C'EST BON A SAVOIR!

Et si vous étiez un bayésien « qui s'ignore » ?

Bruno Lecoutre

U.R.A. 1378, *Analyse et Modèles Stochastiques*,

C.N.R.S. et Université de Rouen - Mathématiques Site Colbert

76821 Mont-Saint-Aignan Cedex France

Courrier électronique : bruno.lecoutre@univ-rouen.fr

L'objet de cet article est de guider le lecteur peu familiarisé dans la découverte de l'inférence bayésienne. Quatre idées pourront motiver cette découverte : l'inférence bayésienne n'est pas récente ; elle apparaît supérieure sur le plan théorique ; elle est une inférence naturelle ; elle va devenir de plus en plus facilement utilisable. L'exposé sera très partiel (et partial), avec tous les oublis et toutes les insuffisances inévitables s'agissant d'un sujet aussi débattu que l'inférence statistique.

Nous prendrons comme point de départ le fait que les interprétations spontanées des résultats des procédures statistiques traditionnelles (seuils de signification, intervalles de confiance), même par des utilisateurs « avertis », sont le plus souvent en termes de probabilités sur les paramètres, qui sont en fait les probabilités *naturelles* : « celles qui vont du connu vers l'inconnu » Ainsi, dans un ouvrage récent d'introduction à la statistique, appartenant à une collection destinée au grand public, dont l'objectif est de permettre au lecteur d'« accéder aux intuitions profondes du domaine », on trouve l'interprétation suivante de l'intervalle de confiance (ou « fourchette ») pour une proportion : «*si dans un sondage de taille* 1000, *on trouve P* [la proportion observée] = 0.613, *la proportion* π_1 *à estimer a une probabilité* 0.95 *de se trouver dans la fourchette* : [0.58,0.64] » (Claudine Robert, 1995, page 221).

Si vous n'êtes pas (encore) bayésien et si votre intuition profonde est que cette interprétation est, soit correcte, soit peut-être incorrecte mais en tout cas souhaitable, vous devez sérieusement vous demander si vous n'êtes pas un bayésien « qui s'ignore ». Pour vous aider à y voir plus clair, nous exposerons d'abord, à partir de cet exemple de l'intervalle de confiance sur une proportion π, la différence essentielle entre l'inférence « classique » et l'inférence bayésienne ; puis nous esquisserons une présentation générale de l'inférence bayésienne.

Tolérer l'erreur ou changer de cadre de justification ?

L'intervalle de confiance (procédure d'échantillonnage)

Dans la conception classique de l'intervalle de confiance, les bornes observées pour l'échantillon (unique) dont on dispose ne sont interprétables qu'en référence à l'ensemble de tous les intervalles qu'on aurait pu observer : formellement, les bornes de l'intervalle de confiance pour le paramètre π sont des grandeurs *aléatoires*, qui varient d'un échantillon à un autre. L'interprétation *correcte* de l'intervalle de confiance 0.95 est alors la suivante : « *95% des intervalles calculés sur l'ensemble des échantillons possibles (tous ceux qu'il est possible de tirer) contiennent la vraie valeur* π ». Mais cet énoncé est *conditionnel à* π : il ne dépend pas des observations et est déterminé avant leur recueil. En fait dans ce cadre de justification, les seules probabilités envisagées sont les probabilités d'échantillonnage conditionnelles à π. En revanche les valeurs possibles du paramètre *ne peuvent pas être probabilisées* : si comme dans l'exemple précédent, les bornes obtenues pour l'échantillon observé sont [0.58,0.64], l'événement « $0.58<\pi<0.64$ » est vrai ou faux (puisque π est fixé), et nous ne pouvons pas lui attribuer de probabilité (sinon 1 ou 0). Il est donc *illégitime* d'écrire « $Pr(0.58<\pi<0.64) = 0.95$ » ou d'énoncer que « *il y a 95% de chances que le paramètre inconnu* π *soit compris entre* 0.58 *et* 0.64 ».

