Estimating Age without Measuring it: A New Method in Paleodemography

Henri Caussinus, **Daniel Courgeau**, **Translated by Jonathan Mandelbaum**

In **Population Volume 65, Issue 1, 2010**, pages 117 to 144

ISSN 0032-4663
ISBN 9782733231012

Henri CAUSSINUS* and Daniel COURGEAU**

Estimating Age without Measuring it: A New Method in Paleodemography

At what age did humans die in the past? To estimate age at death in ancient populations for which no civil records exist, skeletons are often the only information source. Bones and teeth provide indications of the stage of growth or ageing reached by individuals at the moment of their death, but they cannot be used to estimate age with certainty. Skeletons of individuals whose age at death is known, and for whom these biological parameters have been measured, provide a reference population for constructing statistical models to estimate the age distribution at death of persons whose skeletons have been uncovered. But these individuals are not necessarily representative of the general population to which they belonged: skeletal conservation depends on burial conditions and the circumstances of death. One way to resolve this problem is to work on homogeneous groups, such as the convent of nuns whose data are presented here as an illustrative example. How can statistical inferences be drawn from these data? Henri CAUSSINUS *and* Daniel COURGEAU *first describe the methods generally used in paleodemography and demonstrate their limits. They then propose a new method, based on the principle of Bayesian inference, and compare it with standard methods to demonstrate its greater accuracy and flexibility.*

Age, a fundamental concept in demography, cannot be directly measured for most past populations as they did not keep vital statistics. All we can do is to estimate age from biological growth indicators for immature individuals, or ageing for adults, measured on a small number of skeletons belonging to a given population, with the aid of bone or dental remains. Unfortunately, these indicators can give us only a broad range for an individual's age at death because

---

* Institut de mathématiques de Toulouse, UMR CNRS 5219, Université de Toulouse, UPS, 31062 Toulouse, France.
** Institut national d'études démographiques, Paris.

Correspondence: Daniel Courgeau, 114-2 Chemin du Défends, 06250 Mougins,
e-mail: daniel.courgeau@wanadoo.fr

Population-E, 65 (1), 2010, 117-144

there is no precise relationship between age and bone condition, but only a rather weak correlation.

To advance the study of that correlation, paleodemographers have long been using what are known as "reference" data (Masset, 1971), obtained on sites where it has been possible to determine both the age at death and biological indicator(s) for each individual. These data are described in greater detail in Section I, as are data from the "target" site, where only data on indicators are available. The statistical problem consists in estimating the distribution of ages at death on the target site from data observed there and from reference data. Various methods have been proposed for the purpose over the years. We recall the main developments in Section II, focusing on the most common approach: the discrete case where bone characteristics and ages are distributed into classes. We consider the case where only one biological indicator is observed, but the generalization to several indicators is straightforward. The first hypothesis centres on the conditional probabilities of the ages associated with each indicator value. It is less satisfactory than the second, called "invariance hypothesis", which assumes that the conditional distribution of indicators, at a given age, is constant over time in the periods concerned, at least for a first approximation.(1)

Although the performance of methods proposed earlier has gradually improved, it remains disappointing. The results are often visibly aberrant with respect to prior knowledge and plain common sense. In fact, they tend to be general methods, most of which fail to address all random aspects of the data or the specific characteristics of the problem (we shall see, for example, that a widely used method is borrowed from ichthyology, with a basic model that is formally identical but remains far too general). Given that we are dealing with samples that are often small for an estimation problem marked by intrinsically high instability, it is important to set up a methodology that best incorporates the sum of prior paleodemographic knowledge. The most logical means to this end is a Bayesian method – as suggested in Section III.

The method we propose is Bayesian as commonly understood in statistics. By contrast, certain earlier methods are improperly referred to as Bayesian, for the sole reason that – at some point or other – they make use of Bayes' famous formula. Recall that, unlike a frequentist method in which the unknown parameters are assumed to be fixed, a Bayesian method treats these parameters as random variables. We choose a prior distribution, i.e. before observation, for the parameters, and we determine the posterior distribution, i.e. the

---

(1) This hypothesis has been hotly debated by paleodemographers for some fifteen years now: "While the possibility of a secular drift of biological age indicators cannot be ruled out, paleodemographers have tended to neglect it, given their inability to measure it, while hoping that any possible divergences will not be too significant" (Séguy and Buchet, 2010). To guard against this, paleodemographers use pre-industrial reference populations that had not begun their demographic transition, or had done so on a modest scale.

distribution revised for a given target site on the basis of data observed at the site. In the prior distribution, we can introduce information preceding the actual data, notably the fact that the probabilities to be estimated reflect a mortality distribution (in some cases, in a specific known environment). We shall also see that, in our problem, it is logical to assume that the reference data supply a prior distribution for certain parameters that are most often deemed to be known – quite unjustifiably, since the data are affected by sampling errors (not to speak of the necessarily approximate nature of the invariance hypothesis).

As shown in Section IV, our method compares very favourably with earlier ones. In Section V, its application is illustrated with two examples.

I. Data, mathematical formulation, notation

We have a reference population, in which we distinguish *c* age groups and *l* stages for the bone index measured, and an observed population, in which we distinguish the same *l* stages. Table 1 shows the data: number of individuals n_{ij} drawn from the reference population, by age class *j*, and stages *i*, and number of individuals m_i observed for stage *i* on a given site (target population).

Table 1. Matrix of reference population by stage and age class, and observed population by stage

	Reference population										Observed population
	Age classes (j)									Totals by stage	Totals by stage
Stages (i)	n_{11}	.	.	.	n_{1j}	.	.	.	n_{1c}	$n_{1.}$	m_1
	.				.				.	.	
	.				.				.	.	
	.				.				.	.	
	n_{i1}	.	.	.	n_{ij}	.	.	.	n_{ic}	$n_{i.}$	m_i
	.				.				.	.	
	.				.				.	.	
	.				.				.	.	
	n_{l1}	.	.	.	n_{lj}	.	.	.	n_{lc}	$n_{l.}$	m_l
Totals by age	$n_{.1}$	.	.	.	$n_{.j}$	.	.	.	$n_{.c}$	$n_{..}$	m

The left-hand side of Table 1 allows us to compute the frequency of age *j* in the reference population, knowing stage *i*: $f_{j|i} = \frac{n_{ij}}{n_{i.}}$, and the frequency of stage *i*, knowing the age *j*: $f_{i|j} = \frac{n_{ij}}{n_{.j}}$. We can also calculate the frequency by age of the reference population, $f_j = \frac{n_{.j}}{n_{..}}$ and the frequency by stage of the reference population $f_i = \frac{n_{i.}}{n_{..}}$.

The right-hand side of Table 1 gives the frequencies by stage of the observed population: $\frac{m_i}{m}$.

These measured frequencies are associated with various unknown probabilities. We shall write p_{ij} the probability that a random individual in the target population studied belongs to stage i and age class j. The sum on i of the p_{ij} values will be written $p_{.j}$ or simply p_j (the probability that an individual is of age j). The sum on j of the p_{ij} values will be written $p_{i.}$ or simply π_i (the probability that an individual is in stage i). The conditional probability of stage i, knowing age j, will be noted $p_{i|j}$. These probabilities are positive and satisfy the equations $\sum_i \pi_i = \sum_j p_j = 1$, and $\sum_i p_{i|j} = 1$ for all j. They are also linked by the following relationship:

$$\sum_j p_j \; p_{i|j} = \pi_i \qquad \text{for all } i = 1,\ldots,l \qquad (1)$$

Under the assumptions made, several solutions are possible for estimating the probabilities p_j, i.e. for obtaining the age structure of the paleodemographic population observed. Let us examine the main estimates proposed in the past.

II. Methods currently used

We shall not describe certain methods that are less useful today, for which Masset (1973) provides an excellent critical discussion. The only approaches examined here are those currently used by most paleodemographers, as well as their most recent extensions.(2)

1. Tables of minimum distance between each cell

The first method consists in estimating the cells of a matrix of the observed population, each of whose cells is as close as possible to its counterpart in the reference population matrix; the stage frequencies are those of the observed population. The age frequencies obtained will provide the solution to the problem set. Under this option, we need to define a total distance between the

---

(2) We shall not discuss situations where several biological indicators are available, or where the indicators are continuous rather than discrete (Konigsberg and Frankenberg, 1992). Likewise, we shall not examine the use of viability theory to address this problem (Bonneuil, 2005). Viability theory regards the age distribution of deaths, imperfectly observed, as a target in a large space; one looks for the solution that is closest to a stable population (as well as to its confidence interval) and that produces this age distribution of deaths at the end of the observation period.

cells of the two matrixes. The most commonly used distance is a χ^2 distance that divides the squares of each difference by the number of individuals observed in the reference population.(3)

It is easy to show that, in this case, the solution is written:

$$\hat{p}_j = \sum_{i=1}^{l} \frac{m_i}{m} f_{j|i}$$

This method therefore assumes that the probabilities are correctly estimated by the frequencies in both the reference population and the observed population.

