Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 6, n°1; Juin/June 2010

www.jehps.net

La dispersion des mesures démographiques : vue historique

Daniel COURGEAU[1]

Abstract

This paper traces the development of the notions of dispersion, in demography and in probability, distinguishing two main meanings for this term: the one first, to scatter, to cast here and there; in second, to separate the elements, to break the unity of a set. This development was initiated by the work of Pascal and Fermat for probability and Graunt for political arithmetic, during the second half of the seventeenth century. Major progress was made with the probabilistic and demographic work of Laplace, who developed an epistemic approach of these two notions of dispersion. However, with an objectivist probabilistic approach, that took place in the middle of the nineteenth century, and the use of population censuses, these notions entirely disappeared from the demographic field. In the early 1980's the development of event history analysis and of multilevel analysis permitted the reintroduction of these two notions.

Résumé

Cet article suit en parallèle le développement des notions de dispersion, tant en probabilités qu'en démographie, celles-ci étant prises dans les deux sens du terme : d'une part l'action de répandre, de jeter çà et là ; d'autre part l'action de séparer les éléments, de rompre l'unité d'un ensemble. Ce développement fut initié par les travaux de Pascal et Fermat en probabilité et par ceux de Graunt en arithmétique politique, dans la seconde moitié du XVII^{ème} siècle. Il atteint une étape importante avec l'œuvre tant probabiliste que démographique de Laplace, qui développe une approche épistémique originale de ces deux notions de dispersion. Cependant, avec le développement d'une approche probabiliste objectiviste dès le milieu du XIX^{ème} siècle et l'utilisation des recensements de population, ces notions en viennent à disparaître entièrement du champ démographique. Ce n'est que dans les années 1980 que le développement des analyses biographique et multiniveau, réintroduisent ces deux notions.

[1] Institut National d'Etudes Démographiques, 133 boulevard Davout, 75980 Paris cedex 20, courgeau@ined.fr .

1. Introduction

Le terme *dispersion* est issu du verbe latin *dispergere*, qui signifie répandre. Dans la première édition du Dictionnaire de l'Académie Française en 1694 ce terme signifiait uniquement : *action de disperser ou par laquelle on est dispersé*, et était illustré par l'exemple de la dispersion des juifs. Au cours du temps ce mot a désigné des réalités beaucoup plus diverses, en particulier dans le domaine scientifique.

Aujourd'hui les académiciens distinguent deux grandes significations de ce terme, et nous verrons que ces significations sont également utilisées, mais cependant dans des sens plus précis, en démographie.

La première signification est : *action de répandre, de jeter ça et là*, qui a conduit entre autres à la signification statistique de ce terme : étalement des observations autour de leur valeur centrale. Ainsi en démographie on peut parler de la dispersion d'un taux ou d'un indice, dispersion que l'on peut mesurer par divers indicateurs numériques : la variance[2], l'écart type, l'intervalle de confiance, le coefficient de variation, etc., au prix cependant d'une perte d'information (Barbut, 2002).

La seconde signification est : *action de séparer les éléments, de rompre l'unité d'un ensemble*. Ainsi un ensemble de faits considérés comme équivalents dans l'explication d'un phénomène, peuvent ne plus l'être lorsque l'on pousse plus loin son analyse. Nous verrons par exemple qu'en démographie l'approche multiniveau, va considérer séparément l'effet d'une caractéristique sur divers groupes, alors qu'elle était considérée comme jouant uniformément sur l'ensemble d'une population dans les approches antérieurement utilisées.

Nous considérerons également l'opposé de ce terme, l'*homogénéité*, qui correspond à la *concentration* des observations sur une seule valeur, ou à l'homogénéité d'un ensemble dont on ne pense pas utile de distinguer les éléments.

Pour donner une vue historique de l'évolution de ces notions, il nous faut maintenant indiquer les origines aussi bien des probabilités que de la démographie, car les deux disciplines sont intimement liées dès leur apparition.

La notion de hasard a depuis longtemps été présente, mais sa formulation rigoureuse n'est apparue que tardivement, en particulier lorsque Pascal et Fermat (1654) mettent au point leur *géométrie du hasard*, qui deviendra par la suite le *calcul des probabilités*. Pascal (1654)[3] indique clairement les conditions qui permettent ce calcul de probabilités pour les jeux de hasard :

> *Si le jeu est de pur hasard et qu'il y ait autant de hasard pour l'un que pour l'autre et par conséquent non plus de raison de gagner pour l'un que pour l'autre ...*

Cela revient à dire que la dispersion apparente des gains cache la grande régularité de la probabilité de gagner qui, dans ce jeu, est constante et égale à $\frac{1}{2}$. Selon l'état dans lequel le jeu se trouve, les joueurs peuvent le quitter en se partageant équitablement les sommes restant en jeu, ce qu'il appelle le *parti*. Cette notion de probabilité va également jouer en démographie, où sous l'apparente dispersion de l'arrivée des phénomènes, vont se trouver des probabilités sous-jacentes supposées constantes.

[2] Henri Caussinus me signale qu'en grec moderne diaspora (διασπορά) signifie aussi variance.

[3] Il convient de noter ici que si ce *Traité du triangle arithmétique* a été rédigé et imprimé en 1654, il n'a été diffusé qu'en 1665. C'est la présentation qu'en donne Huygens (1657) qui fit connaître ces résultats aux lecteurs intéressés.

Pour la population, même si des recensements ont été réalisés par les Egyptiens dès 2900 avant J.C., la mise en place d'une science de la population n'a eu lieu que beaucoup plus tardivement, en particulier lorsque Graunt (1662), soit huit ans après Pascal et Fermat, présente ses *Observations naturelles et politiques* : c'est la mise en place de l'*arithmétique politique* (Petty, 1690), qui deviendra au XIXe siècle la *démographie*. Il est d'ailleurs intéressant de lire sous sa plume que ses recherches ont été motivées par les estimations fantaisistes qui circulaient sur la population de Londres :

> *Je dois reconnaître que, jusqu'à cette provocation, j'avais eu peur, à l'exemple mal compris de David, de tenter une évaluation quelconque du nombre d'habitants d'un endroit aussi peuplé*[4].

En effet la Bible indique que « Satan se dressa contre Israël et il incita David à dénombrer les Israélites » : cela était un péché contre Dieu, qui infligea trois jours de peste foudroyante à Israël. Cette évaluation est en fait possible à partir des *Bulletins de mortalité*, comme Graunt le démontre dans son ouvrage.

Ce qui est le plus surprenant ici c'est de voir appliqué, comme nous le détaillerons plus loin, le même procédé de calcul que pour les jeux de hasard, alors qu'il est a priori difficile de penser que la probabilité d'un décès ou d'une naissance puisse être la même pour tous les hommes, comme elle peut l'être dans le cas de la probabilité de gain dans chacune des parties d'un jeu de hasard.

2. Dispersion de la mortalité selon l'âge

On trouve déjà chez Graunt les notions de base de la démographie, mais parfois maladroitement développées, avec nombre d'erreurs dans les raisonnements et dans les calculs. Nous laissons ici de coté la construction de sa table de mortalité qui a suscité nombre de commentaires, entre autres de Greenwood (1928), Glass (1950), Vilquin (1977), Le Bras (2000) et Rohrbasser (2002), pour nous attacher ici à un autre calcul vraiment probabiliste, qu'il fait au début du chapitre XI, pour estimer la population des adultes de Londres.

Pour ce faire, il part des *Bulletins de Mortalité* de Londres, qui lui indiquent un nombre de 15000 décès annuels environ. D'après les causes des décès, qui étaient portées sur ces bulletins, il considère que 5000 de ces décès concernent des enfants ou des vieillards. A partir des 10000 décès attribuables aux personnes de 10 à 60 ans, il va estimer leur population, sous certaines hypothèses. Voyons plus précisément comment il les pose.

Il va en premier lieu utiliser la notion de jeu équitable, déjà présentée par Pascal, pour *faire les partis des jeux* (1654). Ainsi il écrit :

> *Ensuite, considérant qu'on estime pouvoir parier à chances égales qu'un Homme quelconque peut vivre encore 10 ans, j'ai supposé équivalent de dire qu'un homme sur 10 pouvait mourir dans l'année.*

Cela signifie en termes actuels que si la probabilité de décès sur 10 ans est de $\frac{1}{2}$, alors la probabilité de décès sur un an peut être estimée à $\frac{1}{20}$. Il fait donc l'hypothèse que s'il connaît la probabilité de décéder sur une période de plusieurs années, il peut en déduire la probabilité annuelle : cela est possible si cette probabilité reste constante tout au long de cette période. Il

[4] Nous utilisons ici et par la suite l'excellente traduction de ce texte par Eric Vilquin (1977).

fait également l'hypothèse sous-jacente que, comme dans les jeux de hasard, on peut appliquer la théorie des probabilités pour estimer les décès d'une population.

