



Henri CAUSSINUS* et Daniel COURGEAU**

Estimer l'âge sans le mesurer en paléodémographie

À quel âge mourraient les hommes autrefois ? Pour estimer l'âge au décès dans les populations du passé, sans état civil, on ne dispose le plus souvent que de squelettes. Les restes osseux et dentaires portent la trace du stade de croissance ou de vieillissement atteint par les individus au moment de leur décès, mais ils ne permettent pas d'estimer l'âge avec certitude. À partir de squelettes de personnes pour lesquelles on a mesuré ces paramètres biologiques et dont on connaît l'âge au décès, qui constituent une population de référence, des modèles statistiques permettent de proposer une estimation de la répartition des âges au décès des personnes dont les squelettes ont été retrouvés. Mais ceux-ci ne sont pas forcément représentatifs de l'ensemble des populations dont ils sont issus : la conservation des squelettes dépend des conditions d'inhumation ou des circonstances des décès. On peut contrôler ce problème en travaillant par exemple sur des groupes homogènes, comme les religieuses dans un couvent dont les données sont présentées ici à titre d'illustration. Reste la question de l'inférence statistique. Henri CAUSSINUS et Daniel COURGEAU décrivent d'abord les méthodes utilisées en paléodémographie et montrent leurs limites. Ils proposent ensuite une nouvelle méthode, fondée sur le principe de l'inférence bayésienne, et la comparent avec les méthodes usuelles pour mettre en évidence sa plus grande précision et sa souplesse d'utilisation.

L'âge, notion fondamentale en démographie, ne peut être directement mesuré pour la plupart des populations du passé car elles ne connaissaient pas l'état civil. On peut seulement l'estimer d'après des indicateurs biologiques de croissance pour les sujets immatures, ou de vieillissement pour les adultes, mesurés sur un petit nombre de squelettes appartenant à une population

* Institut de mathématiques de Toulouse, UMR CNRS 5219, Université de Toulouse, 31162 Toulouse, France.

** Institut national d'études démographiques, Paris.

Correspondance : Daniel Courgeau, 114-2 Chemin du Défends, 06250 Mougins, courriel : daniel.courgeau@wanadoo.fr

donnée, à l'aide de restes osseux ou dentaires. Ces indicateurs ne peuvent malheureusement que nous donner une fourchette assez large pour l'âge au décès d'un sujet, car il n'existe pas de relation liant âge et état osseux avec précision mais simplement une corrélation, malheureusement assez faible.

Pour avancer dans l'étude de cette corrélation, les paléodémographes ont depuis longtemps (Masset, 1971) utilisé des données dites « de référence », obtenues sur des sites où il a été possible d'avoir, pour chaque sujet, à la fois l'âge au décès et le (ou les) indicateur(s) biologique(s). Ces données sont présentées plus en détail dans la partie I, avec celles du site « cible » où l'on dispose seulement de données sur les indicateurs. Le problème statistique consiste à estimer la distribution des âges au décès sur le site cible à partir des données qui y sont observées et des données de référence. Diverses méthodes ont été proposées pour cela dans le passé. Nous rappelons les étapes les plus marquantes dans la partie II en nous focalisant sur l'approche la plus fréquente : le cas discret dans lequel les caractéristiques osseuses et les âges sont répartis en classes. Nous considérons que l'on observe un seul indicateur biologique, la généralisation à plusieurs indicateurs étant dans ce cas immédiate. La première hypothèse, qui se centre sur les probabilités conditionnelles des différents âges associées à chaque valeur de l'indicateur, est moins satisfaisante que la seconde, dite « hypothèse d'invariance », qui suppose que la distribution conditionnelle des indicateurs, à un âge donné, est constante dans le temps pour les périodes concernées, au moins en première approximation⁽¹⁾.

Bien que les performances des méthodes antérieurement proposées aient été progressivement améliorées, elles restent décevantes, donnant souvent des résultats qui sont de toute évidence aberrants par rapport aux connaissances a priori et au simple bon sens. En fait, il s'agit souvent de méthodes générales dont la plupart ne tiennent compte ni de l'ensemble des aspects aléatoires des données, ni des caractéristiques spécifiques du problème (on verra qu'une des méthodes largement utilisée vient de l'ichtyologie, avec un modèle de base qui est certes identique, mais qui reste trop général). Or, avec des échantillons qui sont souvent de petite taille pour un problème d'estimation intrinsèquement très instable, il est important de mettre en place une méthodologie qui incorpore au mieux l'ensemble des connaissances paléodémographiques a priori. Pour ce faire, le plus naturel est de recourir à une méthode bayésienne : c'est ce que nous proposons dans la partie III.

La méthode que nous proposons est bayésienne au sens couramment admis en statistique, alors que certaines méthodes antérieures sont dites bayésiennes de façon impropre, seulement parce qu'on y utilise quelque part la célèbre formule

(1) Cette hypothèse est vivement débattue depuis une quinzaine d'années au sein de la communauté des paléodémographes : « Bien que la possibilité d'une dérive séculaire des indicateurs biologiques d'âge ne puisse être écartée, les paléodémographes ont été conduits à la négliger, faute de pouvoir la mesurer, tout en espérant que les éventuelles divergences ne soient pas trop profondes » (Séguy et Buchet, 2010). Pour s'en prémunir, ils utilisent des populations de référence préindustrielles qui n'ont pas, ou peu, commencé leur transition démographique.

de Bayes. Rappelons en effet que, par opposition à une méthode fréquentiste dans laquelle les paramètres inconnus sont supposés fixés, une méthode bayésienne suppose que ces paramètres sont des variables aléatoires ; on se donne alors une loi *a priori* (c'est-à-dire avant observations) pour ces paramètres et on calcule la loi *a posteriori*, c'est-à-dire la distribution révisée pour un site cible donné en fonction des données observées sur ce site. On peut introduire dans la loi *a priori* des informations antérieures aux données proprement dites, en particulier le fait que les probabilités à estimer correspondent à une loi de mortalité (et, éventuellement, dans un environnement particulier connu). Nous verrons aussi que, dans notre problème, il est naturel de considérer que les données de référence fournissent une loi *a priori* pour certains paramètres qui sont le plus souvent considérés connus, de façon très abusive puisque ces données sont entachées d'erreurs d'échantillonnage (sans compter le caractère nécessairement approximatif de l'hypothèse d'invariance).

Nous verrons dans la partie IV que notre méthode se compare très favorablement aux méthodes antérieures. Nous traiterons enfin deux exemples dans la partie V afin d'illustrer sa mise en œuvre.

I. Données, formulation mathématique, notations

On dispose d'une population de référence, où l'on distingue c groupes d'âges et l stades pour l'indice osseux mesuré, et d'une population observée où l'on distingue ces mêmes l stades. Le tableau 1 présente les données : effectifs n_{ij} tirés de la population de référence, par classe d'âges, j , et stades, i , et effectifs m_i observés pour le stade i sur un site donné (population cible).

Tableau 1. Matrice de la population de référence par stade et par groupe d'âges et de la population observée par stade

	Population de référence									Population observée	
	Groupes d'âges (j)								Totaux par stade	Totaux par stade	
Stades (i)	n_{i1}	.	.	.	n_{ij}	.	.	.	n_{ic}	$n_{i.}$	m_i
	.				.				.		
	.				.				.		
	.				.				.		
	n_{i1}	.	.	.	n_{ij}	.	.	.	n_{ic}	$n_{i.}$	m_i
	.				.				.		
	.				.				.		
	.				.				.		
Totaux par âge	$n_{.1}$.	.	.	$n_{.j}$.	.	.	$n_{.c}$	$n_{..}$	m

La partie gauche du tableau 1 permet de calculer dans la population de référence, la fréquence de l'âge j , connaissant le stade i : $f_{j|i} = \frac{n_{ij}}{n_i}$, et la fréquence du stade i , connaissant l'âge, j : $f_{i|j} = \frac{n_{ij}}{n_j}$. Nous pouvons également calculer la fréquence par âge de la population de référence, $f_j = \frac{n_j}{n}$ et la fréquence par stade de la population de référence : $f_i = \frac{n_i}{n}$.

La partie droite du tableau 1 donne les fréquences par stade de la population observée : $\frac{m_i}{m}$.

À ces fréquences mesurées correspondent diverses probabilités inconnues. On désignera par p_{ij} la probabilités qu'un individu pris au hasard dans la population cible étudiée soit dans le stade i et dans la classe d'âges j ; la somme sur i des p_{ij} sera notée p_j ou simplement p_j (c'est la probabilité qu'un individu soit d'âge j) ; la somme sur j des p_{ij} sera notée p_i ou simplement π_i (c'est la probabilité qu'un individu soit de stade i) ; la probabilité conditionnelle du stade i connaissant l'âge j sera notée $p_{i|j}$. Ces diverses probabilités sont positives

et vérifient les relations : $\sum_i \pi_i = \sum_j p_j = 1$ et $\sum_i p_{i|j} = 1$ pour tout j . Elles sont liées d'autre part par la relation suivante :

$$\sum_j p_j p_{i|j} = \pi_i \quad \text{pour tout } i = 1, \dots, l \quad (1)$$

Selon les hypothèses faites, diverses solutions sont possibles pour estimer les probabilités p_j , c'est-à-dire obtenir la structure par âge de la population paléodémographique observée. Examinons les principales estimations proposées par le passé.

II. Les méthodes actuellement utilisées

Nous ne décrivons pas certaines méthodes de moindre utilité de nos jours, dont on trouvera une excellente présentation et critique dans Masset (1973). Nous ne présenterons ici que les approches actuellement utilisées par la majorité des paléodémographes, ainsi que leurs prolongements plus récents⁽²⁾.

(2) Nous ne traiterons pas le cas où l'on dispose de plusieurs indicateurs biologiques, ni celui où ces indicateurs sont non plus discrets mais continus (Konigsberg et Frankenberg, 1992). De même nous laisserons de côté l'utilisation de la théorie de la viabilité pour traiter ce problème (Bonneuil, 2005), où la distribution, imparfaitement observée, des décès par âge est considérée comme une cible dans un espace de grande dimension, et où on recherche la solution la plus proche d'une population stable (ainsi que son intervalle de confiance) et produisant cette distribution de décès par âge au bout de la période d'observation.