L'intervalle de crédibilité bayésien

Dans l'inférence bayésienne, au contraire, on probabilise explicitement les valeurs possibles du paramètre. A partir d'un état de connaissance *initial* formalisé par une distribution *a priori*, et des données, nous obtenons par la formule de Bayes (voir plus loin) une distribution *a posteriori* qui exprime directement notre incertitude sur le paramètre, *conditionnellement à l'échantillon particulier observé*. La distribution *a posteriori* combine l'information initiale avec l'information apportée par les données.

De la distribution *a posteriori* nous pouvons déduire pour le paramètre une « fourchette », que l'on appelle habituellement intervalle de *crédibilité* pour le distinguer de l'intervalle de confiance. Par définition, nous attribuons à cet intervalle de crédibilité la probabilité considérée (par exemple 0.95) : étant donné les observations, nous avons une probabilité 0.95 (« 95% de chances ») que la proportion vraie soit comprise entre les bornes de l'intervalle.

Il est clair, comme le montrent les figure 1a et 1b toujours pour le même exemple, que cet intervalle de crédibilité dépend de la distribution *a priori*. Dans la figure 1b, la distribution *a priori* est très dispersée et traduit une information initiale relativement *vague* par rapport à l'information apportée par les données. Dans ce cas, l'intervalle de crédibilité a les mêmes bornes que l'intervalle de confiance [0.58,0.64] (tout au moins avec la précision de deux décimales). Dans la figure 1a, la distribution *a priori* est au contraire assez peu dispersée et traduit une information initiale relativement précise ; ceci a pour conséquence, d'une part de décaler la distribution *a posteriori* vers des valeurs plus petites, et d'autre part de diminuer la dispersion de cette distribution (car on a davantage d'informations), d'où l'intervalle de crédibilité [0.55,0.60].

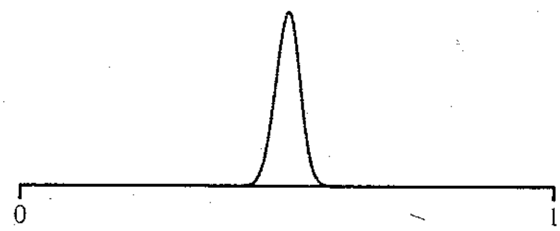
De plus nous pouvons encore calculer la probabilité associée à un intervalle d'intérêt, dont les bornes peuvent être fixées indépendamment des données, tel que [0.50,1]. L'inférence bayésienne apporte donc une analyse probabiliste directe et naturelle sur les valeurs possibles du paramètre, qui peut être regardée comme un prolongement direct des procédures descriptives.

L'interprétation bayésienne de l'intervalle de confiance

Le cadre de justification bayésien fournit une interprétation *légitime* de l'intervalle de confiance en termes de probabilités sur le paramètre. Bien entendu, la probabilité bayésienne associée à cet intervalle (dans l'exemple considéré ici [0.58,0.64]) dépend de la distribution *a priori* et pourra donc différer plus ou moins fortement de la confiance (ici 0.95). Ainsi, si pour la figure 1b nous obtenons bien « $Pr(0.58<\pi<0.64) = 0.95$ », pour la figure 1a, nous obtenons : « $Pr(0.58<\pi<0.64) = 0.36$ ».