The method is derived from many earlier studies on other subjects. Introduced by Fridriksson (1934), who was working on fish statistics, then Kruithof (1937), who was investigating telephone networks, it was taken up by statisticians (Deming and Stephan, 1940), economists (Leontief, 1941), and numerous other researchers including Friedlander, (1961); Thionet, (1963, 1964); Caussinus, (1965); Tugault, (1970); and Willekens, (1977). Many authors now call it the IPFP (Iterative Proportional Fitting Procedure) or ALK (Age Length Key) method.

The method was adopted in paleodemography by Masset (1971), who called it the probability vector method, then by Konigsberg and Frankenberg (1992), who continued to refer to it as the ALK method.

Note that the resulting distribution depends strongly on the age distribution in the reference population and is “flattened by the influence of the reference sample” (Masset, 1995). This is unsurprising, given the assumption that each cell of the estimated matrix must be as close as possible to each cell of the reference matrix.

In reality, this method introduces artefacts by focusing as much on the row probabilities as on the column probabilities deduced from the reference matrix shown in Table 1. We should, instead, only consider the biological uniformity hypothesis (Howell, 1976), also called “invariance hypothesis” (Müller et al., 2002). This states that, for any bone belonging to an individual of a given age at death, the probability that the bone will be classified in a given stage depends only on that age, regardless of the population from which the bone was extracted. In consequence, the reference data are considered solely via the “column profiles”, i.e. the conditional distributions of bone stages for each age class. Hence the search for another estimation method based on that hypothesis alone.

---

(3) Of course, one can use other distances, such as a Euclidean distance, but the results are generally similar.

2. Tables of minimum distance between each column

The second method therefore seeks to estimate the columns of a table that are as close as possible to each column of the reference table. For this, we can use an iterative method that takes a random or uniform initial structure $\hat{p}_j^0$ and estimates the structure of the observed population, $\hat{p}_j$, through successive iterations, with the aid of the recurrence formula:

$$\hat{p}_j^{n+1} = \sum_{i=1}^{l} \frac{m_i}{m} \frac{\hat{p}_j^n f_{i|j}}{\sum_{j=1}^{c} \hat{p}_j^n f_{i|j}}$$

We perform as many iterations as needed for $\hat{p}_j^{n+1}$ to differ from $\hat{p}_j^n$ by as small a quantity as we want. This algorithm is aimed at obtaining maximum-likelihood estimators by assuming fixed $f_{i|j}$ values and a multinomial distribution of on-site observations. We show that the algorithm does indeed supply the estimators, at least in the case of a regular maximum (where the gradient is set to zero).

This method, first introduced by ichthyologists (Hasselblad, 1966; Kimura and Chikuni, 1987) under the name IALK (Iterative Age Length Key), was adopted in paleodemography by Masset (1982) as the method of successive approximations to avoid the overly flat result obtained with the method of probability vectors. It was taken up by Konigsberg and Frankenberg (1992), again under the name IALK. The two approaches – which we shall call American and French for simplicity's sake – provoked much controversy between 1992 and 2002.(4) In the end, however, they proved virtually identical(5) (Konigsberg and Frankenberg, 2002).

The method requires l to be greater than or equal to c, in order to obtain a single solution. Otherwise, system (1) is undetermined, admitting an infinite number of solutions. Unfortunately, some paleodemographers disregard this condition, causing them to obtain unsatisfactory solutions. For example, Jackes (2000) tries to estimate seventeen age classes with only six stages, and obtains a large number of age classes of zero proportions. The same applies to Bocquet-Appel and Bacro (1997), who estimate seven age classes with only six stages (Konigsberg and Frankenberg, 2002).

Moreover, this estimation method can yield clearly unsatisfactory solutions. Masset (1982) takes a reference population comprising seven age classes and seven stages and a stage vector for the observed population without zero elements – yet he obtains an age structure of the observed population with four zero elements.

---

(4) See especially Bocquet-Appel and Masset (1996), who now call the method IPFP despite its being different from what statisticians refer to under that name.

(5) For more details on the differences between these methods, see Courgeau (2010).

3. Methods proposed more recently

At a seminar held in Rostock (Hoppa and Vaupel, 2002), the American school proposed introducing a continuous age rather than a discretized one, and modelling the probability density of the observed population by means of a parametric event-history model such as a Gompertz model (two parameters), a Gompertz-Makeham model (three parameters) or a Siler model (five parameters). This avoids the problem of zero age classes. But such methods still closely resemble the IALK method, as Konigsberg and Herrmann point out in Hoppa and Vaupel's book: "Our current methods fit fairly comfortably within the approaches taken during the Rostock workshop".

However, these methods introduce a number of additional hypotheses that we have hardly any means of verifying. They include: a stationary or stable population to ensure that the event-history model applies to current conditions; and continuity in the age distribution of a given stage, leading to different distributions according to the methods used. Lastly, the Rostock methods continue to assume that the frequencies offer correct estimates of the probabilities.

Bocquet-Appel (2005, 2008a, 2008b) and Bocquet-Appel and Bacro (2008) propose two changes. First, the reference population should no longer be viewed as perfectly estimated by the frequencies. For this purpose, they perform 1,000 draws using the bootstrap procedure in each of the reference population's age classes. Second, they reduce the set of probability vectors $(p_1,\ldots,p_c)$ to a family (a mix of Gompertz-Makeham distributions and extreme values) that they construct in order to represent the largest possible number of cases of mortality in this family of "candidates". They seek the vector that best satisfies system (1) when π_i is replaced by the corresponding observed frequency and $p_{i\,|\,j}$ is replaced by one of the results of the above-mentioned bootstrap draw. The 1,000 draws thus supply 1,000 estimates, each equal to one of the candidate vectors. For the final estimate, the authors can choose either (a) the "best" of the 1,000 vectors obtained, i.e. the one with the minimum distance between the first and second members of equation (1); or (b) the mean of the 1,000 estimates obtained (as we shall do when applying their method in Section IV). This ad hoc estimation is accompanied by "confidence intervals" based on the 1,000 intermediate results. While this method is theoretically worth considering for the ad hoc estimation (we shall examine its performance later), the confidence intervals that it produces are problematic. The authors offer no theoretical validation for it, and the fact that they disregard the inevitable random disturbances in the observed stage frequencies clearly makes these subject to caution, and certainly far too optimistic, as we have been able to verify by simulation.

III. A new estimation method

We propose a new estimation method that is truly Bayesian, being based on the typical ingredients of Bayesian statistics. It therefore contrasts with some earlier proposals sometimes described as Bayesian because they use Bayes' formula or introduce priors in their approach.(6) Our method accordingly features:

- parameters that are assumed to be random, with a prior distribution through which we seek to capture characteristics known independently of the observed data,
- calculation of the probability distribution of these parameters conditional upon the observations; this is known as a posterior distribution and serves as the basis for statistical inference.

1. Model and principle of the method

It is logical to view the stage frequencies m_i ($i = 1,\ldots l$) observed on site as the observed values of a multinomial distribution whose parameters π_i are linked to the p_j and $p_{i|j}$ parameters through system (1). We shall use the latter parameters to continue our modelling.

Let G be the prior density of parameters $p_{i|j}$, $i = 1,\ldots, l$ and $j = 1,\ldots, c$ (we shall see how to express it in the following subsection) and let us suppose that the parameters p_j ($j = 1,\ldots, c$) have a prior density g (also discussed in the following subsection) and are independent of the $p_{i|j}$ parameters.

With M as the m_i vector, P as the $p_{i|j}$ vector (matrix), and p as the p_j vector, the joint density of (M, P, p) will be f, given by:

$$f(M, P, p) = g(p)\, G(P) \frac{m!}{\prod_i m_i!} \prod_i \left(\sum_j p_j \; p_{i|j} \right)^{m_i}$$

The marginal density of the (M, p) pair is:

$$f(M, p) = \int f(M, P, p)\, dP$$

and the marginal density of M is:

$$f(M) = \iint f(M, P, p)\, dp\, dP$$

The integrals are taken from the variation domains of P and/or p, which are a simplex (for p) or a product of simplexes (for P).

The conditional density of p given M is therefore:

$$f(p \mid M) = \frac{\int f(M, P, p)\, dP}{\iint f(M, P, p)\, dp\, dP}$$

---

(6) The Bocquet-Appel and Bacro method (2008) uses "prior" vectors to take into account the fact that the probability distribution to be estimated is a mortality distribution with necessarily specific characteristics. However, this leads them to wisely reduce the parametric space of a method that remains basically frequentist, rather than to adopt a strictly Bayesian approach. To avoid all confusion, we refer to these vectors here as "candidate" vectors, all the more so as we shall propose using them to build a true prior distribution for our Bayesian method.

This is the posterior density of the p_j ($j = 1,\dots c$) parameters – the basic tool for Bayesian estimation.

