Utilisation du théorème de Bayes pour estimer l'âge au décès d'adultes

Marie Tremblay, Aurore Schmitt, Bérangère Saliba-Serre, Laurent Martrille

▶ **To cite this version:**

Marie Tremblay, Aurore Schmitt, Bérangère Saliba-Serre, Laurent Martrille. Utilisation du théorème de Bayes pour estimer l'âge au décès d'adultes. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. inria-00386644

HAL Id: inria-00386644

https://inria.hal.science/inria-00386644

Submitted on 22 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UTILISATION DU THEOREME DE BAYES POUR ESTIMER L'AGE AU DECES D'ADULTES

Marie Tremblay[1], Aurore Schmitt[2], Bérengère Saliba-Serre[2] & Laurent Martrille[1]

[1] Centre Hospitalier Universitaire Lapeyronie, 191 avenue du Doyen Giraud, 34295 Montpellier Cedex 5, France
[2] Unité d'anthropologie Bioculturelle, faculté de Médecine Nord, Boulevard P. Dramard, 13916 Marseille Cedex 15, France

Résumé : L'estimation de l'âge au décès est un sujet majeur en médecine légale, en particulier lors de l'identification d'un cadavre. Les dents, organes les plus solides, sont souvent les seuls restes utilisables. La méthode de Lamendin et collègues (1992), fondée sur un modèle de régression linéaire, relie l'âge au décès à deux indicateurs dentaires (transparence et parodontose). Cependant, cette méthode sur ou sous-estime fortement l'âge et ne prend pas en compte la non-indépendance des données. Le but de cette étude est l'estimation, par méthode Bayésienne, de l'âge au décès à partir des deux indicateurs précédents en utilisant une seule dent par individu. Pour faciliter l'utilisation par les légistes, les trois variables sont catégorisées. A partir d'un échantillon de 214 individus d'âge au décès et de sexe connus, probabilité a priori de l'âge au décès, vraisemblance et probabilité marginale des deux indicateurs sont estimées en utilisant la méthode du Jacknife. Le théorème de Bayes permet alors de calculer la probabilité a posteriori de l'âge au décès. Les performances des deux méthodes, Lamendin (avec ré-estimation des paramètres) et Bayésienne, sont ensuite comparées. Si la méthode Bayésienne permet de réduire le biais (0.35) par rapport à la régression (0.65), l'erreur absolue moyenne par cette méthode est de 16 ans alors qu'elle n'est que de 11 ans avec la régression. La méthode Bayésienne ne permet pas ainsi d'améliorer les résultats obtenus par la méthode de Lamendin. La faible corrélation entre l'âge et les indicateurs, et la discrétisation de l'âge en classes peuvent être à l'origine de ces conclusions.
Mots-clés : médecine légale, estimation âge au décès, régression, théorème de Bayes.

Abstract : In forensic medicine, age-at-death assessment is critical human remains identification. As teeth are the hardest part of the human body, they often remain the only evidence available. Lamendin and colleagues (1992) proposed to assess age-at-death from two dental characteristics (transparency and periodontosis) using linear regression models. However, this method has several biases; in particular, the non-independance between data has not been considered, while several teeth of a same individual were included. The aim of this study was to investigate the performance of a Bayesian method to predict age-at-death from the same two dental characteristics , but using a single tooth. For convenience of use by forensic scientists, variables are converted into classes. Using a sample of 214 cases (Terry collection) with known sex and age-at-death, a priori probability, likelihood and margin probability were estimated using a jacknife resampling strategy. A posteriori probability of age-at-death was estimated using the Bayes theorem. Then, we compared the Lamendin technique (re-estimation of parameters on our sample) with the Bayesian approach. Although the bias reduction was greater using the Bayesian method (0.35 vs. 0.65 for the Lamedin technique), the Lamendin technique provided better estimates of age-at-death with a mean absolute error of 11 years vs. 16 years. The Bayesian approach failed to give better estimates of age-at-death than the regression technique. These disappointing results may be due to age categories and the low correlation between age and dental characteristics.
Key words : forensic medicine, age-at-death assessment, linear regression, Bayes theorem.

1 - Introduction

L'estimation de l'âge au décès est un sujet important en médecine légale, en particulier lors de l'identification de cadavres. Les dents, organe le plus solide chez les humains, sont souvent les seuls restes exploitables.

