Chapitre 12
Statistique bayésienne

Philippe LANOS(1), Henri CAUSSINUS(2)

(1) Archéosciences Bordeaux, Université Bordeaux Montaigne & Géosciences Rennes, Université de Rennes, (2) Institut de Mathématiques, Université Paul Sabatier, Toulouse

Le concept de **statistique bayésienne** est issu des travaux de Thomas Bayes (1702-1761). En 1763 est publié, après lecture devant la « Royal Society », un texte posthume de Thomas Bayes intitulé « Essai en vue de résoudre un problème de la doctrine des chances » (texte disponible *in* Barnard, 1958). Dans cet essai, alors qu'étaient seulement envisagées jusque-là les probabilités des diverses conséquences en fonction de causes données, Thomas Bayes aborde le problème inverse qui a pour but d'analyser les causes (supposées inconnues) à partir d'événements observés. D'où l'idée de **probabilité inverse** qui, redécouverte un peu plus tard par Laplace, va influencer de façon importante les développements de la statistique au XX^{e} siècle. Une grande part du succès actuel des méthodes bayésiennes vient du fait que le calcul des distributions *a posteriori* et prédictives a été rendu possible grâce aux progrès de l'informatique et à l'utilisation de méthodes de simulation basées sur les algorithmes de Metropolis-Hastings et les chaînes de Markov (Droesbeke *et al.*, 2002, chap. 1). De plus, l'approche bayésienne a fini par s'imposer comme méthode inférentielle (*cf.* 8) pertinente parce qu'elle a permis aux praticiens de diverses disciplines de résoudre plus facilement et plus naturellement leurs problèmes. Ainsi, l'analyse statistique bayésienne, qui est un processus d'acquisition de connaissance, offre des outils flexibles et puissants qui permettent d'intégrer des modèles complexes dans une approche d'inférence statistique en prenant en compte l'ensemble des sources de variabilité et d'incertitude : on peut intégrer des informations de plusieurs sources et de nature différente, données bruitées et hétérogènes, information *a priori* sur certains paramètres des modèles. Cette approche permet de quantifier explicitement l'incertitude sans approximations et propose un cadre probabiliste adapté pour la propagation des incertitudes dans les prédictions. L'analyse bayésienne trouve de très nombreuses applications en médecine,

en génomique, en traitement du signal, en économie ou encore en finance, ainsi qu'en archéologie et archéométrie !

La **formule de Bayes** permet de calculer, à partir de données* (conséquences), les probabilités d'événements (causes) basées sur des **informations *a priori*** ou des croyances antérieures sur cet événement ou les conditions qui lui sont liées. Ainsi, dans le cas simple d'un ensemble de données dépendant d'un seul paramètre noté θ, la démarche bayésienne consiste à « probabiliser » ce paramètre inconnu θ en lui associant une loi de probabilité (*cf.* 7) dite **loi *a priori*** car cette loi représente, pour le statisticien bayésien, l'ensemble des informations *a priori* disponibles sur le paramètre θ ainsi que les imprécisions qui s'y rattachent (Robert, 2001). Le choix de cette loi *a priori* est bien entendu crucial pour la suite de l'analyse : deux lois *a priori* différentes conduisent à deux systèmes d'inférence différents, même si l'accumulation de données finit par gommer cette différence. Le paradoxe de l'approche bayésienne est qu'il n'existe pratiquement jamais une unique loi *a priori*, mais plutôt un ensemble de lois compatibles avec les informations *a priori* disponibles et les opinions des modélisateurs ! Tout au plus peut-on chercher à réduire l'influence de ce choix en retenant des lois à faible contenu informatif, comme les lois de maximum d'entropie, ou des lois dites non informatives, comme les lois de Bernardo-Berger, qui minimisent l'information apportée par la loi *a priori* face à l'information apportée par les données. Selon Christian Robert, « *il [...] semble que cette imprécision fondamentale sur la loi a priori, souvent assimilée à la subjectivité de l'approche bayésienne, en est au contraire sa force : reconnaissant que l'ensemble de nos connaissances sur θ est incomplet et imparfait, nous devrions aussi accepter que cette imperfection se transfere à l'inférence résultante après observation des données* ».

Les parties de ce chapitre 12 consistent à expliciter de façon plus formelle la démarche bayésienne, d'abord d'un point de vue général (paragraphes 1 et 2), puis en reprenant la démarche suivie successivement dans les chapitres 1.2, 1.4, 1.5 et 3.5 (paragraphes 3 à 6). Il s'agit là de donner les clés de compréhension des raisonnements et des formules utilisés dans ces applications. Il peut y avoir redondance pour le lecteur mais c'est un choix délibéré du présent manuel, qui permet un apprentissage progressif allant de l'application vers la formalisation.

1 Le paradigme bayésien : formalisation d'un processus d'information

Nous introduisons d'abord les principes de base du calcul des probabilités en considérant le cas simple de deux variables aléatoires continues Y et θ de **densité de probabilité jointe** ou distribution jointe (*cf.* 7) $p(y, \theta)$. En introduisant **densités conditionnelles** et **densités marginales**, on peut écrire :

$$p(y, \theta) = p(y \mid \theta)p(\theta) = p(\theta \mid y)p(y) \qquad (1)$$

où la barre verticale signifie « sachant la variable qui suit fixée » ou « conditionnellement à la variable qui suit » (ce n'est pas un opérateur de calcul, mais un raccourci d'écriture). D'autre part, par abus de langage, nous désignons par p des fonctions

différentes en les distinguant seulement par les variables sur lesquelles elles portent. Ainsi :

$p(\theta \mid y)$ représente la densité de probabilité conditionnelle de θ sachant y fixé.

$p(y \mid \theta)$ représente la densité de probabilité conditionnelle de y sachant θ fixé.

$p(\theta)$ représente la densité de probabilité dite marginale de θ, autrement dit quelle que soit la valeur de y. Cette densité s'obtient en sommant $p(y, \theta)$ sur toutes les valeurs de y, somme usuelle dans le cas où y est une variable discrète, intégrale si y est continue.

$p(y)$ représente la densité de probabilité dite marginale de y, autrement dit quelle que soit la valeur de θ. Elle s'obtient par sommation de $p(y, \theta)$ sur θ.

Cas de l'**indépendance des variables** :

Si $p(y \mid \theta) = p(y)$, la variable Y est dite indépendante de Θ. La connaissance de θ n'apporte rien sur celle de y.

Si $p(\theta \mid y) = p(\theta)$, la variable Θ est dite indépendante de Y. De façon symétrique, la connaissance de y n'apporte rien sur celle de θ.

Ainsi, dans ce cas de l'indépendance, la densité jointe est égale au produit des densités marginales, soit $p(y, \theta) = p(y)p(\theta)$. On dit que les variables Y et Θ sont indépendantes.

Si y est une observation, le plus souvent une série d'observations, mesure(s) issue(s) d'un processus expérimental dont on connaît la loi de probabilité sachant un paramètre θ inconnu, la densité $p(y \mid \theta)$ est appelée **vraisemblance** (***Likelyhood*** en anglais), aussi notée $L(y, \theta)$. L'objectif est d'estimer la valeur de l'inconnue θ grâce à cette série d'observations. L'approche fréquentiste la plus usuelle consiste à chercher la valeur du paramètre θ qui rend maximale l'expression de la vraisemblance, autrement dit qui rend l'observation la plus probable possible.

