \mid \overline{A}) \cdot P(\overline{A})} \, . \tag{18}$$

Or, avec (17) et en posant $P(B \mid A)/P(B \mid \overline{A}) = \alpha$, on peut écrire (18) sous la forme équivalente

$$\theta_2 = \frac{\alpha \cdot \theta_1}{1 - \theta_1 + \alpha\theta_1} \qquad (19) = (J\ 62)$$

qui exprime la relation entre l'ancienne (θ_1) et la nouvelle (θ_2) opinion sur θ. L'astuce d'incorporer l'information B nous a donc permis d'établir une transformation qui projette l'espace du paramètre θ sur lui-même. Il s'agit maintenant de tirer le meilleur profit de (19) pour en déduire la distribution a priori de θ.

---

* Ce "résumé" n'a pas la prétention de contenir toute la richesse des explications données par Jaynes; l'étude de l'original sera indispensable pour tout lecteur désireux de comprendre les idées de base.

Admettons que la population, dans son ensemble, n'ait rien à apprendre de l'information B et qu'elle reste donc dans un état de confusion générale à l'égard de θ. Or, si θ_1 est décrit par la densité de probabilité $f_1(\theta_1)$ et θ_2 par $f_2(\theta_2)$, cet état persistant d'ignorance complète implique que

$$f_1(\theta_1)\, d\theta_1 = f_2(\theta_2)\, d\theta_2 \qquad \text{(20a) = (J 64)}$$

et l'on en déduit que f_1 et f_2 sont les mêmes fonctions, d'où

$$f_1(\theta) = f_2(\theta)\,. \qquad \text{(20b) = (J 65)}$$

Dans ce qui suit tous les détails mathématiques qui complètent la présentation de Jaynes m'ont été fournis par Mme M. Boutillon, à laquelle j'ai pris l'habitude de m'adresser dans les situations désespérées.

A partir de (19) on obtient par différentiation

$$\frac{d\theta_2}{d\theta_1} = \frac{\alpha}{(1-\theta_1+\alpha\,\theta_1)^2} \qquad (21)$$

et il découle de (20a) que

$$f(\theta_2) = f(\theta_1)\cdot(1-\theta_1+\alpha\,\theta_1)^2/\alpha$$

ou aussi

$$\alpha\cdot f(\theta_2) = (1-\theta_1+\alpha\,\theta_1)^2\cdot f(\theta_1)\,. \qquad \text{(22) = (J 66)}$$

La différentiation de (22) par rapport à α donne

$$f(\theta_2) + \alpha\cdot\frac{df(\theta_2)}{d\alpha} = 2\,\theta_1\,(1-\theta_1+\alpha\,\theta_1)\cdot f(\theta_1)\,. \qquad (23)$$

Puisque

$$\frac{df(\theta_2)}{d\alpha} = \frac{df(\theta_2)}{d\theta_2}\cdot\frac{d\theta_2}{d\alpha}$$

et

$$\frac{d\theta_2}{d\alpha} = \frac{\theta_1\,(1-\theta_1)}{(1-\theta_1+\alpha\,\theta_1)^2}\ ,$$

on peut, à l'aide de ces relations, remplacer (23) par

$$f(\theta_1)\cdot\frac{(1-\theta_1+\alpha\,\theta_1)^2}{\alpha} + \alpha\cdot\frac{\theta_1\cdot(1-\theta_1)}{(1-\theta_1+\alpha\,\theta_1)^2}\cdot f'(\theta_1) = 2\,\theta_1\,(1-\theta_1+\alpha\,\theta_1)\cdot f(\theta_1)\ , \qquad (24)$$

où l'on a écrit $f'(\theta_1)$ pour $\frac{df(\theta_2)}{d\theta_2} = \frac{df(\theta_1)}{d\theta_1}$.

