



Colloques du Groupement des Anthropologistes de Langue Française (GALF)

---

## Révision de la méthode proposée par Schmitt (2005) pour estimer l'âge au décès des adultes à partir de la surface sacro-pelvienne iliaque

R. Martins<sup>1</sup>, A. Schmitt<sup>2</sup>, P.E. Oliveira<sup>3</sup>

<sup>1</sup>Centro de Investigação Interdisciplinar Egas Moniz (ciiEM) Caparica – Portugal

<sup>2</sup>Anthropologie bioculturelle (UMR 6578) Marseille - France

<sup>3</sup>CMUC, Departamento de Matemática, Universidade de Coimbra, Coimbra - Portugal

**Auteur Correspondant:** Aurore Schmitt, CNRS, UMR 6578- Anthropologie bioculturelle Faculté de Médecine - Secteur Nord, Université de la Méditerranée, CS80011, Bd Pierre Dramard 13 344 Marseille Cedex 15.

**Mots clés:** Estimation de l'âge au décès, Adulte, «Kernel density», Surface auriculaire de l'ilium.

**Key words:** Age-at-death assessment, Adult, Kernel density, Auricular surface of the ilium.

### Résumé

La surface auriculaire de l'ilium a fait l'objet de nombreuses méthodes pour estimer l'âge au décès des adultes. En 2005, Schmitt proposait un système de cotation simple (4 critères morphologiques avec 2 à 4 modalités), l'utilisation de probabilités, l'aggrégation de plusieurs séries européennes d'âge connu et un intervalle chronologique plutôt qu'un âge précis, pour une estimation sur les échantillons archéologiques. Cette méthode permet d'identifier les individus décédés après 60 ans, ce qui n'est pas le cas des méthodes classiques, et privilégie la fiabilité à la précision. Toutefois, la limite majeure est l'impossibilité d'estimer l'âge de certains sujets. Les données ont été traitées par une autre approche statistique basée sur l'estimateur du noyau de la densité (« kernel density procedure »), tenant compte des observations au voisinage de chaque point. Ce nouveau système permet une approximation de l'âge de tous les individus et des intervalles chronologiques moins larges que ceux proposés dans la méthode initiale.

### Abstract

The auricular surface of the ilium has been largely used as the basis for age-at-death assessment methods. In 2005, Schmitt proposed a simple scoring system (4 features with 2 to 4 phases), the use of a probability estimation procedure, the aggregation of several European series with known age and sex, and the estimation of a credible chronological interval rather than a numerical estimation of age. This method enables identification of individuals deceased after 60 years, but with the drawback of not being able to produce an approximation for many of them. The same data was later processed using another statistical approach, based on smoothing methods. The probability distribution and the conditional distribution are estimated from training data using discretized kernel density procedures. This approach produces an age for all the individuals and the credible intervals are narrower than the initial method.

### Introduction

La surface sacro-pelvienne iliaque a fait l'objet de nombreuses méthodes pour estimer l'âge au décès des adultes (Lovejoy *et al.*, 1985; Buckberry et Chamberlain, 2002, Ygarashi *et al.*, 2005). Chacune de ces méthodes a été élaborée sur une seule série ostéologique d'âge connu et ne tiennent pas compte de l'influence de la population de référence. La méthode proposée par Schmitt (2005) se base sur l'étude de la variabilité des modifications de la surface sacro-pelvienne iliaque sur plusieurs série d'Europe, d'Asie, d'Amérique et d'Afrique. De plus, sa fiabilité a été testée sur des échantillons complètement indépendants du référentiel. Le système de cotation est simple, l'estimation de l'âge au décès est menée par un calcul de probabilités (Lucy *et al.*, 1996), la population de référence consiste en l'agrégation de plusieurs séries européennes d'âge connu. Cette méthode permet d'identifier les individus décédés après 60 ans (ce qui n'est pas le cas des méthodes classiques) et privilégie la fiabilité à la précision. Toutefois, l'impossibilité d'estimer l'âge de certains sujets confère à la méthode un inconvénient majeur. C'est la raison pour laquelle une nouvelle approche statistique a été mise en œuvre.

