

Le principe de la démarche bayésienne sans formule mathématique

Emmanuel Grenier

Reims Management School

emmanuel.grenier@reims-ms.fr

Introduction

La démarche bayésienne est efficace du point de vue opérationnel (pour s'en convaincre, il suffit d'entrer dans un moteur de recherche le mot clé *bayesian*) et naturelle dans le sens où elle autorise une interprétation intuitive de l'estimation et des tests statistiques (Lecoutre B. [3] et [4]).

Cependant, son principe est difficile à comprendre sans une solide formation à la statistique mathématique. Grâce au document joint, ***Estimation bayésienne d'une proportion.xls***, nous allons pouvoir identifier les éléments mis en œuvre par la démarche bayésienne et observer la manière dont ces éléments se combinent. Ceci de manière expérimentale, en examinant l'évolution de graphiques.

Le principe

En premier lieu, il est nécessaire d'exprimer un avis a priori sur la valeur de la grandeur à estimer. L'a priori est ensuite actualisé à partir de l'information provenant des données observées (issues d'enquêtes ou d'expérimentations).

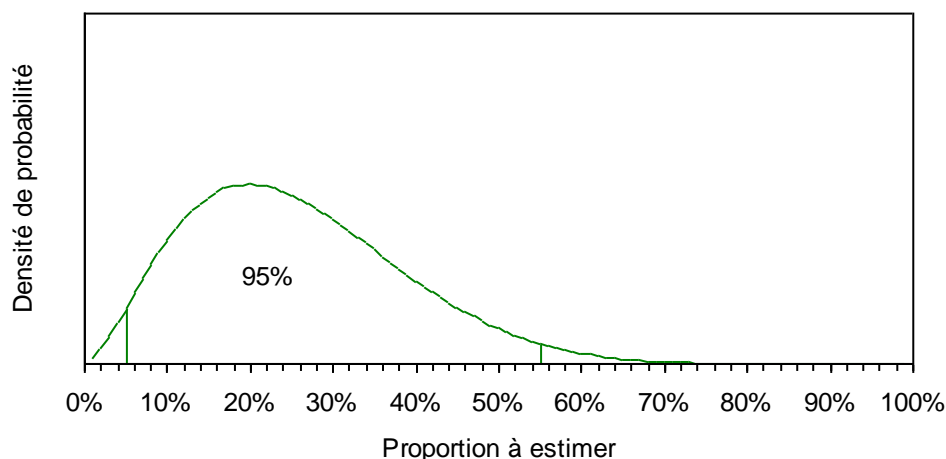
Distribution a priori de la grandeur à estimer

Pour fixer les idées, pensez à votre candidat favori pour les prochaines élections. Quelle est la proportion des électeurs qui partagent votre opinion ?

Ouvrez le document joint ***Estimation bayésienne d'une proportion.xls***.

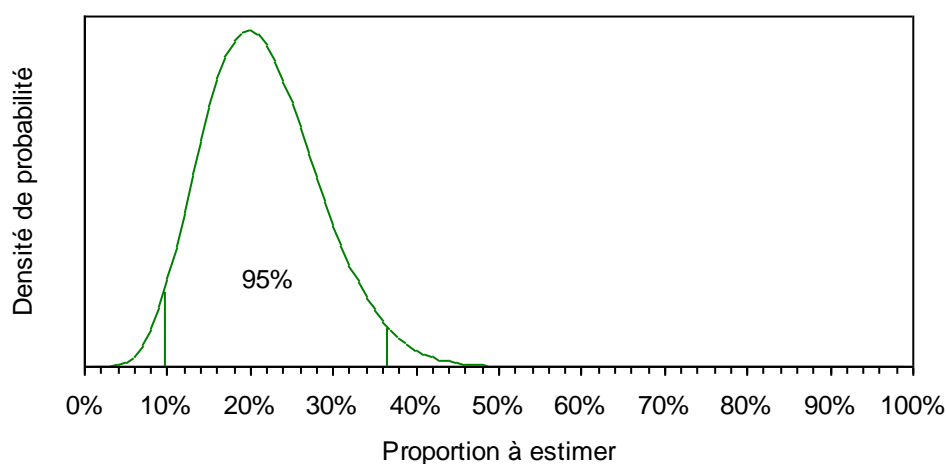
La feuille ***Estimation*** vous permet d'exprimer votre a priori par une courbe, dite densité de probabilité, ici¹ définie par le mode, c'est-à-dire la valeur la plus vraisemblable, et l'incertitude, exprimée par un intervalle de probabilité.