Lamendin et al. (1922) ont proposé l'utilisation de deux critères dentaires facilement mesurables, la parodontose et la translucidité. La méthode est facile à appliquer : utilisation d'une seule dent, sans préparation préalable, et non détruite. La variabilité intra et inter observateurs est par ailleurs faible (Baccino et al., 1999; Foti et al., 2001). Un modèle de régression linéaire est utilisé pour estimer l'âge au décès à partir des deux mesures dentaires. Mais les deux critères n'expliquent qu'une faible partie de la variabilité de l'âge au décès. De plus, l'âge des plus jeunes (plus âgés) est sur-estimé (sous-estimé). Cette erreur systématique est due à la faible corrélation entre l'âge et les critères. Enfin, la méthode ne tient pas compte de la non-indépendance des données, plusieurs dents par individu étant incluses dans l'échantillon. Le but de cette étude est d'évaluer les performances de la méthode Bayésienne pour estimer l'âge au décès à partir des deux critères utilisés dans la méthode de Lamendin.

2 - Matériel et méthode

Un échantillon de 214 individus de la collection de Terry (Musée National d'Histoire Naturelle de Washington) est utilisé. La collection de Terry, crée entre 1900 et 1965, est constituée d'environ 1600 individus désarticulés de sexe, d'âge au décès et, pour la plupart, de cause de décès connus. Une seule dent par individu est incluse dans l'échantillon. Trois mesures sont effectuées. La hauteur de la racine (RH) correspond à la distance entre le sommet de la racine et la jonction amélo-cémentaire. La parodontose (Pa), due à la dégénération des tissus mous entourant la dent, est la distance entre la jonction amélo-cémentaire et le début des tissus mous. La translucidité (Tr) est due à un dépôt de cristaux dans la dentine. Elle correspond à la hauteur maximale entre le sommet et le début de la face vestibulaire. Deux indicateurs sont construits à partir des trois mesures précédentes : T = Tr * 100 / RH et P = Pa * 100 / RH. L'utilisation de ces deux critères permet de compenser des différences dues à la position des dents et / ou à la variation de la hauteur de la racine.

Dans le contexte légal, une classe d'âge à laquelle est associée une probabilité d'appartenance est plus intéressante qu'une estimation ponctuelle avec intervalle de confiance. L'âge a donc été discrétisé en trois classes : jeunes (≤ 40 ans), intermédiaires (40 < âge ≤ 60 ans) et seniors (≥ 61 ans). Les classes fournies par un médecine légiste doivent être assez larges pour ne pas exclure un âge potentiel tout en étant les plus étroites possible pour faciliter l'identification (Osborne et al., 2004). Dans cette première étude, nous avons également discrétiser les indicateurs P et T. Les quartiles respectifs ont été utilisés pour déterminer quatre classes.

La méthode Bayésienne est couramment utilisée en sciences légales (Lucy et al., 1996 et 2002; Aykroyd et al., 1999; Prince et al., 2008; Schmitt, 2008). Le théorème de Bayes combine l'information a priori et l'information contenue dans les données pour calculer la distribution a posteriori :

$$P\left(i \in CA_j \middle| T = t, P = p\right) = \frac{P\left(T = t, P = p \middle| i \in CA_j\right) P\left(i \in CA_j\right)}{\sum_{i=1}^{n} P\left(T = t, P = p \middle| i \in CA_j\right) P\left(i \in CA_j\right)},$$

où i = 1, …, 214, $P\left(i \in CA_j\right)$ est la probabilité a priori d'appartenir à la classe d'âge CA_j, j = 1…4. Elle correspond à la probabilité de décès entre les bornes de CA_j. $P\left(T = t, P = p \middle| i \in CA_j\right)$ est la vraisemblance, probabilité d'observer les valeurs t de T et p de P sachant la classe d'âge. $\sum_{i=1}^{n} P\left(T = t, P = p \middle| i \in CA_j\right) P\left(i \in CA_j\right)$ est la probabilité marginale de T et P. $P\left(i \in CA_j \middle| T = t, P = p\right)$ est la probabilité a posteriori d'appartenir à la classe d'âge CA_j conditionnellement aux valeurs de T et P.

La distribution de probabilité a priori est estimée à partir des données. En effet, aucun avis d'expert ne permet de la définir pour définir une distribution, conjuguée par exemple. Par ailleurs, la

distribution uniforme n'est pas adaptée au contexte : la probabilité de décès n'est pas identique pour tous les âges. Enfin, la représentativité de l'échantillon n'étant pas connue, la distribution des âges dans la population générale n'est pas utilisable.

Pour ne pas sur-estimer les performances de la méthode Bayésienne, la méthode du Jacknife (Efron, 1986), technique de ré-échantillonnage, a été utilisée. Chaque individu est retiré tour à tour de l'échantillon. Les différentes probabilités sont calculées sur la base des n-1 individus restants. Puis la classe la plus probable a posteriori de l'individu retiré est déterminée en fonction de ses valeurs de P et T.

