Approche bayésienne

Analyse de séquences génomiques et phylogénie

Guy Perrière

Laboratoire de Biométrie et Biologie Évolutive
UMR CNRS n° 5558
Université Claude Bernard – Lyon 1

3-6 avril 2017

Historique

- Théorème de Bayes établi au XVIII^{e} siècle :
  - Utilisation courante en probabilités.
- Introduction récente en phylogénie moléculaire :
  - Yang et Rannala (1996).
- Détermination analytique des probabilités postérieures fréquemment impossible :
  - Utilisation d'approximations numériques.

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763. I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

He had, you know, the honour of being a member of that illuſtrious Society, and was much eſteemed by many in it as a very able mathematician. In an introduction which he has writ to this Eſſay, he ſays, that his deſign at firſt in thinking on the ſubject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumſtances, upon ſuppoſition that we know nothing concerning it but that, under the ſame circum-

Théorème de Bayes

- Une définition classique des probabilités conditionnelles est que :

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B) = \mathbb{P}(A)\mathbb{P}(B|A)$$

- En divisant les deux termes de l'équation précédente par $\mathbb{P}(B)$ on obtient la formulation la plus simple du théorème de Bayes, soit :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(B)}$$

avec :

  - $\mathbb{P}(A|B)$, la probabilité *a posteriori* (ou postérieure) de A sachant B.
  - $\mathbb{P}(A)$, la probabilité *a priori* de A.
  - $\mathbb{P}(B|A)$, la *vraisemblance* de A.
  - $\mathbb{P}(B)$, la probabilité *marginale* de B ou *constante de normalisation*.

Généralisation

- Étant donné que :

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(\bar{A} \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(\bar{A})\mathbb{P}(B|\bar{A})$$

  on en déduit :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(\bar{A})\mathbb{P}(B|\bar{A})}$$

- Ce qui peut se généraliser sous la forme :

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\sum_j \mathbb{P}(A_j)\mathbb{P}(B|A_j)}$$

  pour tout élément du s.c.e. $\{A_i\}$, avec i un des éléments de l'ensemble des valeurs possibles de j.

Un exemple classique

- Quelle est la probabilité d'avoir des *faux positifs* lors d'un test de diagnostic ?
- Soit un test de dépistage d'une maladie quelconque :
  - Si un patient a contracté la maladie, le test est positif dans 99% des cas.
  - Si un patient est sain, le test est négatif dans 95% des cas.
  - On estime que la fréquence de la maladie dans la population est de 1‰.
- Quelle est la probabilité qu'un individu testé positif soit effectivement atteint ?

Résolution

- Dans cet exemple, la probabilité *a priori* est égale à la fréquence de la maladie dans la population, soit $\mathbb{P}(A) = 0.001$:
  - On en déduit $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A) = 0.999$.
- Par ailleurs, la probabilité que le test soit positif si le patient est malade est $\mathbb{P}(B|A) = 0.99$.
- Enfin, la probabilité que le test soit négatif si le patient est sain est $\mathbb{P}(\bar{B}|\bar{A}) = 0.95$:
  - On en déduit $\mathbb{P}(B|\bar{A}) = 1 - \mathbb{P}(\bar{B}|\bar{A}) = 0.05$.
- On en déduit la probabilité $\mathbb{P}(A|B)$ qu'un individu soit malade si le test est positif :

$$\mathbb{P}(A|B) = \frac{0.001 \times 0.99}{0.001 \times 0.99 + 0.999 \times 0.05} \simeq 0.019$$

Remarques sur le résultat

- Bien que le test précédent soit apparemment précis, la probabilité d'avoir des faux positifs est très importante (98.1%) :
  - Problème lié au fait que la probabilité *a priori* est faible.
  - Cas fréquent pour les tests de diagnostic :
    - Utilisation de plusieurs tests réalisés de façon séquentielle.
- Dans cet exemple, détermination de l'*a priori* à partir de la fréquence de la pathologie dans la population :
  - L'utilisation du théorème de Bayes ne souffre pas de discussion.
- Dans de nombreux cas, les probabilités *a priori* ne peuvent pas être facilement estimées :
  - Utilisation de valeurs représentant l'appréciation *subjective* de la personne effectuant l'analyse.