Pour $\alpha = 1$ et en posant $\theta_1 = \theta$, on obtient

$$f(\theta) + \theta(1-\theta) \cdot f'(\theta) = 2\,\theta \cdot f(\theta) . \qquad (25) = (J\ 67)$$

Il est facile de vérifier que cette équation différentielle a pour solution

$$f(\theta) = \frac{\text{const.}}{\theta(1-\theta)} , \qquad \text{pour} \quad 0 < \theta < 1. \qquad (26) = (J\ 68)$$

Cette forme de la densité a priori $f(\theta)$ s'écarte très nettement d'une densité rectangulaire de θ (entre les limites 0 et 1) à laquelle on pourrait être amené par le raisonnement fautif qui admet qu'ignorance équivaut à probabilité constante pour tout le domaine des valeurs possibles de θ.

On remarquera que (26) est une densité "impropre" puisqu'elle ne peut pas être normalisée. Nous avons déjà noté la même propriété pour (5) qui est valable pour une distribution de Poisson. Il se pourrait qu'elle soit générale pour des densités exprimant une ignorance totale.

Pour la loi binomiale l'application du théorème de Bayes (3) pour déterminer la densité a posteriori $g(\theta)$ de la probabilité θ donne lieu à des considérations analogues à celles que nous avons faites à l'égard de la loi de Poisson (voir section 3).

Cette fois le résultat d'une mesure ou d'une expérience consiste chaque fois en une paire de valeurs indiquant le nombre k de "succès" remportés pour n "essais". Pour une série de r mesures on dispose donc des couples

$$(k_1 , n_1), (k_2 , n_2), \ldots , (k_r , n_r) .$$

Pour des fonctions appartenant à la "famille" bêta nous utilisons l'écriture

$$B_r(\theta) = \frac{1}{C(\alpha,\beta)} \cdot \theta^{\alpha-1} (1-\theta)^{\beta-1} , \qquad (27)$$

avec $0 \leqslant \theta \leqslant 1$, $\alpha > 0$ et $\beta > 0$,

$$\text{où } C(\alpha,\beta) = \int_0^1 \theta^{\alpha-1} (1-\theta)^{\beta-1}\, d\theta = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

assure la normalisation.

Si l'on écrit comme auparavant

$$K_r = k_o + \sum_{i=1}^{r} k_i \qquad (11)$$

et de façon analogue

$$N_r = n_o + \sum_{i=1}^{r} n_i \ , \tag{11'}$$

il est possible de faire les identifications suivantes

$$\alpha = K_r \ ,$$
$$\beta = N_r - K_r \ . \tag{28}$$

On supposera que la densité a priori contient toute l'information acquise précédemment (donc jusqu'à la mesure numéro r − 1) tandis que la densité a posteriori tient également compte de la nouvelle mesure r.

Par analogie avec la situation considérée pour la loi de Poisson, on peut maintenant exprimer la densité a priori par

$$f(\theta) = B_{r-1}(\theta) \tag{29}$$

tandis que la densité a posteriori est

$$g(\theta) = B_r(\theta) \ . \tag{30}$$

Pour montrer que cette simple identification est valable, il suffit d'insérer (29) dans la formule de Bayes, ce qui est fait dans l'Annexe B.

Par conséquent, les densités de θ sont toujours des membres de la famille des fonctions bêta et seuls les paramètres K et N changent selon l'information dont on dispose à un moment donné.

Considérons maintenant les moments. Pour une variable θ qui suit la densité bêta (27), le moment ordinaire d'ordre s est donné par (voir [4] ou tout traité de statistique)

$$m_s(\theta) = \int_0^1 \theta^s \cdot B_r(\theta) \, d\theta = \prod_{j=1}^{s-1} \frac{\alpha + j}{\alpha + \beta + j} = \frac{\Gamma(\alpha+\beta) \cdot \Gamma(\alpha+s)}{\Gamma(\alpha+\beta+s) \cdot \Gamma(\alpha)} \ . \tag{31}$$

