To cite this paper, please use:

Seneta, E. (1999). Statistical Notes On The First American Census in Methodological Perspective. Anderson, M. and Fienberg (Eds) In "*The History of the First American Census and the Constitutional Language on Censustaking: Report of a Workshop*", pp 27–34.

C Statistical Notes On The First American Census in Methodological Perspective

by
Eugene Seneta[2]
School of Mathematics and Statistics
University of Sydney, NSW 2006, Australia.
E.Seneta@maths.usyd.edu.au

C.1 Questions Addressed

The U.S. Constitution, was drafted in 1787 in Philadelphia and ratified in 1788, with input from **Thomas JEFFERSON (1743-1826)**, who was U.S. Minister to France, 6th August 1784 - 18 September 1789, and a polymath in his interests, with some statistical work already present in his *Notes on the State of Virginia* (1782). He took up duties in New York as Secretary of State on March 21, 1790 and was in this capacity the official in charge of the first American Census of 1790. The U.S. Constitution determines that Congress is responsible for Census enumerations. In the review article "Statistical work of the Federal Government of the United States" by John Cummings, Statistician, United States Bureau of Census [3] Jefferson is not mentioned, although the first census is discussed (pp. 670-671).

The questions addressed in this appendix after Section /refsec:appcsec2 below are:

- What might Jefferson have known of the methodology of statistical estimations, and of enumeration, of populations at the time of ratification of the U.S. Constitution and of the first American Census?
- What was the understanding and application of sampling ideas to populations in France and the Russian Empire?

C.2 Use of Sampling in the U.S. Census in the Year 2000

The U.S. Constitution determines that Congress is responsible for Census enumeration, hence the legal challenges on the basis of U.S. Federal Law to the introduction

[2]Acknowledgements: I thank for advice at short notice: Steve Stigler, Susan Linacre, Bernard Bru, Pierre Crépel, Ian Castles, Jane Barton, Fred Hoppe and Ludmilla Seneta; and Viola Chao for word-processing.

[3]In J. Koren (ed.)(1918) *History of Statistics*, American Statistical Association, pp. 573-689.

of statistical estimation procedures, and the appeal hearing in the U.S. Supreme Court on November 30th, 1998. In Australia regular PES (post-enumeration surveys) using sampling have long been held to determine for each State its share of the total population and hence its share of House of Representatives seats at Federal level. (Elections occur at greater frequency than Censuses.) In contrast to the U.S., the Australian Constitution (of 1901, the time of Federation of separate Colonies) largely influenced by that of the U.S., vests responsibility for this in the Commonwealth Statistician, who heads the Australian Bureau of Statistics. The sampling procedures have not been controversial, or challenged by the States.

C.3 The Statistical Background for the U.S. Constitution and First Census

When writing his *Notes on the State of Virginia* before coming to Paris, Jefferson used the *Histoire naturelle, générale et particuliére* of **George Louis Leclerc BUFFON (1707-1788)**, who asserted that the certainty of a physical truth is to be measured by the probability of the corresponding facts, and in whose *Supplément à l'Histoire naturelle, Vol.IV* the "Essai d'arithmétique morale" of 1777 is concerned with "the measurement of uncertain things". Buffon had already fallen back on statistical methods in his paper "De la vieillesse et de la mort" in Vol. II, 1749, of the *Histoire naturelle* in which he studied human mortality in general, using Dupré de St. Maur's mortality tables. His statistical interests were in life tables rather than population enumeration. Jefferson used Buffon's life table in his letter from Paris of Sept. 6, 1789, to Madison.

Jefferson in his *Notes* took issue with Buffon's conclusions as a naturalist, particularly in reference to the United States, and their later scientific discussions in Paris were on natural history. I have found no evidence that Buffon's probabilistic thinking influenced Jefferson on population enumeration issues. In his own catalogue of his library, the Buffon works are under Chapter 12: Natural-History-Animals.

Chapter VIII ("The number of its inhabitants") of Jefferson's *Notes* of 1782 has two distinct statistical themes. One is projection of population size on the basis of the estimate that the population of Virginia is doubling every 27.25 years, an estimate, as he says, "with a considerable degree of precision". This idea has the spirit of **Thomas Robert MALTHUS (1766-1834)** about it; and indeed Jefferson eventually had Malthus' book (in 2 volumes) in his library, but that was first published in 1798, and a 2-volume edition not till 1806. The second theme is how *to estimate* the total number of inhabitants of Virginia in 1782 (he arrives at a final figure of 567,614), from an enumeration of some kind giving 53,289 free males above the

age of 21, 211,698 slaves of all ages and sexes, and 23,766 "said to be tytheable slaves" - that is, above the age of 16; with "no returns from eight counties". He makes rather crude assumptions on the basis, for example, of "a former experiment" to estimate his final figure, from what one might consider to be a (highly non-random) *sample* of size 288,753, which is just 51% of his final result. [4]