Distribution *a priori* + données → Distribution *a posteriori*

Figure 1a



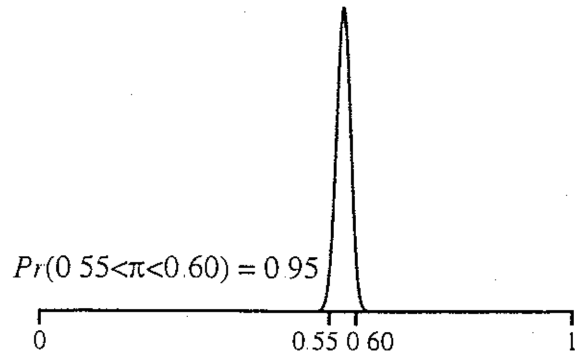
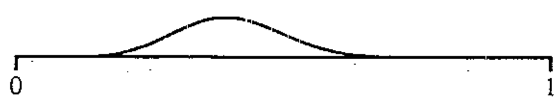
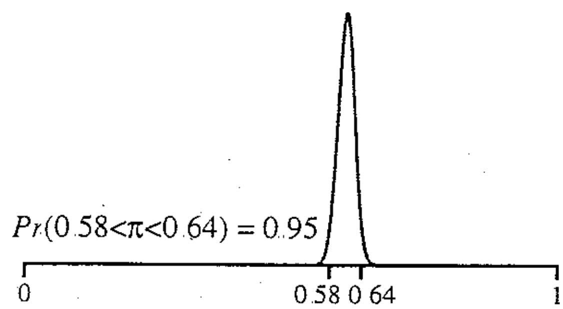


Figure 1b





Et l'utilisateur ?

Dans la pratique, l'utilisateur a donc le choix entre trois attitudes : 1) conserver le cadre de justification classique de l'intervalle de confiance et se satisfaire de l'interprétation « correcte » ; 2) conserver ce cadre tout en adoptant l'interprétation bayésienne alors « erronée » ; 3) adopter explicitement le cadre de justification bayésien.

Actuellement, tout montre que la majorité des utilisateurs adoptent la deuxième attitude. La même situation apparaît dans le cas du test de signification, pour lequel le seuil de signification est souvent interprété comme la probabilité « *que l'hypothèse nulle soit vraie* » alors qu'il est la probabilité *conditionnelle* « de rejeter [à tort] l'hypothèse nulle si cette hypothèse est vraie ». On peut donc penser que ce sont paradoxalement leurs interprétations bayésiennes sauvages qui rendent ces procédures populaires. Mais alors, comme le dit Rouanet (Rouanet *et al.*, 1991, page 43) à ce propos, « tolérer l'erreur, quelle perspective peu exaltante ! ».

Quelques considérations générales

Modèle probabiliste et vraisemblance

Un modèle probabiliste caractérise le comportement des observations x, *conditionnellement* au paramètre θ ; il fournit des probabilités d'échantillonnage $Pr(x|\theta)$. Etant donné l'observation x, $Pr(x|\theta)$ peut être regardée comme une fonction, non pas de x, mais de θ ; considérée ainsi, à la suite de Fisher, elle est appelée la fonction de *vraisemblance* (*likelihood*) de θ pour la valeur de x donnée, et s'écrit $v(\theta|x)$. Formellement, elle permet de réécrire le modèle probabiliste dans le « bon ordre » (dans la mesure où l'objectif est une inférence sur θ) :

$$v(\theta|x) = Pr(x|\theta)$$

(remarque : il suffit ici de considérer des « probabilités discrètes », celles-ci pouvant être remplacées s'il y a lieu par des densités).

Inférence fréquentiste et inférence bayésienne

L'inférence statistique est fondamentalement une démarche d'inversion puisqu'elle vise à « remonter des effets aux causes » : des observations aux paramètres. Seule la théorie bayésienne réalise cette inversion de façon légitime et cohérente : partant d'un état d'incertitude *initial* sur le paramètre, le théorème de Bayes permet d'exprimer la nouvelle incertitude sur les valeurs possibles du paramètre, une fois les données recueillies.

Dans l'inférence statistique traditionnelle (le test de signification, l'intervalle de confiance...), les seules probabilités considérées sont les probabilités d'échantillonnage $Pr(x|\theta)$, conditionnelles au paramètre. Ces probabilités peuvent s'interpréter comme des proportions, ou des fréquences, d'où l'appellation d'*inférence fréquentiste* (on parle également de « procédures d'échantillonnage »). Une remarque sur la terminologie : l'auto-appellation de « classique » que s'attribue l'inférence fréquentiste ne doit pas faire illusion. La théorie bayésienne est loin d'être récente ; les premiers travaux remontent à Thomas Bayes (1701-1761), mais on doit surtout à Laplace (1749-1827) la démonstration du théorème de Bayes dans un cadre plus général (réalisée, semble-t-il, indépendamment des travaux de Bayes).