Il s'ensuit pour la valeur moyenne

$$E(\theta) = m_1(\theta) = \frac{\alpha}{\alpha+\beta} \tag{32a}$$

et pour la variance

$$V(\theta) = m_2(\theta) - m_1^2(\theta) = \frac{\alpha \cdot \beta}{(\alpha+\beta)^2 \, (\alpha+\beta+1)} \ . \tag{33a}$$

A l'aide de (28) on a donc également

$$E(\theta) = \frac{K_r}{N_r} , \tag{32b}$$

$$V(\theta) = \frac{K_r (N_r - K_r)}{N_r^2 (N_r + 1)} . \tag{33b}$$

On remarquera que la distribution a priori (26) qui exprime l'ignorance complète correspond dans la notation de (27) à $\alpha = \beta = 0$, donc à $k_o = n_o = 0$. Pour la première expérience $(r = 1)$ et sans connaissance préalable, on a par conséquent

$$\left.\begin{aligned} E(\theta) &= \frac{k_1}{n_1} \quad \text{et} \\ V(\theta) &= \frac{k_1 (n_1 - k_1)}{n_1^2 (n_1 + 1)} , \end{aligned}\right. \tag{34}$$

tandis que pour $r > 1$ ces deux moments sont donnés par (32b) et (33b), où l'on a maintenant

$$K_r = \sum_{i=1}^{r} k_i \quad \text{et} \quad N_r = \sum_{i=1}^{r} n_i .$$

Notons que ces résultats diffèrent de ceux que l'on obtient en partant d'une densité rectangulaire

$$f^*(\theta) = \begin{cases} 1 & \text{pour } 0 < \theta < 1 \\ 0 & \text{à l'extérieur} \end{cases} \tag{35}$$

qui correspondrait à $\alpha = \beta = 1$, c'est-à-dire, $k_o = 1$ et $n_o = 2$. La forme (35), qui passe traditionnellement pour la densité à utiliser en l'absence de connaissance sur θ, nous aurait amenés, pour une seule mesure $(r = 1)$, aux moments

$$\left.\begin{aligned} E^*(\theta) &= \frac{k_1 + 1}{n_1 + 2} \quad \text{et} \\ V^*(\theta) &= \frac{(k_1 + 1)(n_1 - k_1 + 1)}{(n_1 + 2)^2 (n_1 + 3)} , \end{aligned}\right. \tag{36}$$

expressions qui ont toujours été difficiles à accepter puisqu'elles introduisent un biais incompréhensible dans l'estimation de ces moments.

D'autres propositions, comme par exemple celle qui consiste à prendre [6]

$$f^*(\theta) = 6\,\theta \cdot (1 - \theta), \qquad \text{pour} \quad 0 \leqslant \theta \leqslant 1\,, \tag{37}$$

ce qui correspond à $\alpha = \beta = 2$ ou $k_o = 2$ et $n_o = 4$, mèneraient à des estimations pour les moments qui seraient encore moins acceptables.

Or, de façon générale il est clair qu'une distribution a priori avec $n_o > 0$ correspond au résultat d'une expérience réelle et ne peut donc pas être prise pour l'expression d'une ignorance complète.

On voit maintenant que toutes ces difficultés ne sont que le résultat d'un problème mal posé et qu'elles disparaissent si l'on utilise la formule (26) de Jaynes.

5. Remarques finales

On fait souvent le reproche qu'il n'est pas légitime d'utiliser comme densité a priori une fonction non normalisable; il convient cependant de remarquer que cet inconvénient est purement formel et n'a pas beaucoup de poids. Dans le cas de (27), en particulier, tout est bien défini et clair pour $\alpha > 0$ et $\beta > 0$. Or, rien ne nous empêche d'interpréter (26) comme une écriture simplifiée pour le cas limite $\alpha \longrightarrow 0$ et $\beta \longrightarrow 0$ où l'on s'arrête juste avant d'atteindre zéro. Par ailleurs, il est évident d'après la forme même de (3) que la constante indéterminée dans (26) n'a pas d'effet sur la densité a posteriori $g(\theta)$.