Hacking (2002, *1975*) reconstruit son raisonnement pour arriver à cette conclusion, de la façon suivante :

> *Cela ne sonne guère exact, et pourtant ça l'est en vertu de nombres heureusement choisis. ... Il fait l'hypothèse d'un taux de décès uniforme, c'est-à-dire d'un taux constant, p, de morts par année. La chance de survivre dix ans étant de* $1/2$*, considérons une population de N personnes. Le nombre de survivants après une année est de* $N(1-p)$*. Le nombre de survivants à la fin de la seconde est* $[N(1-p)-pN(1-p)]$ *soit* $N(1-p)^2$*. Le nombre de survivants après dix ans est* $N(1-p)^{10}$*, un nombre que l'on sait égal à* $N/2$*. Maintenant soit q la chance qu'**au moins** un homme par groupe de dix meure dans l'année en cours ; alors la chance qu'aucun ne meure est de* $(1-q)$*. Cela fait justement* $(1-p)^{10}$*, qui vaut comme on le sait d'après l'équation précédente* $1/2$*. Et donc, comme l'écrit Graunt, q vaut aussi* $1/2$.

En fait, ce calcul ne correspond pas à celui fait par Graunt : Graunt calcule la valeur du quotient annuel[5], qu'il estime être égal à $p=\frac{1}{20}=0{,}05$, alors que Hacking nous montre ici simplement que $1-q=(1-p)^{10}$. Il nous faut donc résoudre l'équation précédente, sachant que $q=\frac{1}{2}$, ce qui conduit à $p=1-\sqrt[10]{1-q}=1-\sqrt[10]{0{,}5}=0{,}067$, soit une valeur plus d'un tiers plus élevée que son estimation par Graunt. Contrairement à ce qu'affirme Hacking, le raisonnement tenu par Graunt est donc finalement incorrect. Il est cependant beaucoup plus complexe que celui tenu par Pascal et Fermat, car il cherche à relier une probabilité mesurée sur 10 ans à une probabilité annuelle, en supposant celle-ci égale tout au long de cette durée. Il pose alors l'hypothèse que sur 10 ans on peut considérer la probabilité annuelle de décéder comme invariable selon l'âge, c'est-à-dire sans dispersion ou homogène, selon le second sens du terme.

Il va ensuite plus loin, en posant l'hypothèse que ce quotient de mortalité annuel est constant sur une plus longue durée de la vie s'étalant de 10 à 60 ans. Il considère donc qu'il n'est pas nécessaire de distinguer des quotients de mortalité différents pour chacun des âges, qui forment à nouveau une unité. Il peut alors en déduire la population de Londres soumise à ce risque, à partir des décès observés. En effet, en écrivant ce quotient égal pour chaque âge :

$$p=p_{10}=\frac{D(10,11)}{N_{10}}=p_{11}=\frac{D(11,12)}{N_{11}}=\cdots=p_{60}=\frac{D(60,61)}{N_{60}}=\frac{\sum_{x=10}^{60}D(x,x+1)}{\sum_{x=10}^{60}N_x},$$

où p_x est le quotient de mortalité à l'âge *x*, $D(x,x+1)$ les décès entre les âges *x* et *x*+1 et N_x, les survivants à l'âge *x*. Il en résulte que l'on peut déduire la population âgée de 10 à 60 ans des décès observés et du quotient *p* estimé :

[5] En démographie on a coutume d'appeler ce quotient annuel de mortalité *q* et non *p*, mais nous gardons ici les notations utilisées par Hacking.

$$\sum_{x=10}^{60} N_x = \frac{\sum_{x=10}^{60} D(x, x+1)}{p}.$$

Si l'on suppose ce quotient égal à $\frac{1}{20}$, comme Graunt l'a calculé, on obtient, à partir des 10000 décès observés, une population de 200000 individus âgés de 10 à 60 ans et non de 100000, comme il l'indique incorrectement car il prend un *multiplicateur* de 10 au lieu de 20 (*ce nombre étant multiplié par 10,...*). Mais nous avons également vu que son estimation de p est incorrecte, et qu'il faut prendre en fait $p = 0{,}067$, soit un multiplicateur de 14,925, qui conduit finalement à l'estimation de cette population à 149250, soit environ 150000.

On peut donc en conclure que le raisonnement probabiliste de Graunt est encore très incertain et que ses hypothèses démographiques sont également discutables. Comme le dit en 1669 Lodewijk Huygens dans une lettre à Christiaan Huygens (Huygens, correspondance 1666-1669), alors qu'ils cherchent à estimer la vie moyenne à divers âges : *J'advoüe que mon calcul des aages n'est pas tout à fait juste mais il ij a si peu à dire que cela n'est aucunement considerable, et d'autant moins que la table Angloise, sur laquelle nous nous fondons, n'est pas dans cette derniere justesse aussi bien…*

L'astronome Halley (1693), permit de mettre en place une table de mortalité plus satisfaisante. Il reconnaît les insuffisances des premiers calculs de Graunt : la population soumise au risque manque, les âges au décès sont inconnus et l'immigration dans Londres et Dublin est importante[6]. Il va alors utiliser les données de la ville de Breslau, où la migration est beaucoup plus faible (1238 naissances annuelles contre 1174 décès) : cela lui permet de faire l'hypothèse sous jacente d'une population stationnaire, développée plus tard par Euler (1760), qui n'utilise cependant pas encore ce terme mais qui indique clairement *que s'il nait tous les ans autant d'enfans, qu'il meurt d'hommes, le nombre de tous les hommes demeurera toujours le même, & qu'il n'y aura point alors de multiplication*. Il n'a alors plus aucune raison de supposer ces quotients de mortalité identiques pour tous les âges, puisqu'il dispose maintenant d'une possibilité d'estimer les décès selon l'âge et de fournir une table des survivants, plus précise.

Cependant Halley, comme la plupart des chercheurs du XVII[e] et de la première moitié du XVIII[e] siècle ne disposait que des statistiques des naissances et des décès, insuffisantes pour pouvoir construire une table de mortalité correcte. Il y manque les populations soumises au risque. Il a fallu attendre 1766 pour que l'astronome suédois Wargentin donne une véritable table de mortalité car il dispose dans son pays de registres de population, qui permettent de disposer des populations soumises au risque de mort, et de registres de décès, qui donnent les numérateurs des taux ou quotients à calculer. Les recensements mis en place au cours du XIX[e] siècle ont permis de généraliser le calcul de ces tables.

Ainsi, sans estimation des décès selon l'âge, Graunt fait l'hypothèse de leur homogénéité, au moins de 10 à 60 ans, pour pouvoir estimer la population correspondante. Une fois ces décès mesurés, cette hypothèse devient inutile, car on peut maintenant vérifier sa validité, et montrer la dispersion, selon le second sens du terme, de leurs valeurs selon l'âge.

---

[6] Voici les termes exacts de Halley : But the Deduction from those Bills of *Mortality* seemed even to their Authors to be defective: First, In that the *Number* of the People was wanting. Secondly, That the *Ages* of the People Dying was not to be had. And Lastly, That both *London* and *Dublin* by reason of the great and casual accession of Strangers who die therein, (as appeared in both, by the great Excess of the *Funerals* above the *Births*) rendred them incapable of being Standards for this purpose; …

Qu'en est-il de la dispersion statistique des mesures démographiques, selon le premier sens du terme, tout au long de cette même période. Jacques Bernoulli (1713), dont nous présenterons plus en détail les travaux sur les probabilités épistémiques dans la partie suivante, montre que *l'estimation d'une probabilité peut être cernée par deux limites, aussi précises que l'on peut le souhaiter*. Cela devrait donc permettre l'estimation de la dispersion des indices démographiques, lorsque l'on dispose des effectifs mesurés pour les estimer.