Jefferson's contacts from 1785 with **Marie-Jean-Antoine-Nicholas Caritat de CONDORCET (1743-1794)** were frequent as the two men had similar intellectual interests. Condorcet's (1785) *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* appeared in Paris during Jefferson's appointment, and a copy may have been presented to him there. He possessed a copy on his return to the U.S. Parts of the book are on voting systems and public choice, are now recognized as being pioneering in these areas, but appear to have been beyond the interest or understanding of Jefferson or Madison. There was fruitful exchange of ideas of political and technical kind, but Condorcet's writings on probability and public choice were not among them.

It is worth mentioning that **John ADAMS (1735-1826)** was from 1785 first US Minister to Great Britain and was also writing back to the Constitutional Convention, was on splendid terms with **Richard PRICE (1723-1791)** who was the leading British authority on mortality tables and pensions, as well as a civil libertarian. Adams however was not inclined to quantitative matters; and regarded Condorcet as a "mathematical charlatan".

I have found no evidence that Jefferson during his sojourn in Paris, met with, or knew of the work on the probabilistic estimation of population size by **Pierre-Simon de LAPLACE (1749-1827)**. Condorcet and Laplace detested each other. But Jefferson may have visited **Jean-Baptiste-François de la MICHODIÈRE (1720-1797)** who was, prior to the French Revolution, a leading authority (writing under the pseudonym Messance) on the estimation of population size (though not by probabilistic methods) and knew Laplace.

The work of la Michodière and Laplace is most relevant in addressing the Questions above; it is described by Bernard Bru. [5]

A number of other potential, but not probabilistic, statistical influences on Jefferson in regard to populations include **Pierre-Samuel DUPONT DE NEMOURS (1739-1817)**. A *Comité de l'Agriculture* created in 1782 ordered Dupont (as was then his name) to present to it a summary of the value of the crops of the kingdom. Another member of the then *Comité*, the celebrated chemist **Antoine Lau-**

[4] The number of free males under sixteen should read as 71,052. In at least some printings it reads as 17,052.

[5] Bru, B. Estimations Laplaciennes. Un Exemple: La Recherche de la Population d'un Grand Empire. 1785-1812. *Statistique et Analyse des Données*. 1988. Vol. 13, No.2, pp.3-42.

rent LAVOISIER (1743-1794) prepared a large work by 1784 building on this, which after the Revolution the Constituent Assembly ordered printed in 1791 [6] which alluded to methods of determining the population size. On the basis of this the Constituent Assembly resolved to proceed to a complete enumeration of the French population. (However no census was held till 1801, after the establishment of a Central Statistical Office under another régime.) Dupont was on a *Comité du Commerce* during Jefferson's time in Paris as now *inspecteur général du commerce* and Councillor of State. The task was in part was to study problems of Franco-American trade. Dupont de Nemours emigrated to the U.S. in 1799, returned to France in 1802 helping Jefferson with the Louisiana purchase, fled to the U.S. in 1815, and died there. His economic theories had some influence on U.S. policies. The now multinational chemical company bears his name. His close colleagues Condorcet and Lavoisier died in the Terror.

According to several authorities, Jefferson's expression in mathematics went little beyond arithmetic calculations. His own catalogue [7] accompanying the sale of his books to the Library of Congress reveals an enormous breadth, even for a polymath; we focus on books relevant to the preceding discussion, as he described them: Under "Mathematics. Pure. Arithmetic": De Moivre's Doctrine of Chances; L'analyse de la Probabilité des décisions, par Condorcet; Mémoires Mathématiques de Diderot; Price on Annuities. Under "Political Economy - General - Statistics - Commerce - - - Finance": Petty's Political Arithmetic; Blodget, Statistical Manual for the U.S. of America; Statistique élémentaire de la France, par Peuchet, 1805 [Peuchet was an authoritative and overt critic of the first French Census of 1801 - E.S.]; Malthus on the principles of population, 2v. The only work of Laplace in Jefferson's catalogue is under "Astronomy": Exposition du Système du Monde, par la Place, 2v. The other works of Condorcet, and of Dupont, tend to be linked and there are many, possibly all: for example, under "Modern History. Foreign" and under "Special Governments, Modern". Under "Polygraphical" there is: Encyclopédie de Diderot et D'Alembert, 39v. Lausanne.