Les coefficients de l'équation originale de Lamendin ont été estimés sans tenir compte de la non indépendance des données (plusieurs dents par individu incluses dans l'échantillon). Une nouvelle équation de régression ($y_i = \beta_0 + \beta_1 T_i + \beta_2 P_i + \varepsilon_i$) est donc établie à partir de notre échantillon. L'application de cette méthode nous sert à évaluer les performances de la méthode Bayésienne.

Divers critères ont été utilisés pour comparer les résultats obtenus avec les deux méthodes. Le premier critère est le pourcentage de bien classés. A partir de l'âge observé, les individus sont classés dans l'une des trois classes d'âge. Une classe est également attribuée à partir de l'âge estimé. Pour la méthode Bayésienne, l'âge estimé est égal au centre de la classe la plus probable a posteriori. Le biais - pente de la droite de régression des résidus sur l'âge observé, la corrélation entre âges observés et estimés, et l'erreur moyenne absolue sont également utilisés.

3 - Résultats

Le tableau 1 présente un descriptif de l'échantillon. L'âge est compris entre 26 et 91 ans. La translucidité de la racine se développant tard dans la vie, cet indicateur ne peut pas être utilisé pour estimer l'âge au décès de personnes très jeunes. De plus, la valeur de la constante (25.3 ans) estimée originellement par Lamendin implique que l'âge des individus testés soit supérieur à cette valeur. L'âge minimal d'inclusion dans l'échantillon a donc été fixé à 26 ans. L'âge médian est de 50 ans pour les femmes et de 58 ans pour les hommes. Les trois classes d'âge sont globalement équilibrées en termes d'effectif.

	N (% colonne) Age médian (moyen)	
Classe âge	Sexe	
	Femme	Homme
<= 40 ans	34 (33.33) 35.0 (34.0)	32 (27.83) 32.5 (32.8)
> 40 ans & <= 60 ans	39 (17.97) 51.0 (50.9)	37 (32.17) 52.0 (51.9)
> 60 ans	29 (28.43) 70.0 (70.2)	46 (40.00) 71.5 (73.4)
Total	102 (47.00) 50.0 (50.8)	115 (53.00) 58.0 (55.2)

Tableau 1 : Statistiques descriptives de l'échantillon (n = 214).

Les valeurs moyennes des indicateurs P et T par sexe sont comparées avec un test de Student (Tableau 2). Les moyennes n'étant pas significativement différentes, l'analyse par sexe n'est pas nécessaire.

	Moyenne (écart type)		
Indicateur	Sexe		Test student
	Femme	Homme	p
P	17.9 (7.0)	18.2 (8.6)	0.755
T	52.0 (16.0)	50.5 (18.4)	0.529

Tableau 2 : Valeurs moyennes des indicateurs P et T par sexe

La figure 1 présente l'évolution des indicateurs en fonction de l'âge.

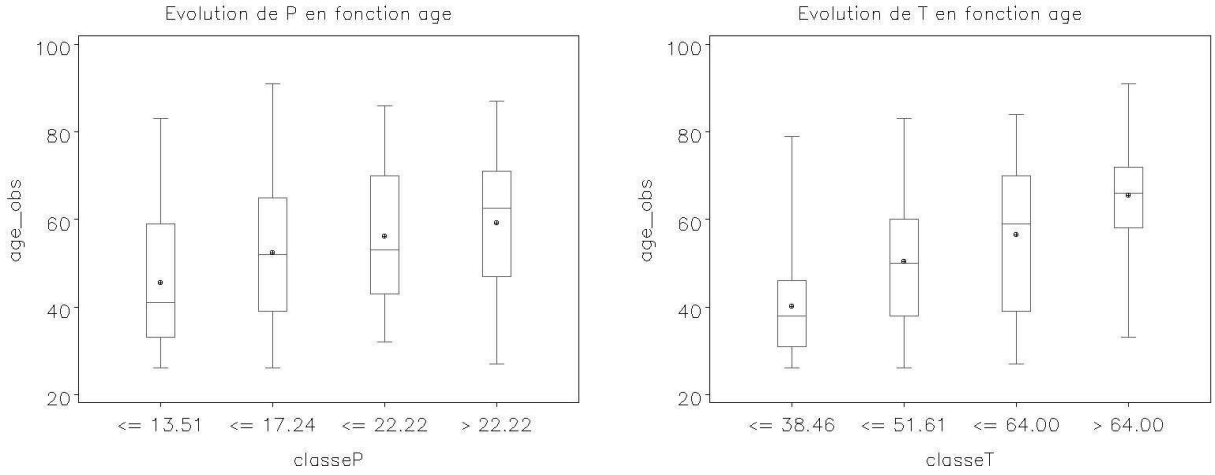

Figure 1 : Evolution des indicateurs dentaires en fonction de l'âge.