Notation en statistiques

- En statistiques, le s.c.e. $\{A_i\}$ correspond à un ensemble d'hypothèses, alors que B correspond aux données observées.
- Dans ce cas, écriture du théorème de Bayes sous la forme :

$$\mathbb{P}(H_i|D) = \frac{\mathbb{P}(H_i)\mathbb{P}(D|H_i)}{\sum_j \mathbb{P}(H_j)\mathbb{P}(D|H_j)}$$

  avec $\mathbb{P}(H_i|D)$, la probabilité conditionnelle d'une hypothèse H_i sous les données D.
- Les différentes hypothèses pouvant correspondre à différentes valeurs pour un paramètre θ, avec $H_1 : \theta = \theta_1$, $H_2 : \theta = \theta_2$, etc.
- Dans le cas où le modèle utilisé comprend plus d'un paramètre, θ correspond alors au vecteur $\boldsymbol{\theta}$ des dits paramètres.

Données continues

- Expression sous la forme de fonctions de densités quand les hypothèses concernent des paramètres *continus* :

$$f(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\boldsymbol{\theta})f(\mathbf{x}|\boldsymbol{\theta})}{f(\mathbf{x})} = \frac{f(\boldsymbol{\theta})f(\mathbf{x}|\boldsymbol{\theta})}{\int f(\mathbf{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}$$

- La constante de normalisation $f(\mathbf{x})$ est obtenue en intégrant la vraisemblance sur la distribution *a priori* de $\boldsymbol{\theta}$:
  - Permet d'avoir $\int f(\boldsymbol{\theta}|\mathbf{x}) = 1$.
  - Si $\boldsymbol{\theta}$ correspond à un vecteur comprenant de nombreux paramètres :
    - Pas de solution analytique au calcul de cette intégrale.
    - Calcul de la probabilité postérieure au moyen d'approximations numériques telles que les *Chaînes de Markov avec technique de Monte-Carlo* (MCMC).

Interprétation des résultats

- Le résultat d'une analyse statistique bayésienne est représenté par la distribution des probabilités postérieures.
- Utilisation de valeurs ponctuelles pour faciliter l'interprétation :
  - Moyenne :
$$\mathbb{E}(\boldsymbol{\theta}|\mathbf{x}) = \int \boldsymbol{\theta} f(\boldsymbol{\theta}|\mathbf{x}) \mathrm{d}\boldsymbol{\theta}$$
  - Médiane.
  - Maximum *a posteriori* :
    - Conceptuellement similaire au maximum de vraisemblance.
- Détermination d'un intervalle de *crédibilité* $[a, b]$ au seuil α tel que :
$$\int_a^b f(\boldsymbol{\theta}|\mathbf{x}) \mathrm{d}\boldsymbol{\theta} = 1 - \alpha$$

Distributions *a priori*

- Conjuguées :
  - Un *a priori* est dit conjugué si $f(\boldsymbol{\theta})$ et $f(\boldsymbol{\theta}|\mathbf{x})$ appartiennent à la même famille de distributions.
  - Permettent de simplifier les calculs (pas de résolution d'intégrales complexes).
- Non informatives ou vagues :
  - $f(\boldsymbol{\theta})$ est non informative si son impact sur $f(\boldsymbol{\theta}|\mathbf{x})$ est faible :
    - Prédominance de la vraisemblance.
  - Utilisées quand aucune information préalable n'est disponible sur les variations du paramètre.
- Informatives :
  - $f(\boldsymbol{\theta})$ est informative si son impact sur $f(\boldsymbol{\theta}|\mathbf{x})$ est fort.
  - Cas de l'analyse bayésienne séquentielle :
    - *A posteriori* d'une étude précédente utilisé comme *a priori* pour l'étude courante.

