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

Dispersion of measurements in demography: a historical view

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

The term *dispersion* comes from the Latin verb *dipsergere*, which means to scatter. In the first edition of the dictionary of the French Academy in 1694 this term meant: *the action to disperse or the action by which something is dispersed*, and was illustrated by the example of the dispersion of the Jews. Over time this word has designated much more diverse realities, especially in science.

Today academicians distinguish two main meanings of this term, and we will see that these meanings are also used, with a more precise signification, in demography.

The first meaning is: *to scatter, to cast here and there*, which led to the statistical significance of the term: spread of observations around their central value. In demography you can talk about the dispersion of a rate or an index, a dispersion which can be measured by diverse numerical indicators: variance[2], standard deviation, confidence intervals, the coefficient of variation, etc., at the price, however, of a loss of information (Barbut, 2002).

The second meaning is: *to separate the elements, to break the unity of a set*. Thus a set of facts considered equivalent in the explanation of a phenomenon can turn out not be when we push the analysis further. We will see for example in demography, that the multilevel approach will separately consider the effect of a characteristic on different groups, while the characteristic was considered to consistently influence the entire population, in the approaches considered previously.

We also consider the opposite of this term, *homogeneity*, which corresponds to the *concentration* of indicators on a single non-random value, or to the homogeneity of a set of which we do not find it useful to distinguish the elements.

In order to give a historical view of the evolution of these concepts we must now indicate the origins of both probability and demography, because the two disciplines were closely related from their beginning.

The concept of hazard has long been present, but its rigorous formulation did not appear until later, especially when Pascal and Fermat (1654) developed their *geometry of chance*, which later became the calculation of probabilities. Pascal (1654)[3] clearly posits the conditions that allow the calculation of probabilities for games of chance:

> *If the game is of pure chance and there is as much chance for one or the other and therefore no more reason to win for one or the other…*

This means that the apparent dispersion of earnings hides the high regularity of the probability of winning, which in this game, is constant and equal to one half. Depending on the state in which the game is, players may leave and share equitably the amounts remaining in the game, what he calls the *bet*. This concept of probability will also play a role in demography, where there will be constant underlying probabilities, under the apparent spread of the arrival of phenomena.

As far as populations are concerned, although censuses were made by the Egyptians around 2900 BC, the establishment of a science of the population occurred only much later. An important step was Graunt's presentation in 1662, that is to say eight years later than Pascal and Fermat, of his *Natural and political observations* corresponding to the

[2] Henri Caussinus told me that in modern Greek διασπορά also means variance.
[3] It should be noted here that if the *Treatise on the Arithmetical Triangle* was written and printed in 1654, it was released only in 1665. It was the presentation of these results by Huygens (1657) which permitted their diffusion to interested readers.

establishment of *political arithmetic*, which will become known as *demography* in the nineteenth century.

It is interesting to hear from him that he engaged his thought on this topic because of the fanciful estimates that were going round on the population of London:

> *I must confess, that until this provocation, I had been frighted with that misunderstood Example of David, from attempting any computation of the people of this populous place; ...*

In fact the Bible says that "Satan stood against Israel and incited David to count the Israelites": it was a sin against God, who inflicted three days of devastating plague upon Israel. This evaluation is made possible from the *Bills of Mortality*, as Graunt showed in his work.

What is most surprising here is to see implemented, as we will discuss in greater detail later, the same method of calculation as for gambling, while it seems *a priori* difficult to believe that the probability of a death or a childbirth may be the same for all human being, although it may be in the case of the probability of winning in each *bet* of a game of chance.

2 Dispersion of mortality rates by age

We can find the bases of demography already in Graunt's work, but sometimes clumsily developed, with many errors in reasoning and calculations. We leave aside the construction of the mortality table that has generated many comments, among others from Greenwood (1928), Glass (1950), Vilquin (1977), Le Bras (2000) and Rohrbasser (2002), so here we will concern ourselves with another really probabilistic calculation, in the beginning of Chapter XI, to estimate the population of adults in London.

To do this, he used the *Bills of Mortality* in London, which showed fewer than 15,000 annual deaths. Based on the causes of death, which were given in these bills, he thought that 5,000 of these deaths are from children or elderly. From the 10,000 deaths of people aged 10 to 60 years, he tries to estimate the population, under certain assumptions. Let us see more precisely how he determined the latter.

First he will use the concept of fair play, already presented by Pascal, for bets in games. So he writes:

> *Next considering, That it is esteemed an even Lay, whether any man lives ten years longer, I suppose it was the same, that one of any 10 might die within one year.*

This means in nowadays terms that if the probability of death over 10 years is $\frac{1}{2}$, then the probability of death over one year can be estimated at $\frac{1}{20}$. He therefore assumes that, if he knows the probability of dying over a period of several years, he can deduce the annual probability, which is possible if this probability remains constant throughout this period. He also makes the underlying assumption that, as in games of chance, probability theory can be applied to estimate the deaths in a population.