L'approche bayésienne est différente : elle consiste à considérer le paramètre θ comme une variable aléatoire qui suit une loi $p(\theta)$ dite *a priori* car elle préexiste à l'observation des valeurs y. Elle est donc définie grâce à des informations issues, héritées de connaissances antérieures. Puisque les mesures y sont reliées à θ grâce au modèle de vraisemblance $L(y, \theta)$, ces observations y vont apporter de nouvelles informations sur θ, autrement dit elles vont permettre de mettre à jour, d'affiner notre connaissance sur θ en considérant, non plus sa loi de probabilité *a priori*, mais sa loi de probabilité conditionnée par les observations y, de densité $p(\theta \mid y)$. Cette densité de probabilité est appelée **densité *a posteriori***. Nous pouvons la calculer explicitement en inversant la formule (1), soit :

$$p(\theta \mid y) = p(y \mid \theta)p(\theta)/p(y) \quad (2)$$

Le nom de formule de Bayes se réfère en général à cette dernière expression. Le point important, qu'il revient à Bayes d'avoir mis en lumière, est que cette loi conditionnelle de θ sachant y ne peut pas se déduire seulement de la loi conditionnelle de y sachant θ mais nécessite la connaissance de la loi *a priori* $p(\theta)$, loi qui précisément constitue le fondement de la statistique dite bayésienne.

Notons que, puisque la densité marginale $p(y)$ ne dépend pas de θ, on peut aussi écrire la formule (2) comme une relation de proportionnalité (en utilisant le symbole $\propto$ pour dire « proportionnel à », ou « égal à un facteur multiplicatif près à ») :

$p(\theta \mid y) \propto p(y \mid \theta)p(\theta)$

Le paradigme bayésien est illustré en figure 1 dans le cas simple de l'estimation *a posteriori* d'un seul paramètre θ. La connaissance *a priori* sur θ est représentée par une distribution de probabilité $p(\theta)$ qui a ici une forme de loi gaussienne étalée, ce qui représente une connaissance plutôt vague sur la valeur du paramètre recherché. Pour préciser notre connaissance sur θ, une mesure y est effectuée en laboratoire, à laquelle on associe une erreur expérimentale σ. Supposons, par exemple, que la vraisemblance $L(y, \theta)$ qui relie y à θ soit une loi normale (*cf.* 7) de moyenne $g(\theta)$ et de variance σ^2. La formule (2) permet de calculer la distribution *a posteriori* de θ sachant la mesure y effectuée et l'erreur associée σ, soit :

$$p(\theta \mid y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-g(\theta)}{\sigma}\right)^2} p(\theta)/p(y) \quad (3)$$

On voit dans l'exemple de la figure 1 que la distribution *a posteriori* apparait mieux définie (la moyenne est décalée vers de plus grandes valeurs de θ) et plus précise (sa variance est moins grande). Ainsi, on peut qualifier le paradigme bayésien de **processus d'information** ou de **processus d'apprentissage** (par acquisition de nouvelles données y) qui permet de réaliser une mise à jour (« updating » en figure 1) de la connaissance que l'on a, *a priori*, sur le paramètre θ. Pour résumer, la connaissance (autrement dit l'incertitude) sur la valeur d'une quantité inconnue θ d'un modèle est représentée sous la forme d'une distribution de probabilité (Parent & Bernier, 2007) qui prend en compte les erreurs mises en jeu dans le processus de mesure. Les probabilités sont interprétées comme des degrés de crédibilité associés à chaque valeur des inconnues (c'est une interprétation subjective des probabilités). La loi *a priori* apporte une connaissance sur le paramètre du modèle, θ, ceci avant d'avoir effectué les observations (avant acquisition des données, avant mesure). Ce dernier point constitue l'originalité de l'approche bayésienne : elle modélise notre connaissance *a priori* sur un ou des paramètres que l'on cherche à estimer, avant de la combiner avec de nouvelles données qui viendront améliorer à leur tour notre connaissance sur ces paramètres. Pour une mise en œuvre pratique dans les logiciels R et OpenBugs, nous recommandons la lecture du livre très pédagogique de M.K. Cowles (2013).

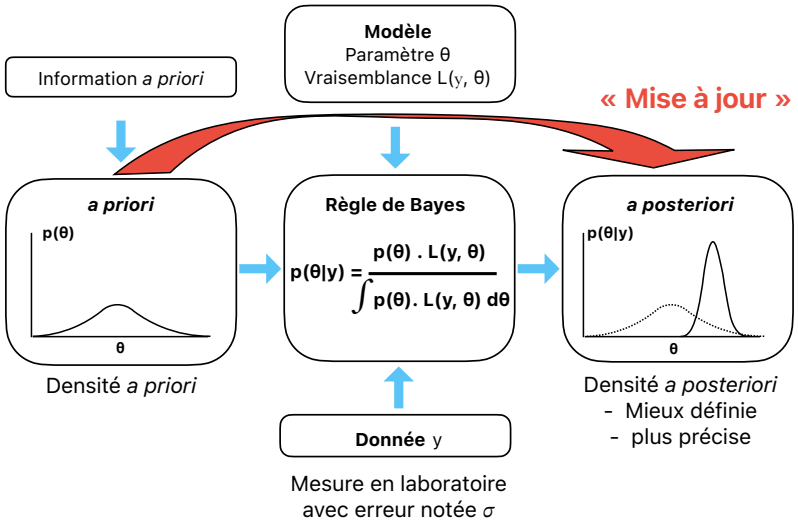


FIG. 1 : Illustration du processus bayésien (schéma d'après Bernier *et al.*, 2000) : expression de la distribution *a posteriori* $p(\theta \mid y)$ en fonction de la distribution *a priori* $p(\theta)$ et de la fonction de vraisemblance $L(y, \theta)$ où θ représente la variable d'intérêt et où y représente la mesure effectuée (appelée ici data, pour « donnée »). Ainsi le processus bayésien opère une « mise à jour » de l'information *a priori* par l'acquisition de nouvelles données.

2 Expression des résultats statistiques *a posteriori*

La distribution de probabilité *a posteriori* d'une variable peut être résumée par une **région** appelée **HPD** pour « *highest posterior density* », c'est-à-dire par une union d'intervalles de plus forte densité de probabilité à un niveau de confiance 1- α fixé, le plus souvent à 95% (valeur habituellement choisie) : cette région correspond par exemple aux deux segments en rouge en figure 2a (Vibet *et al.*, 2016). La probabilité à 95% est représentée par la surface coloriée en gris.

On peut aussi déterminer un intervalle de crédibilité* (CI) qui est défini comme l'intervalle le plus court possible tel que la surface sous la courbe pour cet intervalle, c'est-à-dire la probabilité, soit de 95% : cela correspond au segment en rouge en figure 2b.

Rappelons que la probabilité* est mesurée par la surface sous la courbe de densité entre deux valeurs fixées de la variable. La surface totale sous la courbe est par définition de 100%. Lorsque la distribution est monomodale* (avec un seul maximum), on montre que la région HPD déterminée se réduit à un seul intervalle égal à l'intervalle de crédibilité.