T augmente avec l'âge ($\rho = 0.57$). Par contre la relation entre P et l'âge n'est pas aussi claire ($\rho = 0.28$). Ces valeurs sont conformes à celles trouvées dans d'autres études (corrélation comprise entre 0.4 et 0.9 pour T et entre 0.1 et 0.7 pour P).

Le tableau 3 présente les résultats de la régression linéaire. Les deux indicateurs sont significatifs à 5%. Mais le modèle n'explique qu'une faible partie de la variabilité de l'âge : $R^2 = 0.35$ (R^2 compris entre 0.33 et 0.49 dans des études réalisées sur d'autres populations).

Indicateur	Estimation	Intervalle confiance 95%	p-value
Constante	20.59	[14.128 ; 27.055]	< 0.001
P	0.336	[0.096 ; 0.576]	0.006
T	0.516	[0.407 ; 0.625]	<0.001

Tableau 3 : Modèle de régression linéaire ajusté sur l'échantillon.

La figure 2 présente les résidus de la régression en fonction de l'âge. La figure montre un biais important : l'âge des plus jeunes est sur-estimé de 12.6 ans alors que l'âge des plus âgés est sous-

estimé de 12 ans. Cette tendance a également été mise en évidence dans d'autres études. Aykroyd et al. (1997) ont montré que ce biais est égal à la pente de la droite de régression des résidus sur l'âge, égale à 1 - r^2, corrélation entre l'âge et l'indicateur utilisé dans la régression. Plus la corrélation est faible, plus le biais est important.

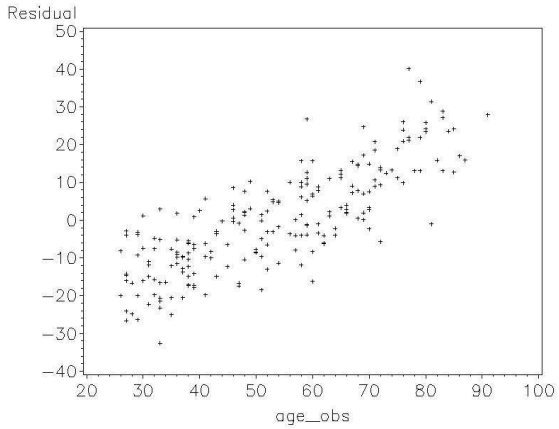

Figure 2 : Résidus de la régression linéaire en fonction de l'âge observé.

Les probabilités a priori des 3 classes, la vraisemblance et les probabilités marginales sont estimées par des proportions. La figure 3 présente les probabilités a posteriori calculées sur l'échantillon global (sans utilisation du Jacknife). Pour chaque combinaison possible des deux indicateurs, la probabilité a posteriori des trois classes d'âge est estimée. La classe d'âge estimée correspond à la probabilité a posteriori la plus grande.

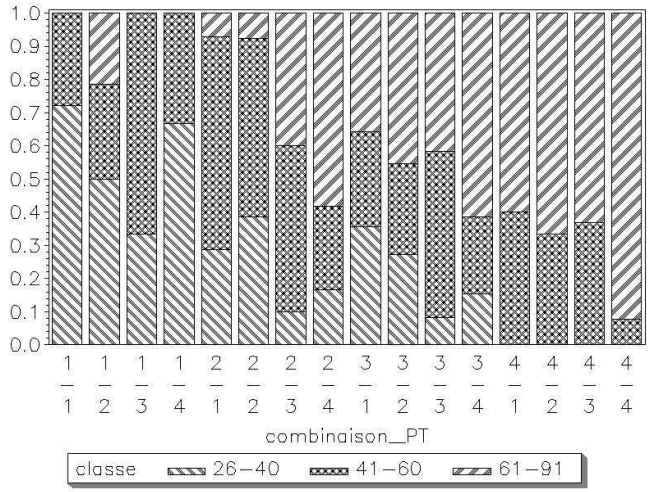

Figure 3 : Probabilité a posteriori des 3 classes d'âge calculées sur l'échantillon global.

Les critères de comparaison sont présentés dans le tableau 4. Seul le biais est amélioré par la méthode Bayésienne. En moyenne, la régression sur / sous estime l'âge d'environ 11 ans alors que la méthode Bayésienne le sur / sous estime de plus de 16 ans.