For example, the posterior mean of p_j will be:

$$E(p_j \mid M) = \frac{\iint p_j f(M, P, p)\, dp\, dP}{\iint f(M, P, p)\, dp\, dP}$$

More generally, the conditional expectation given M of a function φ of p will be:

$$E(\varphi(p) \mid M) = \frac{\iint \varphi(p) f(M, P, p)\, dp\, dP}{\iint f(M, P, p)\, dp\, dP} \qquad (2)$$

We thus obtain, for example, the k^{th}-order moment of p_j with $\varphi(p) = p_j^k$. Taking for $\varphi(p)$ the function that equals 0 for $p_j > x$ and 1 for $p_j < x$ (indicator variable of the event $p_j < x$), we express the posterior distribution function for p_j at point x.

The various integrals of equation (2) can be evaluated using a Monte Carlo method (Robert, 2006) as follows.

Let $X = (X_1,\dots,X_c)$ be a random vector with a density distribution g and Y a family of c vectors $Y = (Y_{1j},\dots,Y_{Ij})$ ($j = 1,\dots,c$), whose joint distribution is independent of X and admits density G. We verify that equation (2) is equivalent to:

$$E(\varphi(p) \mid M) = \frac{E\left(\varphi(X) \prod_i \left[\sum_j X_j Y_{ij}\right]^{m_i}\right)}{E\left(\prod_i \left[\sum_j X_j Y_{ij}\right]^{m_i}\right)}$$

Let us generate S independent sets of such random vectors (X,Y), with s ($s = 1,\dots,S$) representing the repetitions. By virtue of the law of large numbers, if S is large enough, the expression above is approximated by:

$$E(\varphi(p) \mid M) \approx \frac{\sum_{s=1}^{S} \varphi(X_s) \prod_i \left(\sum_j X_{js} Y_{ijs}\right)^{m_i}}{\sum_{s=1}^{S} \prod_i \left(\sum_j X_{js} Y_{ijs}\right)^{m_i}}$$

This notably supplies the posterior expectation of each p_j ($j = 1,\dots,c$) – which can be taken as a point estimate – or the posterior variance useful for characterizing the accuracy of the estimate. The same principle can be applied to evaluate cross-moments, such as the covariance matrix of the posterior distribution of the p_j parameters. Lastly, a p_j parameter's posterior distribution function allows us, for example, to calculate intervals containing that p_j with a given probability.

Called credible intervals in the Bayesian framework, they correspond to confidence intervals in the classic framework.

Some authors have recommended taking the posterior distribution mode rather than posterior expectation as a point estimate of the parameters. We prefer the expectation value for several reasons. First, writing the estimate obtained as $\hat{p}_j$, it minimizes the average cost of the loss function $\sum_j (\hat{p}_j - p_j)^2$, which occurs naturally in our problem since the function penalizes errors in proportion to their amplitudes (by contrast, the mode is optimal for a zero loss if the estimate is extremely close to the true value, and a constant positive loss otherwise, which does not seem suitable here). To this essential reason, we can add the ease of computation and the fact that the posterior density may not be bounded and may therefore exhibit an infinite mode for one or more zero estimated probabilities, yielding a scarcely realistic result.

2. Practical use

Choice of prior distributions

Density G

The only source of information on conditional probabilities $p_{i|j}$ is the reference data. If they are raw data merely obtained by recording the stage frequencies on a sample of skeletons of known ages, we can logically conclude that, for each age class j ($j = 1,\ldots,c$), the frequencies n_{ij} are the observed values of a multinomial distribution with a total n_j and probabilities $p_{i|j}$ ($i = 1,\ldots,l$). Adopting a prior distribution for the $p_{i|j}$ probabilities, we deduce a posterior distribution, conditional upon the reference data. We take this, in turn, as the prior distribution of the $p_{i|j}$ probabilities in the final model. Given the scarcity of supplementary information on these $p_{i|j}$ probabilities beyond what is contained in the reference data, it makes sense to adopt a uniform distribution as the prior distribution of the $p_{i|j}$ probabilities for each j. For a given j, we find a posterior distribution of $p_{i|j}$ probabilities that consists of a Dirichlet distribution of parameters $\alpha_{ij} = n_{ij} + 1$ ($i = 1,\ldots,l$).(7)

---

(7) The random vector $X = (X_1,\ldots,X_k)$ follows a Dirichlet distribution of parameter $a = (a_1,\ldots,a_k)$ if it admits the density $d(x) = \frac{\Gamma(a)}{\prod_{i=1}^{k}\Gamma(a_i)} \prod_{i=1}^{k} x_i^{a_i - 1}$, on the simplex D defined by $x = (x_1, \ldots, x_k) \in D \Leftrightarrow x_i > 0$ for all $i = 1, \ldots, k$ and $\sum_{i=1}^{k} x_i = 1$, Γ being Euler's gamma function. For the properties of this distribution, see, for example, Robert (2006).

The density *G* is the product of these *c* Dirichlet densities:

$$G(P) = \frac{\prod_{j} \Gamma(\alpha_{.j})}{\prod_{i} \prod_{j} \Gamma(\alpha_{ij})} \prod_{i} \prod_{j} p_{i|j}^{\alpha_{ij}-1}$$

The multinomial nature of the reference data is mentioned merely for clarification purposes: it is only notional and not essential to obtaining this prior distribution *G*.

One can refine the choice of *G*, but that does not seem to achieve notable improvements (for fuller details, see Caussinus and Courgeau, 2010); we shall therefore not elaborate on the option here.

Density *g*

The choice of prior distribution for the p_j parameters is more delicate. As there is no clearly designated "class" of distributions from which to select the prior distribution, the most sensible course is to opt for a Dirichlet distribution, which is well-suited to probability vectors. This leaves the problem of choosing the distribution parameters, say $(\beta_1,\ldots,\beta_c)$. In the absence of specific information, we can, as above, choose a uniform distribution and take $\beta_j = 1$ for all *j*. Such a choice allows us to remain "neutral" and may sometimes be justified. It also yields reasonable results on simple examples. However, in paleodemography, other choices would appear to be preferable as certain information is naturally available. We can, for example, take a "standard" mortality distribution and calculate the probabilities of each of its age classes. The class probabilities become the means of the prior distribution. This gives the parameters β_j up to a proportionality coefficient, i.e. the $\beta_j/\beta_\bullet$ values, where $\beta_\bullet$ is the sum of the β_j parameters over $j = 1,\ldots,c$. The remaining step is to choose $\beta_\bullet$, i.e. in practice, the prior distribution variances. Note that the variances need to be relatively large in order to express the fact that the prior means are not very reliable and that the prior distribution should not play a dominant role – in other words, that the family of possibilities envisaged covers a broad field. The $\beta_\bullet$ variances should thus be fairly small, say, below unity or barely above. Some simulations have shown that a simple and efficient criterion is to take $\beta_\bullet = c$, as in the case of the uniform prior distribution (for a concise account, see Caussinus and Courgeau, 2010). That is what we shall do here.

The standard used will be the pre-industrial standard (Séguy et al., 2008; Séguy and Buchet, 2010) for men, women, or the two sexes together, as best suits the circumstances. When using our method with this prior distribution, it will be referred to as BayesPI. In some situations, the paleodemographer may have specific information. For example, regarding the monastic cemetery of Maubuisson (France), we know that the remains are of women presumably in better health than the average population, and not exposed to certain major mortality risks for younger women, notably maternal mortality.

The principle of the above choice of prior distribution can be extended in several ways. For instance, instead of taking a standard mortality distribution as a base for constructing the prior distribution, two "standard" distributions can be combined, giving a mixture of two Dirichlet distributions. This could consist of a combination (in sensibly chosen proportions) of a standard mortality distribution (attrition) and a catastrophic mortality distribution.

A wide variety of approaches are possible, of course. For example, somewhat in the spirit of Bocquet-Appel and Bacro's proposals (2008), the prior distribution can be defined as a uniform distribution over a discrete set of distributions consisting of standard mortality distributions. It is a far more cumbersome solution to implement than the previous one, but it becomes easy to use when the task of building such vector sets has already been performed. We shall use it below (under the name BayesUnif) and compare it with the Bocquet-Appel and Bacro method since it is precisely when the latter method is usable that BayesUnif is easier to apply. From a technical standpoint, note that one of the integrals defining the posterior distribution is now a finite sum. This allows us to simplify the Monte Carlo calculations slightly by applying some mathematics: with the notations of subsection III.1, only Y needs to be simulated, but no longer X.

Posterior distribution and credible intervals

Earlier, we saw how to calculate the posterior distribution function for each p_j probability, point by point. We can thus determine α-credible intervals (Robert, 2006, p. 278) in which a p_j parameter has a probability $1 - \alpha$ conditional upon the observations. The posterior distribution function is calculated laboriously, point by point. Another approach, for heuristic purposes, is to approximate the posterior density of each p_j by a beta density with the same mean and variance; approximation quality can be controlled to a certain extent via higher-order moments: for example, we can check the closeness of the beta distribution's third- and fourth-order moments to the corresponding moments of the "true" posterior distribution, easily calculable by simulation, as seen earlier. In all the examples treated, the posterior distribution functions were calculated exactly and by proxy, and the approximation has generally proved acceptable, indeed excellent in most cases. We can approximate an α-credible interval for each p_j by the interval between the $\alpha/2$ and $1 - \alpha/2$ quantiles of its posterior distribution. In general, it is not the shortest possible (HPD) interval but for practical purposes it is reasonable. It is extremely inadvisable to use an interval of the "mean plus or minus one (or two) standard deviations" type because the posterior distribution is, in most cases, highly dissymmetrical.