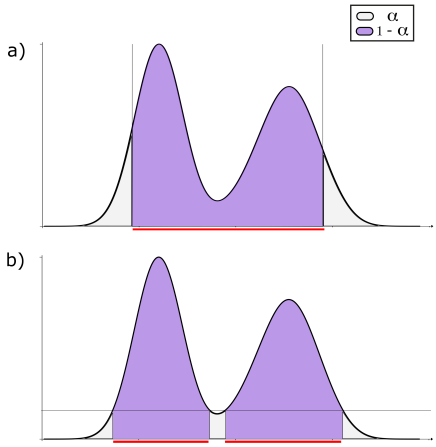


FIG. 2 : Pour une même distribution de date : a) Région HPD (Highest Posterior Density region) composée ici de deux intervalles dont la surface sous la courbe est égale à $1 - \alpha$, avec α choisi égal à $0,05$ (5%). Les intervalles de la région HPD caractérisent les zones de plus forte densité. b) Intervalle de crédibilité : c'est le plus court intervalle tel que la surface sous la courbe soit de 95%.

Nous pouvons aussi résumer la distribution par des paramètres ponctuels tels que le mode *a posteriori* (MAP), la moyenne* ou encore les quantiles* (*cf.* 4), en particulier les quartiles Q_1, Q_2 = médiane et Q_3 ou les quantiles, par exemple 0,025 et 0,975 pour approcher l'intervalle de crédibilité à 95% (comme il est usuel avec l'intervalle de confiance en statistique fréquentiste).

Le **MAP** est la valeur de la variable qui correspond au **mode *a posteriori***, c'est-à-dire au maximum (ou pic) de densité de probabilité. Cette valeur « oublie » l'incertitude associée à la variable et qui est représentée par la distribution toute entière. Dire que ce pic correspond à la valeur la plus probable est un abus de langage. La probabilité de cette valeur ponctuelle est en fait nulle ! Réduire la variable à la valeur du MAP reviendrait à considérer que cette valeur est certaine (représentable par un « Dirac », dont la surface est égale à 1), ce qui est faux vu la distribution *a posteriori*.

Il est déconseillé de considérer la valeur moyenne (avec son écart-type) car celle-ci peut se situer dans des régions de densité de probabilité très faible dans le cas de distributions multimodales. Elle pourrait donc être peu représentative.

Les quartiles Q_1, Q_2 et Q_3 sont les valeurs de la variable qui partagent la distribution en quatre surfaces égales de 25% de probabilité chacune. Q_2 correspond à la définition de la médiane. Ces trois valeurs peuvent servir à tracer les boites à moustaches*, encore appelées « boxplot » (*cf.* 4). Ce graphique est très souvent utilisé dans le domaine bayésien pour exprimer les résultats *a posteriori*. Mais dans le domaine de la datation, l'habitude a été prise, suite aux travaux de C. Buck (1996), d'utiliser plutôt les régions HPD définies plus haut, essentiellement pour la raison qu'elles résument et localisent mieux l'information chronologique.

Insistons cependant sur le fait qu'un MAP, un intervalle de crédibilité IC, une région HPD, etc. sont tous des résumés statistiques qui font perdre une partie de l'information bayésienne *a* posteriori. C'est dans la courbe de densité de probabilité tout entière que cette information reste entièrement contenue.

3 Bayes en action : la calibration d'une mesure en date calendaire.

3.1 Calibration d'un âge 14C

La figure 3 donne un exemple d'application de la formule de Bayes à la résolution d'un problème classique en chronologie : la **calibration** d'un âge ^{14}C en **date calendaire** (*cf.* 1.2 & 1.5), date que l'on notera ici t, car nous réservons la notation θ pour désigner la date d'un événement, *cf.* section 5 infra). Le temps t sera par exemple la date d'abattage d'un arbre si la datation ^{14}C porte sur le dernier cerne sous l'écorce.

Avant tout processus de mesure (datation), l'information *a priori* dont nous disposons sur le temps t est une information dite vague, uniforme sur l'intervalle fixé ici, pour l'exemple, à [5000 ; -3000], intervalle appelé plage d'étude (*study period*) dans le logiciel ChronoModel (Lanos & Dufresne, 2019). Ceci veut traduire l'idée que nous n'avons *a priori* aucune préférence pour une date particulière dans cet intervalle. Nous représentons ce manque d'information par une distribution $p(t)$ qui est une densité de probabilité uniforme sur l'intervalle [-5000 ;-3000]. Par conséquent, pour améliorer notre connaissance *a priori* sur le temps t, nous n'avons pas d'autre moyen que d'effectuer une datation, en l'occurrence une mesure d'âge $y = A$ qu'on peut relier au temps t. La mesure d'âge, obtenue à partir de la détermination du taux résiduel de ^{14}C dans l'échantillon, est représentée par une distribution gaussienne $p(A \mid t)$ (fonction qui correspond à la vraisemblance $L(y, t)$ de la section 1), de moyenne $g(t)$ et variance $S(t)^2 = {\sigma_A}^2 + {\sigma_{g(t)}}^2$. Sa représentation graphique est donnée en grisé sur l'axe vertical de l'âge en figure 3. La fonction $g(.) correspond la courbe de calibration du ^{14}C$ avec son erreur $\sigma_{g(.)}$. L'erreur σ_A correspond à l'erreur expérimentale sur l'âge A fourni par le laboratoire.

En repartant de la formule de Bayes (2), l'expression de la densité *a posteriori* de la date t sachant la mesure d'âge A est donnée par la formule :

$$p(t \mid A) = \frac{1}{S(t)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{A-g(t)}{S(t)}\right)^2} p(t)/p(A) \qquad (4)$$

avec $p(t) = \frac{1}{(t_M - t_m)} 1_{[t_m, t_M]}(t)$: distribution uniforme de t sur l'intervalle $[t_m, t_M]$

En parcourant successivement toutes les valeurs de t dans l'intervalle [-5000 ; -3000], nous pouvons tracer la distribution de date *a posteriori* $p(t \mid A)$, qui est représentée en bleu sur la figure 3. Cette distribution n'est plus gaussienne à cause des variations locales de la courbe de calibration utilisée (ici Intcal13.14C). Toute l'information *a posteriori* est contenue dans cette distribution $p(t \mid A)$. Pour faciliter l'expression de ce résultat, on pourra utiliser les différents résumés statistiques présentés en section 2.

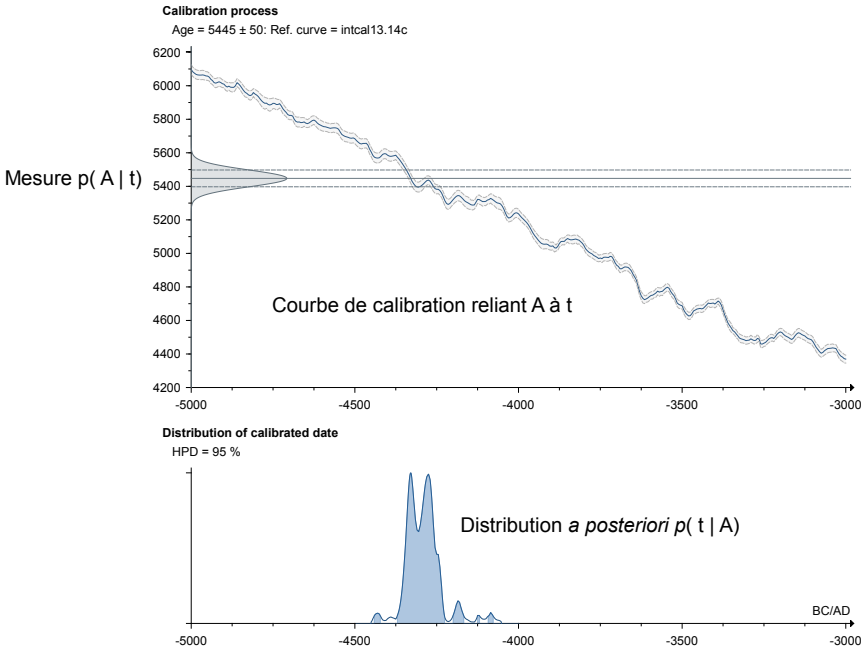


FIG. 3 : Processus de calibration d'une mesure d'âge ^{14}C, notée A, en date calendaire t via la courbe de calibration *IntCal13* (Reimer *et al.*, 2013) qui relie A à t. La distrubution de date *a posteriori* peut présenter de nombreux modes, ceci en fonction de l'allure de la courbe de calibration.

