



HAL
open science

Compte rendu de: BRYANT John, ZHANG Junni L.,
2018, Bayesian Demographic Estimation and
Forecasting, CRC Press: Taylor & Francis Group, Boca
Raton (FL), Statistics in the Social and Behavioural
Sciences Series, 292 p.

Daniel Courgeau

► To cite this version:

Daniel Courgeau. Compte rendu de: BRYANT John, ZHANG Junni L., 2018, Bayesian Demographic Estimation and Forecasting, CRC Press: Taylor & Francis Group, Boca Raton (FL), Statistics in the Social and Behavioural Sciences Series, 292 p.. Population (édition française), 2020, Bibliographie critique, 75 (4), pp.618-620. 10.3917/popu.2004.0618 . hal-04212579

HAL Id: hal-04212579

<https://hal.science/hal-04212579>

Submitted on 21 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BRYANT John, ZHANG Junni L., 2018, *Bayesian Demographic Estimation and Forecasting*, CRC Press: Taylor & Francis Group, Boca Raton (FL), *Statistics in the Social and Behavioural Sciences Series*, 292 p.

S'il existe déjà un grand nombre d'ouvrages présentant l'utilité ou même la nécessité d'utiliser des méthodes bayésiennes en sciences sociales et en statistiques, l'originalité de celui-ci se trouve dans l'application générale qu'il en propose pour la démographie et pour les projections de population. Cependant, de nombreux articles en statistique ont déjà présenté et discuté ce thème, qui prend une importance grandissante dans ce domaine. Arbel et Costemalle (2016)⁽¹⁾ réconcilient notamment deux sources d'estimation des flux d'immigration en France par une approche bayésienne.

Il semble en premier lieu utile de détailler ce qui distingue l'approche fréquentiste ou objectiviste, communément utilisée en démographie, de l'approche bayésienne ou épistémique, encore peu utilisée. En particulier peu d'articles, pour le moment, s'en servent dans *Population*.

Il faut remonter au XVII^e siècle pour assister à la naissance de la notion de probabilité avec le traité de Pascal⁽²⁾ en 1654, puis à la naissance de la démographie avec l'ouvrage de Graunt⁽³⁾ quelques années plus tard, en 1662 : celui-ci utilise l'approche objectiviste des probabilités (relative à l'objet étudié). Ce n'est qu'au XVIII^e siècle, en 1763, que Bayes⁽⁴⁾ propose une notion épistémique de probabilité (relative à la connaissance que l'homme peut en avoir), suivi par Laplace⁽⁵⁾ qui l'utilisa en particulier pour des problèmes démographiques, tels que la comparaison des naissances des garçons et des filles en 1778.

En quoi ces deux notions principales différent-elles ? L'ouvrage sur *Probability and social science* (2012)⁽⁶⁾ en fait une présentation plus complète et détaillée que l'on peut résumer comme suit. L'approche objectiviste suppose que la probabilité d'un événement existe indépendamment du statisticien, et que celui-ci va chercher à l'estimer par des expériences successives : lorsque le nombre d'essais tend vers l'infini, le rapport des cas où l'événement se produit à l'ensemble des observations va tendre vers cette probabilité. Mais l'hypothèse même de l'existence de cette probabilité ne peut guère être clairement démontrée : bien plus encore, Bruno

(1) Arbel J., Costemalle V., 2016, Estimation des flux d'immigration : réconciliation de deux sources par une approche bayésienne, *Économie et statistiques*, n° 483-485, p. 121-149.

(2) Pascal B., 1654, *Traité du triangle arithmétique, avec quelques autres traités sur le même sujet*, Paris, Guillaume Desprez.

(3) Graunt J., 1662, *Natural and political observations mentioned in a following index, and made upon the bills of mortality*. London. (traduction française par Vilquin E., *Observations naturelles et politiques répertoriées dans l'index ci-après et faites sur les bulletins de mortalité*, 1977, Paris, Ined, Classiques de l'économie et de la population).

(4) Bayes T. R., 1763, An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society of London*, 53, p. 370-418.

(5) Laplace P.S., 1778, Mémoire sur les probabilités, *Mémoires de l'Académie royale des sciences de Paris*, 1781, p. 227-332.

(6) Courgeau D., 2012, *Probability and social science*, Dordrecht, Heidelberg, London, New York, Springer.

de Finetti⁽⁷⁾, grand défenseur du bayésien, dit clairement en 1974 que la probabilité n'existe pas objectivement, c'est-à-dire indépendamment de l'esprit humain.