Size of the data table

System (1) described in Section I is undetermined if the number of rows (stages) l is smaller than the number of columns c (ages). In other words, the parameters of interest are not identifiable, given that several values lead to the same distribution of observable samples. The Bayesian method enables us to get around the difficulty since we start with a prior distribution and the aim is simply to make it change by means of the data. The posterior distribution steers us toward a distribution of the unknown parameters, which is wholly compatible with the fact that they are not completely determined. This method can therefore be used with $l < c$. Obviously, the posterior distribution can be somewhat dispersed, which merely reflects the indeterminacy inherent in the situation.

IV. Comparison with earlier methods

When proposing a new method, one must start by comparing its performance with that of the main methods currently used. In addition to the mathematical difficulties that a theoretical study would raise, it will always be hard to compare methods based on different paradigms – frequentist for earlier methods, Bayesian for ours. We shall therefore proceed through simulations under the following conditions in order to mimic reality as closely as possible.

- We take a vector of probabilities p_j ($j = 1,\ldots,c$) as the target value to be estimated and a reference matrix such as the one in Table 1.
- As the reference matrix is a sample rather than a population, we assume that the "true" probabilities $p_{i|j}$ are not exactly the quantities $f_{i|j} = \frac{n_{ij}}{n_{.j}}$ but that, for each j, the $f_{i|j}$ values should be regarded as the probabilities of a multinomial distribution totalling n_{ij}. We therefore draw a new reference matrix at random based on this principle (which is the same as the bootstrap draws used in the method proposed by Bocquet-Appel and Bacro, 2008). At the same time, this procedure allows us to take account of the fact that, even apart from the sampling uncertainties, the real $p_{i|j}$ matrix may not be the reference distribution matrix but may lie only within a certain vicinity of the latter.
- The given p_j values and the $p_{i|j}$ values simulated as above yield the stage probabilities π_i via system (1). We draw the number of individuals m_i in these stages at random under a multinomial distribution with l categories, π_i ($i = 1,\ldots,l$) probabilities, and a given total t.
- We perform R independent repetitions. To each, we apply the various estimation methods to be compared. Naturally, we implement the methods only with the resources actually at our disposal, i.e. in particular, with the observed reference data and not the simulated data. We can choose an identical p_j vector for all R trials if we want to measure the efficacy of the methods in a given case. Alternatively, the vector can vary with each trial so as to cover a reasonable range of possible values, allowing a broader study at minimal cost. For each trial, we give two measures of divergence between the target vectors and the vectors estimated by the various methods being compared. The first divergence is the square of the Euclidean distance, i.e. $\sum_j (\hat{p}_j - p_j)^2$; the second

divergence weights the different terms like a chi-square, i.e. $\sum_j \frac{(\hat{p}_j - p_j)^2}{p_j}$. We average the terms found in the *R* trials and multiply by 100 to obtain "unit percentages". We designate the resulting divergences as a (absolute) and χ (chi-square type).(8) We envisaged other divergence measures, for example by replacing the squares with absolute values or comparing the cumulative frequencies. As the results are entirely similar to those presented here, these alternatives will be omitted so as not to burden the discussion.

The simulation model thus designed differs substantially from the one we used in Caussinus and Courgeau (2010). The model just described is slightly more complex but seems closer to reality. However, the results given below and those of Caussinus and Courgeau (2010) are, on the whole, very consistent.

As seen in Section II, the poor performance of the ALK method was already clearly recognized, and we shall not return to the issue here. Instead, we shall examine the IALK method. It is, in essence, a maximum-likelihood method applied to a multinomial distribution whose probabilities π_i are linked to the parameters to be estimated p_j by the system (1), in which the $p_{i|j}$ values are taken as known (deduced from the reference data). In practice, the original algorithm (which is very slow) was replaced by a classic optimum search algorithm: the constrOptim procedure of the R software package (R Development Core Team, 2008).

Next, we consider Bocquet-Appel and Bacro's recent method (2008). To use it, we need situations in which it is applicable, i.e. ones for which the authors have supplied a set of candidate vectors. We have chosen a division into seven 10-year age classes from 20 to 90 years; the final class can be interpreted as age 80+. The method is used with the set of 756 vectors included in their ProbAtri20-90.txt file (probability distribution models for standard pre-industrial mortality, called attritional mortality(9)).

Lastly, we use our method with two versions of the prior distribution. The first version – easily usable whatever the subdivisions into classes – is a Dirichlet prior distribution with the parameters derived from the pre-industrial standard as noted above. The second version is a uniform prior distribution on the set of 756 vectors of the ProbAtri20-90.txt file. This procedure is easy to follow when the Bocquet-Appel and Bacro method is applicable. Since the aim is to compare our method with frequentist methods, we have consistently reduced its scope to ad hoc estimation using posterior expectation. However, our Bayesian method supplies other elements of analysis, as shown in the examples in Section V.

---

(8) The designation adopted comes from the analogy with a χ^2, for the observed differences are roughly divided by their standard deviations.

(9) See Iterage software at http://www.evolhum.cnrs.fr/bocquet/index.html

We began by performing some simulations with predetermined p_j ($j = 1,\ldots,c$) vectors to be estimated. These represent distributions that can be realistic, but are remote enough from the pre-industrial standard to ensure that our method does not have undue advantage.(10) We verified that the IALK method yielded much less satisfactory results than our method and the Bocquet-Appel and Bacro method. Compared with the latter, our method seemed to give similar, but rather better results (particularly for small samples) when applied with a Dirichlet prior distribution, and distinctly better in every instance when applied with the second prior distribution mentioned above (uniform on the candidate vectors). It is therefore the comparison between our method and the Bocquet-Appel and Bacro method that needed to be performed most carefully. So as not to disadvantage the Bocquet-Appel and Bacro method, we conducted the R repetitions by consistently choosing as the target vector one of the prior vectors of the ProbAtri20-90.txt file through equiprobable random selection. Seven age classes were kept for the reasons given earlier, but we introduced variations for (1) the site sample size m (successively 25, 50, 75, and 100, which are standard orders of magnitude) and (2) the number of classes of bone stages chosen (successively $l = 5, 7$, and 8). In each example, the number of repetitions is $R = 1{,}000$.

The three reference matrixes used are those of "Lisbon Men" deduced from the three Portuguese collections (Ferraz de Macedo in Lisbon, Coimbra, and Bocage Museum of Lisbon), for which cranial suture closure has been measured in 42 stages.(11) All three matrixes comprise seven 10-year age classes from age 20 and the following stage groupings: 0-8, 9-15, 16-21, 22-28, and 29-41 for the first; 0-4, 5-7, 8-12, 13-18, 19-23, 24-34, and 31-41 for the second; and class 13-18 split into 13-15 and 16-18 for the third.

We deduce the β_j vector from the "male pre-industrial" standard:
(0.70 0.77 0.98 1.19 1.47 1.33 0.56).

The values obtained for the two criteria are shown in Figure 1 with, from top to bottom, the criteria a and χ. The IALK method cannot be used for $l < c$, i.e. in the present case, for the subdivision of stages into five classes.

To begin with, the results in Figure 1 show the very poor performance of the IALK method. Note that the divergences were divided by eight for the a values and by four for the χ values to make the figure more legible. In addition to the large divergences in the figure, the estimates supplied are almost always on the boundary (at least one of the $\hat{p}_j$ parameters is zero). Logically enough, all the methods improve when the sample size increases. More unexpectedly, the performance of the Bocquet-Appel and Bacro method improves slightly when more stage classes are considered, whereas the opposite is observed for

---

(10) By way of example, one of the vectors examined is: (0.20 0.15 0.10 0.20 0.20 0.10 0.05).

(11) For a fuller account of these collections and the measurements made, including detailed tables, see Masset (1982) and Séguy and Buchet (2010).