1. Tableaux de distance minimale par case

La première méthode consiste à estimer les cases d'une matrice de la population observée, dont chacune des cases est la plus proche possible de celle de la matrice de la population de référence, et dont les fréquences des stades sont celles de la population observée. Les fréquences par âge obtenues seront alors la solution du problème posé. Il est nécessaire dans ce cas de définir une distance globale entre les cases des deux matrices. La distance la plus couramment utilisée est une distance du χ^2 qui divise les carrés de chaque différence par l'effectif observé dans la population de référence⁽³⁾.

On montre aisément que, dans ce cas, la solution s'écrit :

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

Cette méthode suppose donc que les probabilités sont correctement estimées par les fréquences tant dans la population de référence que dans la population observée.

Elle est issue de nombreux travaux antérieurs sur d'autres sujets : initiée par Fridriksson (1934) qui travaillait sur des statistiques de poissons, puis par Kruithof (1937) qui travaillait sur les réseaux téléphoniques, elle a été reprise par des statisticiens (Deming et Stephan, 1940), des économistes (Leontief, 1941), et de nombreux autres chercheurs, dont Friedlander (1961), Thionet (1963, 1964), Caussin (1965), Tugault (1970), Willekens (1977), etc. Nombre d'auteurs l'appellent actuellement méthode IPFP (*Iterative Proportional Fitting Procedure*) ou méthode ALK (*Age Length Key*).

Cette méthode a été reprise en paléodémographie par Masset (1971), qui l'appelle méthode des vecteurs de probabilité, puis par Konigsberg et Frankenberg (1992), qui l'appellent méthode ALK.

Notons que la distribution ainsi calculée dépend fortement de la distribution des âges dans la population de référence et est « aplatie par l'influence de l'échantillon de référence » comme le dit Masset (1995). Ce qui est évident étant donné l'hypothèse selon laquelle chaque case de la matrice estimée doit être la plus proche possible de chaque case de la matrice de référence.

En fait, cette méthode introduit des artefacts en se centrant autant sur les probabilités en ligne que sur les probabilités en colonnes déduites de la matrice de référence présentée dans le tableau 1. Au contraire, il convient de ne tenir compte que de la seule hypothèse d'uniformité biologique (Howell, 1976), encore appelée « hypothèse d'invariance » (Müller *et al.*, 2002), selon laquelle, pour tout ossement d'âge au décès donné, la probabilité pour qu'un os soit classifié dans un stade donné ne dépend que de cet âge quelle que soit la population d'où cet os a été extrait. Cela conduit à ne considérer les données de

(3) Bien entendu, d'autres distances sont utilisables, par exemple une distance euclidienne, mais les résultats sont généralement proches.

référence qu'au travers des seuls « profils colonnes », c'est-à-dire des distributions conditionnelles des stades osseux pour chaque classe d'âges. D'où la recherche d'une autre méthode d'estimation qui se fonde sur cette seule hypothèse.

2. Tableaux de distance minimale par colonne

La seconde méthode cherche donc à estimer les colonnes d'un tableau qui soient aussi proches que possible de chaque colonne du tableau de référence. Pour ce faire, on peut utiliser une méthode itérative qui, partant d'une structure initiale quelconque ou uniforme \hat{p}_j^0 , estime la structure de la population observée, \hat{p}_j , par itérations successives, à l'aide de la formule de récurrence :

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

On effectue autant d'itérations que nécessaire pour que \hat{p}_j^{n+1} diffère de \hat{p}_j^n

d'une quantité aussi petite que l'on veut. Cet algorithme vise à obtenir des estimateurs de maximum de vraisemblance en supposant les $f_{i|j}$ fixés et les observations sur site de distribution multinomiale ; on montre qu'il fournit effectivement ces estimateurs au moins dans le cas d'un maximum régulier (annulation du gradient).

Cette méthode initialement proposée par les ichtyologues (Hasselblad, 1966 ; Kimura et Chikuni, 1987) sous le nom d'IALK (*Iterative Age Length Key*), a été mise en place en paléodémographie par Masset (1982) sous le nom de méthode des approximations successives pour éviter le résultat exagérément plat de la méthode des vecteurs de probabilité. Elle a été reprise par Konigsberg et Frankenberg (1992) sous le nom d'IALK. Ces deux approches, que nous appellerons pour simplifier américaine et française, ont fait l'objet de nombreuses controverses de 1992 à 2002⁽⁴⁾. Mais finalement elles se révélèrent pratiquement identiques⁽⁵⁾ (Konigsberg et Frankenberg, 2002).

Cette méthode nécessite que l soit supérieur ou égal à c , pour obtenir une solution unique, sinon le système (1) est indéterminé, admettant une infinité de solutions. Malheureusement, certains paléodémographes ne tiennent pas compte de cette condition, ce qui les conduit à des solutions non satisfaisantes. Ainsi Jackes (2000) tente d'estimer dix-sept classes d'âges avec seulement six stades, et obtient un nombre élevé de classes d'âges ayant des proportions nulles. Il en est de même pour Bocquet-Appel et Bacro (1997), qui estiment sept classes d'âges avec seulement six stades (Konigsberg et Frankenberg, 2002).

(4) Voir en particulier Bocquet-Appel et Masset (1996), qui appellent maintenant cette méthode IPFP bien qu'elle soit différente de ce que les statisticiens désignent sous ce nom.

(5) Pour plus de détails sur les différences entre ces méthodes voir Courgeau (2010).

De plus, cette méthode d'estimation peut conduire à des solutions clairement non satisfaisantes. Ainsi Masset (1982), bien que partant d'une population de référence comportant sept classes d'âges et sept stades et d'un vecteur de stades de la population observée sans élément nul, obtient une structure par âge de la population observée ayant quatre éléments nuls.

3. Méthodes plus récemment proposées

L'école américaine, dans un séminaire tenu à Rostock (Hoppa et Vaupel, 2002), propose d'introduire un âge continu plutôt que discrétisé, et d'utiliser une modélisation de la densité de probabilité de la population observée par un modèle biographique paramétrique : modèle de Gompertz (2 paramètres), de Gompertz-Makeham (3 paramètres), de Siler (5 paramètres), etc. Cela évite le problème des classes d'âges nulles. Ces méthodes restent cependant très proches de la méthode IALK, comme l'indiquent dans ce volume Konigsberg et Herrmann (2002) : « *nos méthodes habituelles s'accordent bien avec les approches adoptées pendant le séminaire de Rostock* »⁽⁶⁾.

Cependant ces méthodes introduisent nombre d'hypothèses supplémentaires que l'on n'a guère de moyens de vérifier : celle d'une population stationnaire ou stable pour que le modèle biographique s'applique aux conditions du moment ; celle d'une continuité dans la distribution d'un stade donné selon l'âge, qui conduit à des distributions différentes selon les méthodes utilisées. Enfin ces méthodes considèrent toujours que les probabilités sont correctement estimées par les fréquences.

Bocquet-Appel (2005, 2008a, 2008b) et Bocquet-Appel et Bacro (2008) proposent d'une part de ne plus considérer la population de référence comme parfaitement estimée par les fréquences : ils effectuent pour ce faire 1 000 tirages par la procédure *bootstrap*, dans chacun des groupes d'âges de la population de référence. D'autre part, ils réduisent l'ensemble des vecteurs de probabilités (p_1, \dots, p_c) à une famille (mélange de distributions de Gompertz-Makeham et de valeurs extrêmes) qu'ils construisent afin de représenter le plus possible de cas de mortalité ; dans cette famille de « candidats », ils cherchent le vecteur qui satisfait au mieux le système (1) lorsque π_i est remplacé par la fréquence observée correspondante et les p_{ij} par un des résultats du tirage *bootstrap* mentionné plus haut. Les 1 000 tirages fournissent ainsi 1 000 estimations, chacune égale à l'un des vecteurs candidats. Ils peuvent alors choisir pour estimation finale, soit le « meilleur » des 1 000 vecteurs obtenus, c'est-à-dire celui pour lequel la distance entre premier et second membre de l'équation (1) est minimale, soit plutôt (et c'est ce que nous ferons dans l'application de leur méthode dans la partie IV) la moyenne des 1 000 estimations obtenues. Cette estimation ponctuelle est assortie d'« intervalles de confiance » fondés sur les 1 000 résultats intermédiaires. Si cette méthode est intéressante dans son

(6) « Our current methods fit fairly comfortably within the approaches taken during the Rostock workshop. »

principe pour l'estimation ponctuelle (nous examinerons plus loin son efficacité), elle pose problème en revanche pour les intervalles de confiance qu'elle produit. Les auteurs n'en donnent en effet aucune validation théorique et le fait qu'ils ne tiennent pas compte des aléas incontournables sur les fréquences des stades observées les rend évidemment sujets à caution, certainement bien trop optimistes comme nous avons pu le vérifier par simulation.

III. Une nouvelle méthode d'estimation

Nous proposons une nouvelle méthode d'estimation, de nature proprement bayésienne. Nous voulons dire que, en opposition à certaines propositions antérieures parfois qualifiées de bayésiennes, car elles utilisent la « formule de Bayes » ou introduisent des *a priori* dans leur démarche⁽⁷⁾, la méthode que nous proposons est fondée sur les ingrédients typiques de la statistique bayésienne :

- des paramètres supposés aléatoires ayant une loi *a priori* au travers de laquelle on cherche à prendre en compte des caractéristiques connues indépendamment des données observées,
- le calcul de la loi de probabilité de ces paramètres conditionnellement aux observations, loi dite *a posteriori*, sur laquelle est basée l'inférence statistique.

1. Modèle et principe de la méthode

Il est naturel de considérer que les fréquences m_i ($i = 1, \dots, l$) observées sur le site pour les divers stades, sont les valeurs observées d'une distribution multinomiale dont les paramètres π_i sont liés aux p_j et aux p_{ij} selon le système (1). Nous passerons par ces derniers paramètres pour poursuivre la modélisation.