A notre connaissance, un seul auteur a appliqué ces résultats à des données démographiques : Nicolas Bernoulli (in Montmort, 1713). Il désire réfuter l'argument de la Divine Providence, soutenu par Arbuthnott (1710) à partir de l'observation des enfants nés à Londres entre 1629 et 1710, qu'il présente de la façon suivante :

> *si le hasard gouvernoit le monde, il seroit impossible que les nombres des mâles & des femelles s'approchent de si près pendant plusieurs années de suite qu'ils ont fait depuis 80 ans, …*

Pour ce faire il va prouver :

> *qu'il y a une très grande probabilité que le nombre des mâles & des femelles arrive chaque année entre des limites encore plus petits* [sic] *que ceux qu'on a observés depuis 80 ans de suite*.

Sa démonstration d'ailleurs est proche de celle de Jacques Bernoulli, comme il le reconnaît lui-même :

> *Je me souviens que feu mon oncle a démontré une semblable chose dans son traité De Arte conjectandi, qui s'imprime à présent à Bâle, …*

En revanche aucun des autres auteurs travaillant sur la population jusqu'en 1774, n'eut l'idée de donner des limites dans lesquelles pouvaient se situer ses calculs. On ne trouve chez Kersseboom (1742), Deparcieux (1746), Süßmilch (1741, 1761-1762), etc., aucun essai d'estimation de la dispersion de leurs estimations. Sans doute certains de ces auteurs, comme Süßmilch, pensaient que les causes immuables sous-jacentes à ces phénomènes se trouvent dans l'Ordre divin, qu'une société parfaite montrerait. Dans ce cas la dispersion des indices observée disparaîtrait. Mais, comme Nicolas Bernoulli l'a montré, cet argument est à réfuter : la pensée mythologique, qui vient suppléer l'absence d'explication des phénomènes observés, ne peut pas réellement orienter les recherches (Courgeau, 2010).

3 Vers une estimation de probabilité épistémique

Comme nous l'avons présenté plus haut, Jacques Bernoulli (1713) permet de préciser plus avant le raisonnement sur les degrés de certitude, dans le sens des probabilités épistémiques. Voyons plus en détail comment il a opéré.

Il va considérer les probabilités applicables non seulement aux événements objectifs, comme ceux que l'on rencontre dans les jeux de hasard, mais surtout à d'autres événements quels qu'ils soient. L'exemple démographique qu'il en donne est des plus parlants :

> *Ainsi quand on cherche dans l'abstrait, combien serait plus probable qu'un jeune de vingt ans survive à un vieux sexagénaire, plutôt que celui-ci à lui, il n'est rien que tu puisses considérer outre la différence de l'âge et des années ; mais quand le discours est spécialement au sujet du jeune Pierre individuel, et du vieux Paul, il faut de nouveau prendre garde à leur complexion spéciale et à leur goût, par quoi les deux prennent soin de leur état de santé ; car si Pierre est plus malade, s'il se livre à des*

passions, s'il vit de manière intempérante, on peut concevoir que Paul, quoique plus avancé en âge, soit pourtant capable de concevoir une espérance de vie plus longue.

Ce qui revient à dire qu'un raisonnement démographique classique utilisant des probabilités objectives, sur les membres d'une population, où l'on ne distingue que l'âge et la différence d'âge entre deux individus quelconques, ne vaut plus lorsque l'on s'intéresse à deux individus précis dont on connaît bien d'autres caractéristiques. Notons qu'il précise dans son ouvrage que les caractéristiques sont à adjoindre, *si seulement elles peuvent être acquises*. Elles permettront dans ce cas d'améliorer la recherche sur la survie de l'un des individus par rapport à l'autre. Ainsi un peu plus loin il précise comment, à partir d'expériences faites sur des individus les plus semblables possibles, on peut en extraire une information plus précise sur la probabilité de survie d'un individu donné :

*si par exemple par une expérience faite sur trois cent hommes semblables à Titus, de même âge et de même complexion, tu observais que deux cents d'entre eux ont déjà trouvé la mort avant dix ans exactement, tu pourras conclure plus sûrement qu'il y aura deux fois plus de chances que Titus meure avant dix ans, qu'il puisse franchir ce term*e.

Il pense donc pouvoir extraire de l'expérience sur un grand nombre d'individus, ici trois cent, une estimation approchée de la probabilité subjective inconnue qu'un individu, ici Titus, survive au-delà de dix ans. Leibniz répondit fort justement à cet argument que :

De nouvelles maladies se répandent souvent dans le genre humain et par conséquent quelque soit le nombre de morts dont vous avez fait l'expérience ce n'est pas pour autant que vous avez établi les limites des choses de la nature au point qu'elle ne puisse en varier dans le futur (Bernoulli, Leibniz, 2006).

Bernoulli dut donc reconnaître qu'aucun mortel ne pourra jamais déterminer le nombre de maladies, d'accidents, etc., qui peuvent infliger la mort à un être humain, mais il pense cependant que l'observation d'un grand nombre de cas semblables peut permettre d'extraire cette probabilité, avec une précision croissant avec ce nombre. Il cite d'ailleurs Arnauld et Nicole (1662), comme ayant déjà proposé cette méthode, mais il va la pousser plus avant en posant d'abord, ce que l'on appelé plus tard le *principe de raison insuffisante*[7]. Il indique en effet que pour estimer une probabilité :

Tous les cas sont également possibles, c'est-à-dire que chacun peut se produire aussi facilement que tout autre ;

Cela lui permet d'assigner des probabilités épistémiques a priori pour un fait dont il connaît les divers arguments en faveur ou en défaveur de ce fait.

Poussant plus avant, en vue maintenant de l'estimation d'un intervalle de confiance, le raisonnement tenu par Bernoulli suppose dès le départ que la probabilité de l'événement étudié est connue par l'auteur, mais que l'expérimentateur l'ignore :

je pose dans une certaine urne trois mille jetons blancs et deux mille jetons noirs, ces nombres étant inconnus de toi, et tu retires pour connaître ce nombre par expériences un jeton après l'autre (en replaçant chaque fois celui que tu as retiré, avant de choisir

[7] Cette dénomination permettait d'opposer ce principe au *principe de raison suffisante* de Leibniz, qui pose que pour chaque fait, il y a une raison suffisante pour expliquer pourquoi c'est ce fait qui se produit et pas un autre. Ce principe a ensuite été nommé par Keynes (1921), qui trouvait ce terme peu satisfaisant, le *principe d'indifférence*.

le suivant, de sorte que le nombre de jetons dans l'urne reste constant) et que tu observes combien de fois sortirait un banc et combien de fois un noir.

C'est bien par rapport à cette probabilité inconnue de l'expérimentateur qu'il calcule alors un intervalle *entre deux limites, que l'on peut réduire autant que l'on veut,* proche de ce qu'on appelle de nos jours un intervalle de confiance. En utilisant des notations actuelles, si p est la probabilité inconnue, il peut calculer le nombre d'observations nécessaires n pour obtenir un intervalle de confiance ε pour que la valeur estimée, $\hat{p}_n = \frac{m}{n}$, où m est par exemple le nombre de sorties d'un jeton blanc par rapport au nombre total n d'expériences, soit située dans l'intervalle $[\ p - \varepsilon\ ,\ p + \varepsilon\ ]$. Cela suppose bien l'hypothèse initiale de notre saisie imparfaite d'un monde qui est lui-même absolument déterministe. Une observation de plus en plus précise de ce monde devrait permettre d'en dévoiler tous les rouages et de donner, en reprenant l'exemple précédent, la probabilité que Titus survive au-delà de dix ans, avec une précision croissante. Mais comme l'expérimentateur, dans ce cas, ignore cette valeur de référence p, l'intervalle de confiance ainsi calculé ne lui est guère d'utilité (Courgeau, 2004b).

Le théorème de Bernoulli permet ce que l'on appelle une *approche directe* des probabilités, qui est d'ailleurs celle suivie par ses prédécesseurs, et en permet une quantification précise. Sous cette approche on suppose en effet connue la probabilité de l'événement étudié, et l'on montre comment par épreuves successives la fréquence estimée tend vers cette probabilité. C'est le cas des jeux équitables, où les calculs a priori des probabilités des divers cas envisagés sont possibles. En revanche pour les phénomènes démographiques, dont on n'a *a priori* aucune idée de la valeur d'un quotient par exemple, elle n'est plus envisageable.