A l'opposé, la distribution bayésienne *a posteriori* c'est-à-dire postérieure au recueil des données) probabilise les valeurs possibles du paramètre θ (ce qui n'est pas permis par l'inférence fréquentiste) ; elle exprime donc *directement* l'incertitude sur la vraie valeur de θ par des probabilités $Pr(\theta|x)$, conditionnelles aux données. Le théorème de Bayes permet de dériver ces probabilités à partir des trois sources d'informations disponibles :

- les données x ;
- le modèle probabiliste d'échantillonnage $Pr(x|\theta)$;
- la distribution *a priori*, qui traduit l'état d'incertitude sur θ avant le recueil des données.

Inférence statistique et conceptions de la probabilité

Pour le statisticien, le rôle des probabilités se pose alors en ces termes (Lindley, 1993) : « whether the probabilities should only refer to data and be based on frequency or whether they should *also* apply to hypotheses and be regarded as measures of beliefs ». Les italiques (que nous avons ajoutées) soulignent l'ouverture de la statistique bayésienne à l'égard des différentes conceptions de la probabilité. En particulier elle n'exclut nullement les probabilités fréquentistes, alors que certains voudraient laisser croire qu'elle ne rendrait compte que de probabilités *subjectives*. C'est ainsi par exemple que rien n'empêche un statisticien d'étudier les propriétés fréquentistes des procédures bayésiennes. Mais les travaux théoriques récents conduisent alors à un constat accablant pour une conception exclusivement fréquentiste de la statistique : « bien qu'opposé à tout apport *subjectif*, un fréquentiste devrait aussi n'utiliser que des estimateurs de Bayes ou de Bayes généralisés, car ils sont les plus satisfaisants du point de vue de *sa* théorie » (Robert, 1992, page 337).

De la vraisemblance à la probabilité *a posteriori*
Le rôle fondamental de la vraisemblance dans le raisonnement statistique est mis en avant par l'analyse bayésienne qui établit que la probabilité *a posteriori* (finale, révisée, actualisée, inverse...) $Pr(\theta|x)$ est proportionnelle au produit de la vraisemblance de θ (pour x donné) par la probabilité *a priori* (initiale) $Pr(\theta)$, ce qui s'écrit:

$$Pr(\theta|x) \propto v(\theta|x)Pr(\theta)$$

Cette formule élémentaire, qui est en fait le numérateur de la formule de Bayes complète, est le plus souvent suffisante. Ainsi Lee (1989, page 21) la commente en ces termes : « Often we will find that it is enough to get a result up to a constant of proportionality, but if we need the constant it is very easy to find it because we know that the sum must be one ». De même, Robert (1992, page 30) énonce « Tout en restant rigoureux, les calculs par proportionnalité permettent généralement une plus grande efficacité dans la construction des lois *a posteriori*. Ils seront employés de manière intensive dans cet ouvrage ». Ainsi l'outil de base de la statistique bayésienne est une formule particulièrement simple à utiliser, une fois admis et compris les deux concepts fondamentaux que sont la vraisemblance et la probabilité *a priori*.

Probabilités conjointes relatives aux observations et aux paramètres
Pour le statisticien bayésien, observations et paramètres sont *formellement* des objets identiques, considérés sous des angles différents. On donne donc un sens aux probabilités conjointes $Pr(x \text{ et } \theta) = Pr(x|\theta)Pr(\theta) = v(\theta|x)Pr(\theta)$, dont nous avons vu le rôle essentiel dans la mise en œuvre de l'analyse bayésienne.