Ce type de calcul est implémenté dans différents logiciels de calibration et de modélisation des données chronologiques, par exemple : Calib, OxCal (Bronk Ramsey, 1995, 2009a), BCal (Buck *et al.*, 1999), ChronoModel (Lanos & Dufresne, 2019), Archeophases (Philippe & Vibet, 2020)

3.2 Calibration d'une mesure d'inclinaison en datation par archéomagnétisme

La figure 4 montre un autre exemple de calibration dans le cas de la méthode de datation par archéomagnétisme. On mesure ici l'inclinaison du champ magnétique terrestre ancien porté par une terre cuite prélevée sur les parois d'un four de potier. On datera ainsi la dernière chauffe du four, donc sa dernière utilisation. La courbe de calibration représente cette fois la variation séculaire de l'inclinaison du champ magnétique terrestre en France (à Paris) au cours des trois derniers millénaires. Les points rouges, avec leurs erreurs représentées par les barres verticales, représentent les points de référence bien datés ayant servi à construire cette courbe (en utilisant là encore une méthode bayésienne : Lanos, 2004). La barre en bleu en haut de la figure 4 représente la distribution $p(t)$ *a priori* uniforme sur le temps calendaire t dans l'intervalle $[t_m, t_M]$. La mesure M (ici l'inclinaison) avec son erreur d'estimation σ est représentée par une distribution gaussienne (en bleu) à gauche de l'axe vertical. La distribution de date calendaire calibrée *a posteriori* sera donc donnée sur l'axe du temps horizontal, comme dans le cas du radiocarbone, par $p(t \mid M)$. Nous obtenons

ici une distribution *a posteriori* multimodale du fait que la valeur moyenne M recoupe la courbe de calibration en trois points différents.

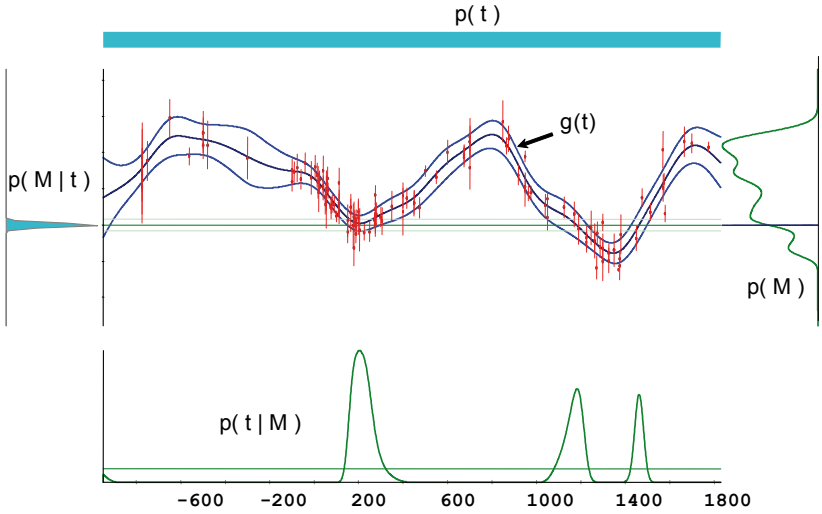


FIG. 4 : Processus de calibration d'une mesure M (ici : inclinaison du champ magnétique ancien en datation par archéomagnétisme) en date calendaire t via la courbe de variation séculaire de l'inclinaison du champ magnétique terrestre en France durant les trois derniers millénaires (Hervé *et al.*, 2013). L'allure oscillante de la courbe de variation de l'inclinaison peut engendrer une distribution de date *a posteriori* multimodale avec des intervalles HPD plus ou moins espacés dans le temps. Le choix d'un intervalle HPD particulier nécessite l'apport d'une information *a priori* supplémentaire.

Nous avons ajouté sur la figure 4 une distribution qui n'est généralement pas donnée dans les publications : c'est la distribution $p(M)$ de la mesure M, qui ne fait pas intervenir le temps t : elle est obtenue par intégration du produit $p(M \mid t)p(t)$ sur le temps t - on parle dans ce cas de distribution « marginale » de M. Ainsi, sur les trois derniers mille ans, cette distribution donne la densité de probabilité des valeurs prises par l'inclinaison. Cette distribution peut servir de critère de fiabilité, de pertinence pour une calibration archéomagnétique : en effet, si la mesure M tombe en dehors du support à 99% (par exemple) de cette distribution, alors la calibration peut être considérée comme non significative, car la valeur de M ne recoupe pas la courbe de calibration. Notons que cette situation n'intervient pas en calibration radiocarbonique car toute mesure d'âge A recoupera la courbe de calibration, à moins d'une grosse erreur sur la mesure !

3.3 Comment formuler un résultat de datation ?

L'expression d'une datation peut se formuler ainsi : « *Au vu des données de datation, et à un niveau de confiance de* 95%, *les fossés ont été creusés entre 4265 et 4075 avant notre ère (selon l'intervalle HPD [4265,* 4075] *BC à* 95% *) et abandonnés entre 4160 et 3975 avant notre ère (selon l'intervalle HPD [4160, 3975] BC à* 95% *)* ». Plus

précisément, dans le contexte bayésien, on dira : « il y a 95% de chances que la date se situe dans cet intervalle, sachant les données de datation observées ». Remarquons que le fait de dire « entre 4265 et 4075 à 95% de confiance » revient à ramener la forme de la distribution de probabilité dans cet intervalle à une distribution uniforme : on « oublie » la forme réelle de la distribution *a posteriori* en résumant l'information à cet intervalle, au niveau de confiance fixé.

Remarque importante : la probabilité d'une date "instant" est nulle, sauf si l'on connait cet instant avec une infinie précision, ce qui n'arrive jamais ! Pour une date entachée d'une erreur, donc représentée par une distribution de probabilité, le mode *a posteriori* (MAP) est la date qui correspond au maximum de la densité de probabilité et non pas au maximum de la "probabilité". Sinon on ferait croire que cette date MAP est de probabilité maximale et que c'est elle qu'il faut retenir. Il est donc recommandé d'utiliser les intervalles de confiance HPD à 95%, qui sont obtenus sachant les données observées, que le nombre de données soit petit ou grand.

4 Bayes en action : modèle de date d'événement (*Event model*)

Le paradigme du logiciel ChronoModel est basé sur le concept de **date d'événement** (***Event date***). L'idée fondamentale à la base de ce concept est de différencier deux variables de temps :

- le temps, noté t, fourni par une datation chronométrique, issu d'une calibration individuelle (*cf.* section 3 supra).

- le temps, noté θ, d'un événement que l'on cherche à dater, appelé événement cible, ou événement d'intérêt (c'est celui qui intéresse l'archéologue ou l'historien), qu'on peut encore appeler « Fait ».