Notons G la densité *a priori* des paramètres p_{ij} , $i = 1, \dots, l$ et $j = 1, \dots, c$ (nous verrons comment l'exprimer au paragraphe suivant) et supposons que les paramètres p_j ($j = 1, \dots, c$) ont une densité *a priori* g (discutée aussi au paragraphe suivant) et sont indépendants des p_{ij} .

En notant M le vecteur des m_i , P le vecteur (matrice) des p_{ij} et p le vecteur des p_j , la densité conjointe de (M, P, p) sera alors f donnée par :

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

(7) La méthode proposée par Bocquet-Appel et Bacro (2008) utilise des vecteurs dits *a priori* pour tenir compte du fait que la loi de probabilité à estimer est une loi de mortalité ayant des caractéristiques nécessairement spécifiques, mais cela les conduit à réduire judicieusement l'espace paramétrique d'une méthode qui reste d'essence fréquentiste, plutôt qu'à adopter un point de vue strictement bayésien. C'est pour éviter toute confusion que nous désignons ici ces vecteurs sous le nom de candidats ; cela d'autant plus que nous proposerons de les utiliser pour construire une vraie loi *a priori* de notre méthode bayésienne.

La densité marginale du couple (M, p) est :

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

et la densité marginale de M est :

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

les intégrales étant prises sur les domaines de variation de P et/ou p , qui sont un simplexe (pour p) ou un produit de simplexes (pour P).

La densité conditionnelle de p sachant M est donc :

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

C'est la densité *a posteriori* des p_j ($j = 1, \dots, c$), outil de base de l'estimation bayésienne.

On aura par exemple pour moyenne *a posteriori* de p_j :

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

De façon plus générale, l'espérance conditionnelle à M d'une fonction φ de p sera donnée par :

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

On obtient ainsi, par exemple, le moment d'ordre k de p_j avec $\varphi(p) = p_j^k$. En prenant pour $\varphi(p)$ la fonction qui vaut 0 pour $p_j > x$ et 1 pour $p_j < x$ (indicatrice de l'événement $p_j < x$), on exprime la fonction de répartition *a posteriori* de p_j au point x .

Les diverses intégrales de l'équation (2) peuvent être évaluées par une méthode de Monte Carlo (Robert, 2006) de la façon suivante :

Notons $X = (X_1, \dots, X_c)$ un vecteur aléatoire suivant une loi de densité g et Y une famille de c vecteurs $Y_j = (Y_{1j}, \dots, Y_{lj})$ ($j = 1, \dots, c$), dont la loi conjointe est indépendante de X et admet la densité G . On vérifie que l'équation (2) est équivalente à :

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

Générons S ensembles indépendants de tels vecteurs aléatoires (X, Y) ; en notant par s ($s = 1, \dots, S$) les différentes répétitions ; en vertu de la loi des grands nombres, si S est assez grand, l'expression ci-dessus est approchée par :

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

Cela fournit, en particulier, l'espérance *a posteriori* de chaque p_j ($j = 1, \dots, c$) qui peut être prise comme estimation ponctuelle, ou la variance *a posteriori* utile pour caractériser la précision de cette estimation. Si on le souhaite, on peut évaluer des moments croisés selon le même principe, par exemple la matrice des covariances de la distribution *a posteriori* des p_j . Enfin, la fonction de répartition *a posteriori* d'un p_j permet, par exemple, de calculer des intervalles contenant ce p_j avec une probabilité donnée, intervalles dits de crédibilité correspondant, dans le cadre bayésien, aux intervalles de confiance du cadre classique.

Il est parfois préconisé de prendre le mode de la distribution *a posteriori* comme estimation ponctuelle des paramètres au lieu de l'espérance *a posteriori*. Nous préférons l'espérance pour plusieurs raisons. La première est que, notant \hat{p}_j , l'estimation ainsi obtenue, elle minimise le coût moyen pour la fonction de perte $\sum_j (\hat{p}_j - p_j)^2$ naturelle dans notre problème puisqu'elle pénalise les erreurs en fonction de leurs amplitudes (au contraire le mode est optimal pour une perte nulle si l'estimation est extrêmement proche de la vraie valeur, constante sinon, ce qui ne semble pas convenir ici) ; à cette raison essentielle on peut ajouter la facilité de calcul et le fait que la densité *a posteriori* peut ne pas être bornée et présenter un mode infini pour une ou plusieurs probabilités estimées nulles, conduisant à un résultat peu réaliste.

2. Utilisation pratique

Choix des lois a priori

La densité G

L'information sur les probabilités conditionnelles p_{ij} vient exclusivement des données de référence ; si celles-ci sont des données brutes simplement obtenues en relevant des fréquences de stades sur un échantillon de squelettes d'âges connus, il est naturel d'admettre que, pour chaque classe d'âges j ($j = 1, \dots, c$), les fréquences n_{ij} sont les valeurs observées d'une distribution multinomiale de total n_j et probabilités p_{ij} ($i = 1, \dots, l$). Adoptant une loi *a priori* pour les p_{ij} , on en déduit une loi *a posteriori*, loi conditionnelle aux données de référence, qui est à son tour prise comme loi *a priori* des p_{ij} dans le modèle final. Comme il y a peu d'information complémentaire sur ces probabilités p_{ij} au-delà de celle qui est contenue dans les données de référence il est naturel d'adopter comme loi *a priori* des p_{ij} , pour chaque j , une loi uniforme. Pour un j donné, on trouve alors pour loi *a posteriori* des p_{ij} une loi de

Dirichlet de paramètres $\alpha_{ij} = n_{ij} + 1 (i = 1, \dots, l)$ ⁽⁸⁾. La densité G est le produit de ces c densités de Dirichlet soit :

$$G(P) = \frac{\prod_j \Gamma(\alpha_j)}{\prod_i \prod_j \Gamma(\alpha_{ij})} \prod_i \prod_j p_{ij}^{\alpha_{ij} - 1}$$

En pratique, le caractère multinomial des données de référence n'est invoqué que pour fixer les idées : il est seulement indicatif et non indispensable pour aboutir à cette distribution *a priori* G .

Il est envisageable de raffiner le choix de G mais cela ne semble pas conduire à des améliorations notables (on peut se reporter à Caussinus et Courgeau (2010) pour quelques indications plus détaillées) ; nous ne développerons donc pas cela ici.

La densité g .

Le choix de la loi *a priori* des paramètres p_j est plus délicat. Comme la « classe » de lois dans laquelle il convient de chercher la loi *a priori* ne semble pas s'imposer de façon particulière, le plus naturel est d'opter pour une loi de Dirichlet qui est bien adaptée à des vecteurs de probabilité. Se pose alors le seul problème du choix des paramètres, disons $(\beta_1, \dots, \beta_c)$, de cette loi. En l'absence d'information particulière, on peut, comme plus haut, opter pour une loi uniforme et prendre $\beta_j = 1$ pour tout j . Un tel choix permet de rester « neutre » et peut se justifier dans certains cas. Il donne d'ailleurs des résultats raisonnables sur des exemples simples. Cependant, en paléodémographie, d'autres choix sont vraisemblablement plus judicieux car certaines indications sont naturellement disponibles. On peut, par exemple, partir d'une loi de mortalité « standard » pour laquelle on calcule les probabilités des diverses classes d'âges. Celles-ci sont alors prises comme moyennes de la loi *a priori*, ce qui donne les paramètres β_j à un coefficient de proportionnalité près, c'est-à-dire les β_j / β_* , où β_* est la somme sur $j = 1, \dots, c$ des β_j . Reste à choisir β_* , c'est-à-dire, en pratique, les variances de la loi *a priori* ; pour cela, il faut noter que ces variances doivent être relativement grandes pour exprimer que les moyennes *a priori* sont peu fiables et que la loi *a priori* ne doit pas jouer un rôle prépondérant ou, autrement dit, que la famille des possibilités envisagées recouvre un large domaine ; il semble donc que les β_* doivent être assez petits,

(8) Le vecteur aléatoire $X = (X_1, \dots, X_k)$ suit une loi de Dirichlet de paramètre

$a = (a_1, \dots, a_k)$ s'il admet la densité $d(x) = \frac{\Gamma(a)}{\prod_{i=1}^k \Gamma(a_i)} \prod_{i=1}^k x_i^{a_i - 1}$, sur le simplexe D défini par

$x = (x_1, \dots, x_k) \in D \Leftrightarrow x_i > 0$ pour tout $i = 1, \dots, k$ et $\sum_{i=1}^k x_i = 1$. Pour les propriétés de cette

loi, voir par exemple Robert (2006).

par exemple inférieurs à 1 ou guère supérieurs. Quelques simulations ont montré (voir Caussin et Courgeau, 2010, pour un compte rendu succinct) qu'un critère simple et efficace consiste à prendre $\beta_j = c$, comme dans le cas de la loi *a priori* uniforme. C'est ce que ferons ici.

Le standard utilisé sera le standard préindustriel (Séguy *et al.*, 2008 ; Séguy et Buchet, 2010) qui peut être celui des hommes, des femmes ou des deux sexes réunis selon les cas. Lorsque l'on utilisera notre méthode avec cette loi *a priori*, nous la noterons BayesPI. Dans certaines situations, il est possible que le paléodémographe possède des informations particulières. Ainsi, dans l'exemple du cimetière monastique de Maubuisson, nous pourrions tenir compte du fait qu'il s'agit de femmes *a priori* en meilleure santé que la moyenne de la population et non soumises à certains risques importants de mortalité pour les plus jeunes, en particulier la mortalité en couche.

Le principe de choix ci-dessus de la loi *a priori* peut être étendu de diverses façons. Par exemple, au lieu de choisir une loi standard de mortalité comme base de construction de la loi *a priori*, on peut choisir le mélange de deux lois « standards », ce qui conduira au mélange de deux lois de Dirichlet ; ce pourrait être le mélange (en proportions judicieusement choisies) d'une loi de mortalité banale (attrition) et d'une loi de mortalité catastrophique.

