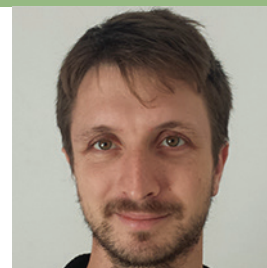


Projections de populations : l'ONU adopte une méthode bayésienne



Vianney COSTEMALLE

Insee¹

Des “projections de population probabilistes” qu’est-ce que ça veut dire ? Comment doit-on s’en servir ? Ces questions concernent désormais le grand public, et pas seulement les spécialistes, depuis que l’ONU a adopté en 2014 une méthodologie “bayésienne” pour ses projections démographiques officielles. Les méthodes statistiques bayésiennes sont encore entourées d’une aura de difficulté, voire d’une certaine suspicion chez certains. Adrian Raftery, statisticien et sociologue Irlandais ayant fait sa thèse à Paris et travaillant actuellement à l’Université de Washington, a joué un rôle majeur dans le développement théorique de ces méthodes et dans leur utilisation pratique sur de nombreux domaines d’application. Il a été la cheville ouvrière de l’élaboration des nouvelles projections de l’ONU, et il a aussi contribué à les expliquer au grand public anglophone, notamment dans la revue *Science*. Sur sa suggestion, *Statistique et Société* a demandé à un jeune statisticien français d’expliquer et de commenter ses travaux pour notre public. Nous remercions Vianney Costemalle d’avoir accepté cette tâche.

Les projections de population sont des prévisions démographiques à base statistique qui permettent d’éclairer le futur d’une région, d’un pays ou de l’humanité entière. Elles sont particulièrement utiles pour anticiper les besoins futurs, en ce qui concerne les infrastructures publiques (écoles, hôpitaux, logements, transports), l’économie (part des actifs, financement des retraites) ou encore la gestion des ressources naturelles.

Tous les deux ans, l’Organisation des Nations Unies (ONU) réalise des projections de populations sur plus de 200 pays et territoires ainsi que sur l’ensemble de la population humaine à l’horizon 2100. Elle a récemment mis en place une nouvelle méthode, reposant sur les statistiques bayésiennes, qui bouleverse notre appréciation des prévisions démographiques.

Des prévisions déterministes aux prévisions probabilistes

En 2014, la prestigieuse revue scientifique *Science* publiait un article dû à Gerland *et al.*, qui présentait les principaux résultats de la nouvelle méthode employée par l’ONU et qui déclarait peu probable la stabilisation de la population mondiale au 21^{ème} siècle contrairement à ce qu’on pensait jusqu’alors. « Peu probable » : c’est bien en termes de probabilité qu’est décrit l’avenir. On ne sait rien de lui avec certitude, mais on estime plus ou moins possibles certains scénarios.

1. Cet article n’engage que son auteur et non l’Institut national de la statistique et des études économiques

Alors que l'homme a longtemps cherché à prédire l'avenir de façon peu rationnelle, en cherchant des signes dans le hasard, des augures de Rome aux chiromanciennes de nos foires, de nouvelles techniques sont apparues dès le 17^{ème} siècle, fondées sur une représentation mathématique du monde. Les premières prévisions issues de ces modèles sont dites déterministes puisque l'état futur du système étudié est entièrement déterminé par l'état initial et les règles d'évolution de ce système ; il n'y a qu'un seul avenir possible : la ronde des planètes dans notre système solaire. Plus tard, avec le développement de la théorie des probabilités au 20^{ème} siècle, ont émergé les prévisions probabilistes où l'avenir apparaît incertain comme c'est le cas des prévisions météorologiques : on annonce par exemple qu'il va pleuvoir demain avec 70% de chance. Dans le cadre de la modélisation mathématique, cette incertitude provient fondamentalement d'une ignorance : soit l'état initial n'est pas connu parfaitement, soit cela provient d'une méconnaissance des mécanismes exacts qui sont à l'œuvre dans l'évolution du système étudié, ce qui est souvent le cas avec des systèmes complexes.