Probabilités prédictives
La statistique bayésienne permet de calculer la probabilité (marginale) prédictive $Pr(x)$ d'observer un événement futur, compte tenu des informations disponibles. Cette probabilité $Pr(x)$, qui n'est autre que le dénominateur de la formule de Bayes, se déduit également des probabilités conjointes $Pr(x \text{ et } \theta)$, en en effectuant la somme (ou l'intégrale) sur toutes les valeurs possibles de θ. Le résultat est assez intuitif, puisqu'il s'agit de la moyenne des probabilités d'échantillonnage $Pr(x|\theta)$, pondérées par les probabilités $Pr(\theta)$. La probabilité prédictive peut apparaître à l'utilisateur facile à appréhender, puisqu'elle est relative aux quantités observables.

Le choix de la distribution *a priori*
Comme nous l'avons vu, il est conceptuellement immédiat de passer de la vraisemblance à la distribution *a posteriori*. En outre on dispose maintenant d'outils appropriés (Robert, 1996) pour résoudre les difficultés techniques éventuelles liées aux calculs numériques. Le problème crucial reste alors le choix de la distribution *a priori*, qui a été souvent la pierre d'achoppement de l'inférence bayésienne.
Suivant la conception bayésienne la plus large, les distributions initiales permettent d'incorporer toutes les connaissances et opinions *a priori* sur les paramètres disponibles avant le recueil des données, notamment des résultats antérieurs, ou encore par exemple l'opinion d'un comité d'experts. Cette possibilité constitue à l'évidence un apport potentiel considérable. Cette conception se situe souvent dans une perspective décisionnelle, dans laquelle l'inférence statistique doit fournir « un critère d'évaluation, qui décrit les conséquences de chaque décision en fonction des paramètres du modèle » (Robert, 1992, page 41).

Procédures bayésiennes standard : la conception « non-informative »

Mais, dans beaucoup de domaines, et en particulier celui de l'analyse des données expérimentales, la conception précédente implique une rupture avec les usages actuels, qui dans la plupart des cas répondent à un besoin d'objectivité nécessaire à la *communication* des résultats. C'est sans doute là une des raisons de la méfiance, voire de l'hostilité, de certains utilisateurs envers la statistique bayésienne.

Il n'est peut-être pas inutile ici de rappeler que, contrairement à certains de leurs successeurs, ni Fisher, ni Neyman et Pearson, ne rejetaient l'utilisation de probabilités sur les paramètres. La situation qui pose problème est celle où ces probabilités sont inconnues et/ou peu fiables. C'est donc dans ces situations que leurs approches diffèrent (*cf.* Poitevineau, 1997).

Pour Fisher, construire des probabilités sur les valeurs possibles du paramètre est essentiel, et c'est à cette fin qu'il a développé la notion de probabilité fiduciaire (du latin *fiducia*, confiance) qui permet cette construction sans faire appel aux probabilités *a priori*. Il exprime ainsi explicitement sa conception (Fisher, 1935/1990, page 198) : « When knowledge *a priori* in the form of mathematically exact probability statements is available, the fiducial argument is not used, but that of Bayes. Usually exact knowledge is absent, and, when the experiment can be so designed that estimation can be exhaustive, similar probability statements *a posteriori* may be inferred by the fiducial argument. »

Neyman et Person pour leur part estiment que les probabilités *a priori* sont en général inconnues, et c'est pour cette raison qu'ils ont recherché des procédures qui en soient indépendantes, en un certain sens (d'où la notion de « test uniformément plus puissant ») : « Yet if it is important to take into account probabilities *a priori* in drawing a final inference from the observations, the practical statistician is nevertheless forced to recognize that the values of φ_i [la probabilité *a priori*] can only rarely be expressed in precise numerical form. It is therefore inevitable from the practical point of view that he should consider in what sense, if any, tests can be employed which are independent of probabilities *a priori*. » (Neyman et Pearson, 1933, page 493).