Des approches assez différentes sont évidemment possibles, par exemple, un peu dans l'esprit des propositions de Bocquet-Appel et Bacro (2008), on peut définir la loi *a priori* comme une loi uniforme sur un ensemble discret de lois correspondant à des lois types de mortalité. C'est une solution beaucoup plus lourde à mettre en œuvre que la précédente, mais elle devient d'utilisation aisée lorsque l'effort de construction de tels ensembles de vecteurs a déjà été fait. Nous l'utiliserons ci-dessous (sous le nom de BayesUnif) et la comparerons à la méthode de Bocquet-Appel et Bacro puisqu'elle est justement facile à mettre en œuvre lorsque cette dernière méthode est utilisable. Sur le plan technique, on peut noter qu'une des intégrales définissant la loi *a posteriori* est maintenant une somme finie, ce qui permet d'alléger les calculs par la méthode de Monte Carlo, moyennant un peu de mathématique : avec les notations du paragraphe III.1, l'élément X n'a plus besoin d'être simulé, mais seulement l'élément Y .

Loi a posteriori et intervalles de crédibilité

On a vu précédemment comment calculer la fonction de répartition *a posteriori* de chaque p_j point par point. Cela permet de déterminer des intervalles α -crédibles (Robert, 2006, p. 278) dans lesquels un paramètre p_j se trouve avec la probabilité $1-\alpha$ conditionnellement aux observations. Le calcul de la fonction de répartition *a posteriori* se fait point par point et s'avère donc laborieux. Une autre façon de procéder, à titre heuristique, est d'approcher la densité *a posteriori* de chaque p_j par une densité Bêta de même moyenne et même variance ; un certain contrôle de la qualité d'approximation peut être exercé au moyen des moments d'ordre supérieur : vérifier par exemple la proximité des moments d'ordre 3 et 4 de cette loi Bêta avec les moments correspondants

de la « vraie » loi *a posteriori*, facilement calculables par simulation comme on l'a vu plus haut. Dans tous les exemples traités, les fonctions de répartition *a posteriori* ont été calculées de façons exacte et approchée et l'approximation s'est toujours avérée convenable, voire excellente. Un intervalle α -crédible approché peut être donné pour chaque p_j par l'intervalle entre le quantile $\alpha/2$ et le quantile $1 - \alpha/2$ de sa loi *a posteriori* ainsi approchée. Ce n'est pas en général l'intervalle le plus court possible (HPD) mais il est raisonnable en pratique. Notons qu'il est formellement déconseillé de prendre un intervalle de type « moyenne plus ou moins écart type (ou 2 écarts types) » car la distribution *a posteriori* est en général très dissymétrique.

Dimension du tableau de données

Le système (1) de la partie I est indéterminé si le nombre l de lignes (stades) est inférieur au nombre c de colonnes (âges). En d'autres termes, les paramètres d'intérêt ne sont pas identifiables en ce sens que plusieurs valeurs conduisent à la même distribution des échantillons observables. La méthode bayésienne permet de tourner la difficulté puisque l'on part d'une loi *a priori* qu'il s'agit seulement de faire évoluer au moyen des données. La loi *a posteriori* oriente alors vers une distribution des paramètres inconnus, ce qui est tout à fait compatible avec le fait qu'ils ne sont pas complètement déterminés. On peut donc utiliser cette méthode avec $l < c$; évidemment, la distribution *a posteriori* pourra être relativement dispersée, ne faisant ainsi que tenir compte de l'indétermination inhérente à la situation.

IV. Comparaison aux principales méthodes antérieures

Proposant une méthode nouvelle, il est d'abord essentiel de comparer son efficacité à celle des principales méthodes actuellement utilisées. Outre les difficultés mathématiques qu'une étude théorique soulèverait, il sera toujours difficile de comparer des techniques fondées sur des paradigmes différents : fréquentiste pour les méthodes antérieures, bayésien pour la nôtre. Nous procéderons donc par simulation. Celles-ci seront effectuées dans les conditions suivantes avec lesquelles nous essayons d'approcher au mieux la réalité.

- On se donne un vecteur de probabilités $p_j (j = 1, \dots, c)$ qui constituera la valeur cible à estimer et une matrice de référence comme celle du tableau 1.

- Cette matrice de référence étant un échantillon plutôt qu'une population, on considère que les « vraies » probabilités p_{ij} ne sont pas exactement les quantités $f_{ij} = \frac{n_{ij}}{n_j}$ mais que, pour chaque j , on doit considérer ces f_{ij} comme les probabilités d'une distribution multinomiale de total $n_{.j}$. On tire donc au hasard une nouvelle matrice de référence selon ce principe (c'est le même principe que les tirages *bootstrap* que l'on trouve dans la méthode proposée par Bocquet-Appel et Bacro, 2008). Cette procédure permet en même temps

de tenir compte du fait que, en dehors même des incertitudes de l'échantillonnage, la matrice des p_{ij} réelle peut ne pas être celle de la distribution de référence mais se trouver seulement dans un certain voisinage de celle-ci.

- Les p_j donnés et les p_{ij} ainsi simulés fournissent les probabilités π_i des divers stades au moyen du système (1). On tire alors les effectifs m_i de ces stades au hasard selon une loi multinomiale à l catégories, de probabilités π_i , $i = (1, \dots, l)$ et de total t que l'on se fixe.

- On fait R répétitions indépendantes pour chacune desquelles on applique les diverses méthodes d'estimation à comparer. Bien entendu, ces méthodes sont appliquées avec les seuls moyens dont on dispose réellement, c'est-à-dire en particulier avec les données de référence observées et non les données simulées. Le vecteur des p_j peut être choisi identique pour les R essais si l'on veut préciser l'efficacité des méthodes dans un cas donné. Il peut aussi varier d'un essai à l'autre pour essayer de couvrir une plage raisonnable de valeurs possibles, ce qui permet une étude plus globale à moindres frais. Pour chaque essai, on donne deux types d'écart entre les vecteurs cibles et les vecteurs estimés par les diverses méthodes comparées. Le premier écart est le carré de la distance euclidienne, soit $\sum_j (\hat{p}_j - p_j)^2$; le deuxième écart pondère les divers

termes à la manière d'un chi-deux, soit $\sum_j \frac{(\hat{p}_j - p_j)^2}{p_j}$. On fait la moyenne des termes obtenus sur les R essais et on multiplie par 100 pour obtenir des « pourcentages unitaires ». Les écarts ainsi calculés seront respectivement appelés a (absolu) et χ (de type chi-deux)⁽⁹⁾. D'autres écarts ont été considérés, en remplaçant par exemple les carrés par des valeurs absolues ou en comparant les fréquences cumulées ; ils fournissent des résultats tout à fait analogues à ceux présentés ici et ne seront donc pas donnés pour ne pas alourdir l'exposé.

Le modèle de simulation ainsi mis en forme est sensiblement différent de celui que nous avons utilisé dans Caussin et Courgeau (2010). Celui que nous venons d'exposer est un peu plus complexe mais semble plus proche de la réalité. Cependant, les résultats que nous donnons ci-dessous et ceux de Caussin et Courgeau (2010) sont dans l'ensemble très concordants.

On a vu dans la partie II que la méthode ALK était déjà bien reconnue comme peu performante. Nous n'y reviendrons pas ici. Nous considérerons en revanche la méthode IALK. En fait, cette méthode est une méthode de maximum de vraisemblance appliquée à une distribution multinomiale dont les probabilités π_i sont liées aux paramètres à estimer p_j par le système (1) dans lequel les p_{ij} sont supposés connus (déduits des données de référence) ; en pratique, nous avons remplacé l'algorithme d'origine (très lent) par un algorithme classique de recherche d'optimum : procédure constrOptim du logiciel R (R

(9) La dénomination adoptée vient de l'analogie avec un χ^2 car les écarts observés sont ici approximativement divisés par leur écart type.

Development Core Team, 2008).

Nous considérons ensuite la méthode récente de Bocquet-Appel et Bacro (2008). Pour pouvoir utiliser cette méthode, il est nécessaire de se placer dans des situations pour lesquelles elle est applicable, c'est-à-dire pour lesquelles un fichier de vecteurs candidats a été fourni par les auteurs. Nous avons choisi une division en 7 classes d'âges décennales de 20 à 90 ans, la dernière classe pouvant s'interpréter comme 80 ans et plus, la méthode étant alors utilisée avec l'ensemble de 756 vecteurs figurant dans leur fichier ProbAtri20-90.txt (modèles de lois de probabilité pour la mort normale préindustrielle dite attritionnelle⁽¹⁰⁾).

Nous utilisons enfin notre méthode avec deux versions de la loi *a priori*. La première version, facilement utilisable quelles que soient les subdivisions en classes est une loi *a priori* de Dirichlet avec les paramètres correspondant au standard préindustriel comme il est indiqué plus haut. La seconde version est une loi *a priori* uniforme sur l'ensemble des 756 vecteurs du fichier ProbAtri20-90.txt ; cette façon de procéder est aisée dès que la méthode de Bocquet-Appel et Bacro est envisageable. Dans chaque cas, puisqu'il s'agit de comparer notre méthode avec des méthodes fréquentistes, nous avons réduit sa portée à l'estimation ponctuelle au moyen de l'espérance *a posteriori*. Il faut cependant noter qu'elle fournit d'autres éléments d'analyse comme on le verra dans les exemples de la partie V.

On a d'abord fait quelques simulations avec des vecteurs à estimer p_j ($j = 1, \dots, c$) fixés, représentant des distributions susceptibles d'être réalistes, mais suffisamment éloignées du standard préindustriel pour ne pas avantager notre méthode de façon inconsidérée⁽¹¹⁾. Il a été vérifié que la méthode IALK donnait des résultats beaucoup moins bons que notre méthode et celle de Bocquet-Appel et Bacro. Par rapport à la méthode de Bocquet-Appel et Bacro, la nôtre semblait donner des résultats proches, mais plutôt meilleurs (particulièrement pour des petits échantillons), lorsqu'elle était appliquée avec *a priori* de Dirichlet, nettement meilleurs dans tous les cas lorsqu'elle était menée avec la seconde loi *a priori* (uniforme sur les vecteurs candidats) mentionnée ci-dessus. C'est donc la comparaison entre ces deux dernières méthodes qu'il importait de mener avec le plus grand soin. Pour être certain de ne pas défavoriser la méthode de Bocquet-Appel et Bacro, nous avons fait les R répétitions en prenant chaque fois pour vecteur cible un des vecteurs *a priori* du fichier ProbAtri20-90.txt par tirage au hasard avec équiprobabilités. Il y a toujours 7 classes d'âges pour les raisons indiquées plus haut, mais nous avons fait varier, d'une part la taille m de l'échantillon de site (successivement 25, 50, 75 et 100 qui sont des ordres de grandeur usuels), d'autre part le nombre de classes de stades osseux retenues (successivement $i = 5, 7$ et 8). Dans chaque exemple, le nombre de répétitions est $R = 1\,000$.