Critiques de l'*a priori*

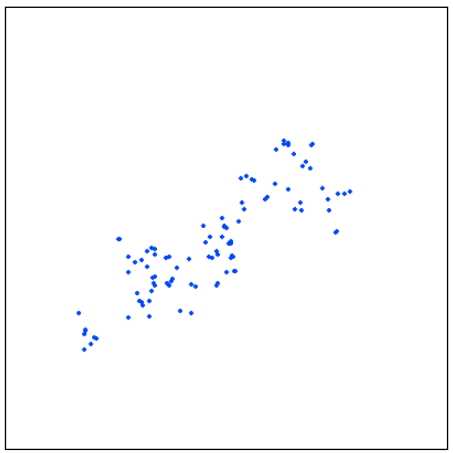
- Depuis le XVIII^{e} siècle, les critiques du bayésien portent essentiellement sur l'*a priori*.
- Résultats différents en fonction d'un *a priori* donné :
  - Rejet par les statisticiens « classiques » de la notion de probabilité subjective.
- Existence d'une école « objective » prônant l'utilisation d'*a priori* les moins informatifs possibles :
  - Distributions uniformes.
  - Loi *a priori* de Jeffreys (1961).
  - Loi de référence de Bernardo (1979).

Principe des MCMC

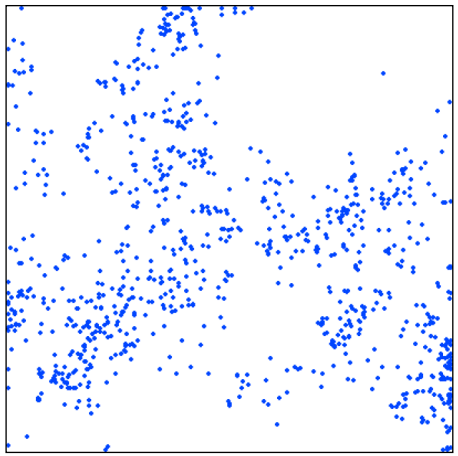
- En analyse bayésienne, impossibilité de déterminer la constante de normalisation si le nombre de paramètres est élevé :
  - Impossibilité de calculer directement la probabilité postérieure.
- Utilisation d'une chaîne de Markov suivant une marche guidée dans l'espace multidimensionnel des paramètres :
  - À la stationnarité, convergence vers les valeurs attendues des probabilités postérieures.

Analogie du randonneur

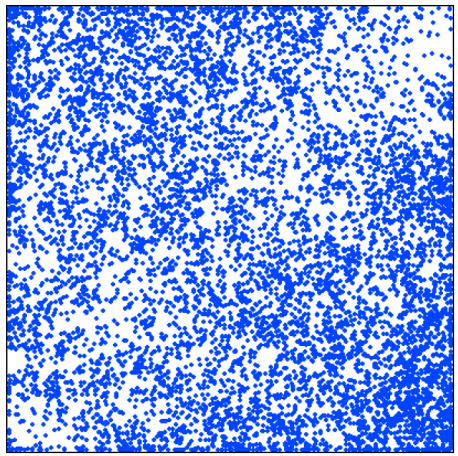
- Soit un randonneur se déplaçant sur une surface plane délimitée en faisant des pas de longueur variable :
  - Amplitude maximale fixée au préalable.
  - Chaque pas est effectué en choisissant aléatoirement une direction quelconque.
  - Rebond si un pas conduit à l'extérieur.
- Au bout d'un certain temps, exploration de l'intégralité de la surface :