¹ On utilise la distribution bêta (voir Denis J.-B. [2])



Nous dirons que la courbe précédente exprime votre a priori : selon vous, la proportion des électeurs qui partagent votre opinion se situe probablement (avec une probabilité de 95%) entre 5% et 55%, la valeur la plus vraisemblable étant égale à 20%.

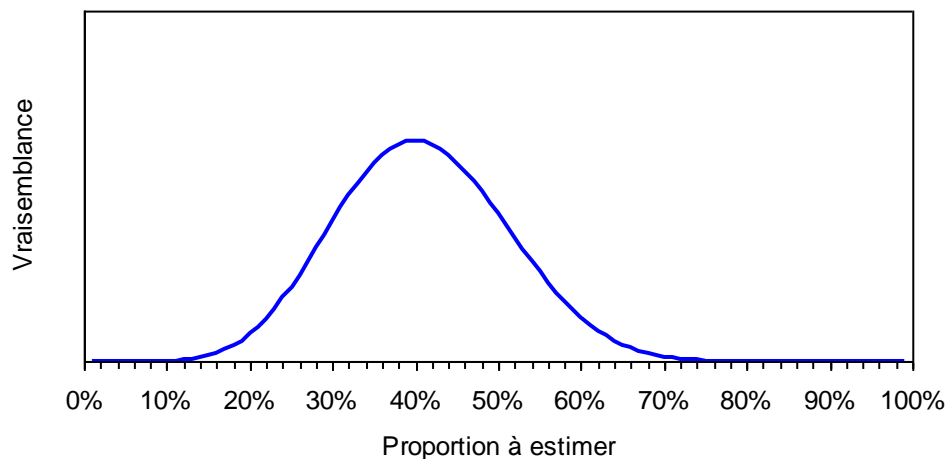
Vous pouvez modifier l'a priori, par exemple réduire l'incertitude (si vous êtes plus sûr de vous) :



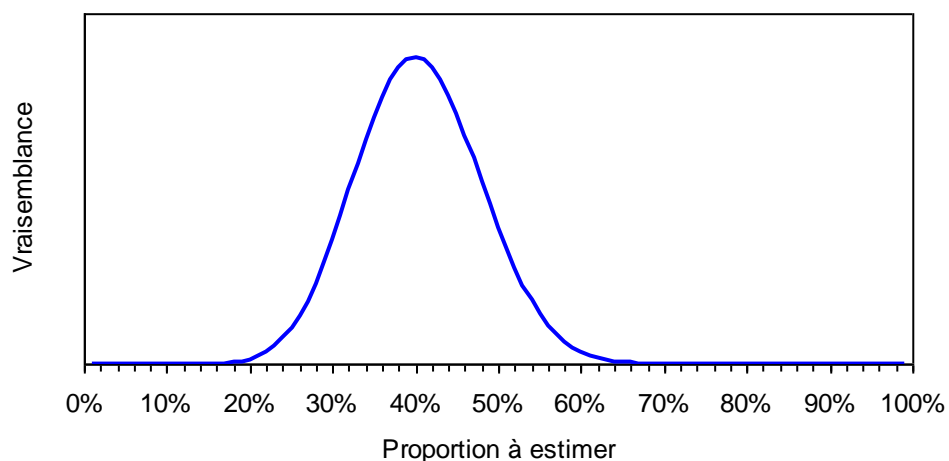
Vraisemblance de l'observation

Un de vos amis, enquêteur professionnel, vous livre cette information : sur les 20 personnes qu'il a interrogées, 8 ont déclaré avoir votre candidat comme favori.