Hacking (1975) reformulates his reasoning for reaching this conclusion, as follows:

> *This does not even sound correct, but it is, thanks to the happy choice of figures ... Graunt assumes a uniform death rate, that is, that there is a constant chance p of dying in a given year. If the chance of living ten years is 0.5, consider a population of*

size N. The number who survive the first year is N(1-p). The number who survive the second is [N(1-p)-pN(1-p)] or N(1-p)^{2}*. The number who survive ten years is N(1-p)*^{10} *= 0.5N. Now let q be the chance that at least one man in a group of ten dies in a given year; then 1-q is the chance that no one dies. This is just (1-p)*^{10}*, which, solving the above equation, is 0.5. So, as Graunt says, q is also 0.5.*

In fact, this calculation does not correspond to the one made by Graunt: Graunt calculates the annual probability of dying[4], which he considers equal to $p = \frac{1}{20} = 0{,}05$, while Hacking shows simply that $1 - q = (1 - p)^{10}$. So we must solve the above equation knowing that $q = \frac{1}{2}$, which leads to $p = 1 - \sqrt[10]{1-q} = 1 - \sqrt[10]{0{,}5} = 0{,}067$, is more than one third higher than its estimate by Graunt. Contrary to Hacking, the reasoning of Graunt is ultimately incorrect. It is however more complex than the one given by Pascal and Fermat, as it seeks to relate a measured probability for 10 years to an annual probability, assuming that it is equal throughout this period. It posits the possibility that in 10 years we can consider the annual probability of dying as being unmodified by age, i.e. without dispersion or uniform, according to the second meaning of this term.

He then goes further by assuming that the annual probability of dying is constant over a longer life span ranging from 10 to 60 years. He therefore considers it unnecessary to distinguish the different probabilities of mortality for each age, which then forms a new unit. He can then infer the population of London subject to this risk, based on observed deaths. Indeed, writing this probability equal for each age he obtains:

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

Where p_x is the probability of dying at age *x*, $D(x, x+1)$ the deaths between the ages *x* and *x*+1 and N_x, the number surviving at age *x*. It follows that we can deduce the population aged 10 to 60 years old from the deaths observed and the estimated probability *p*:

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

Assuming this ratio equal to $\frac{1}{20}$, as estimated by Graunt, we obtain, from the 10,000 reported deaths, a population of 200,000 individuals aged 10 to 60 years, but not of 100,000 as Graunt stated incorrectly while using a multiplier of 10 instead of 20 (*which number being multiplied by 10, …*). But we have also seen that the estimate of *p* is incorrect and this leads to $p = 0.067$, giving a multiplier of 14.925, which ultimately leads to estimating the population to 149,250 individuals, which is around 150,000.

We can therefore conclude that the probabilistic reasoning of Graunt is still very uncertain and that his demographic assumptions are questionable. As stated in 1669 by

[4] In demography, it is customary to call the annual probability of dying *q* and not *p*, but we keep here the notations used by Hacking.

Lodewijk Huygens in a letter to Christiaan Huygens (Huygens, Correspondence 1666-1669), in which they seek to estimate the average life at various ages:

> *I confess that my calculation of ages is not entirely correct, but there is so little to say that this is not too false, especially since the English table, on which we rely, is also not very accurate…*

The astronomer Halley (1693), worked out a set up for a more satisfactory table of mortality more satisfactory. He acknowledged the shortcomings of the first calculations of Graunt: the population at risk lacks, the ages at death are unknown and immigration in London and Dublin is important. He says more precisely:

> *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; …*

He will then use the data from the city of Breslau, where migration is much lower (1238 annual births against 1174 deaths): it allows him to make the underlying assumption of a stationary population, developed later by Euler (1760), who does not yet use this term but clearly indicates *that if every year as many children are born, as many men are dead, then the number of men will always remain the same, and there will not be any increase of the population*. He no longer had any reason to assume the same probability of death for all ages, since he had the opportunity to estimate deaths according to age and provide a more precise survival function.

Halley, however, like most researchers in the seventeenth and first half of the eighteenth century could only use the statistics of births and deaths, insufficient to construct a mortality table correctly. It lacks the populations at risk. It was only in 1766 that the Swedish astronomer Wargentin provided a real life table because in his country there were population registers, which allow measurement of the population subject to the risk of death, and death records, which give the numerators of rates or probabilities to calculate. Finally, the censuses established during the nineteenth century led to generalize the calculation of these tables.

Thus, without any estimation of deaths according to age, Graunt had to make the homogeneity assumption, at least from 10 to 60 years, in order to estimate the corresponding population. Once these deaths have been measured, this assumption becomes useless because you can now check its validity and show dispersion, according to the second sense, of their values depending on age.