100 pas



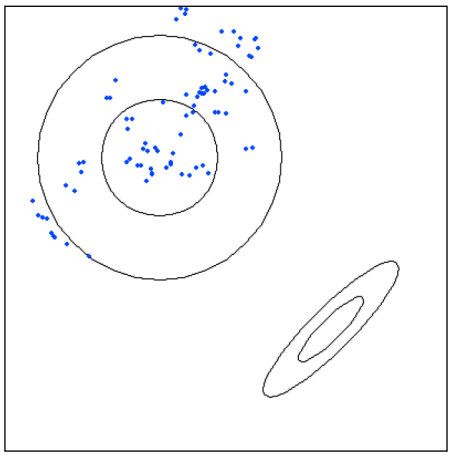
1000 pas



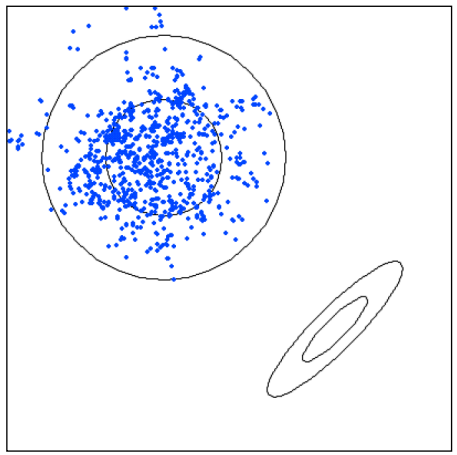
10000 pas

Exploration de reliefs

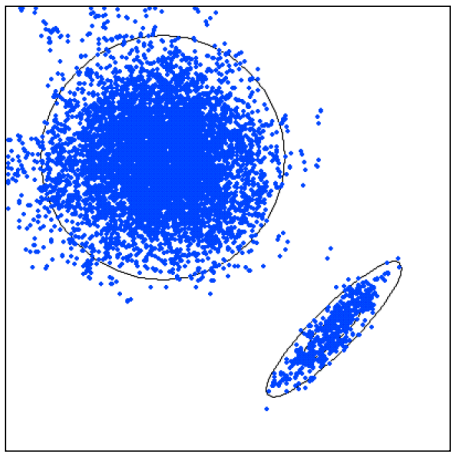
- Introduction de deux règles supplémentaires :
  - Si la direction prise par le randonneur le conduit vers une position plus élevée, il le fait toujours.
  - Si au contraire cette direction est descendante, possibilité de choix :
    - Calcul de $r = h^*/h$, avec h^* la hauteur atteinte en cas de descente et h la hauteur actuelle.
    - Tirage de $u \sim \mathcal{U}(0, 1)$.
    - Si $u < r$, le randonneur descend, sinon il reste où il est.
- Visite préférentielle des points situés en altitude :



100 pas



1000 pas



10000 pas

Problèmes rencontrés

- Nécessité d'éliminer les premiers pas – qui constituent ce que l'on appelle communément la *zone d'approche* ou *burn-in* :
  - Démarrage du trajet en un point sélectionné aléatoirement, point pouvant être situé à une distance importante des reliefs.
- Évitement des maxima locaux :
  - Nécessité d'avoir un nombre de pas suffisamment élevé :
    - Pas toujours suffisant si les pics sont éloignés les uns des autres.
  - Lancement de plusieurs chaînes ayant des points de départ différents :
    - Poursuite de l'exploration jusqu'à convergence des résultats entre les différentes chaînes.

Algorithme de Metropolis-Hastings

1. Soit $\boldsymbol{\theta}_i$, le vecteur des paramètres caractérisant l'état de la chaîne de Markov au temps i.
2. Soit $\boldsymbol{\theta}^*$ le vecteur des paramètres caractérisant un état *candidat* pour constituer le maillon suivant de la chaîne.
3. Calcul de la *probabilité d'acceptation* r, telle que :

$$r = \min\left[1, \frac{f(\boldsymbol{\theta}^*|\mathbf{x})}{f(\boldsymbol{\theta}_i|\mathbf{x})}\right] = \min\left[1, \frac{f(\boldsymbol{\theta}^*)f(\mathbf{x}|\boldsymbol{\theta}^*)}{f(\boldsymbol{\theta}_i)f(\mathbf{x}|\boldsymbol{\theta}_i)}\right]$$

4. Si $r = 1$, alors $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}^*$.
5. Si $r < 1$, tirage de $u \sim \mathcal{U}(0,1)$:
   - Si $u < r$ alors $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}^*$, sinon $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$.
6. Retour à l'étape 1.

Caractéristiques

- Le calcul de r n'implique pas de connaître $f(\mathbf{x})$.
- Initialisation avec un ensemble de paramètres $\boldsymbol{\theta}$ choisis aléatoirement.
- La construction de $\boldsymbol{\theta}^*$ se fait en faisant varier de façon aléatoire les paramètres :
  - Utilisation d'algorithmes générant ce que l'on appelle des *propositions* :
    - Distributions uniformes de type $\mathcal{U}(-w/2, w/2)$, avec w l'amplitude maximale autorisée pour la variation des paramètres.
    - Distributions normales de type $\mathcal{N}(\mu, \sigma^2)$.
- La séquence des états visités forme une chaîne de Markov :
  - Estimation de la probabilité postérieure par la fréquence à laquelle les états sont visités une fois la stationnarité atteinte.