Figure 1. Divergences between estimated and target vectors for IALK method, Bocquet-Appel and Bacro method, Bayes method with pre-industrial (PI) standard prior, and Bayes with uniform prior

Figure 1A. Criteron *a*

Ined 2010

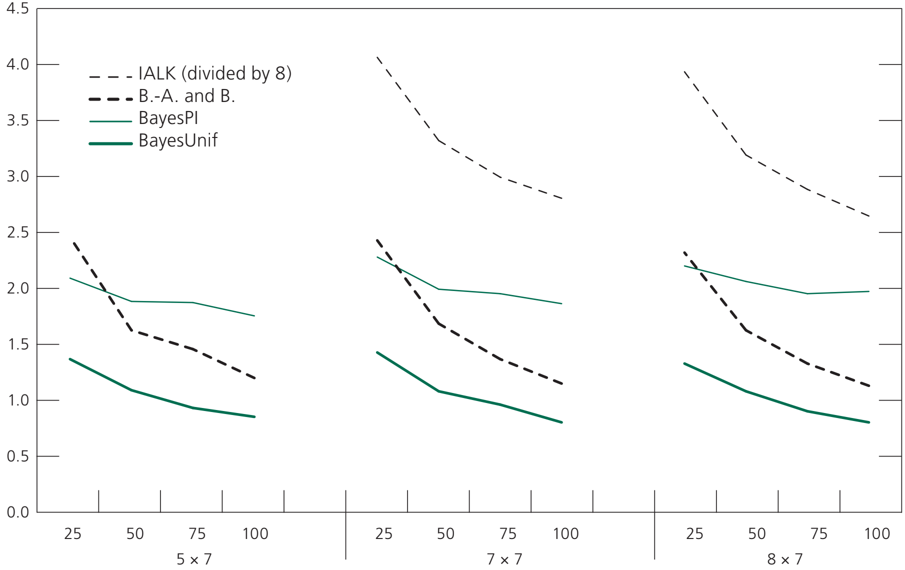
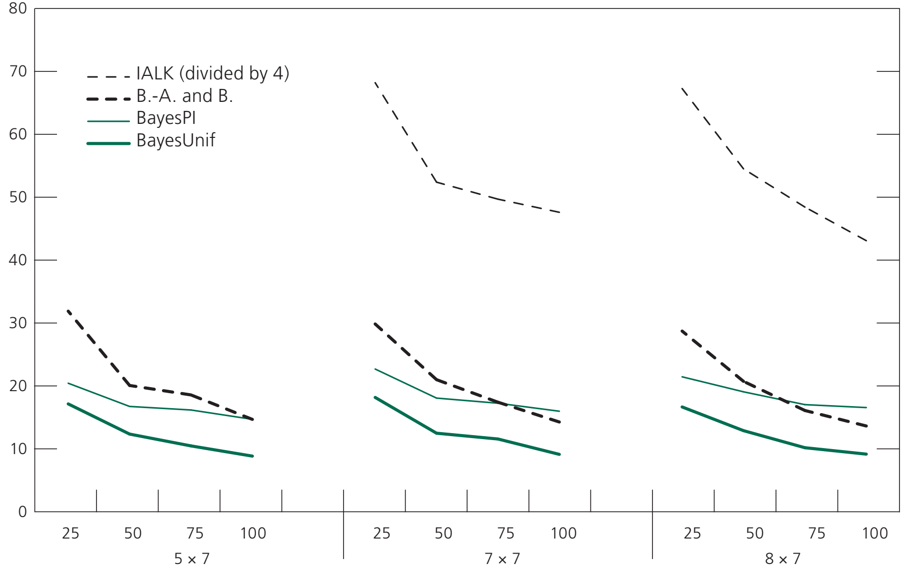


Figure 1B. Criteron χ

Ined 2010



Note: Horizontal axis: sample size m; l × c. For IALK, *a* values have been divided by eight and χ values have been divided by four.

Source: Authors' calculations based on Masset (1982), and Séguy and Buchet (2010).

our method, whatever the prior distribution. With the Dirichlet prior distribution deduced from the pre-industrial standard, the performance of the Bayesian method turns out to be generally similar to that of the Bocquet-Appel and Bacro method – although, as already noted, the choice of target vectors is bound to favour the latter. The comparison depends significantly on the index considered. This suggests that our method tends to be better than the Bocquet-Appel and Bacro method for estimating small probabilities and less so in estimating larger ones. It is also superior for small samples and loses its advantage for larger ones. Lastly, if used with the uniform prior distribution on the candidate vectors of the Bocquet-Appel and Bacro method, the Bayesian method consistently produces distinctly better results. The comparison can be summarized as follows:

- The IALK method should be definitively rejected.
- In a situation where a family of candidate vectors has not been established, our method is very easy to apply with a Dirichlet prior distribution and its performance is comparable to that of the Bocquet-Appel and Bacro method. The considerable task of developing the vector set thus becomes unnecessary, particularly if the sample on the target site is small (in Section V, we shall also see that it is sometimes easy to adjust the parameters of the Dirichlet distribution to further improve performance).
- If a set of candidate vectors has been prepared to apply the Bocquet-Appel and Bacro method, it should be used in preference to determine the prior distribution of our method rather than to apply the Bocquet-Appel and Bacro method itself.

V. Archaeological applications

We shall now discuss two examples to highlight various aspects of our proposed method: its implementation (in particular, the choice of prior distribution) and the information that it can provide for point estimates and beyond.

1. The nuns of Maubuisson (seventeenth-eighteenth centuries)

Our first example concerns the monastic cemetery of the royal abbey of Maubuisson (France), where 162 Cistercian nuns are buried and where 37 skeletons have been exhumed to measure the stages of cranial suture closure.(12) The women, most of whom belonged to the higher nobility, enjoyed very privileged conditions in childhood and adolescence. Their monastic life, while harsh in some respects, sheltered them from the hazards to which their lay contemporaries were exposed during their reproductive years (Séguy and Buchet, 2010).

---

(12) We thank Luc Buchet and Isabelle Séguy for supplying us with the site data. For more details, see Séguy and Buchet (2010).

We adopt a division into seven age classes (ten-year classes from ages 20-29 to 70-79 and one open-ended class for ages 80 and over), and seven skeletal stages. The frequencies observed for these stages on a sample of 37 skulls are (6 2 4 5 3 9 8).

We have important prior information on this site, which is especially useful given the very modest size of the observed sample. The individuals are nuns, therefore women, presumably over 20 years old. This determined our above-mentioned choice of age classes and of a specific reference data set: "Lisbon women" (see Séguy and Buchet, 2010). On the basis of this information, we shall take a Dirichlet prior probability distribution for the seven parameters to be estimated. The β_j parameters of the distribution are proportional to the values of standard pre-industrial mortality (female) and sum to 7, i.e. (0.70 0.77 0.84 1.05 1.47 1.47 0.70). This is the first estimate presented for comparison purposes. But the fact that the population consists of nuns provides additional information. As noted earlier, for many reasons, these women were certainly in better health when they entered the convent than the average population. They were then shielded from several major mortality risks, notably maternal mortality. On the prior assumption that these factors reduced mortality among the 20-29 age group by slightly over 50% and among the 30-39 age group by just under 50%, we replace the prior distribution parameters by: (0.30 0.40 0.84 1.05 1.47 1.47 0.70) or rather by the proportional values (0.337 0.449 0.944 1.180 1.652 1.652 0.786) summing to 7, as recommended in subsection III.2. Using this prior distribution, we offer a second estimate, which, on the evidence, is the one we believe should be adopted in practice. We shall see how the results obtained confirm this assumption, and examine ways of refining the estimate.

Lastly, we have a major item of additional information. Thanks to the abbey registers, the actual ages at death of the 162 nuns who lived at Maubuisson can be determined directly. On these data, we estimate the probabilities of the age classes considered as follows: (0.012 0.025 0.087 0.170 0.289 0.210 0.207).

We therefore have an objective means of gauging the efficacy of the method. Admittedly, some caution is in order – first, because this evaluation is probably no more than approximate, second, because the 37 skulls examined are only a sample (and possibly even biased: was it selected strictly at random? (13)).

Let us begin by an analysis with a prior distribution conforming to the female pre-industrial standard. This gives us the posterior expectations of the seven age classes (0.048 0.067 0.071 0.135 0.301 0.219 0.159) and the

---

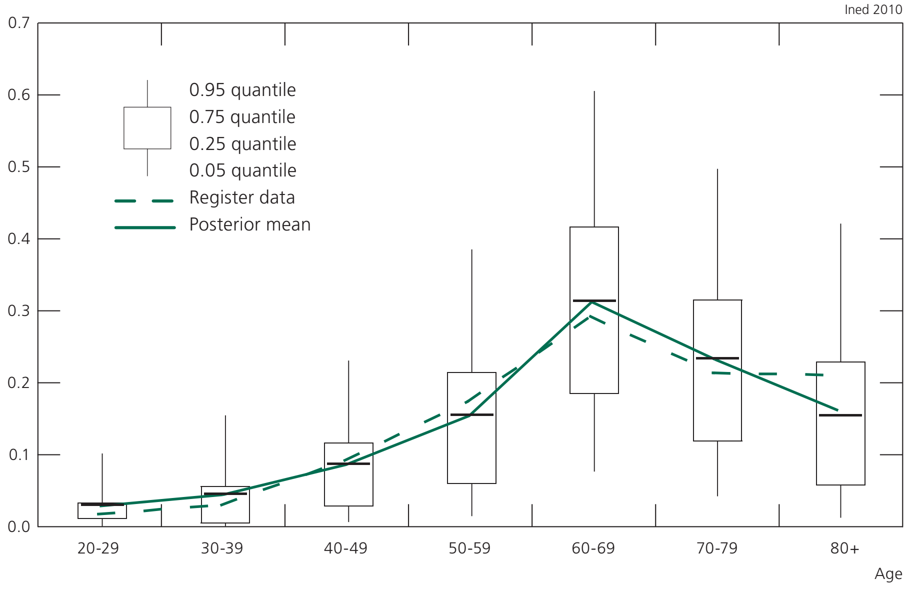
(13) With the reference probabilities used here, the sample is nonetheless highly compatible with the register values. If we calculate theoretical frequencies for the stages using these data and compare them with the values observed by a chi-square, we obtain 1.93 for six degrees of freedom, a value well below the threshold of significance at the 5% level (12.59).

posterior standard deviations (0.050 0,.068 0.069 0.114 0.166 0.142 0.135).