Indicateur	% bien classés	Corrélation âges observés / âges estimés	Biais	Erreur absolue moyenne (années)
Régression	50.2	0.59	0.65	10.9
Bayes	42.5	0.38	0.35	16.1

Tableau 4 : Comparaison des résultats obtenus avec la régression linéaire et la méthode de Bayes.

4 - Conclusion / Perspectives

La comparaison entre la régression et la méthode Bayésienne est difficile. En effet, la régression fournit une estimation ponctuelle alors que la méthode Bayésienne donne une classe d'âge. La discrétisation de l'âge estimé par la régression en classe conduit à une perte d'information. D'autre part, l'affectation du centre de classe comme estimation ponctuelle pour la méthode Bayésienne dépend de la largeur des classes. Or il est habituel en médecine légale d'utiliser des classes larges pour ne pas exclure des âges potentiels. La largeur des classes peut donc expliquer la forte erreur absolue moyenne de la méthode Bayésienne.

Par ailleurs, la faible corrélation entre l'âge et les indicateurs utilisés dans cette étude explique le biais des deux méthodes. Des critères basés sur d'autres parties du corps devrait être ajoutés aux deux utilisés dans cette étude. Saunders et al. (1992) ont montré que plus de critères étaient utilisés, plus l'estimation de l'âge au décès était précise.

L'application de la méthode Bayésienne en conservant les variables continues est une des perspectives de ce travail. La méthode des noyaux ou un modèle de risque de Gompertz (Prince et al. 2008) peuvent être utilisés pour estimer les densités de probabilité utilisées dans le théorème de Bayes.

Bibliographie

[1] Aykroyd, R.G., Lucy, D. Pollard, A.M and Solheim, T. (1997) Technical Note: Regression Analysis in adult age estimation. Am J. Physical Anthropol., 104, 259-265.

[2] Aykroyd, R.G. Lucy, D. and Pollard, M. 1999. Nasty, brutish, but not necessarly short: a reconsideration of the statistical methods used to calculate age-at-death from adult human skeletal and dental age indicators. American Antiquity, 64, 55-70.

[3] Baccino, E. Ubelaker, D.H. Hayek, L-A. and Zerilli, A. (1999) Evaluation of seven methods of estimating age at death from mature human skeletal remains. J Forensic Sci, 44, 931-936.

[4] Efron B. (1986) The Jacknife, the bootstrap and other resampling plans. Philadelphia: Society for Industrial and Applied Mathematics.

[5] Foti, B. Adalian, P. Signoli, M. Ardagna, Y. Dutour, O. and Leonetti, G. (2001) Limits of lamendin methods in age estimation. Forensic Sciences International, 122, 101-106.

[6] Lamendin, H. et Cambrai, J.C. (1980). Etude de la translucidité et des canalicules dentaires pour l'appréciation de l'âge. J. Med. Leg. Droit Med. 24, pp. 489-499.

[7] Lamendin, H. Baccino, E. Humbert, J.F. Tavernier J.C. Nossintchouk, R.M. and Zerilli, A. (1992) A Simple Technique for Age Estimation in Adult Corpses: The Two Criteria Dental Method. J Forensic Sci, 37, 1373-1379.

[8] Lucy, D. Aykroyd, R.G. Pollard, A.M. and Solheim, T. (1996) A Bayesian approach to adult human age estimation from dental observations by Johanson's age changes. J Forensic Sci, 41, 189-194.

[9] Lucy, D. Aykroyd, R.G. and Pollard, A.M. (2002) Non parametric calibration for age estimation. Applied Statistics, 51, 185-196.

[10] Osborne, D.L. Simmons, T.L. and Nawrocki, S.P. (2004) Reconsidering the auricular surface as an indicator of age at death. J Forensic Sci, 49, 905–911.

[11] Prince, D.A. and Konigsberg, L.W. (2008) New formulae for estimating age-at-death in the Balkans utilizing Lamendin's dental technique and Bayesian analysis. J. Forensic Sci., 53, 578-587.

[12] Saunders, S.R. Fitzgerald, C. Rogers, T. Dudar, C. and McKillop, H. (1992) A test of several methods of skeletal age estimation using a documented archaeological sample. Can Soc Forensic Sci J, 2, 97–118.

[13] Schmitt, A. 2008. Une nouvelle méthode pour discriminer les individus décédés avant ou après 40 ans à partir de la symphyse pubienne. J. Med. Leg. Droit Med., 51, 15-24.

[14] Schmitt, A. Saliba-Serre, B. Tremblay, M. and Martrille, L. Looking for an appropriate method to assess age at death from root translucency and periodontosis. Soumis.