Cinquante ans plus tard, Bayes (1763) va résoudre un problème statistique, opposé à l'*approche directe*, et que l'on avait coutume d'appeler l'*approche inverse* des probabilités. Dans ce dernier cas, l'échantillon observé est seul connu mais la population dont il est issu est non seulement inconnue, mais son existence même est une hypothèse : comment peut-on dans ce cas estimer la probabilité de l'événement étudié. Dès le début de son article Bayes pose clairement le problème :

Etant ***donné*** *le nombre de fois qu'un événement inconnu s'est réalisé ou a fait défaut, on* ***demande*** *la chance que la probabilité de sa réalisation lors d'une seule épreuve soit comprise entre deux degrés que l'on puisse assigner.*

On voit qu'il va maintenant prévoir la réalisation d'une épreuve à partir d'un nombre fini d'épreuves semblables, qui peut être très faible.

Nous n'entrerons pas ici dans le détail de sa démonstration (voir en particulier Stigler, 1986). Indiquons qu'il montre finalement qu'ayant estimé la probabilité d'un événement par sa fréquence d'apparition dans n tirages initiaux, $\hat{p}_n$, alors la fréquence avec laquelle un nouvel événement se trouvera dans l'intervalle $[\hat{p}_n - \varepsilon, \hat{p}_n + \varepsilon]$ est égale à[8] :

$$\frac{2(n+1)!}{m!(n-m)!}\hat{p}_n^{\,m}(1-\hat{p}_n)^{n-m}\varepsilon$$

[8] Bayes recherche en fait la probabilité plus complexe qu'elle soit dans un intervalle [b, f] et obtient ainsi une intégrale par rapport à $\hat{p}_n$, entre b et f, de la formule ci-dessous à divisée par 2.

Il obtient bien ainsi un intervalle autour de la probabilité estimée $\hat{p}_n$ où la probabilité recherchée doit se trouver, contrairement à Bernoulli qui construisait un intervalle autour de *p*. Cet intervalle, qui répond à la première signification du terme dispersion, est cette fois-ci parfaitement utilisable par l'expérimentateur. Qui plus est cet intervalle fournit une analyse beaucoup plus pertinente de la dispersion d'une distribution que la variance ou l'écart moyen absolu.

C'est Laplace qui, dans son mémoire de 1774[9] sur la probabilité des causes, a généralisé ce *principe de probabilité inverse* à un nombre quelconque de causes différentes :

> *Si un événement peut être produit par un nombre n de causes différentes, les probabilités de l'existence de ces causes prises de l'événement sont entre elles comme les probabilités de l'événement prises de ces causes, et la probabilité de l'existence de chacune d'elles est égale à la probabilité de l'événement prise de cette cause, divisée par la somme de toutes les probabilités de l'événement prises de chacune de ces causes.*

Ce principe peut se formuler de façon plus succincte. Soit *E* un événement observable et $\{C_1, C_2, \ldots, C_n\}$ l'ensemble de ses causes. Supposons que l'on connaisse les probabilités de *E* pour chacune des causes, C_i, alors si les causes sont considérées comme également vraisemblables, la probabilité de C_i, connaissant E est égale à :

$$p(C_i | E) = \frac{p(E | C_i)}{\sum_{j=1}^{n} p(E | C_j)}$$

C'est exactement ce que démontre Laplace (1786), en indiquant clairement l'hypothèse, déjà citée en note12, comme le *principe de raison insuffisante*[10], que toutes les causes sont également possibles :

> *l'on aura la probabilité d'une cause, prise de l'événement, en divisant la probabilité de l'événement, prise de cette cause, par la somme de toutes les probabilités semblables.*

Ce long détour était nécessaire pour montrer les raisonnements à la base du calcul de la dispersion bayésienne d'une estimation, dans le premier sens du terme. Nous présenterons maintenant l'exemple de son application au taux de masculinité des naissances (Laplace, 1781), tout en rappelant qu'il a abordé nombre d'autres phénomènes démographiques[11] (mortalité, nuptialité, fécondité de diverses populations).

---

[9] Il est intéressant de savoir que Laplace ne semble pas avoir eu connaissance des travaux de Bayes à cette date car l'introduction à son mémoire écrite par Condorcet ne cite pas Bayes, alors que sept ans plus tard (1781) l'introduction du même auteur cite Bayes et Price, qui a présenté ses travaux dans les Philosophical Transactions.

[10] Cette hypothèse est donc différente de celle faite par Bayes selon laquelle c'est le nombre d'épreuves conduisant à l'événement qui est considéré comme uniformément distribué et non sa probabilité. Elle a conduit à des critiques de nombreux auteurs (Edgeworth, 1885 ; Fisher, 1922, 1959) selon lesquelles d'autres distributions monotones de *p* peuvent aussi bien convenir, par exemple $\frac{1}{2} Arc\cos(1-2p)$, conduisant à des résultats différents.

[11] Il est intéressant de voir, qu'en dehors des travaux de Brian et Jaisson (2007), peu d'auteurs actuels se sont intéressés à l'exposé et à la critique des travaux démographiques de Laplace.

Dans ce cas précis la dispersion vient du nombre des mesures effectuées pour réaliser l'estimation proposée. Montrons plus en détail comment il opère.

Il indique dès le départ :

*Lorsqu'on n'a aucune donnée **a priori** sur la possibilité d'un événement, il faut supposer toutes les possibilités depuis zéro jusqu'à l'unité, également probables; ainsi l'observation pouvant seule nous instruire sur le rapport des naissances des garçons & des filles, on doit, à ne considérer la chose qu'en elle-même & abstraction faite des événements, supposer la loi de possibilité des naissances d'un garçon ou d'une fille, constante depuis zéro jusqu'à l'unité, & partir de cette hypothèse dans les différends Problèmes que l'on peut se poser sur cet objet.*

Il pose ainsi clairement la distribution *a priori* dont il va partir pour ensuite, à l'aide des observations, pouvoir estimer une distribution *a posteriori*.

Il va poser la probabilité de la naissance d'un garçon égale à x et celle d'une fille égale à $(1-x)$. Il observe alors $p+q$ naissances, dont p sont des garçons et q sont des filles. La probabilité P pour que la possibilité de la naissance d'un garçon soit comprise entre $\left(\frac{p}{p+q}-\theta\right)$ et $\left(\frac{p}{p+q}+\theta\right)$, où θ est une valeur très petite, est alors égale à :

$$P=\frac{\int_{x=\frac{p}{p+q}-\theta}^{\frac{p}{p+q}+\theta} x^p(1-x)^q\,dx}{\int_{x=0}^{1} x^p(1-x)^q\,dx}.$$

En posant $p=\frac{1}{\alpha}$ et $q=\frac{\mu}{\alpha}$, il arrive, par un calcul complexe, à la valeur approchée suivante de cette probabilité, en négligeant les quantités de l'ordre $p^{-\frac{5}{2}}$ qui dès que les observations sont importantes deviennent très faibles :

$$P'=1-\frac{\sqrt{\alpha\mu}}{\sqrt{2\pi}\,(1+\mu)^{\frac{3}{2}}\theta}\left\{1-\alpha\frac{12\mu^2+(1+\mu)^2(1+\mu+\mu^2)\theta^2}{12\mu(1+\mu)^3\theta^2}\right\}\left\{\begin{array}{l}\left[1-(1+\mu)\theta\right]^{n+1}\left[1+\frac{1+\mu}{\mu}\theta\right]^{n+1}\\ +\left[1+(1+\mu)\theta\right]^{n+1}\left[1-\frac{1+\mu}{\mu}\theta\right]^{n+1}\end{array}\right\}$$

Il en conclut donc, en examinant la valeur de tous les termes de la formule précédente, que la valeur de P' différera d'autant moins de la certitude ou de l'unité, que p et q seront de plus grands nombres.

A titre d'exemple il travaille sur les naissances qui ont eu lieu à Paris de 1745 à 1770. En l'espace de ces vingt six années il est né à Paris $p=251527$ garçons contre $q=241945$ filles. Cela conduit à un taux de masculinité de 50,971%. Une formule voisine de la formule précédente lui permet alors de calculer la probabilité, que la naissance d'un garçon soit égale ou moindre qu'un demi, égale à : $1{,}1521\times10^{-42}$. Il en conclut :

Comme elle est d'une petitesse excessive, on peut regarder comme aussi certain qu'aucune autre vérité morale, que la différence observée à Paris entre les naissances des garçons & celles des filles, est due à une plus grande possibilité dans la naissance des garçons (Laplace, 1781).