L'évolution démographique est un phénomène complexe

Prévoir la taille future de la population mondiale est en apparence simple ; cela revient à prévoir le nombre de naissances et de décès chaque année. Une méthode pour faire cela, la méthode "standard cohort-component projection" développée par Leslie en 1945, est d'appliquer des taux de fécondité et des taux de mortalité chaque année à la population étudiée. Ces taux dépendent de l'âge des personnes, des pays et aussi de l'époque. Le taux de mortalité à 80 ans n'est pas le même en 2015 qu'en 1915. Quelle sera sa valeur en 2100 ? Afin d'anticiper combien d'humains nous serons dans l'avenir, il faut prévoir comment évolueront la fécondité et l'espérance de vie. Or, à ce jour, aucune théorie, aucun mécanisme exact ne permet d'expliquer l'évolution de la fécondité ou de la mortalité. Ces évolutions dépendent de tellement de facteurs, de circonstances, d'aléas (les progrès de la médecine, l'évolution climatique, les décisions politiques, les catastrophes naturelles, les guerres, les comportements de chaque être humain,...) qu'il semble *a priori* impossible de les prévoir.

Une manière d'anticiper est d'observer comment la fécondité et l'espérance de vie ont varié dans les décennies passées. En détectant quelles sont les grandes tendances, les invariants, les motifs qui se répètent, on peut se faire une idée des évolutions futures. On a observé par exemple pour un grand nombre de pays une transition des taux de fécondité ayant toujours la même forme : au départ une fécondité élevée puis une diminution plus ou moins rapide jusqu'en-dessous du seuil de deux enfants par femme, puis dans certains cas un léger regain avec une convergence vers deux enfants par femmes. Cette « transition démographique » est achevée pour les pays développés et est encore en cours pour certains pays en voie de développement (*figure 1*). De même, on observe par exemple que l'espérance de vie augmente de façon régulière partout, et que les femmes vivent plus longtemps que les hommes en moyenne.

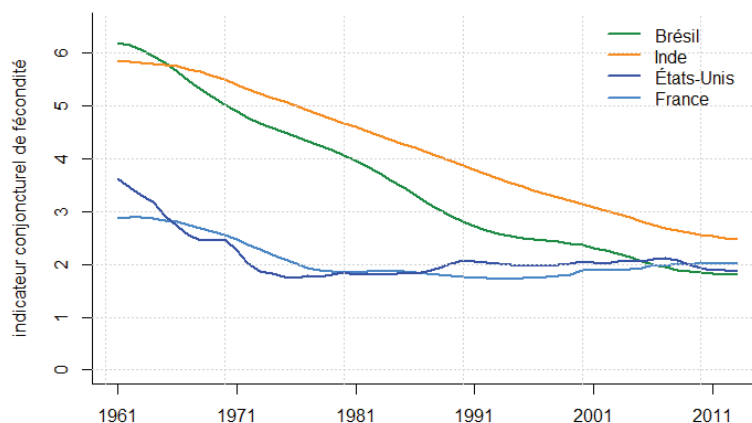


Figure 1. Évolution de l'indicateur conjoncturel de fécondité pour la France, les États-Unis, le Brésil et l'Inde. Source : Banque mondiale.

Il est possible de modéliser ces tendances, ces invariants, ces motifs, par des fonctions mathématiques en introduisant des paramètres qui, comme les rouages d'une horloge, permettent de régler la façon dont peut évoluer la fécondité et la mortalité. Certains paramètres contrôlent, par exemple, à quelle vitesse augmente l'espérance de vie chaque année.

La principale difficulté réside alors dans l'incertitude autour des valeurs de ces paramètres. Jusqu'en 2008, les projections de l'ONU faisaient appel à des savoirs d'experts pour estimer les valeurs les plus vraisemblables. À partir de ces estimations on envisageait alors un scénario médian, un scénario avec une fécondité haute et un scénario avec une fécondité basse correspondant respectivement à plus ou moins 0,5 enfant par femme par rapport au taux du scénario médian. Il s'agissait de projections déterministes : on ne sait pas avec quelle probabilité les scénarios extrêmes peuvent avoir lieu. À partir de 2014, la nouvelle méthode employée permet de réaliser des projections probabilistes à l'aide des méthodes bayésiennes.

Dans le cas de la projection pour un pays, il faut aussi prendre en compte les migrations qui peuvent modifier la taille de la population chaque année : l'émigration fait diminuer cette taille tandis que l'immigration la fait augmenter.