Fréquence d'acceptation

- Proportion du nombre de propositions acceptées dans la chaîne.
- Ne doit être ni trop grande ni trop petite.
- Valeurs optimales :
  - $\approx 50\%$ si $\boldsymbol{\theta}$ ne comprend qu'un seul paramètre.
  - $\approx 26\%$ si $\boldsymbol{\theta}$ comprend plusieurs paramètres.
- Valeurs recommandées :
  - 20-70% si $\boldsymbol{\theta}$ ne comprend qu'un seul paramètre.
  - 15-40% si $\boldsymbol{\theta}$ comprend plusieurs paramètres.

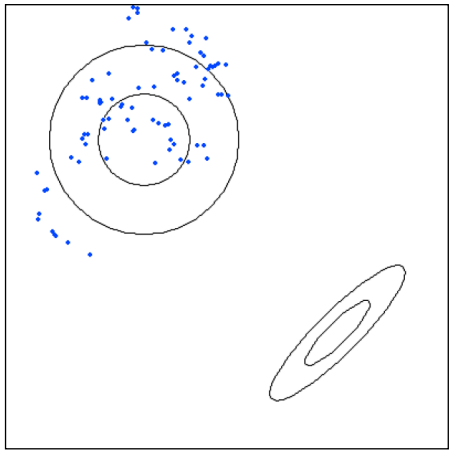
Couplage de Metropolis des MCMC

- Piégeage possible de la chaîne en cas de maximum local.
- Utilisation de plusieurs chaînes au lieu d'une :
  - Couplage de Metropolis des MCMC (MCMCMC ou MC^3).
  - Parmi toutes les chaînes lancées seules les chaînes dites « froides » (faible amplitude des pas) ont besoin de converger :
    - Utilisation de chaînes « chaudes » pour permettre une exploration plus vaste de l'espace des paramètres.
  - Tests à intervalles réguliers pour faire passer une chaîne froide dans une région explorée par une des chaînes chaudes :

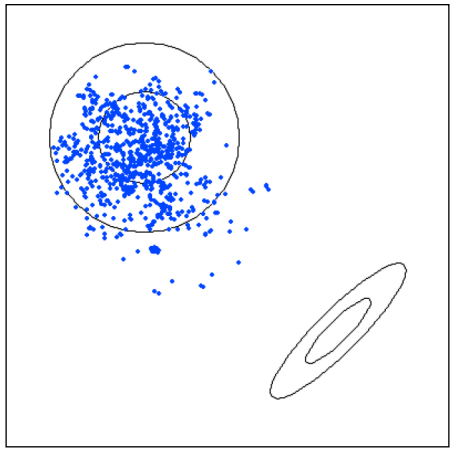
$$r = \min\left[1, \frac{\pi_i(\boldsymbol{\theta}_j)\pi_j(\boldsymbol{\theta}_i)}{\pi_i(\boldsymbol{\theta}_i)\pi_j(\boldsymbol{\theta}_j)}\right]$$

où i et j correspondent aux états de deux chaînes de Markov pour lesquelles la possibilité d'échange est testée.