(10) Voir le logiciel *Iterage* sur le site <http://www.evolhum.cnrs.fr/bocquet/index.html>

(11) À titre d'exemple, un des vecteurs considérés est : (0,20 0,15 0,10 0,20 0,20 0,10 0,05).

Figure 1. Écart entre vecteurs estimés et vecteurs cible pour les méthodes IALK, Bocquet-Appel et Bacro, Bayes avec *a priori* du standard préindustriel (PI) et Bayes avec *a priori* uniforme.

Figure 1A. Critère a

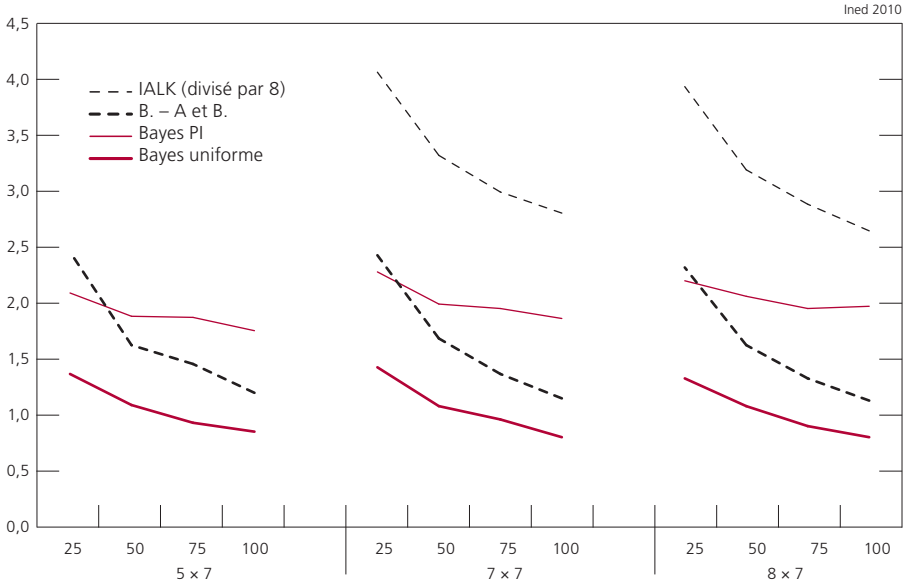
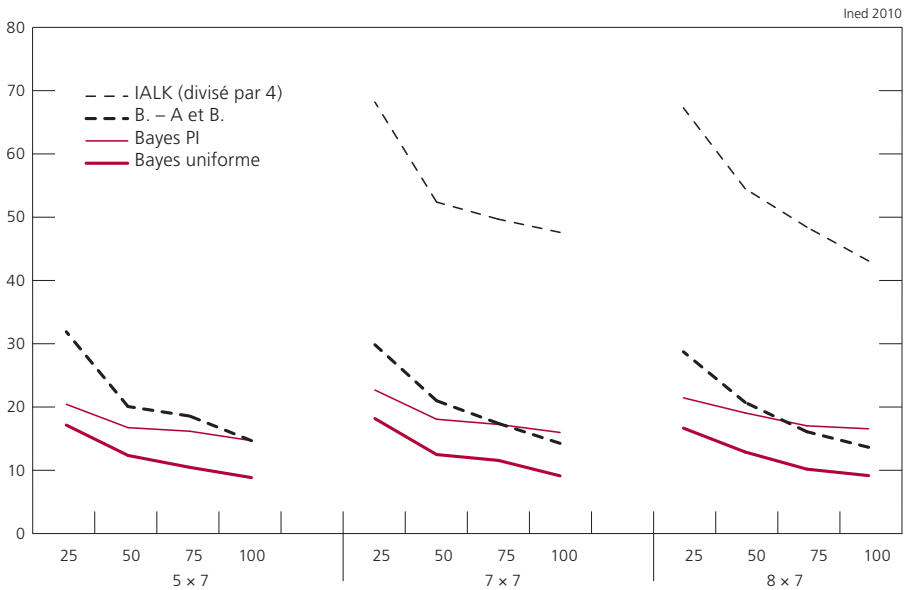


Figure 1B. Critère χ



Note : pour la méthode IALK les valeurs de a ont été divisées par huit, et les valeurs de χ ont été divisées par quatre.

Source : calculs des auteurs à partir de Masset (1982), et Séguéy et Buchet (2010).

Les trois matrices de référence utilisées sont celles de « Lisbonne hommes » déduites des trois collections portugaises (Ferraz de Macedo de Lisbonne, Coïmbra, musée du Bocage de Lisbonne), sur lesquelles on a mesuré l'état de synostose des sutures crâniennes selon 42 stades⁽¹²⁾. Elles comportent toutes 7 groupes d'âges décennaux à partir de 20 ans et les regroupements suivants de stades : 0-8, 9-15, 16-21, 22-28, 29-41 pour la première, 0-4, 5-7, 8-12, 13-18, 19-23, 24-34, 35-41 pour la seconde, avec éclatement de la classe 13-18 en deux : 13-15 et 16-18 pour la troisième.

Le vecteur des β_j est déduit du standard « préindustriel Hommes », soit : (0,70 0,77 0,98 1,19 1,47 1,33 0,56).

Les valeurs obtenues pour les deux critères sont données dans la figure 1 avec, de haut en bas, les critères a et χ . La méthode IALK ne peut pas être utilisée pour $l < c$, soit ici pour la subdivision des stades en 5 classes.

Les résultats de la figure 1 montrent d'abord la très faible efficacité de la méthode IALK, en remarquant que nous avons dû diviser ces écarts par huit pour les valeurs de a et par quatre pour les valeurs de χ pour rendre la figure plus lisible. Outre les écarts importants mentionnés dans la figure, on peut noter que les estimations fournies sont presque toujours sur la frontière (au moins un des \hat{p}_j est nul). Toutes les méthodes s'améliorent de façon logique lorsque la taille de l'échantillon augmente. Il est plus inattendu de voir que les performances de la méthode de Bocquet-Appel et Bacro s'améliorent légèrement lorsque davantage de classes de stades sont considérées, alors que le contraire est observé pour notre méthode, quelle que soit la loi *a priori*. Avec la loi *a priori* de Dirichlet déduite du standard préindustriel, les performances de la méthode bayésienne s'avèrent dans l'ensemble proches de celles de la méthode de Bocquet-Appel et Bacro bien que le choix des vecteurs cibles ne puisse qu'être favorable à cette dernière comme nous l'avons déjà dit. La comparaison dépend sensiblement de l'indice considéré, laissant supposer que notre méthode est plutôt plus performante dans l'estimation des petites probabilités et moins dans l'estimation des plus grandes ; par ailleurs, elle est meilleure pour les petits échantillons et perd son avantage pour les plus grands. Enfin, si la méthode bayésienne est utilisée avec la loi *a priori* uniforme sur les vecteurs candidats de la méthode de Bocquet-Appel et Bacro, elle fournit des résultats nettement meilleurs dans tous les cas considérés. On peut résumer de la façon suivante :

- la méthode IALK doit être définitivement écartée ;
- dans une situation où une famille de vecteurs candidats n'a pas été établie, notre méthode est très facile à appliquer avec une loi *a priori* de Dirichlet et a une efficacité comparable à celle de Bocquet-Appel et Bacro ; elle peut donc éviter le travail considérable d'élaboration de cet ensemble de vecteurs,

(12) Pour plus de détails sur ces collections et les mesures effectuées, voir Masset (1982), et Séguy et Buchet (2010) pour les tableaux détaillés.

particulièrement si l'échantillon sur le site cible est petit (on verra en outre dans la partie suivante que l'on peut parfois jouer facilement sur les paramètres de cette loi de Dirichlet pour en améliorer encore les performances) ;

- si un ensemble de vecteurs candidats a été établi pour appliquer la méthode de Bocquet-Appel et Bacro, il semble préférable de l'utiliser pour déterminer la loi *a priori* de notre méthode que pour appliquer la méthode de ces auteurs.

V. Applications archéologiques

Nous traitons maintenant deux exemples dans le but de montrer divers aspects de la méthode que nous proposons : sa mise en œuvre, avec en particulier le choix de la loi *a priori*, et les informations qu'elle est susceptible de fournir, en matière d'estimation ponctuelle et au-delà.

1. Les religieuses de Maubuisson (XVII^e-XVIII^e siècles)⁽¹³⁾

Nous considérons d'abord l'exemple du cimetière monastique de l'abbaye royale de Maubuisson où 162 religieuses cisterciennes ont été inhumées et les corps de 37 d'entre elles ont été exhumés pour une mesure des stades de synostose crânienne. Ces dames, de haute noblesse pour la plupart, ont bénéficié durant leur enfance et leur adolescence de conditions de vie très privilégiées, et la vie monacale, bien que rude à certains égards, les met également à l'abri des accidents auxquels sont exposées leurs contemporaines laïques au cours de leur vie féconde (Séguy et Buchet, 2010).

Nous adoptons un découpage en 7 tranches d'âges (décennales de 20-29 ans à 70-79 ans et une classe ouverte 80 ans ou plus), et 7 stades osseux. Les fréquences observées pour ces divers stades sur un échantillon de 37 crânes sont (6 2 4 5 3 9 8).