L'inférence bayésienne au service des projections probabilistes

Selon le paradigme bayésien, les paramètres dont on ne connaît pas la valeur exacte sont considérés comme des variables aléatoires. C'est-à-dire que pour chaque intervalle de valeur possible d'un tel paramètre on associe une probabilité. Par exemple, en simplifiant on peut imaginer qu'on ne sait pas de combien augmente exactement l'espérance de vie chaque année mais en revanche on peut savoir avec quelle probabilité cette espérance de vie augmente de 3 mois par an (plus ou moins 1 mois) ou avec quelle probabilité elle augmente de 5 mois par an (plus ou moins 1 mois). L'ensemble de ces probabilités associées à chaque intervalle de valeur possible des paramètres définit ce qu'on appelle la *distribution* de la variable aléatoire. Dans le cadre bayésien il s'agit de déterminer quelle est cette distribution pour l'ensemble des paramètres du modèle. Pour cela on utilise le théorème de Bayes (ecclésiastique - mathématicien britannique du 18^{ème} siècle) qui permet d'actualiser cette distribution à l'aide des données observées sur le passé : avant toute observation on a une distribution *a priori* qui reflète les suppositions minimales de la valeur des paramètres sur lesquelles les experts peuvent s'accorder, et une fois les observations faites et les données recueillies on en déduit une distribution *a posteriori* qui prend en compte toute l'information apportée par les données. L'inférence bayésienne permet de remonter des observations à leurs causes, en attribuant une probabilité à chacune d'elles. Plus le nombre de données observées est grand, plus la distribution *a posteriori* sera resserrée autour d'une seule valeur. Autrement dit, plus on a de données, plus l'incertitude sur la valeur des paramètres est faible.

Les distributions des paramètres contrôlant l'évolution de la fécondité et de la mortalité sont ainsi estimées pour chaque pays en prenant en compte les données de tous les pays. Ces distributions ne sont en effet pas indépendantes entre elles, mais sont reliées dans la modélisation, par ce qui s'appelle un modèle bayésien hiérarchique.

À partir de la modélisation mathématique et de la distribution *a posteriori* des paramètres on peut prévoir les différentes probabilités des évolutions futures de la taille de la population (*figure 2*). Pour une évolution possible, on tire au hasard un jeu de paramètres selon la distribution calculée. Partant de ce jeu de paramètres on fait évoluer jusqu'à l'horizon souhaité la fécondité et l'espérance de vie, de façon mécanique selon les règles retenues lors de la phase de modélisation. On en déduit alors les taux de fécondité et de mortalité par âge pour chaque période future, puis l'évolution de la taille de la population. On obtient ainsi un chemin possible d'évolution de la population mondiale. Si l'on répète ces opérations une deuxième fois, on aura

alors un autre chemin possible (car on aura tiré au départ un autre jeu de paramètres). La répétition de ces opérations un très grand nombre de fois donne toute une palette de chemins possibles. C'est ensuite à partir de l'ensemble de ces chemins qu'on détermine des intervalles de confiance : l'intervalle de confiance à 95% est un intervalle dans lequel 95% des chemins se situent. L'intervalle de confiance à 80% est un intervalle dans lequel huit chemins sur dix se situent. L'intervalle de confiance à 80% est donc lui-même à l'intérieur de l'intervalle de confiance à 95%. En fait, plus le niveau de confiance est élevé, plus l'intervalle, et donc l'incertitude sur le chemin futur, est grand.

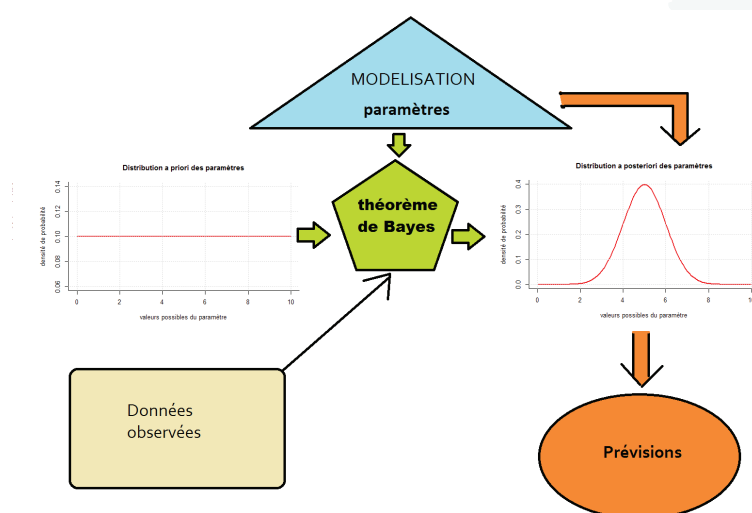


Figure 2. Schéma illustrant les prévisions probabilistes par une approche bayésienne