On peut également calculer grâce à la formule précédente[12] la probabilité pour que la possibilité de la naissance d'un garçon soit comprise dans les limites $0{,}50971 \pm 0{,}001$, qui est approximativement égale à 0,99984, soit très proche de l'unité.

Pour Londres des données équivalentes à celles de Paris donnent, pour la période allant de 1664 à 1758, 737629 garçons contre 698958 filles, conduisant à un taux de masculinité de 51,346%, taux encore plus élevé qu'à Paris. Cela conduit Laplace à se demander si ce taux plus élevé permet de conclure à une probabilité plus élevée.

Soit u la probabilité de naissance d'un garçon à Paris, p le nombre de naissances de garçons et q celui des filles à Paris, $u - x$ la probabilité de naissance d'un garçon à Londres, p' le nombre de naissances de garçons et q' celui des filles à Londres. La probabilité de ce double événement sera alors :

$$K\, u^{p}\, (1-u)^{q}\, (u-x)^{p'}\, (1-u+x)^{q'}$$

Où K est un coefficient constant. Il en résulte que la probabilité que la naissance d'un garçon soit moins probable à Londres qu'à Paris sera égale à :

$$P = \frac{\int\limits_{x=0}^{x=1} \int\limits_{u=0}^{u=x} u^{p}(1-u)^{q}(u-x)^{p'}(1-u+x)^{q'}\, du\, dx}{\int\limits_{x=0}^{x=1} \int\limits_{u=0}^{u=1} u^{p}(1-u)^{q}(u-x)^{p'}(1-u+x)^{q'}\, du\, dx}$$

Après développement en série de cette quantité et en prenant les trois premiers termes de cette série, il obtient la valeur approchée de P :

$$P = \frac{1}{410458}$$

Il peut donc en conclure :

il y a plus de quatre cent mille à parier contre un, que les naissances de garçons sont plus faciles à Londres qu'à Paris ; ainsi l'on peut regarder comme une chose très-probable, qu'il existe dans la première de ces deux villes, une cause de plus que dans la seconde, qui facilite la naissance des garçons, & qui dépend soit du climat, soit de la nourriture & des mœurs (Laplace, 1781).

Il indique dans ce travail l'objectif de ses méthodes :

c'est ici sur-tout qu'il est nécessaire d'avoir une méthode rigoureuse pour distinguer parmi les phénomènes observés, ceux qui peuvent dépendre du hasard, de ceux qui

[12] Pour effectuer ce calcul approché nous avons utilisé le développement en série, donné par Laplace (1781), des derniers termes de la formule précédente :

$$\left[1-(1+\mu)\theta\right]^{n}\left[1+\frac{1+\mu}{\mu}\theta\right]^{n} = e^{-\frac{(1+\mu)^{3}}{2\mu}\frac{\theta^{2}}{\alpha}-\frac{(\mu-1)(1+\mu)^{4}}{3\mu^{2}}\frac{\theta^{3}}{\alpha}-\dots}$$

en remplaçant, pour l'autre terme, θ par $-\theta$.

dépendent de causes particulières, & pour déterminer avec quelle probabilité ces derniers indiquent l'existence des causes.

La distinction qu'il préconise entre phénomènes dépendant de causes précises et ceux dépendant du hasard, nous parait primordiale pour définir les approches démographiques actuelles, tant biographiques que multiniveaux.

Gauss (1809) préconise l'utilisation de la méthode des moindres carrés pour trouver la solution de ce que l'on appelle maintenant un modèle de régression, en utilisant nombre de résultats antérieurs de Laplace. Il applique ses résultats à l'analyse des mouvements des planètes, et il est surprenant de voir qu'il a fallu attendre un siècle pour que ces méthodes soient appliquées aux sciences sociales.

On voit là apparaître ce début d'analyse démographique plus poussée qui fera intervenir diverses caractéristiques individuelles pour chercher à expliquer un phénomène démographique. Il ouvre ainsi la voie vers une analyse de la dispersion des populations étudiées dans le second sens du terme, mais à l'époque son application aux populations humaines reste plus théorique que réelle.

Cependant, après Laplace peu de chercheurs continuèrent dans cette voie (Poisson, 1837 ; Bienaymé, 1838), et une critique de son approche se développa rapidement, comme nous allons le montrer dans la partie suivante.

4 Disparition des notions de dispersion

En effet, dès le milieu du XIXe siècle, de nombreux auteurs rejettent, violemment parfois, cette approche épistémique des probabilités. Les critiques portent sur le fait que ces probabilités entendent traiter de tous les événements, tant objectifs, comme le lancer d'un dé, que subjectifs, comme les jugements humains (Condorcet, 1785 ; Laplace, 1814 ; Poisson, 1837). Ces auteurs préconisent pour le calcul des probabilités une approche purement objectiviste (Cournot, 1843 ; Ellis, 1849 ; Boole, 1854 ; Venn, 1866).

Cournot (1843) fut l'un des premiers à insister sur la distinction entre probabilités objectives et probabilités subjectives. Il indique alors :

Lorsque le nombre des épreuves est peu considérable, les formules données communément pour l'évaluation des probabilités ***à posteriori*** *deviennent illusoires : elles n'indiquent plus que des probabilités subjectives …*

De même, pour Venn (1866), la principale erreur de certains probabilistes passés et non des moindres, Laplace en est selon lui un exemple (page 83), est d'avoir appliqué la théorie des probabilités à des événements auxquels elle n'était pas applicable. Pour lui, la notion de *série* est la plus fondamentale pour décider si la théorie des probabilités est applicable ou non à un événement et cette théorie n'a aucun sens si elle n'est pas reliée à cette notion. Mais il devient alors nécessaire de définir avec précision ce qu'il entend sous le terme *série*. L'exemple démographique qu'il prend va éclairer ce terme.

Considérons la phase suivante : *certains enfants n'atteindront pas l'âge de trente ans*. Si cette phrase est considérée comme une proposition logique, la notion de série lui est tout à fait étrangère. En revanche s'il s'agit d'une proposition qui peut prendre un caractère numérique, en remplaçant le terme certains par une proportion donnée, alors il est difficile de ne plus parler de série. Cela ne revient pas cependant à formuler que, si l'on observe un certain nombre d'enfants, on observera avant trente ans cette proportion exacte de décès, mais que si l'on observe un nombre croissant d'enfants la proportion de décès observés tendra vers cette limite. L'hypothèse sous jacente est alors que cette probabilité, bien que non calculable a priori comme dans le cas des jeux, existe et reste la même tout au long du temps pour l'événement étudié.

Pour l'école objectiviste il est possible de donner un statut objectif à la notion de probabilité, si l'on se cantonne à l'observation d'événements susceptibles de se produire lors d'épreuves répétées. Ainsi von Mises (1939), un des représentants les plus engagés des probabilités objectives, a pu écrire :

> *Nous ne pouvons rien dire au sujet de la probabilité de décéder d'un individu même si nous connaissons le détail de ses conditions de vie et de sa santé. La phrase « probabilité de décès », n'a pour nous absolument aucun sens, quand elle se réfère à un individu unique.*

On voit combien son approche des probabilités semble différente de celle de Bernoulli, mais surtout de celle de Bayes et de Laplace, qui cherchaient au contraire à en préciser l'estimation à partir de l'observation d'un certain nombre d'individus semblables à l'individu concerné. Il n'est possible dans ce cas, que de parler de la probabilité de décéder dans une population dont la taille est aussi grande que l'on veut. Il n'indique pas cependant qu'une population humaine ne peut être considérée comme infinie. De plus parler de la probabilité d'un événement par nature unique ou plus généralement de la probabilité pour qu'une proposition soit vraie, n'a aucun sens pour un objectiviste. L'événement doit faire partie d'une série (un collectif au sens de von Mises), dont il n'est qu'un élément parmi une infinité d'autres

Ils vont donc rejeter l'utilisation de la formule de Bayes lorsqu'elle part, par exemple, d'une probabilité a priori uniforme, pour estimer ensuite à partir des observations la probabilité de l'événement étudié. C'est en fait une hypothèse, ici l'hypothèse d'une distribution uniforme, qu'ils rejettent comme n'ayant rigoureusement aucun sens, et la notion de probabilité décrivant un état de notre connaissance. Pour eux le terme probabilité ne signifie que la fréquence d'un événement dans une expérience donnée.

Ce rejet dura presque un siècle, avec cependant certains chercheurs qui continuaient à maintenir la position de Laplace (Pearson, 1920, 1925) sans cependant l'utiliser dans des applications précises.