Nous avons des informations *a priori* importantes sur ce site, d'autant plus utiles que l'échantillon observé est de taille très modeste. Il s'agit de religieuses, donc de femmes, en principe âgées de plus de 20 ans. Cela conduit au choix des classes d'âges mentionné plus haut et au choix de données de référence spécifiques, ici « Lisbonne femmes » voir (Séguy et Buchet, 2010). Si nous nous en tenons à cette information, nous prendrons pour les 7 paramètres à estimer une loi de probabilité *a priori* de Dirichlet, dont les paramètres β_j sont proportionnels aux valeurs du standard de mortalité préindustrielle (femmes) et de somme 7, soit (0,70 0,77 0,84 1,05 1,47 1,47 0,70). C'est la première estimation que nous présenterons, à titre de comparaison. Mais le fait qu'il s'agisse de religieuses donne une information supplémentaire : comme nous l'avons dit plus haut, lors de leur admission, pour de multiples raisons ces femmes sont certainement en meilleure santé que la moyenne de la population ; elles

(13) Nous remercions Luc Buchet et Isabelle Séguy de nous avoir fourni les données de site. Voir pour plus de détails Séguy et Buchet (2010).

sont ensuite préservées de plusieurs risques importants de mortalité, en particulier de la mortalité en couche. Si nous évaluons *a priori* l'influence de ces facteurs comme réduisant la mortalité des 20-29 ans d'un peu plus de 50 % et la mortalité des 30-39 ans d'un peu moins de 50 %, nous sommes amenés à remplacer les paramètres de la loi *a priori* par : (0,30 0,40 0,84 1,05 1,47 1,47 0,70), ou plutôt par les valeurs proportionnelles (0,337 0,449 0,944 1,180 1,652 1,652 0,786) de somme 7 conformément aux recommandations du paragraphe III.2.1. À partir de cette loi *a priori*, nous proposerons une seconde estimation. De prime abord, c'est celle qui nous semble devoir être retenue dans la pratique, nous verrons en quoi les résultats obtenus le confirment ; nous verrons aussi comment il est possible de l'affiner.

Finalement, il y a ici une information supplémentaire majeure : les registres permettent d'apprécier directement les âges effectifs de décès des 162 religieuses ayant vécu à Maubuisson ; ils conduisent à évaluer les probabilités des classes d'âges considérées de la façon suivante :

(0,012 0,025 0,087 0,170 0,289 0,210 0,207).

Nous aurons donc une façon objective de juger l'efficacité de la méthode, même si une certaine prudence est de mise car, d'une part cette évaluation n'est sans doute qu'approximative, d'autre part les 37 crânes considérés ne constituent qu'un échantillon (et peut-être même biaisé : a-t-il été strictement tiré au hasard ? ⁽¹⁴⁾).

Commençons par l'analyse avec une loi *a priori* conforme au standard préindustriel des femmes. Nous avons alors obtenu les espérances *a posteriori* des sept classes d'âges suivantes

(0,048 0,067 0,071 0,135 0,301 0,219 0,159), et les écarts types *a posteriori* (0,050 0,068 0,069 0,114 0,166 0,142 0,135).

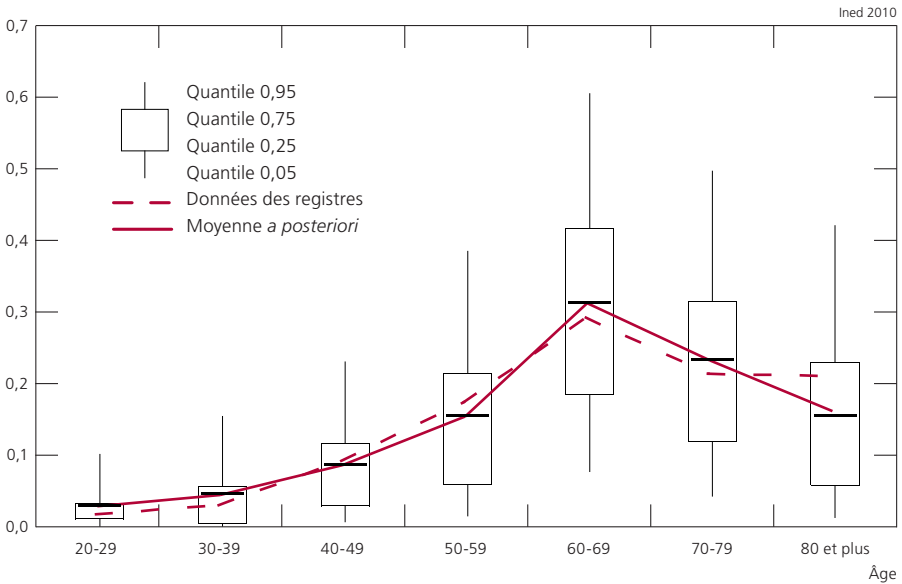
Il est intéressant de comparer les moyennes *a posteriori* aux moyennes *a priori* qui sont ici : (0,10 0,11 0,12 0,15 0,21 0,21 0,10). On voit que les données nous font nettement réviser à la baisse les probabilités des classes « jeunes », et à la hausse les probabilités des seules deux dernières classes ; ce qui est conforme à la discussion développée plus haut. Nous ne poursuivrons donc pas davantage la présente analyse et passons à l'analyse bayésienne avec une loi *a priori* de paramètres (0,337 0,449 0,944 1,180 1,652 1,652 0,786) correspondant à un standard préindustriel modifié (PIM) conformément à la discussion précédente. Nous obtenons pour moyennes *a posteriori* (0,025 0,041 0,083 0,151 0,311 0,230 0,159) et écarts types *a posteriori* (0,037 0,054 0,074 0,119 0,163 0,142 0,132). On peut noter en passant que, en remplaçant les p_j du système (1) par les valeurs ainsi estimées et en estimant les p_{ij} par les

(14) On peut cependant noter que, avec les probabilités de référence que nous utilisons, l'échantillon est très compatible avec les valeurs des registres. En effet, si nous calculons des fréquences théoriques pour les stades au moyen de ces données et les comparons aux valeurs observées par un chi-deux, nous obtenons 1,93 pour 6 degrés de liberté, valeur bien inférieure à la limite de signification pour un niveau de 5 % qui est 12,59.

données de référence, on obtient (7,5 2,5 4,3 5,6 4,3 6,8 5,9) pour fréquences « théoriques » des stades, ce qui est très proche des valeurs observées.

Pour aller plus loin, nous pouvons passer aux densités *a posteriori*. Nous avons d'abord constaté, en comparant les fonctions de répartition *a posteriori* exacte et approchée par la loi Bêta, que l'approximation par des densités Bêta est excellente. C'est donc à partir de ces approximations qu'ont été calculés les intervalles de crédibilité donnés dans la figure 2.

**Figure 2. Exemple de Maubuisson (loi *a priori* PIM).
Estimations des probabilités par la moyenne *a posteriori*
et quantiles donnant les intervalles de crédibilité à 90 % et 50 %.**



Source : calculs des auteurs à partir de Séguy et Buchet (2010).

Il est clair d'abord que, avec un échantillon aussi petit, il n'est pas possible d'obtenir des estimations très précises, comme en témoigne la largeur des intervalles de crédibilité. On peut cependant tirer des informations pertinentes de l'analyse des données. On voit que celle-ci conduit encore à réviser à la baisse les probabilités des trois premières classes et à la hausse les probabilités des cinquième et septième classes : les moyennes *a priori* sont ici

(0,048 0,064 0,135 0,169 0,236 0,236 0,112).

Les estimations que nous obtenons s'avèrent cependant très proches des valeurs tirées des registres comme on le voit sur la figure 2 ; les différences les plus sensibles (surtout en valeur relative) concernent les deux premières probabilités dont la valeur tirée des registres est faible à un point qui semble difficile à imaginer *a priori*, et à un moindre degré la dernière valeur qui est

corrélativement plus élevée. En fait, avec les distributions très dissymétriques correspondant aux très petites probabilités, la moyenne peut être trompeuse ; si nous examinons, par exemple, la première classe, nous voyons que 50 % de la probabilité *a posteriori* est dans l'intervalle [0,001 0,032] dont le milieu 0,017 se rapproche de la valeur cible ; il en va de même pour la seconde classe. Il pourrait donc être légitime de penser que, pour ces deux classes, les moyennes *a posteriori* surestiment les vraies valeurs, indication qui s'avère aller dans le sens de la réalité. De façon plus générale, on voit que les intervalles de crédibilité les plus courts (à 50 %) encadrent bien les valeurs cibles qui sont même très proches de leur centre. Au total, dans une situation usuelle « aveugle », on peut dire que les moyennes *a posteriori* donnent une information de base qui peut être utilement complétée par un ensemble de considérations, comme leur évolution par rapport aux moyennes *a priori* et les intervalles de crédibilité ; il faut souligner en particulier l'importance de ce type de considérations dans le cas de très petites probabilités : nous retrouverons cela dans l'exemple suivant.

Pour compléter cet exemple dans la ligne de nos comparaisons de la partie IV, la méthode IALK donne ici des résultats particulièrement aberrants avec plusieurs probabilités nulles ; la méthode de Bocquet-Appel et Bacro appliquée avec les vecteurs candidats du fichier ProbAttri20-90.txt donne les estimations (0,025 0,036 0,073 0,133 0,209 0,268 0,255) qui sont dans l'ensemble cohérentes avec les registres, mais moins bonnes que les nôtres : par exemple la somme des carrés des erreurs (respectivement l'erreur maximale) vaut ici 0,014 (respectivement 0,080) contre 0,004 (respectivement 0,048) avec notre méthode. Finalement, nous pouvons ici appliquer notre méthode avec la loi *a priori* uniforme sur les vecteurs du fichier ProbAttri20-90.txt ; nous obtenons alors pour les probabilités des diverses classes d'âge les estimations (0,059 0,057 0,096 0,149 0,204 0,232 0,203) et les critères ci-dessus de distance aux valeurs des registres sont respectivement 0,011 et 0,085 ; cette seconde loi *a priori* conduit donc à de moins bonnes estimations que la première ; elles sont assez voisines des estimations fournies par le programme *Iterage* de Bocquet-Appel. En fait, si cette loi *a priori* conduisait à de bonnes estimations dans les simulations de la partie IV, son défaut est de ne pas pouvoir prendre en compte les connaissances *a priori* spécifiques à cette population, alors que la première version de notre méthode le fait de façon extrêmement simple et, on l'a vu, efficace.