Très probablement entre 9,5 et 13,3 milliards d'êtres humains en 2100

La *figure 3* présente les résultats de la projection probabiliste pour la taille de la population mondiale, ainsi que les résultats de la projection déterministe classiquement effectuée par l'ONU. L'intervalle de confiance à 95% est beaucoup plus étroit que l'intervalle entre le scénario « haut » et le scénario « bas » de la projection déterministe. Le scénario médian se trouve quant à lui presque au milieu des intervalles de confiance à 95% et 80%. Selon ces projections probabilistes, le monde comptera entre 10 milliards et 12,5 milliards d'habitants en 2100 avec une probabilité de 80% et entre 9,5 milliards et 13,3 milliards avec une probabilité de 95%. Les intervalles de confiances obtenus avec la méthode probabiliste peuvent parfois être plus grands que l'intervalle entre le scénario haut et bas des projections déterministes, ce qui est le cas pour beaucoup de pays d'Afrique. Cela reflète que l'incertitude sur la fécondité dans ces pays est plus grande que plus ou moins 0.5 enfant par femme. Il y a en particulier une très grande incertitude sur l'évolution de la population au Nigéria qui, selon ces estimations, a 80% de chance d'atteindre un niveau entre 440 millions et 1,2 milliard d'habitants d'ici la fin du siècle, soit un écart de près de 800 millions de personnes ! Par ailleurs, le milieu de cet intervalle (820 millions) est bien supérieur au scénario médian de la projection déterministe pour ce pays (750 millions). Cela provient de l'incertitude sur la fécondité qui ne diminue pas aussi vite que cela avait été observé dans les pays d'Asie ou d'Amérique Latine il y a quelques années. Il reste néanmoins que la population du Nigéria va très probablement augmenter très fortement et contribuer pour beaucoup à la forte augmentation de la population de l'Afrique. L'ONU prévoit de plus que le rapport du nombre de personnes de 20 à 64 ans sur le nombre de personnes de 65 ans et plus va chuter dans beaucoup de pays : selon ces prévisions il va passer en Chine de 7,8 aujourd'hui à 1,8 en 2100 et de 8,6 à 1,5 pour le Brésil, atteignant ainsi des niveaux bien inférieurs au niveau actuel du Japon (2,6).

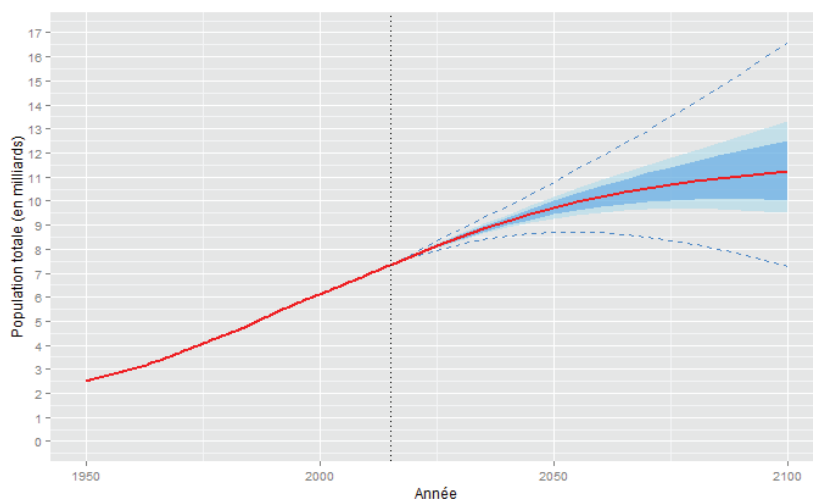


Figure 3. Évolution passée et future de la taille de la population mondiale. En rouge le scénario médian obtenu par projections déterministes et en pointillés les scénarios « haut » et « bas » correspondant à une fécondité de plus ou moins 0,5 enfant par femme. La zone en bleu très clair indique l'intervalle de confiance à 95% et celle en bleu plus foncé indique l'intervalle de confiance à 80% obtenu par projections probabilistes.

Source : Nations Unies, Département d'économie et des affaires sociales, Division de la population.