L'approche épistémique va au contraire se centrer sur la connaissance que nous pouvons avoir du même phénomène. Lors d'une expérience, enquête sur échantillon ou recensement plus ou moins exhaustif en démographie, nous disposons de nouvelles informations sur le phénomène étudié. Le statisticien épistémique peut alors en profiter pour améliorer l'opinion qu'il peut avoir *a priori* de sa probabilité, en utilisant le théorème de Bayes : celui-ci permet de calculer sa probabilité *a posteriori*. Bien sûr, cette estimation va maintenant dépendre de la probabilité *a priori* choisie, mais lorsque celle-ci est mûrement réfléchie, le résultat s'en trouvera largement amélioré par rapport à une probabilité objective.

Lorsqu'il s'agit d'utiliser ces notions pour prendre une décision, les deux approches vont encore plus différer. Lorsque l'objectiviste fournit un intervalle de confiance de 95 % pour son estimation, il peut seulement dire que s'il tirait un grand nombre d'échantillons de même taille, on peut s'attendre à ce que 95 % des intervalles de confiance obtenus contiennent le paramètre estimé. On voit combien cette définition complexe ne correspond pas à ce que l'on pourrait en attendre. En revanche le bayésien, à partir de ses hypothèses initiales, peut clairement dire qu'un intervalle de confiance bayésien, ou de crédibilité, de 95 % fournit bien la probabilité que le paramètre estimé s'y trouve.

On peut dès lors se demander pourquoi l'approche bayésienne, qui semble mieux adaptée aux sciences sociales, ait mis tant de temps à s'imposer auprès de ses chercheurs. Une raison importante se trouve dans la complexité des calculs auxquels elle conduit, que seule la puissance des ordinateurs actuels permet d'effectuer. Il suffit pour s'en rendre compte de relire l'excellent article de Laplace (1778) qui, pour un problème simple portant sur la comparaison des naissances de filles et de garçons, présente les calculs complexes et les approximations qu'il doit opérer pour y arriver. Une seconde raison se trouve dans le souhait d'une démographie objective, qui ne devrait pas faire appel aux jugements personnels. Cette notion d'une démographie objective se trouve de plus en plus mise en doute devant les analyses faisant intervenir des interactions de plus en plus complexes entre événements, et des projections de sous-populations basées sur de faibles effectifs.

L'ouvrage de John Bryant et J.L. Zhang, didactique, guide le lecteur de façon progressive vers l'utilisation de ces méthodes bayésiennes. Il en fournit d'abord les principaux éléments avec des exemples démographiques : échangeabilité, distributions *a priori* informatives ou non, hiérarchies, calculs de la distribution *a posteriori*, différents tests de validité possibles, etc.

Il analyse ensuite de façon très détaillée différentes situations démographiques avec les données existantes : mortalité infantile dans les comtés suédois de 1995

(7) Finetti de B., 1974, *Theory of probability*, 2 vol., London, New York, John Wiley & Sons.

à 2015 et projections jusqu'en 2025 ; espérance de vie au Portugal de 1990 à 2015 et projections jusqu'en 2035 ; dépenses de santé en Hollande de 2003 à 2011 et projections jusqu'en 2020 ; migrations internes en Islande de 2000 à 2014 et projections jusqu'en 2020 ; fécondité des provinces cambodgiennes estimée à l'aide du recensement de 2008 et de l'enquête de 2010 ; évolution de la population néo-zélandaise par régions de 2008 à 2016 et estimation des migrations internes ; évolution de la population chinoise de 1990 à 2015. Il montre ainsi que l'estimation des populations passées et la projection de ces populations constituent un même problème, que les méthodes bayésiennes permettent de résoudre au mieux.

Il manque toutefois à cet ouvrage une comparaison plus précise des estimations et des projections utilisant l'une et l'autre approche. Il aurait été par exemple fort utile de voir combien les méthodes bayésiennes permettent d'améliorer les projections faites avec des méthodes fréquentistes. Seul, à notre connaissance, l'ouvrage de Bijak (2011)⁽⁸⁾ a pu comparer le résultat de telles prévisions non seulement aux observations, mais également aux prévisions fréquentistes. Ses résultats montrent clairement la supériorité des méthodes bayésiennes. La fréquence des observations empiriques qui tombent à l'intérieur des intervalles de confiance prévus par la projection bayésienne est toujours très supérieure à celles prévues par la projection fréquentiste.

En dépit de ce manque, la lecture de cet ouvrage est vivement conseillée, car les démographes et les statisticiens y trouveront des estimations et des projections de population bien mieux justifiées et beaucoup plus précises que ne le permettent les méthodes classiques. D'ailleurs, l'ONU a adopté en 2014 une méthodologie bayésienne pour ses projections démographiques officielles, par sexe et âge. N'est-ce pas là un encouragement pour les démographes à utiliser plus fréquemment les estimations bayésiennes ?

Daniel COURGEAU

(8) Bijak J., 2011, *Forecasting international migration*, Springer Series on Demographic Methods and Population Analysis, 24, Dordrecht / Heidelberg / London / New York, Springer.