We will now look at the statistical dispersion of demographic measures, in the first sense, throughout the same period. Jacques Bernoulli (1713), which work on epistemic probabilities we will present in more detail his in the next section, shows that *the estimation of a probability may be delimited by two boundaries, as accurate as one could wish*. This should enable the estimation of the dispersion of demographic indices, when we have the numbers measured in order to calculate it.

To our knowledge, only one author has applied these results to demography: Nicolas Bernoulli (in Montmort, 1713). He wants to refute the argument for Divine Providence, supported by Arbuthnott (1710), from his observation of children born in London between 1629 and 1710, which was presented as follows:

if chance ruled the world, it would be impossible that the number of boys and girls are so close for several years in a row as it was for 80 years.

In order to show that this result is incorrect, he had to prove that:

there is a high probability that the number of boys and girls is every year among more narrow limits than those that were observed during the following 80 years …

His demonstration is very close to that of Jacques Bernoulli, as he himself admits:

I remember my late uncle has demonstrated a similar thing in his treatise De Arte conjectandi, which is now in print in Basel.

However none of the other authors working on the population until 1774, had the idea to give the limits within which his calculations could be contained. We find in Kersseboom (1742), Deparcieux (1746), Süβmilch (1741, 1761-1762), etc., no attempt to estimate the dispersion of their estimates. No doubt some of these authors, like Süβmilch, thought that the immutable causes underlying these phenomena were to be found in the *Divine Order*, which would come to expression in a perfect society. In this case the observed dispersion indices would disappear. But, as Nicolas Bernoulli has shown, this argument is to be refuted: the mythological thinking, which makes up for the lack of explanation of observed phenomena, can not really direct research (Courgeau, 2010).

3 Towards an estimation of epistemic probability

As we said earlier, Jacques Bernoulli (1713) can further clarify the reasoning about the degrees of certainty in the sense of epistemic probabilities. Let see more in detail how he operated. He will consider the probabilities applicable not only to objective events, such as those found in games of chance, but mostly to other arbitrary events. The demographic example he gives is most eloquent:

For instance when we search, in the abstract, how much more likely it would be for a youth of twenty to outlive an old sexagenarian, rather than the latter outliving the former, there is nothing you can take into account apart from their difference in age and their years; but when the discussion specifically concerns young man Peter and old man Paul, you need once again to pay careful attention to their particular constitution and their likings, which determine how the two take care of their health; for if Peter is more ill, if he indulges in passions, if he lives an intemperate life, it is conceivable that Paul, despite his older age, may yet be able to contemplate a longer life expectancy.[5]

In the classic demographic approach using objective probability, the only criteria for distinguishing between any two members of a population were their ages and their age gap. Bernoulli's example above implies that this classic approach ceases to apply when we examine two specific persons for which many other characteristics—apart from age—are very familiar to us. However, Bernoulli indicates that the characteristics are to be included in the

[5] Ita cùm qæritur in abstracto, quantò sit probabilius, juvenem vigenti annorum senem sexagenario fore superstitem, quàm verò hunc illi, præter discrimen ætatis & annorum nihil is, quod considerare possis; sed ubi specialiter sermo is de individuis Petri juvenis & Pauli senis, attendere insuper opportet ad specialem eorum complexionem & studium, quo uterque valetudinem suam curat; nam si Petrus sit valetudinarius, if infectibus indulgeat, if intepemperanter vivat, fieri potest, ut Paulus, etsi ætate provectior, optima tamen ratione longioris spem vitæ concipere valeat.

analysis *only if they can be acquired*.[6] If so, they will improve our estimation of the chances that one individual will outlive the other. A few pages later, Bernoulli describes how, from trials performed on people who resemble one another as closely as possible, we can extract more specific information on a given person's probability of survival:

> *if, for example, in a test conducted on three hundred men resembling Titius, of identical age and constitution, you observed that two hundred of them had already died before the exact age of ten, you could conclude more surely that Titius is twice as likely to die before age ten as he is of living beyond that limit*.[7]

Bernoulli therefore believed that, by testing a large number of individuals (here, three hundred) he could obtain a rough estimate of the unknown subjective probability that an individual (here, Titius) will survive beyond age ten. Leibniz rightly replied to this argument:

> *New diseases are often spread in the human race and therefore no matter how many deaths you have experienced this does not mean you have established the entirety of the laws of nature about this fact, so that it could no more change in the future* (Bernoulli, Leibniz, 2006).

Bernoulli must therefore recognize that no mortal can ever determine the number of diseases, accidents, etc., which can result in the death to a human being, but he nevertheless believes that the observation of a large number of similar cases may help extract this probability with an accuracy increasing with their number. He also cites Arnauld and Nicole (1662), as having already proposed this method, and he pushes it further by putting first, what was later called the principle of *non-sufficient reason*[8]. It states that to estimate a probability:

> *All cases are equally possible, i.e. everyone can occur as easily as any other;…*[9]

This allows him to assign epistemic *a priori* probabilities for a fact when he knows the various arguments for or against it.