Dans le même temps la démographie ne s'est plus guère préoccupée d'une analyse aussi fine que celle de Laplace, des phénomènes qu'elle étudiait. Il faut bien voir que la mise en place des recensements de population a mis à l'écart certaines des préoccupations antérieures, en fournissant en particulier des populations soumises au risque exhaustives, qui évitaient l'utilisation des données de l'état civil pour estimer ces populations par la méthode du *multiplicateur*, qui nous l'avons vu était déjà utilisée par Graunt.

Ainsi ces recensements de la population, qui apparaissent au XVIII^{e} siècle et se mettent en place tout au long du XIX^{e} siècle en Europe, couplés avec les données exhaustives de l'état civil, vont changer l'utilisation des probabilités. En recueillant des données sur l'ensemble de la population à un moment donné le démographe, peut travailler avec une approche objectiviste, dès lors que ces effectifs sont très grands. Il ne calculera même plus la variance de ses quotients, tant celle-ci est faible. L'approche transversale, avec la méthode des variations concomitantes (Durkheim, 1895; Landry, 1949), aussi bien que longitudinale (Pressat, 1966; Henry, 1972), n'envisage en fait jamais ce calcul.

Pour en montrer les raisons, prenons en exemple les effectifs de la génération masculine française atteignant 60 ans en 1962 (Pressat, 1966) : étant donnés le nombre de décès entre 60 et 61 ans, $D(60, 61)$, et la population de 60 ans, $N(60)$, le quotient annuel de mortalité[13] est estimé à $q_{60} = \dfrac{D(60,61)}{N(60)} = \dfrac{6210}{265344} = 0{,}0234$, soit 23,4 pour mille, et l'auteur ne calcule même pas sa variance qui peut pourtant être estimée (Smith, 1992) comme égale à :

$$Var(q_{60}) = \frac{[N(60) - D(60,61)]D(60,61)}{N^3(60)} = 0{,}000000085,$$

soit 8,5 pour cent millions, en supposant une distribution binomiale des décès, c'est-à-dire issus d'une population homogène. Cette variance est en effet si faible, qu'elle ne présente plus aucun intérêt. Ce n'est que très rarement, lorsqu'il calculera des quotients sur de plus courtes périodes (mensuelles par exemple) que le démographe devra tenir compte de ces variances (Hoem, 1983), car les effectifs, même exhaustifs, connaissant l'événement seront beaucoup plus réduits. Cela explique parfaitement pourquoi la démographie classique, bien qu'ayant gardé sa signification probabiliste, ait laissé de coté toute mesure de la dispersion au premier sens du terme. Notons cependant ici que cette hypothèse d'une population dans laquelle la probabilité de décéder à un âge donné est identique pour tous ses membres, est totalement irréaliste comme nous le montrerons dans la partie suivante.

Nous allons maintenant voir qu'également la dispersion, au second sens du terme, a pratiquement disparu de l'approche démographique longitudinale. En effet, sous cette approche, seule l'analyse des phénomènes démographiques considérés comme *indépendants* les uns des autres et se produisant dans une population *homogène* est réellement possible (Blayo, 1995). Il en résulte de sérieuses difficultés, et même une impossibilité complète, de prendre en compte la dispersion de ces phénomènes, tant dans leur *interaction* réciproque que dans l'*hétérogénéité* des populations. Nous renvoyons le lecteur à Courgeau (2003, 2004, 2007) pour une analyse critique de cette approche. Il en résulte une impossibilité de tenir compte de la dispersion, au second sens du terme, tant celle créée par les autres phénomènes démographiques sur celui que l'on étudie, que celle créée par la diversité des membres de cette population.

Nous pouvons en conclure que lors de cette période qui s'étend jusqu'au début des années 1980, la démographie a pu presque totalement laisser de coté les deux aspects de la dispersion des phénomènes qu'elle étudie, ne gardant que la différence, selon l'âge, des quotients.

5 Réapparition de la dispersion dans l'approche biographique et multiniveau

Pour répondre aux critiques faites à l'approche longitudinale, l'approche biographique va considérer un ensemble de trajectoires individuelles dans toute leur complexité, généralement saisies lors d'enquêtes détaillées. L'unité d'analyse ne sera plus l'événement, comme dans l'analyse longitudinale, mais la biographie individuelle, considérée comme un processus stochastique complexe. Elle ne va plus considérer les divers événements étudiés comme indépendants, mais au contraire va analyser les *dépendances* qui existent entre eux. De même elle ne va plus considérer la population comme homogène, mais au contraire étudier l'*hétérogénéité* existant dans cette population. Cela permet de lever la plupart des critiques faites à l'analyse longitudinale. Nous renvoyons le lecteur à Courgeau et Lelièvre (1996) pour une présentation critique plus poussée de cette approche.

[13] Nous reprenons ici la notation d'un quotient par q, habituelle en démographie.

Pour ce qui est de la dispersion, au premier sens du terme, la variance d'un quotient sera indispensable à considérer pour pouvoir conclure à un effet d'interaction entre phénomènes ou de dépendance entre un phénomène étudié et diverses caractéristiques prises en compte. De plus si cette approche a pu d'abord se développer dans une approche statistique essentiellement objectiviste (Kalbfleisch et Prentice, 1980; Cox et Oakes, 1984; Courgeau et Lelièvre, 1989, 1992 2001; Andersen *et al.*, 1993), une approche bayésienne a plus récemment permis de résoudre de nombreuses difficultés d'estimation et correspond mieux à l'esprit dans lequel cette analyse est effectuée (Ibrahim *et al.*, 2001).

Ainsi, cette approche bayésienne permet en particulier d'incorporer toute information a priori utile pour la question étudiée, ce que l'approche objectiviste ne permettait pas. De même ces méthodes permettent maintenant, grâce à l'échantillonnage de Gibbs et aux Méthodes de Monte Carlo par Chaînes de Markov (MCMC) (Robert, 2006), de résoudre beaucoup plus facilement des problèmes complexes, sans faire appel à des calculs objectivistes asymptotiques. Elle a également de nombreux autres avantages par rapport à l'approche objective, grâce à la disponibilité et à la flexibilité d'outils pour l'élaboration de modèles et l'analyse des données.

Pour ce qui est de la dispersion, au second sens du terme, cette approche en tient également compte, en introduisant l'estimation de l'hétérogénéité d'une population et de la dépendance entre les phénomènes étudiés. Les raisons à la fois internes (dépendance entre phénomènes) et externes à la démographie (hétérogénéité d'une population), peuvent ainsi être mises en évidence et leurs effets sur les comportements individuels peuvent être analysés avec un grand détail. On risque cependant de commettre, dans ce cas, ce que l'on appelle l'*erreur atomiste*, car en faisant seulement intervenir les caractéristiques de l'individu, on ignore le contexte dans lequel les conduites humaines se produisent. Ce risque d'erreur s'oppose au risque d'*erreur écologique* de l'approche transversale, qui avait été indiqué par des sociologues (Robinson, 1950), et qui risquait d'attribuer à l'individu des raisons plus collectives, liées aux groupes utilisés pour réaliser l'analyse.

Pour échapper à ces risques d'erreur, l'approche contextuelle et multiniveau permettent d'expliquer un comportement individuel en faisant simultanément intervenir divers regroupements d'individus. Ainsi l'approche contextuelle permet d'associer le comportement d'un individu à la fois à ses propres caractéristiques (mesure individuelle) et aux caractéristiques des groupes dont il fait partie (mesure agrégée). L'approche multiniveau permet d'aller plus loin en introduisant une dépendance interne aux divers groupes simultanément aux caractéristiques individuelles et contextuelles. Ces approches permettent donc d'échapper à la fois à l'erreur écologiste, car les caractéristiques agrégées ne sont plus considérées comme un substitut des caractéristiques individuelles, et à l'erreur atomiste à partir du moment où l'on fait intervenir correctement le milieu dans lequel les individus vivent (Courgeau, 2003, 2004a, 2007a).

Bien entendu l'approche bayésienne va permettre une analyse multiniveau encore plus satisfaisante (Goldstein, 2003; Courgeau, 2007b, Draper, 2008), comme dans le cas de l'analyse biographique. En particulier lorsque le nombre d'unités présentes dans un des niveaux d'agrégation est faible, l'estimation par le maximum de vraisemblance des écarts-types asymptotiques peut s'en trouver fortement biaisée, et peut même être négative (Draper, 2008). La situation devient encore plus délicate dans le cas de variables binaires ou plus généralement discrètes : dans certains cas les méthodes de vraisemblance classiques ne peuvent même pas conduire à une évaluation des paramètres du modèle. L'utilisation de méthodes bayésiennes est alors nécessaire.