Dans cette approche, mise en œuvre dans le chapitre 1.2, on considère que l'événement est « ponctuel » dans le temps, qu'il n'a donc pas de durée. La date de l'événement θ sera estimée par une ou plusieurs dates chronométriques t_i (l'indice i désignant une date chronométrique) que l'on suppose *a priori* contemporaines entre elles et donc de l'événement cible θ. Ce point constitue le cœur du modèle de date d'événement : l'archéologue doit définir l'événement cible θ ainsi que l'ensemble des dates t_i supposées contemporaines de l'événement à dater (Dean, 1978). C'est un modèle à effet aléatoire qui à chaque date t_i associe une erreur aléatoire individuelle σ_i pour prendre en compte la dispersion des dates t_i par rapport à la date cible θ (Lanos & Philippe, 2017). Cette erreur, appelée aussi **erreur irréductible**, représente les effets des erreurs humaines dans le processus de datation, les erreurs d'échantillonnage, les problèmes taphonomiques, de pollution ou de variabilité des conditions de conservation, ou tout autre facteur d'origine inconnue et qui font que les datations individuelles t_i ne donnent pas toutes le même résultat. Le modèle statistique de date d'événement possède une structure bayésienne hiérarchique qui est représentée par le **graphe acyclique orienté** (**DAG** : *directed acyclic graph*) en figure 5 . Nous passons de la mesure M_i observée (avec son erreur expérimentale S_i) à la mesure vraie inconnue μ_i (variable éliminée par intégration), puis au temps t_i via le processus de calibration (*cf.* section 3), et enfin à la date cible θ et à l'erreur individuelle σ_i via un modèle statistique

défini *a priori* gaussien qui relie t_i à θ_j et σ_{ji}. Enfin, l'approche bayésienne nécessite aussi de définir une densité de probabilité *a priori* pour chacune des variables θ et σ_i. Les motivations dans le choix de ces densités *a priori* et les détails mathématiques sont exposés dans Lanos & Philippe (2017).

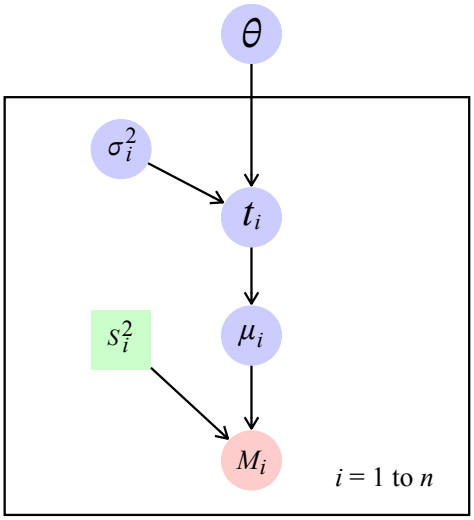


FIG. 5 : Graphe acyclique orienté (*Directed Acyclic Graph* : *DAG*) du modèle d'événement (*Event*) dans ChronoModel (Lanos & Philippe, 2017), qui décrit les variables mises en jeu dans les noeuds (cercles en bleu), les données (cercle orange et carré vert), ainsi que les relations de dépendance biérarchique probabiliste (flèches). Chaque noeud (variable) est conditionnellement indépendant de ses non-descendants sachant ses parents.

La densité de probabilité jointe de ce modèle d'événement pour n dates observées s'écrit alors :

$$p\left(M_1,\ldots,M_n,\mu_1,\ldots,\mu_n,t_1,\ldots,t_n,\sigma_1^2,\ldots,\sigma_n^2,\theta\right) = p(\theta)\prod_{i=1}^{n}\left\{p\left(M_i \mid \mu_i\right)\mathrm{p}\left(\mu_i \mid t_i\right)\mathrm{p}\left(t_i \mid \sigma_i^2,\theta\right)p\left(\sigma_i^2\right)\right\} \quad (5)$$

où les distributions conditionnelles qui apparaissent dans la décomposition hiérarchique sont données par :

$M_i \mid \mu_i \sim N\left(\mu_i, s_i^2\right)$ loi de l'observation (vraisemblance) sur les mesures expérimentales M_i (datation en laboratoire)

$\mu_i \mid t_i \sim N\left(g_i\left(t_i\right), \sigma_{g_i}^2\left(t_i\right)\right)$ loi de la mesure inconnue μ_i (variable latente)

$t_i \mid \theta, \sigma_i^2 \sim N\left(\theta, \sigma_i^2\right)$ loi *a priori* gaussienne sur t_i (modèle d'événement)

$\sigma_i^2 \sim \text{Shrinkage}\left(s_0^2\right)$ loi *a priori* « shrinkage » sur la variance irréductible σ_i^2, de paramètres fixé s_0^2

$\theta \sim \text{Unif}(T)$ loi *a priori* uniforme de l'événement θ sur l'intervalle T : période d'étude définie *a priori* par l'archéologue.

Une des propriétés remarquables de ce modèle d'événement est qu'il est structurellement simple et qu'il permet de gérer de façon quasi automatique la présence de **dates hors tendance ou aberrantes** (***outliers***). Il n'est donc pas nécessaire de construire un modèle d'*outlier* comme c'est le cas dans OxCal (Bronk Ramsey, 2009b). En effet, l'introduction d'erreurs aléatoires individuelles σ_i permet d'opérer une pondération automatique des dates t_i. Le calcul de la date cible θ apparait alors comme un calcul particulier de moyenne basé sur le principe d'une loi de la majorité : autrement dit, les dates qui s'éloignent d'un groupement principal de dates t_i ne contribueront pas de façon significative au calcul de la date cible θ et seront par conséquent caractérisées par une erreur individuelle σ_i *a posteriori* d'autant plus forte que son caractère hors tendance sera accentué. Par conséquent, le modèle de date d'événement s'avère être un modèle statistique robuste, c'est-à-dire peu sensible aux dates aberrantes. Toute date incohérente sera rendue cohérente grâce à une variance individuelle qui deviendra grande. Des essais de contamination montrent que la date d'un événement reste stable jusqu'à une proportion de 30% de dates aberrantes (Lanos & Philippe, 2017). Cette propriété de robustesse intervient aussi dans le cas de séquences stratigraphiques où des dates chronométriques t_i peuvent être en contradiction avec l'ordre stratigraphique observé. Les erreurs individuelles σ_i permettent alors de « gérer » ces incohérences afin qu'une série de dates cibles $\theta_1, \ldots, \theta_r$ respecte *a posteriori* l'ordre stratigraphique imposé.

5 Bayes en action : modèle de combinaison de mesures en datation par luminescence

Nous exposons ici le modèle de combinaison de mesures en partant de l'exemple de la datation par thermoluminescence exposé au chapitre 1.4. Dans ce cas, la mesure est une mesure d'âge notée A.

On rappelle l'équation d'âge de la luminescence :

$A = \frac{D_a}{\dot{d}}$ (6)

où D_a est la dose archéologique, ou quantité d'énergie (ramenée à l'unité de masse de matière, unité SI : le Gray, $1\text{Gy} = 1\ \text{J}\cdot\text{kg}^{-1}$) accumulée depuis l'instant zéro à dater, et où $\dot{d}$ est le débit de dose d'irradiation annuelle ou dose annuelle d'énergie absorbée par le matériau qui a servi à la mesure de luminescence. Notons que le point sur la lettre d fait référence à la « dérivée » en mathématiques (notation souvent utilisée par les physiciens à la place de $d'(t)$), pour dire qu'il s'agit d'un débit de dose, autrement dit d'un taux de variation de la dose d par rapport au temps t.