### Matériel et méthode

Le système de cotation est publié dans l'article princeps (Schmitt, 2005). Trois séries utilisées dans la méthode initiale ont été utilisées dans la présente étude. Elles sont issues de la collection identifiée de l'Université de Coimbra au Portugal (Rocha, 1995), de la collection de Spitalfields en Angleterre (Molleson et Cox, 1993) et de la collection Hammann-Todd à Cleveland aux Etats-Unis (Mensforth et Latimer, 1989). La composition du corpus est présentée dans le tableau 1. Les deux sexes sont représentés et la distribution par âge est homogène.

Séries	Hommes	Femmes
Conchada-Portugal	39	51
Spitafields-Royaume Uni	46	50
Hammann-Todd-USA	58	62
Total	145	163

**Tableau 1.** Effectif des individus par collection ostéologique.

L'approche statistique étant développée par ailleurs (Martins *et al.*, 2011), nous la présentons brièvement. Le traitement des données commence par une décomposition Bayésienne de la distribution. L'estimation se base sur la méthode du noyau, avec des adaptations appropriées pour les variables discrètes, inspirée par Lucy *et al.* (1996). Représentons par  $X$  l'âge de l'individu et par  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$  les  $m$  caractères morphologiques. D'après le Théorème de Bayes, la distribution par âge connaissant les caractères observés  $\mathbf{y} = (y_1, \dots, y_m)$  est:

$$f(x|\mathbf{y}) = \frac{f(\mathbf{y}|x)f(x)}{f(\mathbf{y})} \quad (1)$$

où  $f(\mathbf{y}|x)$  est la vraisemblance des caractères quand l'âge est connu  $X=x$ , et  $f(x)$  est la distribution de  $X$ . La distribution marginale des caractères  $f(\mathbf{y})$ , peut s'estimer en additionnant par rapport à  $x$  le numérateur de (1).

La caractérisation de  $f(\mathbf{y}|x)$  est délicate car cette fonction dépend des relations entre les variables. Une première approche, très simplifiée, suppose que lorsque que l'âge est connu, les caractères sont statistiquement indépendants. On appelle ce modèle *indépendance conditionnelle*. La vraisemblance peut donc s'écrire ainsi:

$$f(\mathbf{y}|x) = \prod_{i=1}^m f(y_i|x) \quad (2)$$

où  $f(y_i|x)$  est la vraisemblance individuelle de la distribution du caractère  $Y_i$  admettant l'âge connue  $X=x$ . Un modèle plus réaliste, appelé *dépendance conditionnelle*, doit inclure la dépendance entre les caractères. Pour mesurer cette dépendance nous utilisons une méthode proposée par Chow and Liu (1968): il s'agit de construire un *arbre de dépendance* décrivant le parcours reliant les caractères. On désigne, aléatoirement, un caractère pour point de départ, tout caractère  $i$  a un progéniteur désigné par  $j(i)$ . Ainsi, la fonction de vraisemblance peut se décomposer de la façon suivante:

$$f(\mathbf{y}|x) = \prod_{i=1}^m f(y_i|y_{j(i)},x) \quad (3)$$

où  $y_{j(i)}$  est la valeur observée pour le progéniteur de  $i$ . Pour le point de départ,  $f(y_\ell|y_{j(\ell)},x) = f(y_\ell|x)$ .

En ce qui concerne la construction d'estimateurs pour les distributions considérées, nous utilisons la méthode du noyau (Silverman, 1986; Wand and Jones, 1995). Pour la densité de l'âge  $X$ , on calcule:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} K\left(\frac{x-X_i}{h_n}\right), \quad x \in R, \quad (4)$$

où les  $h_n > 0$ ,  $\lim_{n \rightarrow +\infty} h_n = 0$ , et  $K(t)$  est une fonction non-négative telle que  $\int K(t)dt = 1$ . Pour la fonction  $K$  nous prenons le noyau Gaussien, bien connu pour ses propriétés quasi-optimales et la simplification des aspects calculatoires. Le choix de la fenêtre a plus d'impact sur les estimations. Pour cet aspect, nous avons suivi la procédure de Silverman (1986):