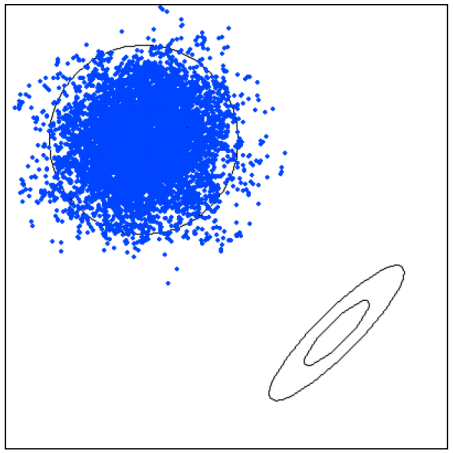
Application au problème du randonneur



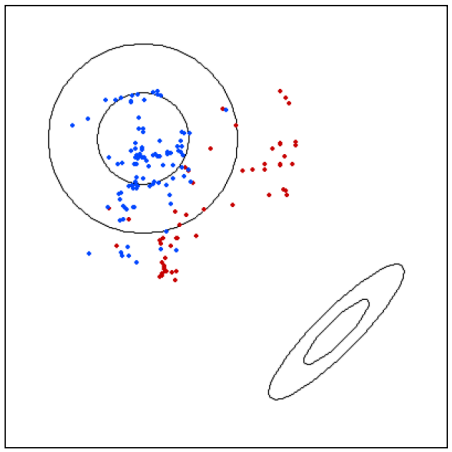
100 pas

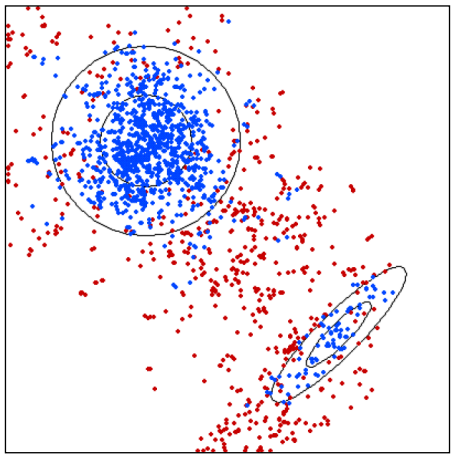


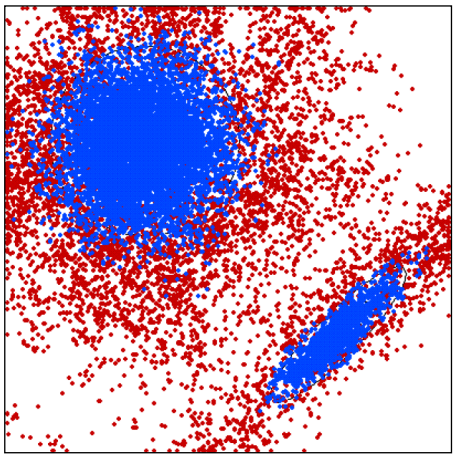
1000 pas



10000 pas







Détermination de la convergence

- Quand faut-il interrompre une MCMC ?
  - A-t-on atteint la distribution stationnaire de la chaîne ?
- Outils disponibles :
  - Inspection visuelle du graphe montrant les déplacements dans l'espace des paramètres.
  - Étude de la variation des valeurs de vraisemblance :
    - Pas de tendances particulières attendues à la stationnarité.
  - Mesure de l'autocorrélation des valeurs successives des paramètres :
    - Absence d'autocorrélation si convergence.
  - Tests statistiques :
    - Test de Gelman et Rubin (1992), ou *Potential Scale Reduction Factor* (PSRF) dans MrBayes.

Probabilité *a priori*

- Estimation par approche bayésienne de la distance évolutive entre deux séquences d'ADN sous le modèle de Jukes et Cantor.
- Calcul de la probabilité *a priori* :
  - Choix d'une distribution exponentielle :

$$f(d) = \frac{1}{\mu} \mathrm{e}^{-d/\mu}$$

    avec μ la moyenne de cette distribution et d la distance évolutive :
    - La probabilité d'obtenir des distances importantes tend rapidement vers 0.
  - D'autres choix sont possibles :
    - Distribution uniforme.

Vraisemblance

- Le calcul de la distance évolutive entre deux séquences au moyen du modèle de Jukes et Cantor est donnée par la formule :

$$d = -\frac{3}{4}\ln\left(1 - \frac{4}{3}p\right) \Leftrightarrow p = \frac{3}{4} - \frac{3}{4}\mathrm{e}^{-4d/3}$$