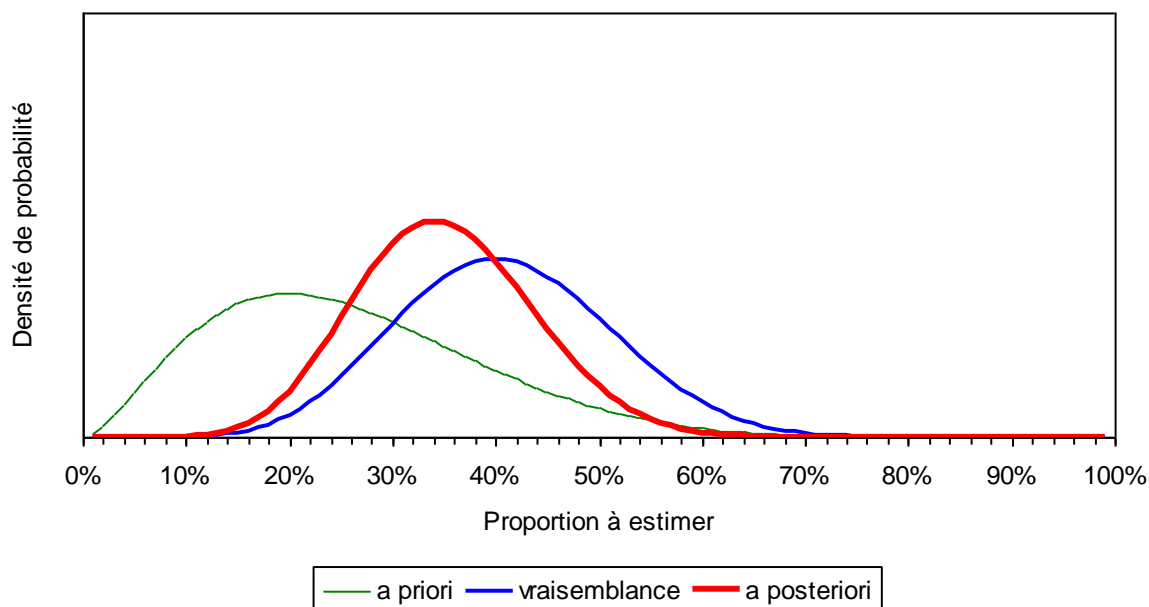
Ce résultat vous rassure : au moins, la proportion des électeurs qui partagent votre opinion n'est pas nulle ! Certes, elle pourrait être faible, par exemple égale à 10%, mais, dans ce cas, il serait très peu vraisemblable d'observer 8 personnes favorables sur les 20 personnes interrogées. En fait, le résultat observé est vraisemblable quand on suppose que la proportion chez les électeurs est proche de la proportion dans l'échantillon, soit 40%.



La vraisemblance du résultat se resserre autour de la proportion observée quand la taille de l'échantillon augmente. Si votre ami avait observé la même proportion sur 40 personnes, soit 16 personnes favorables sur 40, ce résultat n'aurait eu aucune vraisemblance dans la situation où 10% des électeurs partagent votre opinion :



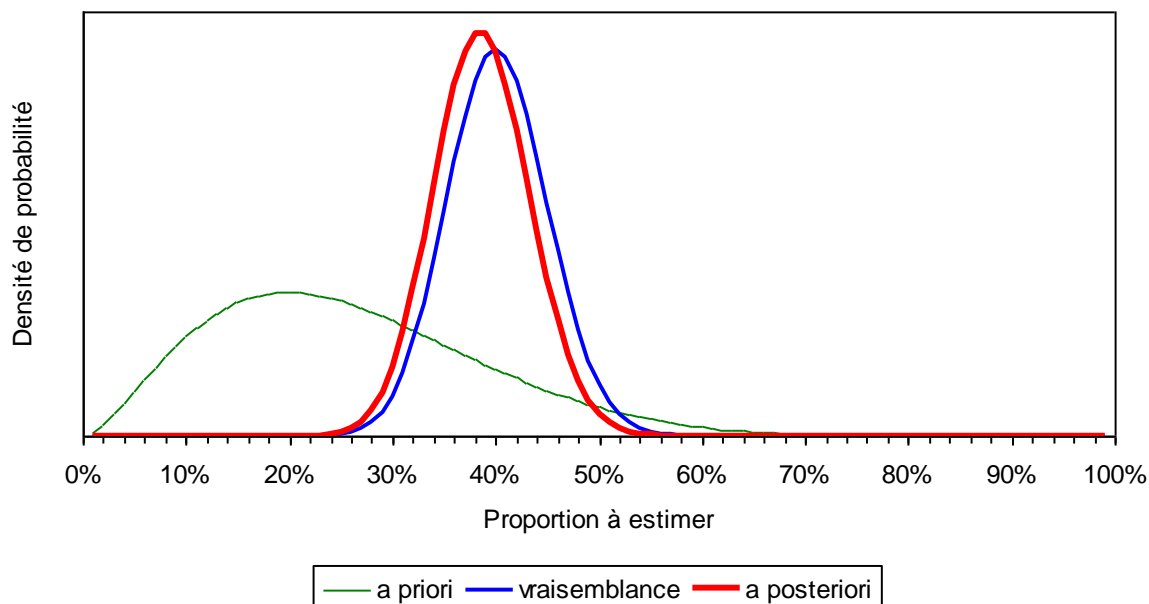
Distribution a posteriori de la grandeur à estimer



La distribution a posteriori est la combinaison de l'avis a priori et de l'information issue de l'observation. Cette information est intégrée par l'intermédiaire de la vraisemblance de l'observation.

