Approximation d'a priori impropres par des a priori vagues

Christèle Bioche [1] & Pierre Druilhet [2]

Laboratoire de mathmatiques, Universit Blaise Pascal, 63177 Aubière CEDEX, France
Christele.Bioche@math.univ-bpclermont.fr [1]
Pierre.DRUILHET@univ-bpclermont.fr [2]

Résumé. Nous proposons un mode de convergence pour les distributions a priori qui autorise une suite de mesures de probabilité à avoir pour limite une mesure impropre. Nous définissons une suite d'a priori vagues comme une suite de mesures de probabilité qui converge vers un a priori non-informatif. Nous considérons des cas où les a priori vagues ont de grandes variances mais remarquons que ce n'est pas obligatoire. Nous donnons des constructions d'a priori vagues approximant la mesure de Haar ou l'a priori de Jeffreys. Puis, nous étudions les conséquences de la convergence des distributions a priori sur l'analyse a posteriori. Enfin, nous revisitons le paradoxe de Jeffreys-Lindley.

Mots-clés. Approximation d'a priori impropre, convergence de mesures, a priori non-informatif, a priori vague, a priori de Jeffreys, paradoxe de Jeffreys-Lindley.

Abstract. We propose a convergence mode for prior distributions which allows a sequence of probability measures to have an improper limiting measure. We define a sequence of vague priors as a sequence of probability measures that converges to a non-informative prior. We consider some cases where vague priors have necessarily large variances and other cases where they have not. We give some constructions of vague priors that approximate the Haar measures or the Jeffreys priors. Then, we study the consequences of the convergence of prior distributions on the posterior analysis. We also revisit the Jeffreys-Lindley paradox.

Keywords. Approximation of improper prior, convergence of measure, noninformative prior, vague prior, Jeffreys prior, Jeffreys-Lindley paradox.

1 Introduction

Les statistiques bayésiennes reposent sur la formule de Bayes qui nécessite un a priori sur le paramètre et une distribution conditionnelle pour les observations. Lorsqu'on ne dispose pas d'informations a priori assez précises, différentes approches consistent à utiliser des a priori non-informatifs tels que l'a priori de Jeffreys (1946), les a priori de référence comme dans l'article de Berger et. al. (2009) ou les mesures de Haar comme le propose Eaton (1989). Ces a priori sont souvent impropres, c'est-à-dire de masse infinie. Formellement, la formule de Bayes peut-être étendue aux a priori impropres. Cependant, l'utilisation d'a priori impropres peut entrainer des problèmes tels que des a posteriori impropres ou des comportements non-désirables dans les tests d'hypothèse. C'est pourquoi, certains auteurs préférent remplacer ces a priori impropres par des a priori vagues, c'est-à-dire des a priori fournissant très peu d'information sur le paramètre. Dans la littérature, il existe plusieurs définitions pour les a priori vagues. Parfois on considère que c'est un a priori avec une grande variance sans forcément faire référence à un a priori impropre. Parfois on considère qu'un a priori dont l'estimateur a posteriori est proche de l'estimateur posteriori obtenu en considérant un a priori impropre est un a priori vague.

Considérons par exemple le modèle standard gaussien $X|\theta \sim \mathcal{N}(\theta, 1)$. Sans information a priori sur θ, on pourrait choisir l'a priori $\Pi = \lambda_{\mathbb{R}}$ ce qui correspond à la mesure de Haar, à l'a priori de Jeffreys mais aussi à la mesure uniforme sur $\mathbb{R}$. Dans ce cas, l'estimateur de Bayes est aussi l'estimateur du maximum de vraisemblance en fréquentiste. Cet a priori impropre peut-être remplacé par l'a priori vague $\Pi_n = \mathcal{N}(0, n)$ avec n grand. Cet a priori vérifie les deux définitions proposées: il a une grande variance et l'estimateur de Bayes associé converge vers celui obtenu en considérant la mesure de Lebesgue sur $\mathbb{R}$ comme a priori. Il semblerait alors que la distribution $\mathcal{N}(0, n)$ converge vers la mesure de Lebesgue mais cette convergence n'est possible dans aucune des topologies usuelles car la masse totale d'une mesure limite d'une suite de probabilité ne peut être strictement suprieur à 1. D'autres questions émanent de cet exemple: est-ce-que tous les a priori avec de grandes variances fournissent le même estimateur limite ? Est-ce-que cette "convergence" vers la mesure de Lebesgue dépend du modèle statistique ou est-elle plus intrinsèque, c'est-à-dire indépendante du modèle statistique ?

Le but de cet exposé est de donner des réponses à ces questions en proposant un nouveau mode de convergence pour les distributions a priori. On remarque sur la formule de Bayes que deux a priori Π et $\widetilde{\Pi}$ fournissent le même a posteriori s'il existe un réel α strictement positif tel que $\Pi = \alpha\widetilde{\Pi}$. Ainsi, il apparait naturellement une relation d'équivalence sur l'espace des mesures de Radon strictement positives. Le mode de convergence que nous proposons correspond à la topologie quotient.