L'âge A est relié à la date calendaire t par un processus de « calibration » simple qui se résume à la transformation linéaire suivante :

$A = g(t) = t_0 - t$ (7)

où t_0 correspond à l'année de détermination de l'âge au laboratoire.

La détermination de l'âge est effectuée sur plusieurs échantillons différents mais qui sont supposés contemporains de l'événement t. On choisit un modèle bayésien hiérarchique gaussien avec un premier niveau correspondant aux erreurs expérimentales

$\sigma_{i,stat}$, dites statistiques, affectant chaque échantillon i. Le second niveau correspond à une erreur systématique σ_{syst} partagée par l'ensemble des échantillons.

La densité de probabilité jointe s'écrit alors :

$$\begin{aligned} p\left(A_1,\ldots,A_n,\mu,t\right) &= \prod_{i=1}^{n}\left\{p\left(A_i \mid \mu,t\right)\right\}\mathrm{p}(\mu \mid t)\mathrm{p}(t) \\ &= \prod_{i=1}^{n}\left\{\frac{1}{\sigma_{i,\text{ stat}}\sqrt{2\pi}}\right\} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{A_i-\mu}{\sigma_{i,\text{ stat}}}\right)^2} \frac{1}{\sigma_{\text{syst}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu-g(t)}{\sigma_{\text{syst}}}\right)^2} p(t) \end{aligned} \tag{8}$$

où la variable μ est la valeur de l'âge commun qui est ici inconnu. La loi *a priori* du temps t est supposée uniforme sur un intervalle T, faute d'information plus précise, soit $p(t) = 1_T(t)$, où $1_T(.)dsignelafonctionindicatricequivaut1$ sur l'intervalle T, et 0 en dehors. Autrement dit, toutes les dates t sont *a priori* également probables sur cet intervalle.

Le graphe acyclique orienté (DAG) de ce modèle est donné en figure 6.

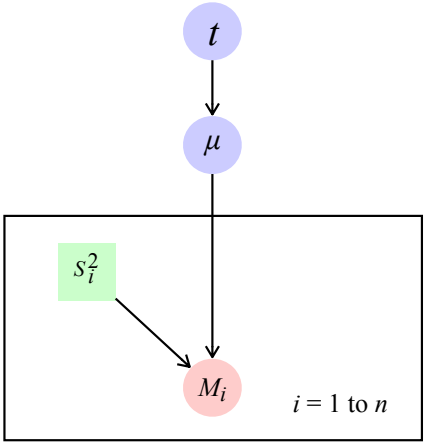


FIG. 6 : Graphe acyclique orienté (*Directed Acyclic Graph* : DAG) du modèle de combinaison (*Cf.* légende de la figure 5).

On montre que la moyenne $\bar{A}$ des n âges $A_1,\ldots,A_n$ observés définie par :

$$\bar{A} = \frac{1}{W}\sum_{i=1}^{n} w_i \cdot A_i \tag{9}$$

suit une loi gaussienne de moyenne μ et de variance $\sigma^2_{\text{stat}} = \frac{1}{w}$

$$\text{avec} \quad w_i = \frac{1}{\sigma^2_{i,\text{ stat}}} \quad \text{et} \quad W = \sum_{i=1}^{n} w_i \tag{10}$$

On peut noter que la variance σ^2_{stat} s'exprime comme l'inverse de la moyenne harmonique des variances statistiques individuelles $\sigma^2_{i,stat}$ divisée par le nombre n d'échantillons (*cf.* section 3.1 du chapitre 4). Il s'agit donc bien d'une erreur, qui diminue avec le nombre d'échantillons.

La loi jointe des variables $\bar{A}, \mu$ et t s'écrit alors :

$$p(\bar{A}, \mu, t) = \frac{1}{\sigma_{\text{stat}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\bar{A}-\mu}{\sigma_{\text{stat}}}\right)^2} \frac{1}{\sigma_{\text{syst}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu-g(t)}{\sigma_{\text{syst}}}\right)^2} 1_T(t) \qquad (11)$$

La variable μ (dite **variable latente** car elle ne peut être mesurée, observée directement) peut être éliminée par intégration entre $-\infty$ à $+\infty$, d'où la densité de probabilité des deux variables $\bar{A}$ et t :

$$p(\bar{A}, t) = \frac{1}{\sigma_{\text{total}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\bar{A}-g(t)}{\sigma_{\text{total}}}\right)^2} 1_T(t) \quad (12)$$

L'expression de la variance totale est donnée par :

$$\sigma^2_{\text{total}} = \sigma^2_{\text{stat}} + \sigma^2_{\text{syst}} = \frac{1}{\sum_{i=1}^{n}\left(\frac{1}{\sigma^2_{i,stat}}\right)} + \sigma^2_{\text{syst}} \quad (13)$$

Plus le nombre n d'échantillons augmente, plus la variance statistique σ^2_{stat} diminue, tandis que la variance systématique σ^2_{syst} reste constante et commune à tous les échantillons.

En utilisant la formule d'inversion de Bayes, on en déduit que la densité de probabilité *a posteriori* de la date t est donnée, à une constante près, par :

$$p(t \mid \bar{A}) \propto \frac{1}{\sigma_{\text{total}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\bar{A}-g(t)}{\sigma_{\text{total}}}\right)^2} 1_T(t) \quad (14)$$

Outre la présente application à la thermoluminescence, ce modèle de combinaison est aussi implémenté dans le logiciel ChronoModel (fonction « combine ») ainsi que dans le logiciel OxCal (fonction « R-Combine », Bronk Ramsey, 2009a). Il est adapté à la combinaison de mesures répétées sur des échantillons ayant la même variable latente μ elle-même générée par une erreur systématique commune. C'est un modèle qui se situe entre deux autres modèles utilisés en datation par luminescence (Galbraith *et al.*, 1999) : le « common age model » (avec erreur systématique supposée nulle, autrement dit la mesure μ est directement égale à l'âge commun) et le « central age model » où la variable latente μ dépend de chaque échantillon, soit μ_i, mais partageant toujours une seule et même erreur systématique.

Dans la formule (8), la forme quadratique $\sum_{i=1}^{n}\left(\frac{A_i-\mu}{\sigma_{i,stat}}\right)^2$ peut se décomposer en deux parties :

$$\sum_{i=1}^{n}\left(\frac{A_i-\mu}{\sigma_{i,\ \text{stat}}}\right)^2 = w(\bar{A}-\mu)^2 + wS^2 \quad (15)$$

La moyenne pondérée $\bar{A}$ est donnée par la formule (8) et la variance S^2 des observations A_i est donnée par :

$$S^2 = \frac{1}{w}\sum_{i=1}^{n} w_i \left(A_i - \bar{A}\right)^2 \quad (16)$$

On montre que la variable $D^2 = wS^2$ suit une loi du Khi-2 à $n-1$ degrés de liberté (*cf.* 8). Ceci permet de tester l'homogénéité des déterminations A_i sous l'hypothèse d'une dispersion due aux seules erreurs statistiques $\sigma_{i,stat}$.

Ce sont les formules (9), (10), (13) et (16) qui sont mises en œuvre dans l'exemple de datation par thermoluminescence en section 4.3 du chapitre 1, l'âge étant alors converti en date calendaire selon la relation linéaire exprimée par la formule (6).