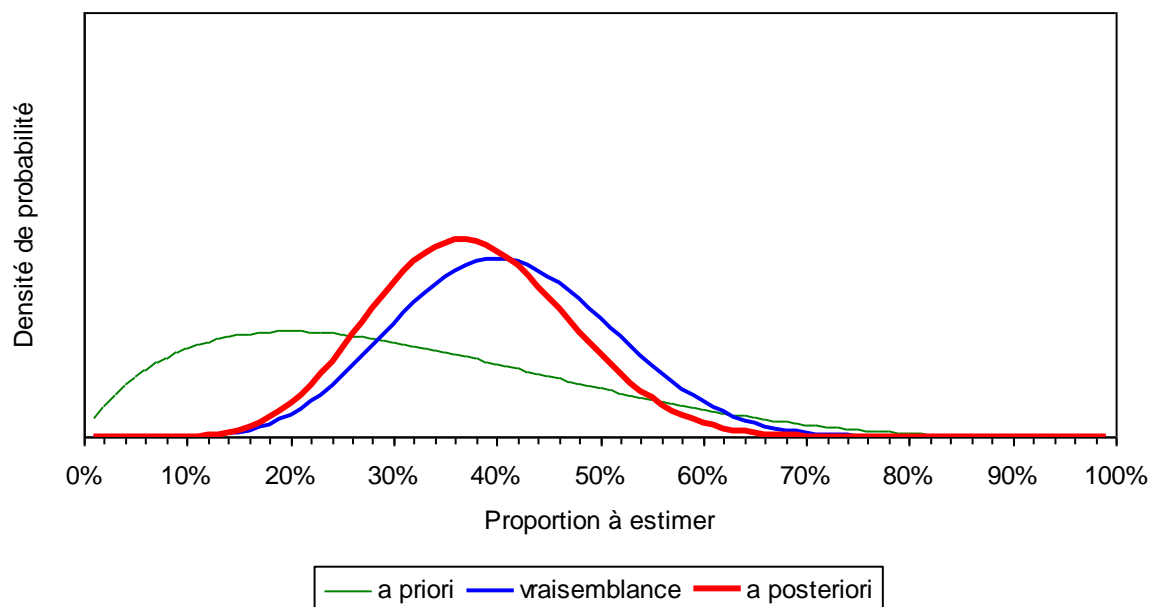
Ici l'observation l'emporte sur l'a priori : le mode de la distribution a posteriori (34%) est plus proche du mode de la vraisemblance (40%) que de celui de l'a priori (20%).

Gardons la même proportion observée mais augmentons la taille de l'échantillon, par exemple, 40 personnes favorables sur un échantillon de 100 personnes :