It is interesting to compare the posterior means with the prior means, which are (0.10 0.11 0.12 0.15 0.21 0.10). We see that the data make it necessary to revise the probabilities of the "young" classes sharply downwards, and the probabilities of two of the three oldest classes sharply upwards. This is consistent with our earlier discussion. We shall therefore end the present analysis here and move on to the Bayesian analysis with a prior distribution of parameters (0.337 0.449 0.944 1.180 1.652 1.652 0.786) corresponding to a modified pre-industrial standard (MPI), in keeping with the previous discussion. We obtain the posterior means (0.025 0.041 0.083 0.151 0.311 0.230 0.159) and the posterior standard deviations (0.037 0.054 0.074 0.119 0.163 0.142 0.132). Note, in passing, that by replacing the p_j values of system (1) with the values just estimated, and by estimating the $p_{i|j}$ values from the reference data, we obtain (7.5 2.5 4.3 5.6 4.3 6.8 5.9) as the "theoretical" stage frequencies, which are very close to the observed values.

We can now move on to the posterior densities. By comparing the exact posterior distribution function and the function approximated by the beta distribution, we found that the beta densities offer an excellent approximation. These proxies were therefore used to calculate the credible intervals given in Figure 2.

Figure 2. Maubuisson example (MPI prior distribution). Probability estimates using posterior mean and quantiles giving 90% and 50% credible intervals



Source: Authors' calculations based on Séguy and Buchet (2010).

Clearly, with such a small sample, it is not possible to obtain very precise estimates, as evidenced by the width of the credible intervals. We can, however, obtain relevant information by analysing the data. The analysis leads us, once again, to revise the probabilities of the first three classes downwards and those of the fifth and seventh classes upwards: the prior means here are (0.048 0.064 0.135 0.169 0.236 0.236 0.112).

The estimates obtained prove very similar to the values yielded by the registers, as Figure 2 shows. The most significant differences, especially in terms of relative value, concern the first two probabilities (whose register-based value is so low as to be hard to imagine *a priori*) and, to a lesser extent, the last value, which is correlatively higher. In fact, given the highly dissymmetrical distributions for the very low probabilities, the mean can be misleading. In the first class, for example, we see that 50% of the posterior probability lies in the interval [0.001 – 0.032], whose midpoint 0.017 comes close to the target value. The same is true of the second class. We may legitimately assume, therefore, that for these two classes, the posterior means overestimate the true values – a pattern that turns out to be fairly consistent with reality. More generally, we see that the shortest credible intervals (at 50%) effectively bracket the target values, which are even very close to their centres. Overall, in a standard "blind" situation, the posterior means give basic information that can be usefully supplemented by a set of other factors, such as their change from the prior means and the credible intervals. In particular, we should stress the importance of this type of factor in the case of very small probabilities, as will again be apparent in the Frénouville example below.

To conclude the analysis of the Maubuisson example in keeping with our comparisons in Section IV, the IALK method yields highly aberrant results here, with several zero probabilities. The Bocquet-Appel and Bacro method applied with the candidate vectors of the ProbAttri20-90 file gives the estimates (0.025 0.036 0.073 0.133 0.209 0.268 0.255), which are broadly consistent with the registers, but inferior to ours: for example, the sum of the squares of errors is 0.014 versus 0.004 using our method; the maximum error over the seven classes is 0.080 versus 0.048 with our method. Lastly, we can apply our method here with the uniform prior distribution for the ProbAttri20-90 file vectors. The estimated age-class probabilities obtained are (0.059 0.057 0.096 0.149 0.204 0.232 0.203), and the above criteria for distance from register values are 0.011 and 0.085 respectively. This second prior distribution therefore yields poorer estimates than the first. They are fairly close to the estimate supplied by Bocquet-Appel's Iterage programme. While this prior distribution provided good estimates in the simulations in Section IV, it cannot incorporate the prior knowledge specific to the population concerned. By contrast, the first version of our method does so extremely simply and, as we have seen, effectively.

2. The Frénouville cemetery (Merovingian period)

In this example, the bone stages of 200 skulls were distributed into five classes (the same as in the five-class example of Section IV), whose observed frequencies are (92 29 22 27 30).(14) The age distribution was performed for two subdivisions: eight and fourteen classes. In both alternatives, the first class is age 18-19, the last class is age 80+. The 20-79s were divided into ten-year classes (for a total of eight classes) and five-year classes (for a total of fourteen classes) respectively.

As no indication of sex is taken into account, the Lisbon reference data were used for both sexes combined. The Bayesian method was applied with a Dirichlet prior distribution, which makes it easy to examine any subdivision into age classes. There is no specific information here on the population concerned. We therefore chose the β_j values proportional to the probabilities of the pre-industrial standard (both sexes combined), i.e.:

- for eight classes: (0.02 0.10 0.11 0.13 0.16 0.20 0.19 0.09)
- for fourteen classes: (0.02 0.05 0.05 0.05 0.06 0.06 0.07 0.007 0.09 0.10 0.11 0.11 0.09 0.09). The sum of β_j values is equal to the number of classes.

With eight age classes, we obtained the results reported in Table 2 and Figure 3 (the quantiles are calculated on the approximation of posterior densities through beta densities: the result is not as good as the one obtained in the first example, but quite sufficient in practice). Recall that the 5% and 95% quantiles are the extremes of the 90% credible intervals, while the 25% and 75% quantiles are the extremes of the 50% credible intervals. The interquartile range (IQR) is a means of measuring dispersion that usefully complements the standard deviation, in particular for highly dissymmetrical distributions.

Table 2. Frénouville example, eight age classes.
Selected characteristics of posterior distribution

Age class	18-19	20-29	30-39	40-49	50-59	60-69	70-79	80+
Mean	0.173	0.312	0.070	0.056	0.071	0.138	0.092	0.088
Standard deviation	0.174	0.192	0.072	0.053	0.059	0.080	0.065	0.067
Inter-quartile range	0.225	0.283	0.082	0.062	0.073	0.106	0.083	0.085

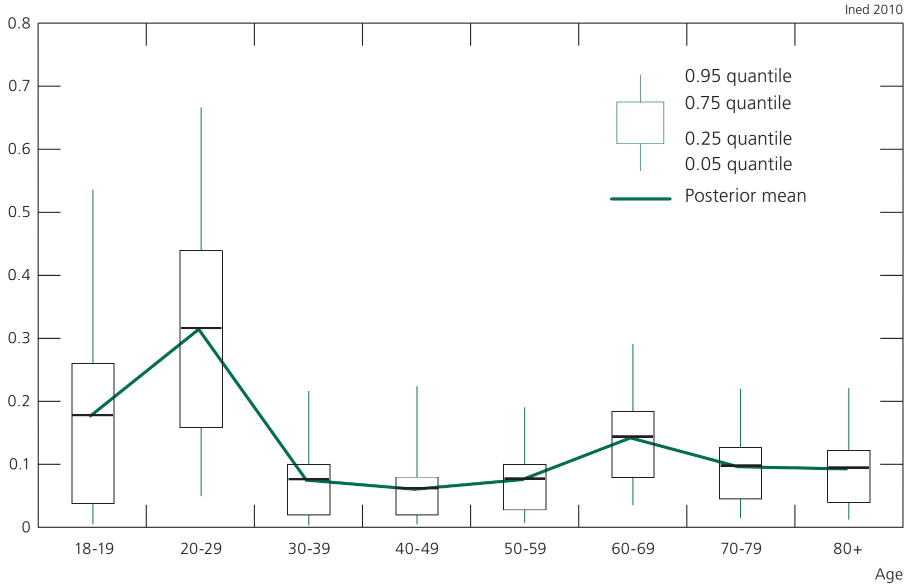
Source: Authors' calculations based on Séguy and Buchet (2010).

The posterior expectations ("Mean" Table 2) warrant two revisions. The first is a sharp upward revision of the mortality of the two youngest classes

---

(14) We thank Luc Buchet and Isabelle Séguy for supplying us with the data for this site. For more details, see Buchet (1978).

Figure 3. Frénouville example. Probability estimates using posterior mean and quantiles giving 90% and 50% credible intervals



Source: Authors' calculations based on Séguy and Buchet (2010).