It can be said with a high degree of certainty that the Founding Fathers of the U.S. Constitution knew nothing of sampling as such, and could not have rejected its use. There is a once much-quoted opinion [8] of **Moreau de Jonnes** " ... that American founders looked only to practical ends ... a careful search through the 'Madison papers' has failed to reveal that any member of the Convention considered the census in its scientific bearings".

[6] *Résultats extraits d'un ouvrage intitulé: de la Richesse territoriale du royaume de France.*
[7] Gilreath, J. and Wilson, D.L. (Eds.) (1989) *Thomas Jefferson's Library*: L.C., Washington.
[8] Cummings, p.670.

Further, from the forgoing discussion, one could presume that Jefferson, with his scientific and inventive turn of mind, had he known of and understood la Michodière's and Laplace's work in the U.S. context, would not have been averse to the cautious use of probabilistic sampling and estimation. His *Notes* of 1782 already support this view.

C.4 France: Sampling

Lavoisier (1791) refers to the books of la Michodière (Messance) and de Montyon (Moheau) regarding the number of inhabitants, and estimates without details the population of France for about 1784 at about 25 million. These two were the most remarkable of the 32 *intendants* of the provinces (*généralités*) towards the end of the *ancien régime*. The *intendant* was the supreme administrative head of each province. Their archives form principal sources for statistical studies of pre-revolutionary France. Only a few provinces attempted head-by-head enumeration, twelve used a census of households, some borrowed figures from tax registers. Enumeration was regarded so difficult and expensive as being unreasonable; and an approximation to population size of France was obtained by synthesizing figures from the provinces. By *mid 18th century* it was agreed to adopt as an aid "the least uncertain index", the number of births, since registers of these were carefully kept.

The procedure was as follows for a province. A number of parishes was "carefully selected", and in these a complete enumeration of persons made. This number was divided by the mean number of births over the six preceding years in the parishes. Because of the variability in this ratio (it fluctuated wildly between provinces), an "average figure" was obtained to be used in every province, which was then multiplied by the total number of births in the province to give an estimate of total population size in the province. Thus in modern terms the essence of this was to calculate, on the basis of a *sample* (hardly random) of parishes, a *ratio estimator* of the population size. For the whole of France, these province totals could be added; but, it seems, de Montyon, Necker, and des Pommelles multiplied the total number of births in France (an average over a number of years) by a different ratio to get the ratio estimate (one thinks in modern terms of the "combined" and "separate" *ratio estimators from a stratified sample*). The estimates obtained were 24,000,000 (1774); 24,802,500 (1781); 25,065,883 (1784).

Laplace's involvement in the question of determination of population size appears to begin with a paper of 1786 in which he uses la Michodière's ratio estimate of 26 for the population size to the number of births in France and multiplies it by the average yearly number of births for 1781-1782, 973,054.5 to obtain 25,299,417. The innovation is that the ratio estimator is complemented by *proba-*

bility limits on its accuracy, and his primary purpose is the theoretical derivation of these. He determines the sample size of total number of persons required (771,469) to give an absolute deviation of at most a given size (500,000) with high probability (1000/1001). The prediction methodology is, as usual for the times, Bayes' Theorem with uniform prior; and the model used presupposes randomness in the sampling.

It is unlikely that even Laplace's mathematical colleagues understood at the time the significance of what he was trying to do; nor Jefferson, already in Paris, who might just have heard of it, as of the earlier ratio-estimation-sampling work, from la Michodière if he had met him.

The Bastille fell on July 14, 1789 while Jefferson was still in Paris. Laplace adjusted to the several changes of régime, and for a short period was Minister of Interior, thus, one supposes, with an intimate knowledge of demographic statistics of his time, before being replaced as Minister by Lucien Bonaparte who ordered the first census of the French population in 1801. This met with difficulties and the results were not in for some two years. In the meantime, Laplace persuaded the Government to carry out a *sample survey* to estimate the population size. This was carried out on September 22, 1802; about 2,000,000 persons were involved in the sample. The number of births used was for the 3 years preceding, September 22, 1799 to September 22, 1802. The results were reported, rather casually, in the editions 1812-1820 of Laplace's *Théorie analytique des probabilités*. Another unreliable Census had been held in 1806; then not repeated till 1821. If anything, Laplace's work had cast *doubts on the accuracy of a complete enumeration*. Laplace's own sampling procedure had the practical form of *a two-stage cluster sample*, but the random (representative sample) character was lacking: thirty *départements* were chosen and in each of these a number of places where the mayor was intelligent and zealous.

By this time the first American Census of 1790 and its successors were well and truly over. Jefferson resigned as Secretary of State on December 31, 1793, was then Vice-President (1797-1801), and in retirement in the interim.