6 Bayes en action : modèle paléodémographique

Le problème archéologique traité dans le chapitre 3.5 permet d'illustrer la démarche d'une communauté scientifique (ici les paléodémographes) dans la recherche de l'outil statistique le plus adéquat, lequel s'est finalement avéré être l'approche bayésienne. Après formalisation du problème, nous ferons un bref historique de la démarche (pour plus de détails, *cf.* Caussinus *et al.*, 2017 ; Caussinus, 2021) et nous verrons comment les résultats très généraux obtenus peuvent apporter des réponses à des questions d'intérêt peu ou pas abordées antérieurement.

6.1 Position du problème et historique succinct des solutions proposées

La question de base est le passage d'âges biologiques, définis sur des ossements, à l'âge chronologique au décès du sujet concerné. Ces âges sont discrétisés, c'est-à-dire rapportés à des classes d'âge, r classes pour l'âge biologique, c classes pour l'âge chronologique.

Les données disponibles sont :

- Les effectifs n_{ij} de classe biologique $i(i = 1, 2, \ldots, r)$ et de classe chronologique $j(j = 1, 2, \ldots, c)$ qui ont été observés sur des sujets pour lesquels ces deux aspects étaient disponibles : ils sont rapportés dans un tableau (matrice $r \times c$) dit de référence. Toutes les précautions ont été prises pour que les sujets considérés aient vécu dans des conditions semblables aux sujets ayant vécu sur le site dont il est question ci-dessous.

- Les effectifs des divers stades de l'indicateur biologique choisi sur un échantillon de taille m issu d'un site funéraire donné, soit $(m_1, \ldots, m_i, \ldots, m_r)$.

Les paramètres « naturels » du problème sont :

- Les probabilités conditionnelles $\left(p_{1/j}, \ldots, p_{i/j}, \ldots, p_{r/j}\right)$ que le stade osseux soit le stade i pour un sujet dont la classe d'âge chronologique est j ; ces probabilités sont naturellement estimées pour chaque couple i, j par $n_{ij}/n_{.j}$, où $n_{.j} = \sum_{i=1}^{r} n_{ij}$ est la somme sur i des effectifs n_{ij}.

- Les probabilités marginales $(p_1, \ldots, p_j, \ldots, p_c)$ qu'un sujet du site étudié soit décédé à l'âge chronologique j. Le vecteur $\boldsymbol{p}$ des p_j est appelé structure par âges (chronologiques).

Notons que, si les $p_{i/j}$ peuvent s'estimer comme mentionné plus haut, il ne saurait en être de même des $p_{j/i}$ obtenus en permutant les rôles des deux types d'âge car le tableau de référence a été obtenu pour une distribution quelconque des âges chronologiques. En fait, les c colonnes de cette matrice doivent être considérées comme des échantillons indépendants. Ce fut d'ailleurs là un tout premier point de discussion, voire de conflits, entre les pionniers de la paléodémographie à la fin du XX$^{\text{e}}$ siècle.

Si nous notons $(\pi_1, \ldots, \pi_i, \ldots, \pi_r)$ les probabilités des classes d'âge biologiques sur le site considéré, les probabilités conjointes $p_{i,j}$ qu'un individu soit de classe biologique i et d'âge au décès j sont :

$p_{i,j} = p_{i|j} p_j = p_{j|i} \pi_i$

En sommant sur les colonnes j, nous pouvons alors en déduire la relation :

$\sum_{j=1}^{c} p_{i|j} p_j = \pi_i \quad (17)$

Si les $p_{i/j}$ et les π_i sont remplacés par leurs estimations naturelles, respectivement par $n_{i/j}/n_j$ et par m_i/m, fournies par les données de référence et les données du site considéré, le système d'équations (17) apparaît comme une régression linéaire multiple où il s'agit d'estimer les paramètres p_j. Cependant, on se heurte là à une double difficulté dans le cadre fréquentiste :

- Il y a une grande instabilité des estimations et on obtient assez souvent des estimations totalement aberrantes (on conçoit alors qu'évaluer la précision de ces estimations est sans espoir... et n'a guère de sens),

- Il faut que c soit inférieur à r pour que la solution de moindres carrés soit unique.

Plutôt que chercher une solution naïve à l'équation (17) par moindres carrés, on a suggéré une solution fréquentiste plus élaborée en supposant que l'échantillon $(m_1, m_i, \ldots, m_r)$ avait une distribution multinomiale (ce qui est peu contestable) de paramètres π_i dont l'expression était donnée par (17) en remplaçant chaque $p_{i/j}$ par son estimation $n_{ij}/n_{.j}$. L'estimation des p_j se faisait alors par maximum de vraisemblance en tenant compte de la contrainte de positivité de ces probabilités. Les estimations obtenues étaient encore souvent bizarres et il est clair qu'aucun calcul d'incertitude n'était légitime puisque la part liée au côté aléatoire des $n_{ij}/n_{.j}$ était scotomisée (Il est possible de tenir compte du caractère aléatoire des $n_{ij}/n_{.j}$ tout en restant dans le cadre fréquentiste mais cette complication, théoriquement plus réaliste, ne résout pas grand-chose, sans doute parce que le bruit ainsi accru perturbe encore davantage la situation). Restait aussi la difficulté de la contrainte $c \leq r$. Une façon de la contourner était de supposer que les p_j correspondaient à une structure par âges suivant une loi utilisée en démographie, par exemple la loi de Gompertz qui les expriment en fonction de seulement deux paramètres. Cette approche est cependant décevante (en fait il a été montré plus tard qu'elle est moins efficace que la méthode bayésienne, même pour estimer les paramètres d'une éventuelle loi de Gompertz sous-jacente).

Cette introduction d'une loi de Gompertz avait au moins un mérite : tenir compte du fait que le vecteur des p_j ne saurait être une suite quelconque dans la mesure où il représente une loi de mortalité. Cette idée est prise en compte par Bocquet-Appel et Bacro (2008) qui proposent de chercher une solution au système (17) dans une famille de vecteurs $(p_1, \ldots, p_j, \ldots, p_c)$ représentant toute une série de lois de mortalité réalistes (ils proposent deux telles séries, chacune de 1000 lois). De plus, afin de tenir compte du caractère aléatoire des données de références, ils utilisent une technique de bootstrap conduisant à plusieurs vecteurs $(p_1, \ldots, p_j, \ldots, p_c)$ dont on peut prendre la moyenne comme estimation ponctuelle et utiliser la dispersion pour évaluer l'incertitude de cette estimation. Ce dernier point reste discutable pour plusieurs raisons,

essentiellement : les auteurs remplacent les π_i par m_i/m sans prendre en compte le caractère aléatoire des m_i (ce que l'on peut néanmoins ajouter facilement à leur méthode), une liste de lois de mortalité, aussi large soit-elle, exclut totalement toute solution ne se trouvant pas parmi elles (ou plutôt à l'intérieur de leur combinaison convexe). Afin de surmonter ces déficiences, ne serait-il donc pas mieux de nous placer dans une vision résolument bayésienne ? (Car, quoiqu'en disent les auteurs, leur méthode n'est pas réellement bayésienne au sens utilisé dans le présent ouvrage.) C'est l'approche qui a été suivie au chapitre 3.5., les ingrédients étant les suivants :

- On admet que les données de référence et de site correspondent à des échantillons multinomiaux,

- Aux paramètres, respectivement $p_{i/j}$ et p_j, de ces distributions sont affectées des lois *a* priori discutées plus bas.