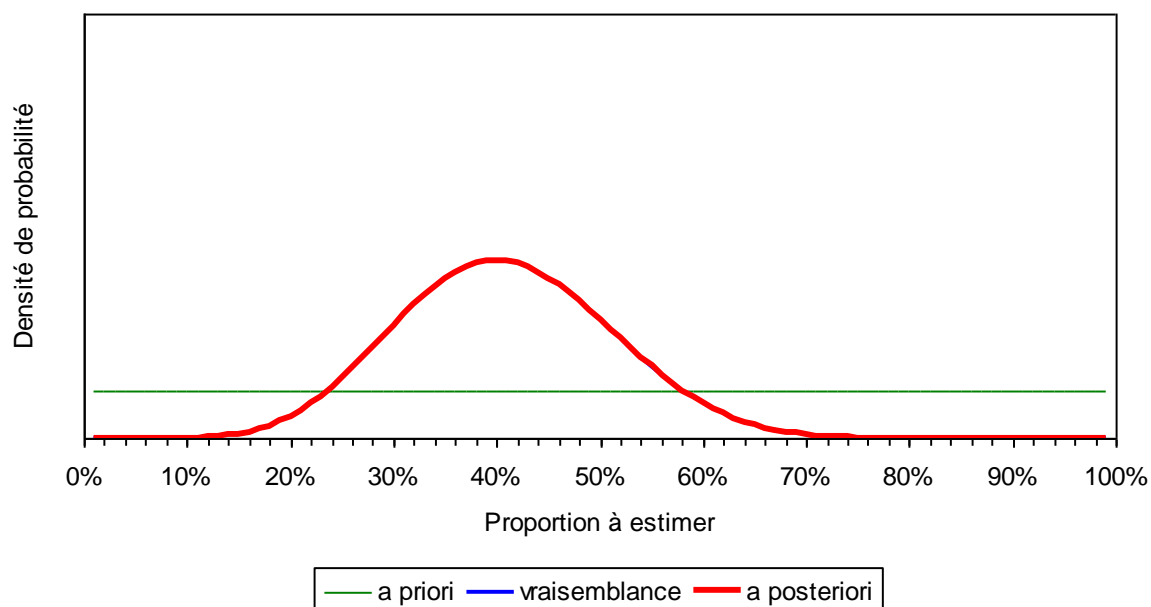
L'influence de l'a priori se réduit quand le poids de l'information (la taille de l'échantillon) augmente.

Augmentons l'incertitude de l'a priori.



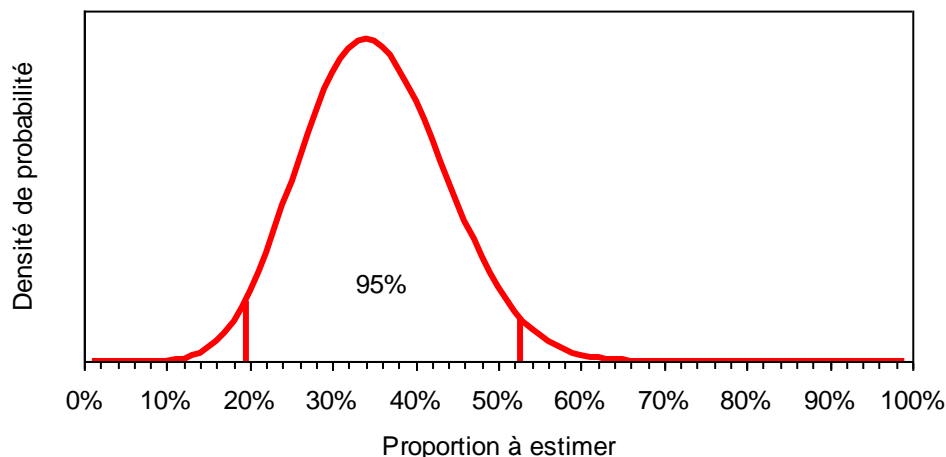
Le poids de l'a priori diminue quand l'incertitude augmente.

Il est nul ici dans le cas d'une distribution a priori uniforme entre 0 et 100% (la distribution a posteriori se confond avec la vraisemblance) :



Interprétation

De la distribution a posteriori, on peut tirer des estimations ponctuelles : le mode de distribution, la médiane ou la moyenne. On peut également donner une estimation par intervalle.



Il s'agit alors d'un intervalle de probabilité qu'on interprète comme tel : on peut dire que le paramètre à estimer a une forte probabilité d'appartenir à l'intervalle.

Remerciements

Merci à Henry Aubert, autant pour ses encouragements que pour ses conseils sur la mise en forme de la feuille d'Excel.

Références

- [1] Boreux J.-J., Parent E., Bernier J. - Pratique du calcul bayésien. Springer, 2010
- [2] Denis J.-B., Ritz M. - Notions de base utiles en modélisation stochastique de phénomènes complexes. Disponible sur : <http://w3.jouy.inra.fr/unites/miaj/public/matrisq/jbdenis/notes/welcome.html> (consulté le 25/01/2011)
- [3] Lecoutre B. - Et si vous étiez un bayésien qui s'ignore ? Revue MODULAD n°32, 2005
- [4] Lecoutre B. - Former les étudiants et les chercheurs aux méthodes bayésiennes pour l'analyse des données expérimentales. Revue MODULAD n°33, 2005
- [5] Parent E., Bernier J. - Le raisonnement bayésien : modélisation et inférence. Springer, 2007
- [6] Robert C. - Le choix bayésien. Principes et pratique. Springer, 2006