Pushed further, in view of estimating a confidence interval, the reasoning given by Bernoulli assumed from the outset that the probability of the event studied is known by the author, but is ignored by the experimenter:

> *in a given urn I place three thousand white tokens and two thousand black ones, these numbers being unknown to you, and to determine the number by experiment you remove one token after another (replacing each token as you remove it, before choosing the next one, so that the number of tokens in the urn remains constant) and you observe how many times a white token comes out and how many times a black one comes out.*[10]

---

[6] si modo haberi possunt.

[7] si ex. gr. facto olim experimento in tercentis hominibus ejusdem, cujus nunc Titius is, ætatis & complexionis, observaveris ducentos eorum ante exactum decennium mortem oppetiisse, reliquos ultravitam protraxisse, satis tu colligere poteris, duplo plures casus esse, quibus & Titio intra decennium proximum naturae debitutm solvendum sit, quàm quibus terminium hunc transgredi possit.

[8] This denomination permits to set this principle against the *principle of sufficient reason* given by Leibniz, which says that for each fact, there is a sufficient reason in order to explain why this fact occurred against another one. This principle was further named by Keynes (1921), who found this term clumsy and unsatisfactory, *principle of indifference*.

[9] binis limitibus conclusam, sed qui tam arcti constitui possunt, quam quis voluerit.

[10] pono in urna quadem te inscio reconditos esse ter thousand calculos albos & bis thousand nigros, teque eorum nyumerum experimentis exploraturum educere calculum unum post alternum (reponendo tamen singulis vicibus

It is precisely in view of this probability—unknown to the experimenter—that Bernoulli then determines what we now call a confidence interval *between two limits, which we can reduce as much as we want*.[11] Using current notations, if p is the unknown probability, he can calculate the number of observations n needed to obtain a confidence interval ε such that the estimated value, $\hat{p}_n = \frac{m}{n}$ (where m is, for example, the number of draws of a white token divided by the total number of trials n), lies within the interval $\left[\, p - \varepsilon \,,\; p + \varepsilon \,\right]$. This relies indeed on the initial hypothesis that we have an imperfect grasp of a world that is however totally deterministic. An increasingly precise observation of that world should enable us to reveal all of its mechanisms and—returning to the previous example—to compute with a growing accuracy the probability that Titius will live beyond ten years. But as the experimenter, in this case, does not know the reference value p, the confidence interval thus determined is of little use to him (Courgeau, 2004b).

Bernoulli's theorem allows what we call a *direct approach* to probability—which is, in fact, the one adopted by his predecessors—and allows an accurate quantification of probability. The approach assumes that the probability of the event studied is known, and shows how through successive trials the estimated frequency tends toward that probability. One example is fair games, where we can determine *a priori* the probability of the various outcomes considered. By contrast, the approach is not applicable to subjective phenomena.

Fifty years later, Bayes (1763) will solve a statistical problem, opposed to the *direct approach*, which was called the *inverse approach* of probabilities. In the latter case, the observed sample is only known but the population from which it is derived is not only unknown, but its existence is a hypothesis: how can we in this case estimate the probability of the event studied? At the very start of his paper, he states the problem clearly:

> ***Given*** *the number of times in which an unknown event has happened and failed:* ***Required*** *the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.*

He thus sets out to predict the occurrence of a trial on the basis of a finite number of similar trials, which can be very small. We shall not go into the details of his demonstration (see in particular Stigler, 1986). Let us just say that having finally estimated probability of an event by its initial frequency of occurrence in n trials, $\hat{p}_n$, then the frequency with which a new event will lie in the interval $\left[\hat{p}_n - \varepsilon, \hat{p}_n + \varepsilon\right]$ is equal to:[12]

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

Bayes, therefore, effectively obtains an interval around the estimated probability $\hat{p}_n$, in which the sought-for probability must lie, while Bernoulli built an interval around p. This time, the interval is perfectly usable by the experimenter.

---

illum quem eduxisti, priusquam sequentem eligas, ne numerus calculorum in urna minuatur) & observare, quoties albus & quoties ater exeat.

[11] binis limitibus conclusam, sed qui tam arcti constitui possunt, quam quis voluerit.

[12] Bayes, in fact, seeks the more complex probability that the sought-for probability lies in an interval [*b*, *f*]. He thus obtains an integral relative to $\hat{p}_n$, between *b* and *f*, of the equation below divided by 2.