L'exposé sera composé de trois parties. Dans un premier temps nous définirons ce mode de convergence et nous en donnerons les principales propriétés. Nous étudierons ensuite les conséquences de la convergence des a priori sur l'analyse a posteriori. Enfin, nous revisiterons le paradoxe de Jeffreys-Lindley au vu de notre mode de convergence.

2 Définition, propriétés et exemples de convergence q-vague

Dans cette partie, nous définirons le mode de convergence qui nous intéresse, nous le nommerons convergence q-vague. Nous montrerons qu'une suite d'a priori vagues admet au plus une mesure limite, et ce, indépendamment du modèle statistique. Nous montrerons aussi qu'il n'y a pas de discontinuité entre les mesures propres et les mesures impropres au sens où toute mesure impropre peut être approximée par une suite de mesures propres et inversement, toute mesure propre peut être approximée par une suite de mesures impropres. Nous donnerons des conditions suffisantes sur les suites de densités pour obtenir la convergence q-vague des suites de mesures associées. Nous regarderons sous quelles conditions la suite des variances associées aux a priori diverge. Nous étudierons aussi les conditions pour conserver la convergence q-vague des a priori lors d'un changement de paramétrisation. Enfin, nous donnerons quelques exemples de constructuion de suites d'a priori vagues vers la mesure de Haar de certains groupes et vers l'a priori de Jeffreys.

3 Conséquences sur l'analyse a posteriori

Dans cette partie, nous étudierons les conditions sous lesquelles la convergence q-vague des a priori entraine la convergence q-vague des a posteriori et la convergence des estimateurs. Nous traiterons aussi quelques exemples variés, notamment celui des lois $\beta(\frac{1}{n}, \frac{1}{n})$ qui apparait souvent dans la littérature (par exemple dans Tuyl et. al. (2009), mais aussi dans Bernardo (1979), Lane et Sudderth (1983) ou encore Lehmann et Casella (1998)).

4 Paradoxe de Jeffreys-Lindley

Dans cette partie, nous verrons que la convergence q-vague permet d'expliquer les convergences surprenantes du paradoxe de Jeffreys-Lindley.

Bibliographie

[1] Barndorff-Nielsen, O. (1978), *Information and exponential families in statistical theory*, John Wiley and Sons Ltd., Chichester.
[2] Berger, J. O. (1985), *Statistical decision theory and Bayesian analysis*, second ed. Springer Series in Statistics, Springer-Verlag, New York.
[3] Berger, J. O., Bernardo, J. M. and Sun, D. (2009), The formal definition of reference priors, *Ann. Statist*, 37 905938.
[4] Bernardo, J.-M. (1979), Reference posterior distributions for Bayesian inference, *J. Roy. Statist. Soc. Ser.B* 41 113147.
[5] Billingsley, P. (1968), *Convergence of probability measures*, John Wiley and Sons Inc., New York.
[6] Billingsley, P. (1986), *Probability and measure*, second ed. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*, John Wiley and Sons Inc., New York.

[7] Dauxois, J.-Y., Druilhet, P. and Pommeret, D. (2006), A Bayesian choice between Poisson, binomial and negative binomial models, *Test*, 15 423432.
[8] Diaconis, P. and Ylvisaker, D. (1979), Conjugate priors for exponential families, *Ann. Statist*, 7 269281.
[9] Druilhet, P. and Pommeret, D. (2012), Invariant conjugate analysis for exponential families, *Bayesian Anal*, 7 903916.
[10] Eaton, M. L. (1989), *Group invariance applications in statistics. NSF-CBMS Regional Conference Series in Probability and Statistics*, Institute of Mathematical Statistics, Hayward, CA.
[11] Hartigan, J. A. (1996), Locally uniform prior distributions, *Ann. Statist*, 24 160173.
[12] Lane, D. A. and Sudderth, W. D. (1983), Coherent and continuous inference, *Ann. Statist*, 11 114120.
[13] Lehmann, E. L. and Casella, G. (1998), *Theory of point estimation*, second ed, Springer Texts in Statistics. Springer-Verlag, New York.
[14] Malliavin, P. (1982), *Intgration et probabilits, Analyse de Fourier et analyse spectrale*, Masson, Paris.
[15] Morris, C. N. (1983), Natural exponential families with quadratic variance functions: statistical theory, *Ann. Statist*, 11 515529.
[16] Robert, C. P. (2001), *The Bayesian choice*, second ed., Springer Texts in Statistics. Springer-Verlag, New York.
[17] Tuyl, F., Gerlach, R. and Mengersen, K. (2009), Posterior predictive arguments in favor of the Bayes-Laplace prior as the consensus prior for binomial and multinomial parameters, *Bayesian Anal*, 4 151158.