Dans le cadre bayésien, la réponse est apportée par l'utilisation des distributions *a priori* « non-informatives », conçues pour exprimer un « état d'ignorance » (nous dirions plutôt un état d'indifférence) sur les paramètres, en ne privilégiant pas *a priori* de valeurs particulières (Jeffreys, 1931). Les probabilités *a posteriori* correspondantes expriment alors l'apport propre des données (« ce que les données ont à dire »). Cette approche de l'inférence bayésienne est maintenant reconnue comme un standard : « We should indeed argue that noninformative prior Bayesian analysis is the single most powerful method of statistical analysis » (Berger, 1985, page 90). Bien entendu ces probabilités « standard » pourront différer des probabilités « personnelles » qu'on pourra obtenir en incorporant des connaissances (voire des préjugés) étrangères aux données traitées.

Cette conception de l'inférence bayésienne permet une évolution progressive des pratiques actuelles n'impliquant pas nécessairement leur abandon : tout en reconnaissant de manière lucide les insuffisances de ces pratiques, on peut les *prolonger* par des méthodes bayésiennes (Lecoutre, 1996). A l'heure actuelle, face à l'usage si répandu du test de signification qui est devenu une norme quasi incontournable dans de nombreuses publications scientifiques, c'est sans doute la seule attitude réaliste.

La souplesse de l'approche bayésienne

Clairement, l'approche bayésienne apporte une plus grande souplesse dans la méthodologie statistique de l'analyse des données. D'une part les procédures bayésiennes standard, qui peuvent maintenant être mises en œuvre aussi facilement que les tests traditionnels, ont le statut privilégié d'objectivité nécessaire à la communication scientifique. D'autre part différentes distributions *a priori*, exprimant des résultats antérieurs ou des opinions d'experts, *favorables* ou *défavorables* à la conclusion recherchée, peuvent être utilisées pour éprouver la « robustesse » des conclusions et prendre ainsi des décisions « personnelles ». En outre les probabilités prédictives permettent d'évaluer les chances d'obtenir une conclusion donnée pour des observations futures, sur la base d'une étude « pilote » ou à partir des résultats partiels d'une étude en cours ; elles fournissent notamment un outil efficace pour planifier (« quel effectif choisir?") et conduire (« quand s'arrêter?") une expérimentation.

Références

Berger J. (1985) - *Statistical Decision Theory and Bayesian Analysis*. New York : Springer Verlag.

Fisher R.A. (1990) - *Statistical Methods, Experimental Design and Scientific Inference*. Oxford University Press (réédition).

Jeffreys (1931) - *Theory of probability*. Oxford : Clarendon.

Lecoutre B. (1996) - *Traitement statistique des données expérimentales : Des pratiques traditionnelles aux pratiques bayésiennes - Avec programmes Windows® par B. Lecoutre et J. Poitevineau*. Saint-Mandé : C.I.S.I.A. [Ces programmes sont également disponibles sur Internet à l'adresse : http://epeire.univ-rouen.fr/labos/eris/pac.html].

Lee P. (1989) - *Bayesian Statistics : An Introduction*. Oxford : Oxford University Press.

Lindley, D.V. (1993). Discussion de l'article de J. Whitehead, the case for frequentism in clinical trials. *Statistics in Medicine* **12**, 1419.

Neyman J., Pearson E.S. (1933) - The testing of statistical Hypotheses in Relation to Probabilities *a priori*. *Proceedings of the Cambridge Philosophical Society* **29**, 492-510.

Poitevineau (1997) - *Méthodologie de l'analyse des données expérimentales : Etude de la pratique des tests statistiques chez les chercheurs en psychologie*. Thèse du doctorat de psychologie, soutenance prévue en 1997.

Robert Christian (1992) - *L'Analyse statistique bayésienne*. Paris : Economica.

Robert Claudine (1995) - *L'Empereur et la girafe*. Paris : Diderot Editeur.

Robert Christian (1996) - *Méthodes de Monte Carlo par chaînes de Markov*. Paris : Economica.

Rouanet H., Lecoutre M.-P., Bert M.-C., Lecoutre B., Bernard J.-M.(1991) - *L'inférence statistique dans la démarche du chercheur*. Publications Universitaires Européennes. Berne : Peter Lang.