Laplace, in his 1774 paper[13] on the probability of causes, generalized this *principle of inverse probability* to any given number of different causes:

> *If an event can be produced by a number* n *of different causes, the probabilities of the existence of these causes given the event stand with respect to one another as the probabilities of the event given these causes, and the probability of the existence of each is equal to the probability of the event given this cause, divided by the sum of all the probabilities of the event given each of these causes.*

We can express this principle more concisely. Let E be an observable event and $\{C_1, C_2, \ldots, C_n\}$ the set of its causes. Let us assume that we know the probabilities of E for each cause, C_i. If we view all the causes as equally likely, the probability of C_i, knowing E, is:

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

That is exactly what Laplace demonstrated (1786), clearly designating the hypothesis that all the causes are equally possible (this hypothesis was already mentioned in note 7 as the *principle of non-sufficient reason or principle of indifference*[14]):

> *we shall obtain the probability of a cause, determined from the event, by dividing the probability of the event, given this cause, by the sum of all the similar probabilities.*

This long detour was necessary to show the reasoning behind the calculation of the dispersion of Bayesian estimation in the first sense. We now present an example of its application to sex ratios at birth (Laplace, 1781), while recalling that he has addressed many other demographic phenomena (death, marriage, fertility of different populations).

In this case dispersion comes from the number of measurements effectuated to achieve the proposed estimate. Let us show in more detail how it operates.

He said at the outset:

> *When no prior is given for the possibility of an event, we must assume all the possibilities from zero to unity, equally probable; so the observation can only give us information on the ratio of births of boys & girls; we must, to consider the thing in itself and excluding events, assume the law of possibility of the birth of a boy or a girl, constant from zero to the unit, and use this hypothesis in the different problems that may arise on this subject.*

He clearly posits the *prior* distribution from which he will depart and then, using observations, he can estimate a *posterior* distribution.

---

[13] Interestingly, Laplace does not seem to have been aware of Bayes's work at that date, for the introduction to his paper (written by Condorcet) does not mention Bayes. By contrast, seven years later (1781), Laplace's introduction quotes Bayes and Price, who published his results in the *Philosophical Transactions*.

[14] This hypothesis is therefore different from Bayes's hypothesis, namely, that it is the number of trials leading to the event that is regarded as uniformly distributed and not its probability. Many authors criticized Laplace's hypothesis (Edgeworth, 1885; Fisher, 1922, 1959), arguing that other monotonic distributions of p, for example $\frac{1}{2} Arc\cos(1-2p)$, could be equally suitable and yield different results. We shall discuss the hypothesis in greater detail at the end of the chapter.

He introduces the probability of birth of a boy equal to x and that of a girl equal to $(1-x)$. He then observed $p+q$ births, p for boys and q for girls. The probability P that the possibility of the birth of a boy is between $\left(\frac{p}{p+q}-\theta\right)$ and $\left(\frac{p}{p+q}+\theta\right)$, where θ is a very small value, is then equal to:

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

Setting down $p=\frac{1}{\alpha}$ and $q=\frac{\mu}{\alpha}$, he obtains, by an intricate calculation, a rough value of this probability, overlooking the quantities in the order of $p^{-\frac{5}{2}}$ which as the observations are in great number are extremely small, equal to:

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

He concludes, by examining the value of all the terms of the previous formula, that the value of P' will vary less from certainty or unity, the greater the numbers p and q.

For example, he works on births that took place in Paris from 1745 to 1770. Over the period of twenty six years it was born in Paris 251,527 males against 241,945 females. This leads to a sex ratio of 50.971%. A formula similar to the previous one can then be used to calculate the probability that the birth of a boy is equal to or less than one half, which is equal to: 1.1521×10^{-42}. He concludes:

> *As it is a very small value, it can be considered to be as certain as any other moral truth, that the difference observed in Paris between the births of those boys & girls, is due to a greater opportunity in the birth of boys* (Laplace, 1781).

You can also calculate using the previous formula[15] for the probability, that the possibility of the birth of a boy will be included in the confidence interval 0.50971 ± 0.001, is approximately equal to 0.99984, very close to unity.

For London equivalent data to those of Paris, for the period from 1664 to 1758, give 737,629 boys against 698,958 girls, leading to a sex ratio of 51.346%, an even higher rate

[15] In order to undertake this approximate calculation we used the expansion in series, given by Laplace (1781), of the last terms of the preceding formula:

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

replacing, for the second term θ by – θ.

than in Paris. This led Laplace to wonder whether we can conclude from this higher rate a higher probability.

Let u be the probability of birth of a boy in Paris, p the number of births of boys and q of girls in Paris, $u-x$ the probability of birth of a boy in London, p' the number of births of boys and q' the number for girls in London. The probability of these two events will be:

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

where K is a constant coefficient. It follows that the probability that the birth of a boy to be less likely in London than in Paris will be equal to:

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

The series obtained by expanding this formula leads, while taking the three first terms, to the following approximate value for P:

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

He can then conclude:

> *There are more than four hundred thousand chances against one, that male births are easier in London than in Paris, so we can view as very likely that there is, in the first of these two cities, more causes than in the second, which facilitates the birth of boys, and which is either the climate or food and manners* (Laplace, 1781).