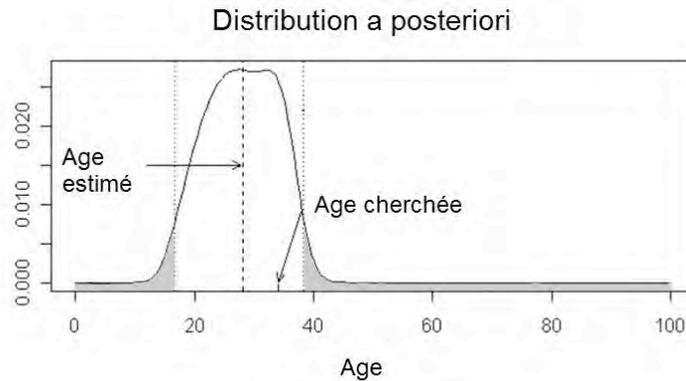
$$h = 0.9An^{\frac{1}{5}}, \quad \text{où } A = \min(\text{deviation standard}, \text{interval interquartil}/1.34). \quad (5)$$

Ce choix est approprié du point de vue calculatoire et donne des résultats satisfaisants pour une grande variété de distributions.

Le traitement de la distribution des caractères est un peu plus délicate car elle est discrète, donc les versions multi-dimensionnelles de l'estimateur (4) doivent prendre cet aspect en considération. En admettant l'âge  $X$  continu et les caractères discrets, nous estimons la distribution conjointe avec:

$$\hat{f}(x,y) = \frac{1}{n(y)} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \delta(y-y_i), \quad (6)$$

où  $K$  est un noyau (Gaussien),  $h$  une fenêtre indépendante de  $y$ ,  $\delta(u) = 1$  si  $u = 0$  et  $\delta(u) = 0$  pour toutes les autres valeurs de  $u$ , et  $n(y) = \sum_{i=1}^n \delta(y-y_i)$ , fait le comptage du nombre d'observations avec les caractères égaux à  $y$ . Cette expression correspond à l'estimation de plusieurs distributions conditionnelles (figure 1).



**Figure 1.** Distribution *a posteriori* de l'âge estimé d'un individu dont l'âge réel est 28 ans.

Ce traitement a été appliqué sur chacune des trois échantillons séparément puis sur le corpus total. L'analyse a été menée avec le Logiciel R.

### Résultats

Les résultats sont présentés dans le tableau 2. Quel que soit le set de données considéré, l'application de l'hypothèse d'une dépendance conditionnelle des caractères observés permet d'obtenir des écarts absolus moyen (EAM), des biais et des intervalles d'âge estimé (IE) moins élevés et donc plus fiables que l'utilisation de l'hypothèse d'indépendance conditionnelle. Le pourcentage d'intervalle fiable (incluant l'âge réel des individus) est plus élevé pour cette dernière, mais ce résultat est lié au fait que les intervalles sont beaucoup plus grands. Les biais oscillent entre 0.3 et 0.45.

	calibration	EAM	Biais	IE	PI
Spitfields					
sexes combinés	dép.cond.	10.08	0.37	43.87	0.88
	indép. cond.	12.43	0.77	71.01	0.97
femmes	dép.cond.	10.43	0.30	47.6	0.88
	indép. cond.	12.58	0.75	74.57	0.96
hommes	dép.cond.	10.78	0.43	45.44	0.87
	indép. cond.	13.03	0.78	74.8	0.93
Conchada					
sexes combinés	dép.cond.	10.24	0.43	42.32	0.86
	indép. cond.	11.17	0.71	65.38	0.94
femmes	dép.cond.	11.8	0.45	39.24	0.82
	indép. cond.	12.79	0.78	70.01	1,00
hommes	dép.cond.	9.54	0.43	38.83	0.77
	indép. cond.	11.31	0.74	60.82	0.97
Hammann-Todd					
sexes combinés	dép.cond.	10.86	0.4	40.02	0.91
	indép. cond.	11.91	0.73	61.91	0.98
femmes	dép.cond.	10.51	0.44	44.42	0.87
	indép. cond.	13.23	0.85	63.75	0.97
hommes	dép.cond.	10.57	0.43	40.86	0.9
	indép. cond.	13.42	0.72	65.94	0.97
Corpus total					
sexes combinés	dép.cond.	10.31	0.38	41.42	0.87
	indép. cond.	11.34	0.68	63.21	0.97
femmes	dép.cond.	10.23	0.38	41.69	0.88
	indép. cond.	11.24	0.69	63.66	0.98
hommes	dép.cond.	10.98	0.43	41.83	0.86
	indép. cond.	11.25	0.66	66.38	0.97

**Tableau 2.** Résultats pour les 3 échantillons séparés et le corpus complet selon le sexe. EAM: écart absolu moyen; IE: taille de l'intervalle estimé; PI: pourcentage des intervalles (à 95%) qui incluent l'âge réel.