Les résultats présentés dans ce rapport ont également un certain lien avec le problème tant discuté de la représentation des incertitudes expérimentales qui sont du type souvent appelé "systématique". En admettant que l'on ne puisse leur attribuer que des limites et que notre ignorance sur la valeur "vraie" à l'intérieur de cet intervalle corresponde à une probabilité constante, on a proposé de les décrire par une densité de forme rectangulaire. La validité de cette conclusion a été mise en doute [7] pour des raisons qui nous semblent toujours valables. Le travail de Jaynes montre maintenant que la notion d'ignorance complète, qui paraissait trop vague, peut avoir dans un contexte donné un sens bien défini. Le paramètre θ dans la loi binomiale donne finalement un exemple qui ressemble beaucoup au problème posé par les incertitudes "systématiques" où l'on ne connaît non plus que les limites.
Le résultat (26) ne peut donc que renforcer les doutes déjà émis d'associer une densité de probabilité de forme rectangulaire à ce type d'incertitude.

C'est grâce à Francis G. Perey, Oak Ridge National Laboratory, qui a visité le BIPM le 23 octobre 1979, que j'ai eu connaissance du travail important de E.T. Jaynes. Je suis également très reconnaissant à Mme M. Boutillon dont la perspicacité mathématique n'est dépassée que par sa modestie.

Ce petit rapport aurait pleinement atteint son but s'il contribuait à mieux faire connaître les idées du Prof. Jaynes parmi les physiciens et les statisticiens qui, dans leur majorité, continuent à les ignorer.

Note ajoutée aux épreuves

Nous venons de découvrir un nouvel et très important article de Jaynes [8] qui reprend les problèmes discutés dans [1] et les généralise de manière remarquable. Pour les questions abordées ici rien n'est à changer, bien que les fondements de l'approche soient maintenant mieux consolidés.

ANNEXES

A. Dérivation de la relation (13)

Nous voulons montrer que pour une distribution de Poisson et une densité a priori $f(\mu)$ du type gamma, l'emploi de la formule de Bayes donne pour la densité a posteriori $g(\mu)$ une fonction qui appartient également à cette famille. Par conséquent, la mise à jour de la nouvelle information peut se faire par un simple changement du paramètre K_r qui détermine la fonction gamma.

Pour démontrer ce fait nous calculons d'abord le dénominateur de (3). Avec (4) et (12) on a

$$\int_0^\infty P(k_r | \mu) \cdot G_{r-1}(\mu)\, d\mu = \frac{1}{k_r!} \int_0^\infty \mu^{k_r} \cdot e^{-\mu} \cdot \frac{\mu^{K_{r-1}-1}}{(K_{r-1}-1)!} \cdot e^{-\mu}\, d\mu$$

$$= \frac{1}{(K_r - k_r - 1)!\, k_r!} \int_0^\infty \mu^{K_{r-1}+k_r-1} \cdot e^{-2\mu}\, d\mu . \qquad \text{(A1)}$$

Puisque $K_{r-1}+k_r = K_r$, on obtient, en posant $2\mu = \lambda$,

$$\binom{K_r-1}{k_r} \frac{1}{(K_r-1)!} \int_0^\infty \left(\frac{\lambda}{2}\right)^{K_r-1} \cdot e^{-\lambda} \frac{d\lambda}{2} = \binom{K_r-1}{k_r} \cdot 2^{-K_r} . \qquad \text{(A2)}$$

Ceci donne pour la densité a posteriori

$$g(\mu) = \frac{\mu^{k_r}}{k_r!} \cdot e^{-\mu} \cdot \frac{\mu^{K_{r-1}-1}}{(K_{r-1}-1)!} \cdot e^{-\mu} \cdot \frac{2^{K_r}}{\binom{K_r-1}{k_r}} = \frac{\mu^{K_r-1}}{(K_r-1)!} \cdot e^{-2\mu} \cdot 2^{K_r} . \qquad \text{(A3)}$$