Ainsi depuis près de trente ans la démographie qui avait complètement laissé de coté la dispersion de ses mesures, est revenue à la vue que la population est hétérogène et que les phénomènes sont dépendants entre eux, qui permet d'introduire une analyse plus poussée de la dispersion.

Plus récemment une vue bayésienne s'est imposée dans d'autres domaines de la démographie, tels que la paléodémographie. En effet l'observation d'indices sur un petit nombre de squelettes est la seule possibilité d'estimer la structure par âge d'une population ancienne. Les méthodes objectivistes d'estimation conduisent souvent à des solutions situées hors du domaine [0,1] de validité de telles probabilités par âge. Dès lors, une solution pleinement bayésienne s'impose et permet une solution optimale du problème (Caussinus et Courgeau, 2010).

6 Conclusion

L'examen de la dispersion des mesures démographiques nous a montré un va et vient entre homogénéité et dispersion qui, cependant, est intervenu entre des unités différentes, selon la période considérée. En effet, comme dans toute discipline scientifique, ce n'est pas la population dans toute la complexité des individus la composant, qui fait l'objet de la démographie, mais certains de ses aspects, qui peuvent se complexifier au cours du temps, mais qui sont toujours caractérisés par un petit nombre de mesures, jugées indispensables pour la compréhension des phénomènes qui l'affectent.

Au départ, on n'a considéré que des probabilités concernant l'ensemble d'une population et l'hypothèse de leur dispersion, au second sens du terme, selon les âges a été testée : la conclusion est qu'il est nécessaire de considérer des probabilités différentes pour chaque âge. Également la dispersion, au premier sens du terme, de certains indices a été testée : pour le taux de masculinité des naissances, l'absence de dispersion a été observée au cours de 80 ans à Londres. Cependant, peu d'exemples de son utilisation, en particulier parmi les ouvrages d'arithmétique politique du XVIII^{e} siècle, peuvent être trouvés.

Laplace continue d'observer une population dans son ensemble mais, cette fois-ci, d'un point de vue bayésien. Il part de l'hypothèse de probabilités a priori uniformément réparties sur l'intervalle [0,1], pour obtenir des probabilités a posteriori, dont il peut estimer avec précision la dispersion, au premier sens du terme. Il indique également l'intérêt de traiter de la dispersion, au second sens du terme, en faisant intervenir des causes qui joueraient sur certaines sous-population et non sur d'autres (mœurs, climat, nourriture). Cependant les méthodes par régression, proposées par Gauss, ne furent guère utilisées en démographie à cette époque.

La diffusion des recensements au cours du XIX^{e} siècle et une critique des bases du calcul bayésien entraînèrent un rejet de l'approche de Laplace. L'analyse transversale et longitudinale utilisées jusqu'au début des années 1980, ont laissé de coté toute mesure de la dispersion des mesures démographiques, dans les deux sens du terme.

Puis la dispersion refait irruption en démographie, avec l'approche biographique et multiniveau, où la prise en compte de caractéristiques individuelles et agrégées fait éclater le cadre d'analyse. La dispersion intervient dans le premier sens du terme, car il devient indispensable d'estimer la variance des effets estimés pour juger de leur validité. La dispersion intervient dans le second sens du terme, du fait des unités de niveau différent introduites simultanément dans l'analyse et des caractéristiques individuelles qui affectent des sous-populations chaque fois différentes.

Un dernier point reste encore à régler. En effet, tout au long de cet article, nous avons contrasté la vue de ceux qui appliquent la probabilité à un cas individuel, et celle de ceux qui rejettent entièrement cette possibilité : pour simplifier, d'un coté Jacques Bernoulli, lorsqu'il parle des chances de décès de Titus, de l'autre von Mises, lorsqu'il dit que la probabilité de décès n'a aucun sens pour lui dès lors qu'elle se réfère à un individu unique. Il s'agit là en fait de la distinction entre *probabilité subjective* et *probabilité objective*, appliquée aux données démographiques.

Ainsi de Finetti (1937), l'un des représentants les plus importants de l'approche subjective, indique clairement que :

> *le degré de probabilité attribué par un individu à un événement donné est révélé par les conditions dans lesquelles il serait disposé de parier sur cet événement.*

Il précise bien plus loin que pour lui un *événement est toujours un fait singulier*. A l'inverse von Mises, l'un des représentants les plus importants de l'approche objective refuse de parler de la probabilité d'un fait singulier, qui pour lui n'existe pas. Il importe de voir comment la démographie se situe vis-à-vis de ces deux positions extrêmes?

Pour la démographie classique, la probabilité objective semble parfaitement convenir, sous l'hypothèse que la population observée exhaustivement peut être considérée comme tirée d'une population théorique infinie ayant les mêmes probabilités qu'elle de connaître les divers événements démographiques. Les variances des probabilités estimées sous cette hypothèse sont suffisamment faibles, comme nous l'avons montré, pour la soutenir parfaitement.

Mais dès que l'on passe aux approches biographique ou multiniveau, qui utilisent souvent des données non exhaustives d'enquêtes, le maintien d'une probabilité objective, bien que toujours possible et utilisé, peut être remis en cause. Une probabilité subjective semble mieux apte à incorporer toute l'information a priori utile sur les phénomènes étudiés, que la probabilité objective ne permettait pas. La dispersion extrême des probabilités, au second sens du terme, selon les caractéristiques individuelles et selon les interactions entre les phénomènes étudiés rend d'ailleurs souvent cette utilisation indispensable. Mais les prévisions individuelles que cette analyse permet, ne sont que très approchées, car un individu a bien d'autres caractéristiques que celles considérées dans l'analyse, qui peuvent dès lors modifier fortement cette prévision (Courgeau, 2007b).

En dépit de ce dernier point les approches biographiques et multiniveau, en introduisant la dispersion dans les deux sens du terme en démographie, ont permis des avancées considérables dans la compréhension des comportements humains.

Remerciements

Je tiens à remercier ici Marc Barbut et Henri Caussinus pour leurs commentaires et avis sur une première version de cet article, qui m'ont permis de l'améliorer. Je désire aussi remercier Emily Tanimura pour une attentive et efficace relecture d'une première traduction en anglais que j'avais faite initialement. Je garde évidemment l'entière responsabilité de mes propos.

Bibliographie

Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N. (1993). *Statistical models based on counting processes*. New York / Berlin / Heidelberg: Springer-Verlag.

Arnauld, A., Nicole, P. (1662). *La logique ou l'Art de penser*. Paris.

Arbuthnott, J. (1694). An argument for divine providence, taken from the constant regularity observ'd in the birth of both sexes. *Philosophical Transactions of the Royal Society of London*, **27**, 186-190.

Barbut, M. (2002). Une définition fonctionnelle de la dispersion en statistique et en calcul des probabilités : les fonctions de concentration de Paul Lévy. *Mathématiques et Sciences Humaines*, **40**, 31-57.

Bayes, T. R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53**, 370-418. (Traduction française de Clero J.P. (1998). *Cahiers d'Histoire et de Philosophie des Sciences*, n° 18).

Bernoulli, J. (1713). *Ars conjectandi.* Bâle: Impensis Thurnisiorum fratrum. (Traduction française de Peyroux J. (1998). *L'art de conjecturer*. Paris : Blanchard).

Bernoulli, J., Leibniz, G. (2006). Quelques échanges? *Journ@l électronique d'Histoire des Probabilité et de la Statistique*, **2**, 1-12.

Bienaymé, J. (1838). Mémoire sur la probabilité des résultats moyens des observations; démonstration directe de la règle de Laplace. *Mémoires Présentés à l'Académie Royale des Sciences de l'Institut de France*, **5**, 513-558.

Boole G. (1854). *An investigation of the laws of thought: on which are founded the mathematical theories of logic and probability*. London: Walton and Maberly.

Brian, E., Jaisson, M. (2007). *The descent of human sex ratio at birth.* Methodos Series 4, Dordrecht: Springer.