In fact the use of sampling to provide information of demographic and social kind went into severe decline as the method of total enumeration gained ground. **Georg von MAYR (1841-1925)** and his school objected to *indirect methods* such as sampling, and it began to gain ground only in the early years of the 20th century.

C.5 The Russian Empire: Sampling

Relevant Census and sampling activity here began considerably later than in France or the U.S.[9]

The first and only general enumeration, prior to Soviet times, of population, according to the authoritative **A.A. KAUFMANN** [10] **(1864-1919)**, took place in 1897. Before this time Russian official statistics had to rely on means such family registers kept for the purpose of recording persons eligible for military service. Administrative and police estimates "were made through the aid of local statistical offices, ... , and the district community administration". The data was "compared with the data in the Central Office [Central Statistical Office, est.1858, within the Ministry of Interior] and published after having been carefully tested". Plans for a complete enumeration of population date to the 1860's. There was no use of sampling methodology in this context.

There were, however, studies dating from 1877 of the agricultural and the rural economy coordinated by the Central Statistical Office. A Russian Agricultural Census took place in 1916. Some 18,000,000 landholding cards were obtained. As part of the investigation of variability within the rural economy it was decided to *sample* these cards. Only the first (methodological) part of the investigation was published (in 1917) due to the Russian Revolution. Its author **S.S. KOHN (1888-1933)**, Assistant Director, Russian Agricultural Census, ascribes the formula for *variance of sample mean for a simple random sample without replacement* to his former teacher at the St. Petersburg Polytechnic Institute, **A.A. CHUPROV** (or **TSCHUPROW) (1874-1926)**.

Sampling had been used earlier in statistical investigations of the rural economy in Russia within institutions of local government called *zemstvos*, established from 1864. (There as an analogy with the provinces of the *ancien régime* in France of a century before.) However the measure-of-precision approach akin to Laplace's in France, but now of a truly random sample without replacement, is credited to S.S. Kohn and A.A. Chuprov.

Within the zemstvo investigations, the idea developed that it was not necessary to put "certain of the questions contained in the community schedule in every instance... it would be ordered that every tenth or twentieth person in alphabetical or

[9] We asked Professor Seneta to describe this particular component of the development of scientific sampling primarily to illustrate how remote the developments actually were from the events surrounding the writing of the U.S. Constitution and the taking of the the first U.S. Census. Other notable contributtors to the development of sampling at the end of the 19th and beginning of the 20th centuries include: Kiaer (Norway), Bowley (England), Neyman (England and Poland). *The Editors.*

[10] In Koren (ed.)(1918), pp. 469-534; *Novyi Entsiklopedicheskii Slovar* **18**, [1911-1916], cols. 617-632.

other mechanical order should be questioned". In 1896 **A.V. PESHEKHONOV (PJESCHECHONOW) (1867-1933)** took a mechanical (that is, as we now say, *systematic*) sample of every 10th landholding, and **V.G. GROMAN** (or **GROHMANN)** used systematic sampling extensively in 1911-1913. It should be kept in mind that this method amounts to the *random selection of a cluster*.

The above account alludes to the use of the systematic sample and purely random sample as independent methods of statistical investigation, dictated by necessity in the presence of vast data sets. The more general tendency in Russia as elsewhere was "monographic" sampling. This pervaded early attempts at representative sampling. The "monographic method" amounted to selecting a cluster of elements "typical" of a population in relevant respects, and their extensive investigation. The method was generally used as an adjunct to a foregoing (local) complete enumeration, which might be used to determine the "typicality" of the elements to be sampled. Complete enumeration would be the only method used to check the precision of estimates obtained from such monographic samples. **A.I. CHUPROV (1842-1908)** proposed the idea in 1894 of *monographic sampling as a useful independent* means of investigation (not merely as an adjunct to complete enumeration) but the idea of a *random sample* did not begin to take root till a paper of his son A.A. Chuprov in 1910, and was extinguished by the aftermath of the Russian Revolution.

Kaufmann's article (of 1918) makes comparison with the American system (p.528: "The prevailing method ... in collecting zemstvo statistics is modelled on the American correspondent system") whereby a village representative fills in a schedule. (p.531: A very special technique of enumeration has been devised which at all points departs from the west European forms and in many respects approach the American). There was therefore an awareness of what was happening in the U.S., and, pre-Revolution, good progress in sampling methodology in the Russian Empire. It had no influence outside of the Russian Empire, however, and was implemented in just a few zemstvos. Although A.I. Chuprov tried to popularize among zemstvo statisticians the work on sample surveys of Kiaer, the influence of this work in Russia was minimal.

The idea of using some kind of sampling to supplement complete enumeration, however, was significant in the late 1800's.