Soit M le vecteur des observations m_i, qui suit une loi multinomiale de paramètres π_i reliés aux $p_{i/j}$ et p_j, par la formule (17), soit P désignant la matrice des $p_{i/j}$ et p désignant le vecteur des p_j, soit G la densité de probabilité *a priori* de la matrice P et g la densité *a priori* du vecteur p, la densité jointe f de (M, P, p) est alors donnée par :

$$f(M, P, p) = \text{ Multinomiale } (M \mid p, P) g(p) G(P)$$

La densité d'intérêt que nous recherchons est la densité conditionnelle (ou *a posteriori*) $h(p \mid M)$ qui est donnée par la formule de Bayes, après intégration sur la variable P :

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

6.2 Lois *a priori*

Pour les probabilités conditionnelles, il n'y a pas d'autre information que celle contenue dans les données de référence. La loi G de P a donc été choisie uniforme (loi de Dirichlet de paramètres tous égaux à 1) pour chacune des c colonnes de référence, les c distributions étant indépendantes.

Le point clé de la démarche bayésienne est le choix de la loi *a priori* g du vecteur $(p_1, \ldots, p_j, \ldots, p_c)$. On est pour cela parti d'une loi de Dirichlet, conjuguée de la loi multinomiale, basique pour un vecteur aléatoire à éléments positifs de somme un. La moyenne a été prise égale à ce que Séguy et Buchet (2011) appellent le standard préindustriel, défini comme la moyenne des diverses lois de mortalité préindustrielles publiées dans la littérature, la raison essentielle étant que l'intérêt est le plus souvent porté aux écarts par rapport à ce standard, lequel joue en quelque sorte un rôle qui peut être considéré analogue à l'hypothèse nulle de la statistique classique (mais il est loisible de changer de moyenne dans certains cas particuliers).

Le point important est maintenant le suivant : sans être démographe il semble clair que la probabilité de décès ne saurait varier trop brutalement d'une classe d'âge à la

classe voisine, ne serait-ce qu'à cause du caractère arbitraire des limites de classes. Cette idée est exprimée de diverses façons dans les « modèles adaptés aux populations préindustrielles » (titre du chapitre VI de Séguy et Buchet, 2011). Empiriquement, on peut voir, par exemple, que dans les 167 lois de mortalité préindustrielle extraites de la littérature compilée par ces auteurs, les coefficients de corrélation empiriques entre les probabilités de classes voisines sont élevés (par exemple, entre la classe 25-29 ans et les classes 18-24, 30-34 et 35-39 ans ils valent respectivement 0.890, 0.865 et 0.779). Cela suggère un lissage des probabilités *a priori* des classes d'âge, qui a été réalisé de la façon suivante :

- on part d'une loi de Dirichlet de paramètres $\beta_j (j = 1, 2, \ldots, c)$,

- on effectue ensuite une moyenne mobile* afin de « lisser » ces valeurs,

Les paramètres β_j sont choisis de sorte que les moyennes *a priori* des classes d'âge soient égales au standard préindustriel et la somme des β_j soit égale à c.

Illustrons à quoi conduit ce choix sur la distribution *a priori* du couple (p_1, p_2) des probabilités *a priori* des classes 18-24 ans et 25-29 ans. Dans la figure 7 ci-après les points rouges correspondent aux 167 lois empiriques de mortalité présentées dans Séguy et Buchet (2011) ; la densité *a priori* est figurée par le nuage d'un bleu d'autant plus sombre qu'elle est plus élevée. On voit que cette densité correspond à un certain « élargissement » des possibles observés avec les 167 lois compilées. Celles-ci se trouvent essentiellement dans sa partie dense montrant une cohérence bienvenue, mais des sites à la distribution éloignée des valeurs centrales ont une probabilité *a priori* raisonnable permettant une révision *a posteriori* relativement substantielle même pour des échantillons de petite taille (voir exemples du chapitre 3.5.).

6.3 Comparaison de sites

A côté de la loi *a posteriori* de la structure par âges $(p_1, \ldots, p_j, \ldots, p_c)$, premier paramètre d'intérêt pour les paléodémographes, l'approche présentée fournit aisément la loi *a posteriori* de toute fonction de l'ensemble des paramètres p_i et $p_{j/i}$. Par exemple, comme le nombre de classes d'âge peut être assez grand (les exemples présentés ont été traités avec 13 classes, une classe 18-24 ans et une classe ≥ 80 ans encadrant 11 classes quinquennales), l'âge moyen au décès $\bar{a}$ d'un sujet ayant atteint l'âge adulte est assez bien approché par :

$$\bar{a} = \sum_{j=1}^{c} p_j a_j$$

où a_j est le centre de la jème classe d'âge, et l'on peut calculer aisément la densité *a posteriori* de ce nouveau paramètre $\bar{a}$.

Ce paramètre synthétique permet une comparaison graphique aisée de deux sites au travers des densités *a posteriori* correspondantes comme cela est montré dans les figures 2(b), 2(d) et 3 en 3.5. Mais on peut aller plus loin et quantifier la différence observée. En effet, si X et Y désignent deux variables aléatoires réelles indépendantes de densités respectives φ et ψ, la probabilité de l'événement $X > Y$ est :

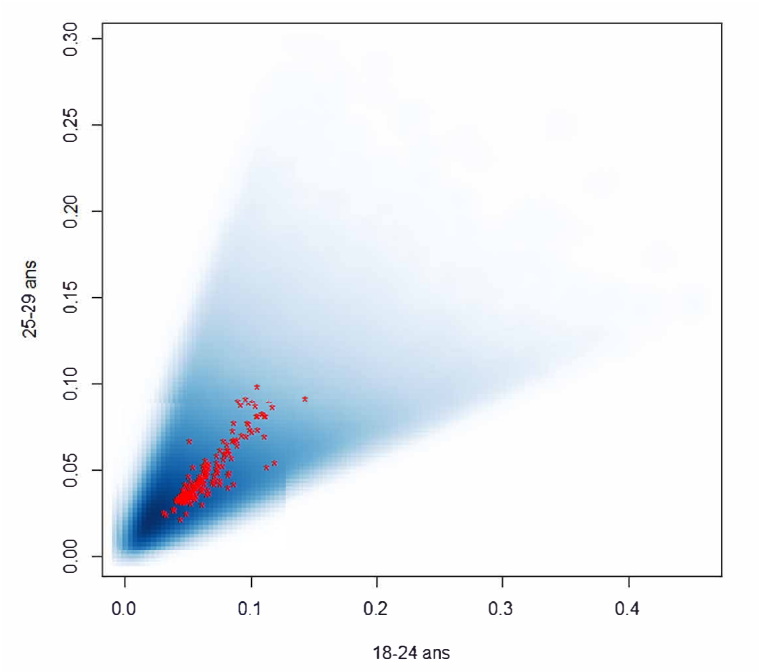


FIG. 7 : Distribution empirique du couple (p_1, p_2) pour les structures d'âge empiriques de Séguy et Buchet (2011) (points rouges) et densité conjointe de ces paramètres pour la loi *a priori* proposée (nuage bleu).

$\iint_D \varphi(x)\psi(y)dxdy$ où le domaine d'intégration est $D = \{(x, y) : x > y\}$.

Les échantillons observés sur les deux sites étant indépendants, il suffira d'évaluer (numériquement) cette intégrale avec les densités *a posteriori* φ et ψ des âges moyens au décès sur chacun des sites. Quelques cas dont a vu l'utilité sont présentés en 3.5.