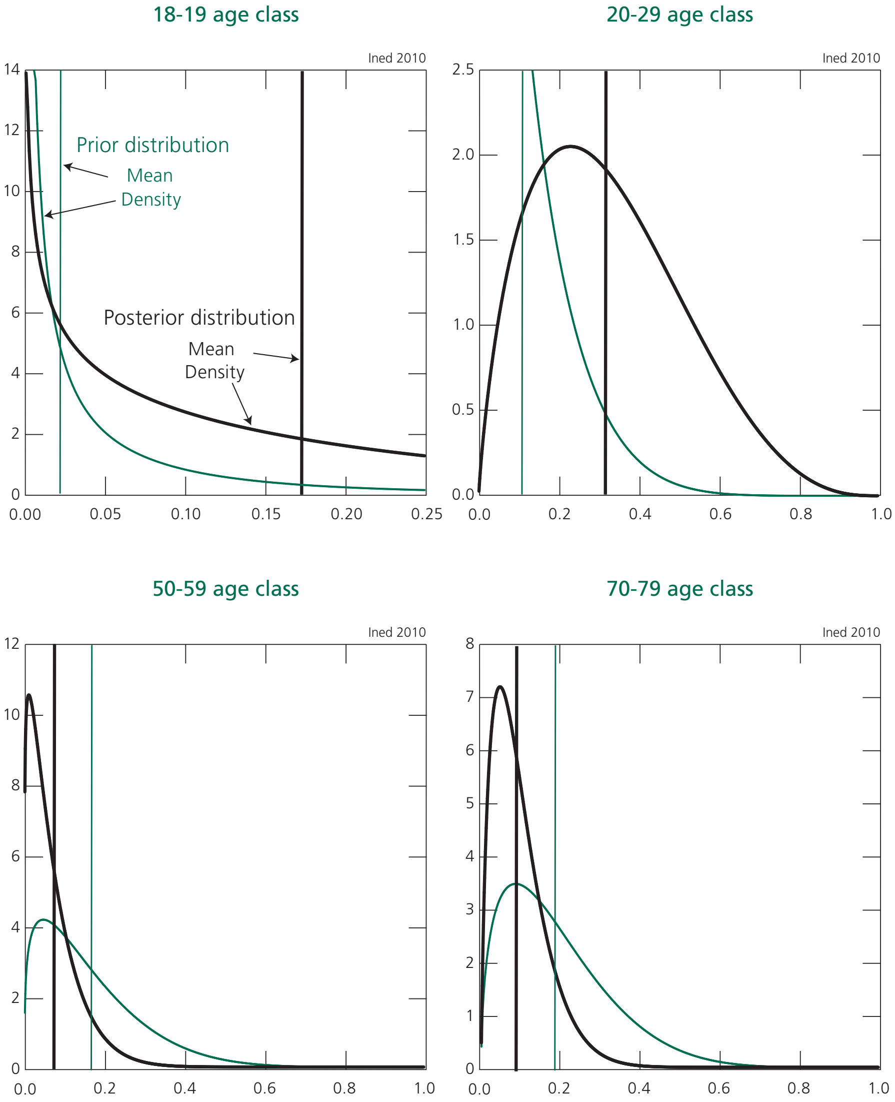
relative to the prior values; however, a high uncertainty persists, as witnessed by the standard deviations and the interquartile ranges of Table 2, and the credible intervals of Figure 3. The second is a downward revision for the other classes, with a lesser uncertainty.

These results can be explained by examining the charts in Figure 4, which give the following details for four of the age classes: beta approximation of posterior density (in black), prior density (in green), and posterior and prior means shown by black and green vertical lines, respectively. For the first class (ages 18-19), the densities are hard to read, as the distributions are concentrated on small values; clearly, however, the estimated probability significantly exceeds that of the pre-industrial standard, although the dissymmetry of the distribution may overstate the phenomenon when the estimation is based on the mean. The centre of the 50% credible interval (0.148) is below that mean, and the posterior distribution median is even more so (0.116). For the second class (ages 20-29), the estimate greatly exceeds the pre-industrial standard, with a major imprecision and a nearly symmetrical posterior distribution (hence the closeness of the mean, median, and centre of the inter-quartile interval).(15)

---

(15) For this class, and to a lesser extent for the first, note that the approximation via the beta distribution is relatively poor (an exception among all the cases we have examined). We have, however, kept it for simplicity's sake, as the discussion based on the exact distribution would have been entirely similar regarding the aspects addressed here.

Figure 4. Frénouville example. Prior distributions (thin green lines) and posterior distributions (thick black lines) for four age classes, with respective means (vertical lines)



Note: For the 18-19 class, the values on the x-axis range from 0 to 0.25 in order to improve legibility. The legends are the same for all four graphs.

Source: Authors' calculations based on Séguy and Buchet (2010).

For the oldest class (age 80+, not shown in Figure 4), the estimate is practically identical to the pre-industrial standard (the prior and posterior distributions are very close). For all the other classes, the estimated probabilities are smaller than the probabilities of the pre-industrial standard. For the 50-59, 60-69, and

70-79 classes (of which two are shown in Figure 4), the posterior densities are quite sharply “pointed”, a sign that the estimates using the posterior mean are reliable.

To conclude our analysis, here are the means and posterior standard deviations for the 14 age classes:

- means: (0.130 0.231 0.070 0.049 0.034 0.045 0.041 0.050 0.066 0.073 0.061 0.045 0.071)
- standard deviations: (0.135 0.135 0.078 0.057 0.038 0.036 0.043 0.035 0.039 0.047 0.055 0.044 0.036 0.051)

From these values, we can recalculate estimates for eight classes through groupings. To this end, we use the additivity of means and the classic formula for the variance of a sum: $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$ (in our presentation of the method, we saw that posterior covariance was easily calculable as well). We obtain:

- means: (0.130 0.301 0.084 0.079 0.090 0.139 0.105 0.071)
- standard deviations: (0.135 0.127 0.068 0.057 0.051 0.061 0.054 0.051)

Comparing these results with the estimates obtained directly for eight classes, we find a high level of consistency. The most distinct gap concerns the first class, whose estimated value is lower here (as noted earlier, there are grounds for concluding that this estimate may have been artificially high). For standard deviations, we can see that – apart from one instance of near-equality – this second estimation approach yields lower values, sometimes significantly so. (An interesting possible extension of this study would be to investigate whether this property is contingent or systematic and, if the latter, for what reason.)

Conclusion

Our simulations and the proposed archaeological applications have highlighted several advantages of our suggested method for estimating the age structure of past populations in cases where no recorded data on age at death are available, but where they are replaced by measures of biological indicators. Let us summarize these main advantages.

First, our method is simple to use(16) and very flexible: it is valid whatever the divisions into age classes and bone stages, and users’ prior knowledge can be very easily introduced. Yet despite its generality and simplicity, the method is effective: only some of the results obtained with the more elaborate Iterage algorithm can match it when it is employed with an “all-purpose” prior distribution. However, the latter method – developed by Bocquet-Appel and Bacro – is more complex to apply than ours and more limited in scope. Moreover,

---

(16) An R calculation programme (R Development Core Team, 2008) is available from the authors.

in cases where it is usable, the inputs needed to apply it can be incorporated into our own method in the form of a new prior distribution, making our method preferable.

Second, our method fits into a clear statistical environment that allows a proper validation of its theoretical properties. We would mention, in particular, the reliability of the credible intervals provided.

However, like any new method, it needs to be refined in the light of user feedback. Above all, therefore, we hope that paleodemographers will test it so as to complement our experience with their own, explore the application procedures in detail, and promote the necessary improvements.

Acknowledgements. We wish to thank Isabelle Séguy and Luc Buchet, as well as the three anonymous referees whose comments have been useful in preparing this final version. Of course, the authors remain wholly responsible for the content of this article and for any errors and omissions.

REFERENCES

BOCQUET-APPEL, J.-P., 2005, "La paléodémographie", in *Objets et méthodes en paléoanthropologie*, Dutour, O., Hubler, J.-J., and Vandermersh, B. (eds.), Paris: Comité des travaux historiques et scientifiques, pp. 271-313.

BOCQUET-APPEL, J.-P., 2008a, *La paléodémographie*, Paris: Éditions Errance, 192 pp.

BOCQUET-APPEL, J.-P. (ed.), 2008b, *Recent Advances in Paleodemography*, Dordrecht, Springer, 294 pp.

BOCQUET-APPEL, J.-P., BACRO, J.-N., 1997, "Brief communication: estimates of some demographic parameters in a Neolithic rock-cut chamber (approximately 2000 BC) using iterative techniques for aging and demographic estimators", *American Journal of Physical Anthropology*, 102, pp. 569-575.

BOCQUET-APPEL, J.-P., BACRO, J.-N., 2008, "Estimation of an age distribution with its confidence intervals using an iterative bayesian procedure and a bootstrap sampling approach", in *Recent Advances in Paleodemography*, Bocquet-Appel, J.-P. (ed.), Dordrecht, Springer, pp. 63-82.