Avec $2\mu = \lambda$ nous avons

$$g(\mu) = \left(\frac{\lambda}{2}\right)^{K_r-1} \cdot \frac{e^{-\lambda}}{(K_r-1)!} \cdot 2^{K_r} = 2 \cdot \frac{\lambda^{K_r-1}}{(K_r-1)!} \cdot e^{-\lambda} ,$$

ou bien, puisque $d\mu/d\lambda = 1/2$,

$$g(\lambda) = g(\mu) \cdot 1/2 .$$

En écrivant λ au lieu de μ , on arrive finalement à la formule

$$g(\mu) = \frac{\mu^{K_r-1}}{(K_r-1)!} \cdot e^{-\mu} = G_r(\mu) . \qquad \text{(13)}$$

B. Vérification de la relation (30)

Pour une loi binomiale, le théorème de Bayes (3) prend pour la mesure numéro r la forme

$$g(\theta) = \frac{\binom{n_r}{k_r}\, \theta^{k_r} (1-\theta)^{n_r-k_r} \cdot f(\theta)}{\binom{n_r}{k_r} \int_0^1 \theta^{k_r} (1-\theta)^{n_r-k_r} \cdot f(\theta)\, d\theta} \; . \tag{B1}$$

Si la densité a priori est donnée par (29), donc avec (27) et (28) par

$$f(\theta) = \frac{1}{C(\alpha,\beta)}\, \theta^{K_{r-1}-1} (1-\theta)^{N_{r-1}-K_{r-1}-1} \; , \tag{B2}$$

on obtient pour (B1)

$$\begin{aligned} g(\theta) &= \frac{\theta^{k_r} (1-\theta)^{n_r-k_r} \cdot \theta^{K_{r-1}-1} (1-\theta)^{N_{r-1}-K_{r-1}-1}}{\int_0^1 \theta^{k_r} (1-\theta)^{n_r-k_r} \cdot \theta^{K_{r-1}-1} (1-\theta)^{N_{r-1}-K_{r-1}-1}\, d\theta} \\ &= \frac{\theta^{K_r-1} (1-\theta)^{N_r-K_r-1}}{C(K_r, N_r-K_r)} \; , \end{aligned} \tag{B3}$$

puisque $K_{r-1} + k_r = K_r$ et $N_{r-1} + n_r = N_r$.

La comparaison avec (27) et (28) montre immédiatement que l'on a en effet

$$g(\theta) = B_r(\theta) \; . \tag{30}$$

Références

[1] E.T. Jaynes: "Prior probabilities", IEEE Trans. Syst. Sci. and Cyb. SSC-4, 227 (1968)

[2] H. Jeffreys: "Theory of Probability" (Clarendon, Oxford, 1939^1, 1961^3)

[3] L.J. Rainwater, C.S. Wu: "Applications of probability theory to nuclear particle detection", Nucleonics, Oct. 1947, p. 60

[4] J.W. Müller: "Traitement statistique des résultats de mesure", Rapport BIPM-108 (1969 ff.), p. 111 ff.

[5] D. Morgenstern: "Einführung in die Wahrscheinlichkeitsrechnung und mathematische Statistik" (Springer, Berlin, 1968^2), p. 191

[6] A.M. Mood, F.A. Graybill: "Introduction to the Theory of Statistics" (McGraw-Hill, New York, 1963^2), p. 196

[7] J.W. Müller: "Some second thoughts on error statements", Nucl. Instr. and Meth. 163, 241 (1979)

[8] E.T. Jaynes: "Marginalization and prior probabilities", in "Bayesian Analysis in Econometrics and Statistics", éd. par A. Zellner (North-Holland, Amsterdam, 1980), p. 43 ff.

(Août 1980)