Le cas de la France : les prévisions de l'ONU et celles de l'Insee.

Les dernières projections démographiques réalisées en 2010 par l'Institut national de la statistique et des études économiques (Insee) vont jusqu'en 2060. Ce sont des projections déterministes déclinées selon plusieurs scénarios de fécondité, mortalité et soldes migratoires. Le scénario central retenu donne une population de 73,6 millions habitants en France métropolitaine au 1er janvier 2060. Selon le scénario de fécondité haute, la population pourrait atteindre 77,6 millions d'habitants d'ici 2060 et selon le scénario de fécondité basse, 69,9 millions (la différence de fécondité entre ces deux scénarios est de 0,3 enfant par femme). Les prévisions probabilistes de 2014 de l'ONU indiquent que la population de France métropolitaine se situera entre 68,4 et 76,1 millions d'habitants en 2060 avec une probabilité de 80% et entre 66,5 et 78,5 millions avec une probabilité de 95%.

Les progrès informatiques à la source des avancements méthodologiques

Les prévisions probabilistes fondées sur l'inférence bayésienne ont été rendues possibles grâce à l'essor exponentiel de la puissance de calcul informatique. Il faut en effet explorer l'ensemble du champ des possibles un très grand nombre de fois pour déterminer les probabilités d'occurrences des événements. Les prévisions probabilistes permettent de mieux appréhender l'incertitude inhérente aux phénomènes complexes pour lesquels d'innombrables facteurs entrent en jeu. Quantifier la probabilité de réalisation des différents avenir possibles apporte plus d'information sur nos connaissances réelles qu'une unique prévision. Dans le cas des projections de population de l'ONU, cela a permis de réduire considérablement l'incertitude. De plus, une des forces des méthodes bayésiennes est que l'ensemble des données est utilisé pour estimer un paramètre précis. Par exemple, l'évolution future de l'espérance de vie en

France est estimée à l'aide des données françaises mais aussi à l'aide des données de l'ensemble des autres pays. Enfin, avec l'inférence bayésienne, on peut intégrer des savoirs extérieurs à la modélisation, par le biais des distributions a priori. Ces distributions reflètent notre connaissance (ou méconnaissance) des paramètres avant toute observation. Le statisticien a en effet le choix de la distribution a priori qu'il va utiliser. Il peut alors prendre en compte des savoirs divers qui permettent d'avoir une idée de cette distribution. C'est souvent cette liberté qui est la source la plus vive des critiques émises à l'encontre des méthodes bayésiennes car, pour certains, liberté est alors synonyme d'arbitraire.

Références

- Alkema, L., Raftery, A.E., Gerland, P., Clark, S.J., Pelletier, F., Buettner, T. and Helig, G. (2011). *Probabilistic Projections of the Total Fertility Rate for All Countries*. *Demography* vol. 48, p.815-839.
- Attali J., « Peut-on prévoir l'avenir ? », Fayard, 2015.
- Blanpain N. et Chardon O., « Projections de population 2007-2060 pour la France métropolitaine : méthode et principaux résultats », *Document de travail*, Direction des Statistiques Démographiques et Sociales, Insee, 2010.
- Efron B., « Bayes' Theorem in the Twenty-First Century », *Science*, 2013.
- Gerland, P., Raftery, A.E. ; Ševčíková , H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B.K., Chunn, J.L., Lalic, N., Bay, G., Buettner, T., Heilig, G.K. and Wilmoth, J. (2014). World Population Stabilization Unlikely This Century., *Science*, 2014, n°346, p 234-237.
- Leslie P.H., « On the use of matrices in certain population mathematics », *Biometrika*, 1945, Vol. 33, p 183-212.
- Raftery A., « Use and communication of probabilistic forecasts », arXiv 1408.4812v1, 2014.
- Raftery A., Alkema L. et Gerland P., « Bayesian Population Projections for the United Nations », *Statistical Science*, 2014, Vol. 29, n°1, p 58-68.
- Raftery, A.E., Chunn, J.L., Gerland, P. and Ševčíková , H. (2013). *Bayesian Probabilistic Projections of Life Expectancy for All Countries*. *Demography*, vol. 50, p.777-801.
- Raftery A., Lalic N., Gerland P., « Joint probabilistic projection of female and male life expectancy », *Demographic Research*, 2014, Vol. 30, p 795-822.
- Robert C., *Le Choix Bayésien*, 2006, Springer.