Dans les trois séries considérées séparément, les intervalles d'âge estimé sont plus larges pour les femmes et pour la série de Spitafields, mais pas nécessairement plus fiables (PI). Lorsque les trois échantillons sont réunis dans l'analyse, IE et PI sont sensiblement les mêmes quelle que soit la modalité de la variable sexe (homme/femme, indéterminé), mais moins fiables que lorsque les séries sont analysées séparément.

### Discussion

L'analyse statistique montre que les caractères observés ne sont pas indépendants entre eux et qu'en tenant compte de ce paramètre, la fiabilité est accrue.

Les biais sont globalement élevés ce qui montre que l'âge des individus jeunes est surestimé et celui des plus âgés sous-estimé.

L'approche statistique proposée permet d'estimer l'âge au décès (sous forme d'intervalle chronologique fiable) de tous les individus ce qui n'était pas le cas de la méthode initiale dont la technique statistique paramétrique était conditionnée par des classes d'âge prédéfinies.

Globalement, l'estimation de l'âge si le sexe de l'individu est connu est plus fiable.

Le script R permettant d'appliquer ce traitement statistique est accessible sur demande ([aurore.schmitt@univmed.fr](mailto:aurore.schmitt@univmed.fr)). Il permet d'estimer l'âge en fonction du sexe (une modalité indéterminée est prévue) et de la population d'origine (si elle est connue). Pour l'application sur du matériel archéologique, la compilation des 3 séries compose le corpus de référence.

### Références bibliographiques

- Buckberry J.L., Chamberlain A., 2002, Age Estimation from the auricular surface of the ilium: a revised method. *American Journal of Physical Anthropology*, 119, 231-329.
- Chow CK., Liu CN., 1968, Approximating Discrete Probability Distributions with Dependence Trees. *IEEE Transactions on Information Theory*, 14, 462-467.
- Igarashi Y., Kagumi U., Tetsuaki W., Kanazawa E., 2005, New method for estimation of adult skeletal age at death from the morphology of the auricular surface of the ilium. *American Journal of Physical Anthropology*, 128, 324-339.
- Lovejoy C.O., Meindl R.S., Prysbeck T.R., Mensforth R.P., 1985, Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology*, 68, 15-28.
- Lucy D., Aykroyd R.G., Pollard A.M., Solheim T., 1996, A Bayesian approach to adult human age estimation from dental observations by Johanson's age changes. *Journal of Forensic Sciences*, 41, 189-194.
- Martins R., Oliveira P. E., Schmitt A., Estimation of age at death from the pubic symphysis and the auricular surface of the ilium using a smoothing procedure. *Forensic Science International*, in press.
- Mensforth R.P., Latimer B.M., 1989, Hamann-Todd Collection aging studies: osteoporosis fracture syndrome. *American Journal of Physical Anthropology*, 80, 461-479.
- Molleson T., Cox M., 1993, The Spitafields project, volume 2, the Anthropology. Council for British Archaeology Research Report, 86, 167-179.
- Rocha M.A., 1985, Les collections ostéologiques humaines identifiées du Musée Anthropologique de l'Université de Coimbra. *Antropologia Portuguesa*, 13, 7-38.
- Schmitt A., 2005, Une nouvelle méthode pour estimer l'âge au décès des adultes à partir de la surface sacro-pelvienne iliaque. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, 17, 89-101.
- Silverman B.W., 1986, *Density estimation for statistics and data Analysis*. (Londres: Chapman and Hall).
- Wand M.P., Jones M.C., 1995, *Kernel Smoothing* (Londres: Chapman and Hall).