2. Le cimetière de Frénoville (époque mérovingienne)⁽¹⁵⁾

Dans cet exemple, les stades osseux de 200 crânes ont été répartis en 5 classes (les mêmes que dans le cas à 5 classes de la section IV) dont les fréquences observées sont (92 29 22 27 30). La répartition par âge a été effectuée

(15) Nous remercions Luc Buchet et Isabelle Séguy de nous avoir fourni les données de ce site. Pour plus de détails, voir Buchet (1978).

pour deux subdivisions : 8 et 14 classes. Dans les deux situations, la première classe est 18-19 ans, la dernière classe est 80 ans et plus, la population des 20-79 ans étant respectivement divisée en classes décennales (il y a alors 8 classes au total) ou quinquennales (il y a alors 14 classes en tout).

Aucune indication de sexe n'étant considérée, nous utilisons les données de référence de Lisbonne sexes réunis. La méthode bayésienne a été utilisée avec une loi *a priori* de Dirichlet qui permet de considérer facilement toute répartition en classes d'âges. Il n'y a ici aucune indication particulière sur la population concernée : les β_j ont donc été choisis proportionnels aux probabilités du standard préindustriel (hommes et femmes réunis), à savoir :

- pour 8 classes :

(0,02 0,10 0,11 0,13 0,16 0,20 0,19 0,09),

- pour 14 classes :

(0,02 0,05 0,05 0,05 0,06 0,06 0,07 0,07 0,09 0,10 0,11 0,11 0,09 0,09),

la somme des β_j étant égale au nombre de classes.

Avec 8 classes d'âges nous avons obtenu les résultats indiqués dans le tableau 2 et la figure 3 (les quantiles sont calculés sur l'approximation des densités *a posteriori* par des densités Bêta, approximation un peu moins bonne que dans le premier exemple, mais tout à fait suffisante en pratique). Rappelons que les quantiles 5 % et 95 % (respectivement 25 % et 75 %) sont les extrémités des intervalles de crédibilité à 90 % (respectivement 50 %). L'écart interquartile est une façon de mesurer la dispersion qui complète utilement l'écart type, surtout dans les cas de distributions très dissymétriques.

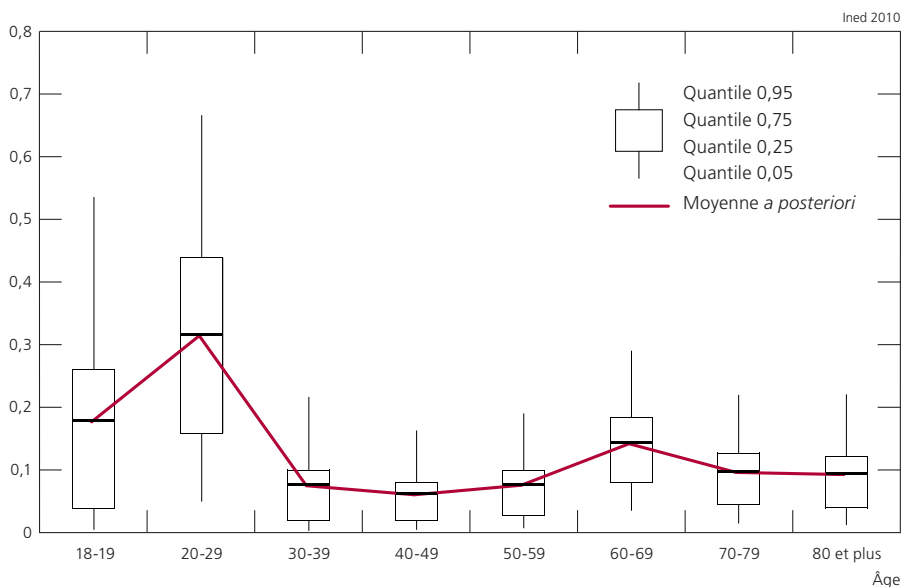
**Tableau 2. Exemple de Frénouville, 8 classes d'âges.
Quelques caractéristiques de la loi *a posteriori*.**

Classe d'âges	18-19 ans	20-29 ans	30-39 ans	40-49 ans	50-59 ans	60-69 ans	70-79 ans	80 ans et plus
Moyenne	0,173	0,312	0,070	0,056	0,071	0,138	0,092	0,088
Écart type	0,174	0,192	0,072	0,053	0,059	0,080	0,065	0,067
Intervalle interquartile	0,225	0,283	0,082	0,062	0,073	0,106	0,083	0,085

Source : calculs des auteurs à partir de Séguy et Buchet (2010).

Les espérances *a posteriori* (ligne « Moyenne » du tableau 2) conduisent à une forte révision à la hausse de la mortalité des deux classes les plus jeunes par rapport aux valeurs *a priori* (avec cependant une forte incertitude dont témoignent les écarts types aussi bien que les écarts interquartiles du tableau 2 ainsi que les intervalles de crédibilité de la figure 3) et une révision à la baisse pour les autres classes, avec une incertitude plus faible.

Figure 3. Exemple de Frénouville : estimation des probabilités par la moyenne *a posteriori* et quantiles donnant les intervalles de crédibilité à 90 % et à 50 %.



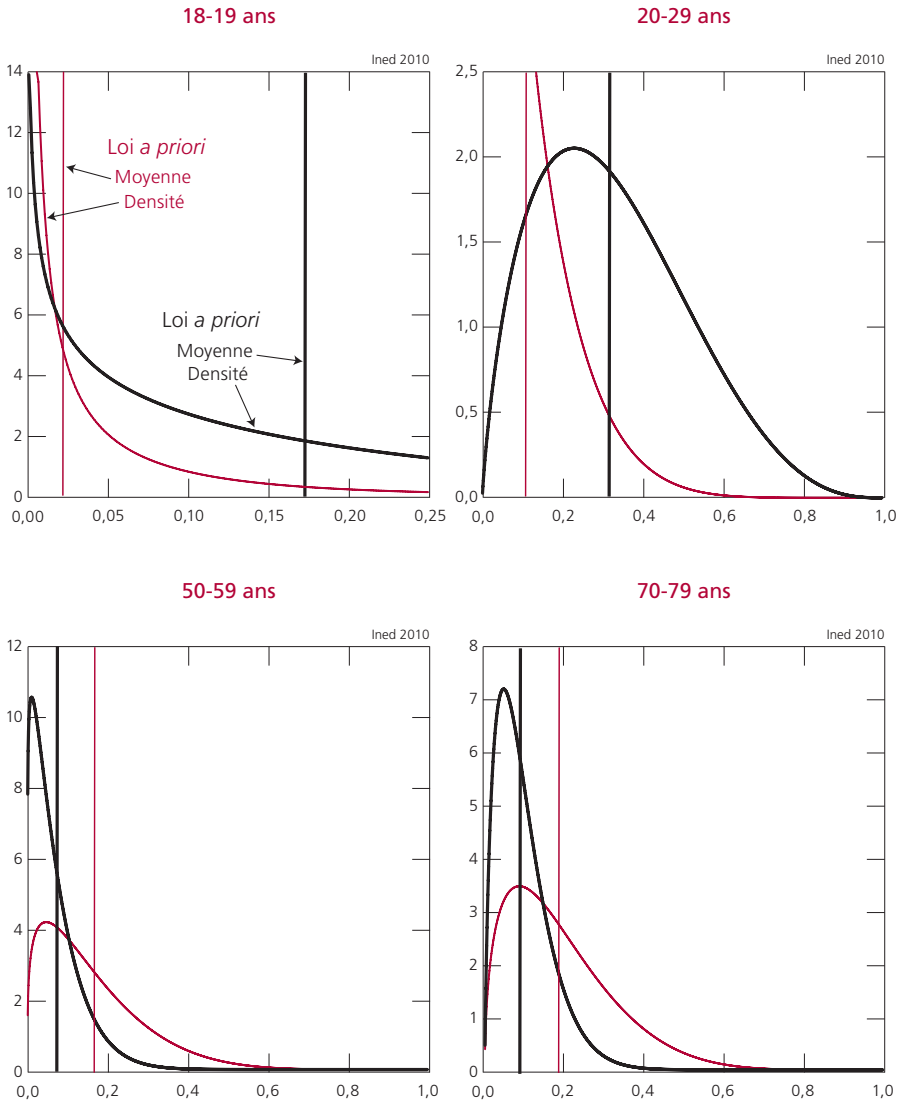
Source : calculs des auteurs à partir de Séguy et Buchet (2010).

Cela peut être précisé en examinant les graphiques de la figure 4 qui donnent pour quatre des classes d'âges : l'approximation Bêta de la densité *a posteriori* (en noir), la densité *a priori* (en rouge) et les moyennes *a posteriori* et *a priori* schématisées par des traits verticaux respectivement noirs et rouges. Pour la première classe (18-19 ans), les densités sont difficilement lisibles car les lois sont concentrées sur de faibles valeurs ; il est clair cependant que la probabilité estimée est nettement plus élevée que celle du standard préindustriel, même si la dissymétrie de la distribution peut exagérer le phénomène quand l'estimation est faite par la moyenne ; on voit que le milieu de l'intervalle de crédibilité à 50 % (0,148) est en dessous de cette moyenne, et encore davantage la médiane de la loi *a posteriori* (0,116). Pour la seconde classe (20-29 ans), l'estimation est nettement au-delà du standard préindustriel, avec une imprécision importante et une loi *a posteriori* presque symétrique (d'où la proximité de la moyenne, de la médiane et du milieu de l'intervalle interquartile)⁽¹⁶⁾. Pour la dernière classe (80 ans et plus, non représentée dans la figure 4), l'estimation est pratiquement celle du standard préindustriel (les lois *a priori* et *a posteriori* sont très proches). Pour toutes les autres classes on obtient des probabilités estimées plus faibles que les probabilités du standard préindustriel ; on peut

(16) Pour cette classe, et dans une moindre mesure pour la première, il convient de noter que l'approximation par une loi Bêta est relativement mauvaise (une exception parmi tous les cas que nous avons traités). Nous l'avons cependant conservée par souci de simplicité car la discussion à partir de la loi exacte aurait été tout à fait analogue pour les aspects relevés ici.

noter que, pour les classes 50-59, 60-69, 70-79 (dont deux sont représentées dans la figure 4), les densités *a posteriori* sont relativement « pointues », montrant une bonne fiabilité des estimations par la moyenne *a posteriori*.