Caussinus, H., Courgeau, D. (2010). Estimation de la structure par âge des décès : nouvelles propositions. In *Manuel de paléodémographie*, Séguy, I., Buchet, L. eds, Paris : Ined,

Condorcet (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris : Imprimerie Nationale.

Courgeau, D. (Ed.) (2003). *Methodology and epistemology of multilevel analysis. Approaches from different social sciences.* Methodos Series, vol. 2, Dordrecht / Boston / London: Kluwer Academic Publishers.

Courgeau, D. (2004a). *Du groupe à l'individu. Synthèse multiniveau.* Paris : Ined.

Courgeau, D. (2004b). Probabilités, démographie et sciences sociales. *Mathématiques et Sciences Humaines*, **42**, 27-50.

Courgeau, D. (2007a). *Multilevel Synthesis. From the group to the individual*. Dordrecht: Springer.

Courgeau, D. (2007b). Inférence statistique, échangeabilité et approche multiniveau. Mathématiques et Sciences Humaines, **45**, 5-19.

Courgeau, D. (2010). Paradigmes démographiques et cumulativité. In *La cumulativité du savoir en sciences sociales*, Walliser, B. ed., Paris : Editions de l'EHESS, 243-276.

Courgeau, D., Lelièvre, E. (1989). *Analyse démographique des biographies*. Paris : Ined

Courgeau, D., Lelièvre, E. (1992). *Event history analysis in demography.* Oxford: Clarendon Press.

Courgeau, D., Lelièvre, E. (1996). Changement de paradigme en démographie. *Population*, **51**, pp. 645-654.

Courgeau, D., Lelièvre, E. (2001). *Análisis demográfico de las biografías*, México: El Collegio de México.

Cox, D.R., Oakes, D. (1984). *Analysis of survival data*. London / New York: Chapman and Hall.

Cournot, A.A. (1843). *Exposition de la théorie des chances et des probabilités*. Paris : Hachette.

Deparcieux, A. (1746). *Essai sur les probabilités de la durée de la vie humaine*. Paris : Frères Guerin. (Réédité en 2003 , Paris : Ined).

Draper, D. (2008). Bayesian multilevel analysis and MCMC. In *Handbook of multilevel models*, de Leeuw, J., Meyer, E. eds., New York: Springer-Verlag, 77-140.

Durkheim, E. (1895). *Les règles de la méthode sociologique*. Paris : Alcan.

Ellis, R.L. (1849). *On the Foundations of the theory of probability*. Cambridge: Cambridge Philosophical Society, vol. VIII.

Euler, L. (1767). Recherches générales sur la mortalité et la multiplication du genre humain. *Histoire de l'Académie Royale des Sciences et des Belles Lettres de Berlin*, pour l'année 1760, vol. 16.

Fermat, P. (1679). *Varia opera mathematica*. Toulouse: Joannen Pech.

Finetti de, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Poincaré*, **7**, 1-68.

Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society* (A), **222**, 309-368.

Gauss, C.F. (1809). *Theoria motus corporum celestium*. Hamburg: Perthes et Besser.

Glass, D.V. (1950). Graunt's life table. *Journal of the Institute of Actuaries*, **76**, 60-64.

Goldstein, H. (2003). *Multilevel statistical models.* London: Edward Arnold.

Graunt, J. (1662). *Natural and political observations mentioned in a following index, and made upon the bills of mortality.* London. (Traduction française de Vilquin E., *Observations Naturelles et Politiques répertoriées dans l'index ci-après et faites sur les bulletins de mortalité*. (1977). Paris: Ined.).

Greenwood, M. (1928). Graunt and Petty. *Journal of the Royal Statistical Society*, **91**, 79-85.

Hacking, I. (1975). *The emergence of probability.* Cambridge: Cambridge University Press.

Hacking, I. (2002). *L'émergence de la probabilité*. Paris : Seuil.

Halley, E. (1693). An estimate of the degrees of the Mortality of Mankind, drawn from curious Tables of the Births and Funeral's at the City of Breslau; with an Attempt to ascertain the price of the Annuities upon Lives. *Philosophical Transactions Giving some Accounts of the Present Undertaking, Studies and Labour of the Ingenious in many Considerable Parts of the World*, vol. XVII, n° 196, 596-610.

Henry, L. (1972). *Démographie. Analyse et modèles*. Paris: Larousse.

Hoem, J. (1983). Multistate mathematical demography should adopt the notions of event history analysis. *Stockholm Research Reports in Demography*, 10.

Huygens, C. (1657). *De ratiociniis in ludo aleae*. Leyde: Elzevier.

Huygens, C. (1666-1669). Correspondance. In *Œuvres complètes, Tome Sixième*, La Haye : Martinus Nijhoff, 1895.

Ibrahim, J.G., Chen, M.-H., Sinha, D. (2001). *Bayesian survival analysis*. New York / Berlin / Heidelberg: Springer-Verlag.

Kalbfleish, J.D., Prentice, R.L. (1980). *The statistical analysis of failure time data*. New York / Chichester / Brisbane / Toronto: John Wiley and Sons.

Kersseboom, W. (1742). *Essais d'arithmétique politique*. La Haye : Jan van der Bergh (traduit en français en 1970, Paris : Ined).

Keynes, J.M. (1921). *A treatise on probability*. London : Macmillan and Company Limited.

Landry, A. (1949). *Traité de démographie*. Paris : Payot.

Laplace, P.S. (1774). Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Académie Royale des Sciences de Paris*, Tome VI, 621-656.

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

Laplace, P.S. (1786). Suite du mémoire sur les approximations des formules qui sont fonctions de très-grands nombres. *Mémoires de l'Académie Royale des sciences de Paris*, 1783, 423-467.

Laplace, P.S. (1812). *Théorie analytique des Probabilités, 2 vols*. Paris: Courcier Imprimeur.

Laplace, P.S. (1814). *Essai philosophique sur les probabilités*. Paris : Courcier Imprimeur.

Le Bras, H. (2000). *Naissance de la mortalité. L'origine politique de la statistique et de la démographie*. Paris : Seuil/Gallimard.

Mises von, R. (1928). *Wahrsheinlichkeit, statistik und wahrheit.* Wien: Springer (traduction anglaise (1957), *Probability, Statistics and truth*, London: George Allen & Unwin Ltd.).

Montmort de, P.R. (1713). *Essay d'analyse sur les jeux de hazard*, Seconde Édition. Paris : Jacques Quillau.

Pascal, B. (1654). *Traité du triangle arithmétique, avec quelques autres traités sur le même sujet*. Paris : Guillaume Desprez. (Voir en particulier : III Usage du triangle arithmétique pour déterminer les partis que l'on doit faire entre deux joueurs qui jouent plusieurs parties).

Pascal, B. (1654). *Celeberrimæ mathesos academiæ Pariensi*. Paris : Académie Parisienne.

Pascal, B. (1922). *Les lettres de Blaise Pascal accompagnées de lettres de ses correspondants*. Paris : Les Éditions G. Crès. (Voir le courrier échangé avec Pierre de Fermat en 1654, 188-229).

Pearson, K. (1920). The fundamental problem of practical statistics. *Biometrika*, **13**, 1-16.

Pearson, K. (1925). Bayes' theorem, examined in the light of experimental sampling. *Biometrika*, **17**, 388-442.

Petty, W. (1690). *Political arithmetick*. London: Robert Clavel & Hen. Mortlock.

Poisson, S.-D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Paris : Bachelier.

Pressat, R. (1966). *Principes d'analysedémographique*. Paris : INED.

Robert, C.P. (2006). *Le choix bayésien*. Paris : Springer-Verlag France.

Robinson,W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, **15**, 351-357.

Rohrbasser, J.-M. (2002). Qui a peur de l'arithmétique? Les premiers essais de calcul sur les populations dans la seconde moitié du XVIIe siècle. *Mathématiques et Sciences Humaines*, **159**, 7-41.

Süβmilch J.P. (1741). *Die götliche Ordnung*, Berlin: J.C. Spenner. (traduit en français en 1998 par J.M. Rohrbasser, Paris: Ined).

Süβmilch J.P. (1761-1762). *Die götliche Ordnung*, Berlin: G.L. Winter. (traduit en français en 1979 par J. Hecht, Paris, Ined).

Venn, J. (1866). *The logic of chance*. London : Macmillan.

Vilquin, E. (1977). Introduction. In *Observations Naturelles et Politiques répertoriées dans l'index ci-après et faites sur les bulletins de mortalité*, J. Graunt, Paris: Ined, 7-31.