He gives also in this work the objective of his methods:

> *it is here especially that it is necessary to have a rigorous method so as to distinguish among the observed phenomena those which may depend on chance, from those which depend on particular causes, and to determine how likely they indicate the existence of causes.*

The distinction between the phenomena that he refers to as those depending on specific causes and those depending on chance seems essential for us, in order to define the forthcoming demographic approaches, both biographical and multilevel.

Gauss (1809) advocates the use of the method of least squares for the solution of what is now called a regression model, using a variety of previous results of Laplace. He applies its findings to the analysis of the movements of planets, and it is surprising to see that it took a century for these methods to be applied to social sciences.

We can see the beginnings of a more developed demographic analysis, which will further involve individual characteristics, as an attempt to explain a demographic phenomenon. It opens the way towards an analysis of the dispersion of the populations studied in the second sense, but at the time its application to human populations is more theoretical than real.

However, after Laplace few researchers continued in this way (Poisson, 1837; Bienaymé, 1838), and criticisms of this approach developed rapidly, as we shall show in the next section.

4 Disappearance of the concepts of dispersion

Indeed, from the mid-nineteenth century, many authors disagreed, sometimes violently, with this epistemic approach of probabilities. Criticisms focus on the fact that these probabilities are trying to deal with all events, both objective ones such as the throw of a dice, and subjective ones, such as human judgments (Condorcet, 1785; Laplace, 1814; Poisson, 1837), while these critics advocate an entirely objectivist approach for the definition of probabilities (Cournot, 1843; Ellis, 1849; Boole, 1854; Venn, 1866).

Cournot (1843) was one of the first to insist on the distinction between objective probabilities and subjective probabilities. He said:

> *When the number of trials is small, the formulas usually given for the assessment of* **posterior** *probabilities become illusory: they only indicate subjective probabilities...*

Similarly, for Venn (1866), the main error of some past and not least authors on probability, among whom we find for example Laplace is an example (page 83), is to have applied the theory of probabilities to events for which it was not applicable. For him, the concept of the *series* is more fundamental for deciding whether the theory of probability is applicable or not to an event and such a theory has no meaning unless it is linked to that concept. It then becomes necessary to define precisely what is meant by the term *series*. The demographic example that he used allows this term to be clarified.

Consider the following sentence: *some children will not live to thirty*. If this sentence is regarded as a logical proposition, the concept of series is quite foreign to it. However if it is a proposal that can take a numeric character, replacing the term *some* by a given proportion, then it is difficult not to speak of a series. This does not mean however that, if we observe a number of children, we will observe before thirty years this exact proportion of deaths, but only that, if there is a growing number of children, the proportion of observed deaths will tend towards this limit. The underlying assumption is then that this probability, although not calculable *a priori* as in the case of games, exists and remains the same throughout time for the event studied.

For the objectivist school, it is possible to give an objective status to the concept of probability, if one is confined to observing events that may occur during repeated trials. As von Mises (1939), one of the most committed representatives of objective probabilities, could write:

> *We can say nothing about the probability of death of an individual even if we know his condition of life and health in detail. The phrase 'probability of death', when it refers to a single person, has no meaning at all for us.*

We see how his approach to probability is different from that of Bernoulli, and especially that of Bayes and Laplace, who were seeking instead to clarify its estimation from the observation of a number of individuals similar to the one concerned. It is possible in this case to speak of the probability of dying in a population whose size is as large as one want. It does not mean however that a human population can be considered as infinite. Also to talk about the probability of a unique event in nature or more generally about the probability for a proposition to be true, has no meaning for an objectivist. The event must be part of a series (a *collective* in the sense of von Mises), which is one element among a multitude of others.

They therefore reject the use of Bayes' formula when it uses, for example, a uniform prior probability, in order to estimate the probability of the event studied from observations. This is actually a hypothesis, the hypothesis of a uniform distribution, they reject as having absolutely no sense, and the concept of probability as describing a state of our knowledge. For them the word probability means only the frequency of an event in a given experiment.

This rejection lasted almost a century, but with some researchers who still maintained the position of Laplace (Pearson, 1920, 1925) without using it in specific applications.

At the same time demography did no more use as fine an analysis as that of Laplace, for its studied phenomena. It must be understood that the implementation of population censuses has sidelined some of the earlier concerns, providing in particular exhaustive populations at risk, which avoided the use of data from vital statistics to estimate these populations by the multiplier method, which we saw was already used by Graunt.

Thus the population censuses, which appeared in the eighteenth century and were put in place throughout the nineteenth century in Europe, coupled with comprehensive vital statistics, would change the use of probabilities. By collecting data on the entire population at a given moment the demographer can work with an objectivist approach, since these numbers are very large. He does not even calculate the variance of the rates, as it is very low. The cross-sectional approach, with the method of concomitant variations (Durkheim, 1895; Landry, 1949), as well as the cohort analysis (Pressat 1966, Henry, 1972), never proposed this calculation.