BOCQUET-APPEL, J.-P., MASSET, C., 1996, " Paleodemography: expectancy and false hope", *American Journal of Physical Anthropology*, 99, pp. 571-583.

BONNEUIL, N. (2005), "Fitting to a distribution of deaths by age with applications to paleodemography. The route closest to a stable population", *Current Anthropology*, 46, pp. S29-S45.

BUCHET, L. (1978), "La nécropole gallo-romaine et mérovingienne de Frénouville (Calvados). Étude anthropologique", *Archéologie médiévale*, 8, pp. 5-54.

CAUSSINUS, H., 1965, "Contribution à l'analyse statistique des tableaux de corrélation", *Annales de la Faculté des Sciences de l'Université de Toulouse*, 29, pp. 77-182.

CAUSSINUS, H., COURGEAU, D., 2010, "Une nouvelle méthode d'estimation de la structure par âge au décès", in *Manuel de paléodémographie*, Séguy, I. and Buchet, L. (eds.), Paris, INED.

COURGEAU, D., 2010, "Critique des méthodes actuellement utilisées", in *Manuel de paléodémographie*, Séguy, I. and Buchet, L. (eds.), Paris, INED.

DEMING, W.E., STEPHAN, F.F., 1940, "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known", *Annals of Mathematical Statistics*, 11, pp. 427-444.

FRIDRIKSSON, A., 1934, "On the calculation of age distribution within a stock of cod by means of relatively few age-determinations as a key to measurements on a large scale", in *Rapports et procès-verbaux des réunions du Conseil permanent international pour l'exploration de la mer*, 86, pp. 1-14.

HASSELBLAD, V., 1966, "Estimation of parameters for a mixture of normal distribution", *Technometrics*, 8 (3), pp. 431-444.

HOPPA, R.D., VAUPEL, J.W. (eds.), 2002, *Paleodemography: Age Distributions from Skeletal Samples*, Cambridge (U.K.), Cambridge University Press, 260 pp.

HOWELL, N., 1976, "Toward a uniformitarian theory of human paleodemography", *Journal of Human Evolution*, 5, pp. 25-40.

JACKES, M., 2000, "Building the bases for paleodemographic analysis: adult age estimation", in *Biological Anthropology of the Human Skeleton*, Katzenberg, M. and Saunders, S. (eds.), New York, Wiley-Liss, pp. 417-466.

KIMURA, D.K., CHIKUNI, S., 1987, "Mixture of empirical distributions: an iterative application of the age-length key", *Biometrics*, 43, pp. 23-35.

KONIGSBERG, L.W., FRANKENBERG, S.R., 1992, "Estimation of age structure in anthropological demography", *American Journal of Physical Anthropology*, 89, pp. 235-256.

KONISGBERG, L.W., FRANKENBERG, S.R., 2002, "Deconstructing death in paleodemography", *American Journal of Physical Anthropology*, 117, pp. 297-309.

KONIGSBERG, L.W., HERRMANN, N.P., 2002, “Markov chain Monte Carlo estimation of hazard model parameters in paleodemography”, in *Paleodemography: Age Distributions from Skeletal Samples*, Hoppa, R.D. and Vaupel, J.W. (eds.), pp. 222-242.

KRUITHOF, R., 1934, “Telefoonverkeersreking”, *De Ingenieur*, 52 (8), pp. E15-E25.

LEONTIEF, W.W., 1949, *The Structure of the American Economy 1919-1939: An Empirical Application of Equilibrium Analysis*, New York, Oxford University Press, 264 pp.

MASSET, C., 1971, “Erreurs systématiques dans la détermination de l’âge par les sutures crâniennes”, *Bulletins et mémoires de la Société d’anthropologie de Paris*, 12 (7), pp. 85-105.

MASSET, C., 1973, “La démographie des populations inhumées. Essai de paléodémographie”, *L’Homme*, 7, pp. 95-131.

MASSET, C., 1982, *Estimation de l’âge au décès par les sutures crâniennes*, doctoral dissertation, Université de Paris VII, 302 pp.

MASSET, C., 1995, “Paléodémographie: problèmes méthodologiques”, *Cahiers d’anthropologie et biométrie humaine*, XII (1-2), pp. 27-38.

MÜLLER, H.-G., LOVE, B., HOPPA, R.D., 2002, “Semiparametric methods for estimating paleodemographic profiles from age indicator data”, *American Journal of Physical Anthropology,* 117, pp. 1-14.

R DEVELOPMENT CORE TEAM (2008), *R: A Language and Environment for Statistical Computing*, Vienna (Austria), R Foundation for Statistical Computing, ISBN 3-900051-07-0, URL http://www.R-project.org.

ROBERT, C., 2006, *Le choix bayésien. Principes et pratique*, Paris, Springer, 638 pp.

SÉGUY, I. AND BUCHET, L. (eds.), 2010, *Manuel de paléodémographie*, Paris, INED.

SÉGUY, I., BUCHET, L., BRINGÉ, A. et al., 2008, “Model life tables for pre-industrial populations. First applications in paleodemography”, in Bocquet-Appel, J.-P. (ed.), *Recent Advances in Paleodemography. Data, Techniques, Patterns*, Dordrecht, Springer, pp. 109-141.

THIONET, P., 1963, “Sur certaines variantes des projections du tableau d’échanges inter-industriels”, *Bulletin de l’Institut international de statistique*, XL(1), pp. 431-446.

THIONET, P., 1964, “Note sur le remplissage d’un tableau à double entrée”, *Journal de la Société de statistique de Paris*, 10-11-12, pp. 228-247.

TUGAULT, Y., 1970, “Méthode d’analyse d’un tableau ‘origine-destination’ de migrations”, *Population*, 25, pp. 59-68.

WILLEKENS, F., 1977, *The Recovery of Detailed Migration Patterns from Aggregate Data: An Entropy Maximizing Approach*, RM-77-58, Laxenburg (Austria), IIASA, 112 pp.

Henri CAUSSINUS, Daniel COURGEAU • **ESTIMATING AGE WITHOUT MEASURING IT: A NEW METHOD IN PALEODEMOGRAPHY**

To estimate the structure of past populations by age at death, with only biological indicators available, paleodemographers have developed several methods that rely on a reference population whose biological indicators and ages at death are known. First, we examine these approaches with their underlying assumptions, and show their weaknesses. To remedy these shortcomings, we propose a new statistical method that provides a more reliable estimate of the age distribution of deaths. It is a Bayesian method, whose principle and practical use involve choosing a prior distribution, determining a posterior distribution, and applying credibility intervals. A simulation-based comparison with earlier methods shows the clear superiority of our approach, which we then apply to actual archaeological data. The article concludes with an overview of the main advantages of the proposed method: flexibility and efficiency.

Henri CAUSSINUS, Daniel COURGEAU • **ESTIMER L'ÂGE SANS LE MESURER EN PALÉODÉMOGRAPHIE**

Pour estimer la structure par âge au décès des populations du passé, alors qu'ils ne disposent que d'indicateurs biologiques, les paléodémographes ont développé un certain nombre de méthodes, qui s'appuient sur une population de référence pour laquelle indicateurs biologiques et âges au décès sont connus. Nous présentons d'abord ici ces différentes approches avec leurs hypothèses sous-jacentes, et montrons leurs points faibles. Pour pallier ces inconvénients nous développons une nouvelle méthode statistique qui permet une estimation mieux assurée de la structure des décès par âge. Il s'agit d'une méthode bayésienne dont nous présentons d'abord le principe et l'utilisation pratique : choix de la loi *a priori*, calcul de la loi *a posteriori*, intervalles de crédibilité. La comparaison avec les méthodes antérieures, à l'aide de simulations, montre un net avantage de cette approche, que nous appliquons ensuite à des données archéologiques observées. Une vue synthétique des principaux avantages de la méthode proposée, souplesse et efficacité, conclut cet article.

Henri CAUSSINUS, Daniel COURGEAU • **ESTIMAR LA EDAD SIN MEDIRLA EN PALEODEMOGRAFÍA**

Para estimar la edad al morir en las poblaciones del pasado, no disponiendo que de indicadores biológicos, los paleodemógrafos han desarrollado un cierto número de métodos que se apoyan en una población de referencia para la cual los indicadores biológicos y la edad al morir son conocidos. En primer lugar, presentamos estos diferentes enfoques con sus hipótesis subyacentes, y mostramos sus puntos flacos. Para corregir estos inconvenientes desarrollamos después un nuevo método estadístico que permite una estimación más segura de la estructura de los muertos por edad. Se trata de un método bayesiano del cual presentamos el principio y su utilización práctica : elección de la ley *a priori*, cálculo de la ley *a posteriori*, intervalos de credibilidad. La comparación con los métodos anteriores, gracias a simulaciones, muestra claramente la ventaja de este enfoque, que aplicamos después a datos arqueológicos observados. Concluimos con una síntesis de las principales ventajas del método propuesto, es decir su flexibilidad y su eficacia.

Translated by Jonathan Mandelbaum.