Figure 4. Exemple de Frénouville. Lois *a priori* (rouge) et lois *a posteriori* (noir) pour quatre des classes d'âges et leurs moyennes respectives.



Note : pour la classe 18-19 ans, l'échelle des abscisses couvre les valeurs 0 à 0,25, pour améliorer la lisibilité de la figure. Les légendes sont identiques pour les 4 figures.

Source : calculs des auteurs à partir de Séguy et Buchet (2010).

Pour compléter cet exemple, voici enfin les moyennes et écarts types *a posteriori* pour 14 classes d'âges :

moyennes :

(0,130 0,231 0,070 0,049 0,034 0,034 0,045 0,041 0,050 0,066 0,073 0,061 0,045 0,071)

écarts types :

(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)

À partir de ces valeurs, on peut recalculer des estimations pour 8 classes par regroupement. Pour cela, on utilise l'additivité des moyennes et la formule classique concernant la variance d'une somme :

$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$ (on a vu dans la présentation de la méthode que la covariance *a posteriori* était aussi aisément calculable).

On obtient :

moyennes : (0,130 0,301 0,084 0,079 0,090 0,139 0,105 0,071)

écarts types : (0,135 0,127 0,068 0,057 0,051 0,061 0,054 0,051)

Si l'on compare aux estimations obtenues directement pour 8 classes, on constate une bonne cohérence, avec l'écart le plus net pour la première classe dont l'estimation est ici plus faible (notons que nous avons donné plus haut quelques raisons de penser que cette estimation était éventuellement surévaluée) ; pour les écarts types, on voit que, avec cette seconde façon d'estimer, à l'exception d'un cas de quasi égalité, ils sont plus faibles, et parfois de façon notable. (Un des prolongements intéressants de ce travail serait d'étudier si cette propriété est contingente ou systématique et, dans ce dernier cas, quelle en est la cause).

Conclusion

Les simulations effectuées, ainsi que les applications archéologiques proposées, ont permis de montrer nombre d'avantages de la méthode que nous proposons pour estimer la structure par âge des populations du passé, lorsqu'on ne dispose d'aucun enregistrement de l'âge au décès mais où les mesures d'indicateurs biologiques viennent remplacer cet enregistrement. Résumons ces principaux avantages.

En premier lieu, notre méthode est d'utilisation simple⁽¹⁷⁾ et très souple : d'une part elle est valable quel que soit le découpage en classes d'âges et de stades osseux, d'autre part elle permet d'introduire très facilement les connaissances *a priori* de l'utilisateur. Malgré sa généralité et sa simplicité, la méthode s'avère efficace : seuls certains résultats obtenus avec l'algorithme *Iterage*, de construction plus complexe, arrivent à la concurrencer quand elle est employée avec une loi *a priori* « passe-partout » ; cependant, cette méthode de Bocquet-Appel et Bacro est d'utilisation plus complexe et plus limitée que la nôtre et,

(17) Un programme de calcul en R (R Development Core Team, 2008) est disponible auprès des auteurs.

lorsqu'elle est utilisable, les éléments nécessaires à son usage peuvent être intégrés à la nôtre sous forme d'une nouvelle loi *a priori* qui la rend préférable.

En second lieu, la méthode que nous proposons s'insère dans un contexte statistique clair qui permet de valider correctement ses propriétés théoriques ; nous pensons en particulier à la fiabilité des intervalles de crédibilité qui s'en déduisent.

Cependant, comme toute méthode nouvelle, elle demande sans doute à être affinée à la lueur des remarques de ses utilisateurs. Nous espérons donc avant tout que de nombreux paléodémographes l'utiliseront afin que leur expérience complète la nôtre, précise l'usage de la méthode et suscite les améliorations nécessaires.

Remerciements : Nous tenons à remercier ici Isabelle Séguy et Luc Buchet, ainsi que les trois lecteurs anonymes dont les commentaires ont été utiles pour cette version définitive. Bien entendu les auteurs restent entièrement responsables des propos tenus ici.



RÉFÉRENCES

- BOCQUET-APPEL J.-P.**, 2005, « La paléodémographie », in Dutour O., Hubler J.-J., Vandermerch B. (dir.), *Objets et méthodes en paléanthropologie*, Paris, Comité des travaux historiques et scientifiques, p. 271-313.
- BOCQUET-APPEL J.-P.**, 2008a, *La paléodémographie*, Paris, Éditions Errance, 192 p.
- BOCQUET-APPEL J.-P. (dir.)**, 2008b, *Recent Advances in Paleodemography*, Dordrecht, Springer-Verlag, 294 p.
- 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, p. 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 Bocquet-Appel J.-P. (dir.), *Recent Advances in Paleodemography*, Dordrecht, Springer-Verlag, p. 63-82.
- BOCQUET-APPEL J.-P., MASSET C.**, 1996, « Paleodemography: Expectancy and false hope », *American Journal of Physical Anthropology*, 99, p. 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, p. S29-S45.
- BUCHET L.**, 1978, « La nécropole gallo-romaine et mérovingienne de Frénouville (Calvados). Étude anthropologique », *Archéologie médiévale*, 8, p. 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, p. 77-182.
- CAUSSINUS H., COURGEAU D.**, 2010, « Estimation de la structure par âge des décès : nouvelles propositions », in Séguy I., Buchet L. (dir.), *Manuel de paléodémographie*, Paris, Ined, Collection Les Manuels.
- COURGEAU D.**, 2010, « Critiques sur les méthodes actuellement utilisées pour estimer la structure par âge au décès d'une population archéologique », in Séguy I., Buchet L. (dir.), *Manuel de paléodémographie*, Paris, Ined, Collection Les Manuels.
- DEMING W. E., STEPHAN F. F.**, 1940, « On a least squares adjustment of a sampled frequency table when the expected marginal totals are known », *The Annals of Mathematical Statistics*, 11, p. 427-444.
- FRIEDLANDER D.**, 1961, « A technique for estimating contingency tables, given marginal totals and some supplemental data », *Journal of the Royal Statistical Society, Series A*, 124, p. 412-420.
- 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 », *Rapports et procès-verbaux des réunions du Conseil permanent international pour l'exploration des mers*, 86, p. 1-14.
- HASSELBLAD V.**, 1966, « Estimation of parameters for a mixture of normal distribution », *Technometrics*, 8(3), p. 431-444.
- HOPPA R. D., VAUPEL J. W. (eds)**, 2002, *Paleodemography: Age Distributions from Skeletal Samples*, Cambridge, Cambridge University Press, 260 p.

- HOWELL N., 1976, « Towards an uniformitarian theory of human paleodemography », *Journal of Human Evolution*, 5, p. 25-40.
- JACKES M., 2000, « Building the bases for paleodemographic analysis: Adult age estimation », in Katzenberg M., Saunders S. (eds), *Biological Anthropology of the Human Skeleton*, New York, Wiley-Liss, p. 417-466.
- KIMURA D. K., CHIKUNI S., 1987, « Mixture of empirical distributions: An iterative application of the age-length key », *Biometrics*, 43, p. 23-35.
- KONIGSBERG L. W., FRANKENBERG S. R., 1992, « Estimation of age structure in anthropological demography », *American Journal of Physical Anthropology*, 89, p. 235-256.
- KONIGSBERG L. W., FRANKENBERG S. R., 2002, « Deconstructing death in paleodemography », *American Journal of Physical Anthropology*, 117, p. 297-309.
- KONIGSBERG L. W., HERRMANN N. P., 2002, « Markov chain Monte Carlo estimation of hazard model parameters in paleodemography », in Hoppa R. D., Vaupel J. (eds), *Paleodemography: Age Distributions from Skeletal Samples*, p. 222-242.
- KRUITHOF R., 1937, Telefoonverkeersreking, *De Ingenieur*, 52(8), p. E15-E25.
- LEONTIEF W. W., 1941, *The Structure of the American Economy 1919-1929*, Cambridge, Harvard University Press, 181 p.
- 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), p. 85-105.
- MASSET C., 1973, « La démographie des populations inhumées. Essai de paléodémographie », *L'Homme*, 7, p. 95-131.
- MASSET C., 1982, *Estimation de l'âge au décès par les sutures crâniennes*, Thèse, Université de Paris VII, 302 p.
- MASSET C., 1995, « Paléodémographie : problèmes méthodologiques », *Cahiers d'anthropologie et biométrie humaine*, XII(1-2), p. 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, p. 1-14.
- R DEVELOPMENT CORE TEAM, 2008, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- ROBERT C., 2006, *Le choix bayésien. Principes et pratique*, Paris, Springer-Verlag, 638 p.
- SÉGUY I., BUCHET L. (dir.), 2010, *Manuel de paléodémographie*, Paris, Ined, Collection Les Manuels.
- 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. (dir.), *Recent Advances in Paleodemography. Data, techniques, Patterns*, Dordrecht, Springer-Verlag, p. 109-141.
- THONET P., 1963, « Sur certaines variantes des projections du tableau d'échanges inter-industriels », *Bulletin de l'Institut international de statistique*, XL(1^{ère} livraison), p. 431-446.
- THONET P., 1964, Note sur le remplissage d'un tableau à double entrée, *Journal de la Société de statistique de Paris*, 10-11-12, p. 228-247.
- TUGAULT Y., 1970, « Méthode d'analyse d'un tableau « origine-destination » de migrations », *Population*, 25(1), p. 59-68.
- WILLEKENS F., 1977, *The Recovery of Detailed Migration Patterns from Aggregate Data: An Entropy Maximizing Approach*, RM-77-58, IIASA, Laxenburg, Austria, 112 p.

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 • 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 offset these drawbacks, 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 whose 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: its flexibility and efficiency.

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 paleodémógrafos 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.