To show why, consider the example of the actual generation of French men reaching 60 years in 1962 (Pressat 1966): given the number of deaths between 60 and 61 years, $D(60,61)$, and the population 60 years old, $N(60)$, the annual probability of death is estimated at $q_{60} = \frac{D(60,61)}{N(60)} = \frac{6210}{265344} = 0.0234$, or 23.4 per thousand, and the author does not even calculate its variance which can be estimated (Smith, 1992) as equal to:

$$Var(q_{60}) = \frac{[N(60) - D(60,61)]D(60,61)}{N^3(60)} = 0.000000085,$$

or 8.5 in 100 million, assuming a binomial distribution of deaths, that is to say issued from a homogeneous population. This variance is in fact so weak that it no longer has any interest. Only rarely, when calculating the probabilities over shorter periods (e.g. monthly) the demographers should consider these variances (Hoem, 1983) because the numbers, even exhaustive, who suffer the event will be much smaller. This perfectly explains why classical demography, although keeping the notion of probability, have left aside any measure of dispersion in the first sense. Note here however that this assumption of a population in which the probability of dying at a given age is the same for all its members, is totally unrealistic as we will see in the next section.

We will now show that the dispersion in the second sense of this term has also virtually disappeared from the longitudinal demographic approach. Indeed, under this approach, only an analysis of demographic phenomena considered independently from each other and appearing in a homogeneous population is actually possible (Blayo, 1995). This causes serious difficulties, and even a complete inability to take into account the dispersion of these phenomena, both in their mutual interaction and in heterogeneous populations. We refer the reader to Courgeau (2003, 2004, 2007) for a critique of this approach. This results in an inability to take into account the dispersion in the second sense, the one introduced by other demographic phenomena that we studied, as well as the other introduced by the diversity of the members of this population.

We can conclude that during this period from the middle of the nineteenth century to the early 1980s, demography has almost completely left aside the two aspects of the dispersion phenomena it studies, keeping only the differences between rates, according to age.

5 Reappearance of the dispersion in the biographical and multilevel approaches

In response to the previous criticisms, the biographical approach will consider a set of individual trajectories in all their complexity, usually observed by detailed surveys. The unit of analysis is no longer the event, as in the longitudinal analysis, but the biography, seen as a complex stochastic process. This approach will no longer consider the various events studied as independent, but rather will analyze the dependencies between them. Similarly the population will no longer be considered as a homogeneous one, but the heterogeneity existing in it will be studied. This removes most of the criticisms of the longitudinal analysis. We refer the reader to Courgeau and Lelièvre (1996) for a further critical presentation of this approach.

Regarding the dispersion in the first sense, it is essential to consider the variance of a quotient in order to conclude an interaction between phenomena or a dependency between a phenomenon and various features taken into account. Furthermore, while this event history analysis initially used an essentially objectivist statistical approach (Kalbfleisch and Prentice, 1980, Cox and Oakes, 1984; Courgeau and Lelièvre, 1989, 1992 2001, Andersen et al., 1993), a Bayesian approach has recently helped to solve many problems of estimation and corresponds better to the spirit in which this analysis is performed (Ibrahim *et al*., 2001).

Thus, the Bayesian approach allows in particular incorporating any prior information useful to the research question, what the objectivist approach does not allow. Also these methods permit now, through Gibbs sampling and Monte Carlo Markov Chains (MCMC) (Robert, 2006), to solve much easily complex problems without resorting to asymptotic objectivist calculations. It also has many other advantages over the objective approach, thanks to the availability and flexibility of tools for modelling and data analysis.

Regarding the dispersion in the second sense of the term, this approach also takes it into account, by introducing the estimation of the heterogeneity of a population and dependence between the phenomena studied. The reasons both internal (dependency between events) and external to demography (heterogeneous population), can thus be identified and their effects on individual behaviour can be analyzed with great detail. One risk is committing, in this case, the so-called *atomistic fallacy*, because by introducing only the characteristics of the individual, it ignores the context in which human behaviour occurs. This risk precludes the risk of *ecological fallacy* in the transversal approach, which had already been shown by sociologists (Robinson, 1950), and which might attribute to the individual more collective reasons, related to the groups used to perform this analysis.

To avoid these risks of error, the contextual and multilevel approach will have to explain individual behaviour by simultaneously introducing various groupings of individuals. Thus the contextual approach can associate the behaviour of an individual both to his own characteristics (individual measure) and to characteristics of the groups he belongs to (aggregate measure). The multilevel approach can go further by introducing a dependency internal to the various groups, simultaneously at the individual and contextual level. These approaches can thus escape from both the ecological fallacy, because the aggregate characteristics are no longer considered a substitute for individual characteristics, and the atomistic fallacy as long as it involves the proper environment in which people live (Courgeau, 2003, 2004a, 2007a).

Naturally, the Bayesian approach will enable an even more satisfactory multilevel analysis (Goldstein, 2003; Courgeau 2007b, Draper, 2008), as in the case of event history analysis. Especially when the number of units in one level of aggregation is low, estimated asymptotic standard deviations by the maximum likelihood method can be severely biased, and may even be negative (Draper, 2008). The situation becomes even more difficult in the case of binary variables or more generally discrete ones: in some cases the methods of objective probability can not even conduct an evaluation of model parameters. The use of Bayesian methods is in this case necessary.

Thus for nearly thirty years demography, which had in the past completely ignored the dispersion of its measures, returned to the view that the population is heterogeneous and that the phenomena are interdependent, which allows a more elaborated analysis of dispersion to be introduced.

More recently a Bayesian view has emerged in other demographic topics, such as paleodemography. Indeed the observation of age indicators on a small number of skeletons is the only way to estimate the age structure of archaeological populations. The objectivist methods often lead to solutions outside the domain [0,1] of validity for such probabilities by age. Therefore, a fully Bayesian solution is needed and allows an optimal solution of the problem (Caussinus and Courgeau, 2010).

6 Conclusion

Examination of the dispersion of demographic measures showed us a continuous movement back and forth between homogeneity and dispersion which, however, was made between different units, depending on the period. Indeed, as in any scientific discipline, it is not the full complexity of the phenomena implicating the individuals who compose a population, which are the object of demography, but some of their aspects, which may become more complex in the course of time, but which are always characterized by a small number of parameters considered indispensable for understanding demographic phenomena that affect the population.

Initially, we considered only the probabilities for the entire population and the possibility of their dispersion in the second sense of this term, according to ages was tested: the conclusion was that it is necessary to consider different probabilities for each age. Also the dispersion in the first sense of this term was tested for some indices: for the sex ratio at birth, lack of dispersal was observed during 80 years in London. However, few examples of its use, particularly among political arithmetic authors of the eighteenth century, can be found.

Laplace continues to observe a population as a whole, but this time from a Bayesian perspective. He assumes a prior distribution of probability uniformly distributed on the interval [0,1], to obtain posterior probabilities, which permit an accurate estimation of the dispersion in the first sense. He also indicates the importance of tackling the dispersion in the second sense, involving cases that would play on some sub-populations and not on others (climate, food and manners). However, the regression methods proposed by Gauss were hardly used in population studies at his time.

The dissemination of Censuses during the nineteenth century and a critique of the bases of calculation led to a rejection of the Bayesian approach of Laplace. The cross-sectional and cohort analyses used until the early 1980s have left aside any measure of dispersion of demographic rates and probabilities, in both senses of the term.

At that time the re-emergence of dispersion occurs in demography, with the event history and the multilevel approaches, where consideration of individual characteristics and aggregate ones splits the analytical framework. Dispersion is again considered in the first sense of this term, as it becomes necessary to estimate the variance of the estimated effects to assess their validity. Dispersion occurs in the second sense of the term, because units of different levels are simultaneously introduced in the analysis and because of the individual characteristics that affect different sub-populations each time.

One last point remains to be solved. Indeed, throughout this article we have contrasted the views of those who apply the probability on individual cases, and the views of those who entirely reject this possibility: to simplify, on one side Jacques Bernoulli, when he speaks on the probability of survival of Titus, on the other side von Mises, when he says that the probability of death has no meaning for him since it refers to a single individual. This is in fact the distinction between subjective probability and objective probability, applied to population data.

Thus de Finetti (1937), one of the most important representatives of the subjective approach, clearly indicates that:

> *the degree of probability attributed by an individual in a given event is given by the conditions under which he would be willing to bet on this event.*

He says much further than *an event for him is always a singular fact*. Conversely von Mises, one of the most important representatives of the objective approach refused to talk about the probability of a singular fact, which for him does not exist. It is important to see how demography is vis-à-vis these two extremes?

For classical demography, an objective probability seems perfectly appropriate, under the assumption that the observed population can be considered as a sample from a theoretical infinite population having the same probability of being subject to the various demographic events. The variances of the estimated probabilities in this case are sufficiently low, as we have shown, to permit such a use.

But when we turn to event history or multilevel approaches, often using data from not exhaustive surveys, maintaining an objective probability approach, although still possible and used, can be questioned. A subjective probability seems better able to incorporate any prior information relevant to the phenomena studied, which the objective probability does not allow. The extreme dispersion of probabilities, in the second sense of the term, according to individual characteristics and to interactions between the phenomena studied often makes such use necessary. However the individual predictions that this analysis allows are only very approximate, because such a person has many more features than those considered in the analysis, something that may significantly change the prediction (Courgeau, 2007b).

Despite this last point the event history and multilevel approaches, introducing the dispersion in both senses of the term in demography, have permitted considerable progress in order to understand human behaviour.

Acknowledgment

I wish to thank Marc Barbut and Henry Caussinus for their comments and advice on an earlier version of this paper, which helped me to improve it. I want also to thank Emily Tanimura for a very careful and efficient rewriting of my first translation. I still retain obviously full responsibility for my words.

References

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.