- Soit ℓ le nombre de sites dans l'alignement et n le nombre de sites pour lesquels il y a une substitution entre les deux séquences :
  - Dans ce cas, la fonction de vraisemblance pour d est donnée par la distribution binomiale $\mathcal{B}(\ell, p)$ telle que :

$$\begin{aligned} L(d) = f(p|d) &= \binom{\ell}{n} p^n (1-p)^{\ell-n} \\ &= \frac{\ell!}{n!(\ell-n)!}\left(\frac{3}{4} - \frac{3}{4}\mathrm{e}^{-4d/3}\right)^n \left(\frac{1}{4} + \frac{3}{4}\mathrm{e}^{-4d/3}\right)^{\ell-n} \end{aligned}$$

Probabilité postérieure

- Probabilité postérieure, sans la constante de normalisation :

$$f(d|p) \propto f(d)f(p|d)$$
$$\propto \frac{1}{\mu}\mathrm{e}^{-d/\mu}\left(\frac{3}{4}-\frac{3}{4}\mathrm{e}^{-4d/3}\right)^{n}\left(\frac{1}{4}+\frac{3}{4}\mathrm{e}^{-4d/3}\right)^{\ell-n}$$

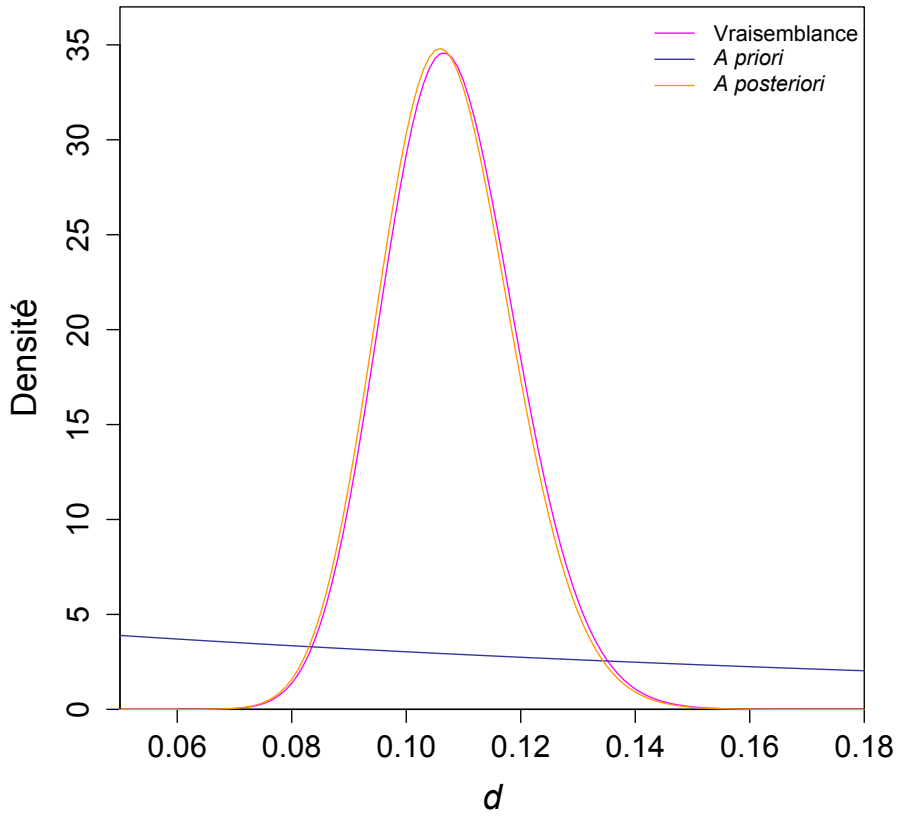
  - Le coefficient binomial étant lui aussi une constante, il peut être omis de cette expression.
- Valeur de la constante de normalisation donnée par :

$$f(p) = \int_0^\infty f(d)f(p|d)\mathrm{d}d$$

  - Solution analytique ou intégration numérique.

Application numérique

- Paire Homme-Gorille du jeu de données de Brown *et al.* (1982) :
  - $\ell = 896$
  - $n = 89$
- Moyenne de la distribution *a priori* fixée à $\mu = 0.2$.
- Estimation au maximum de vraisemblance :
  - $d \simeq 0.1066$
- Estimation bayésienne via la moyenne :
  - $\mathbb{E}(d|p) \simeq 0.1072$



Approximation par MCMC

- Calcul de la probabilité d'acceptation :

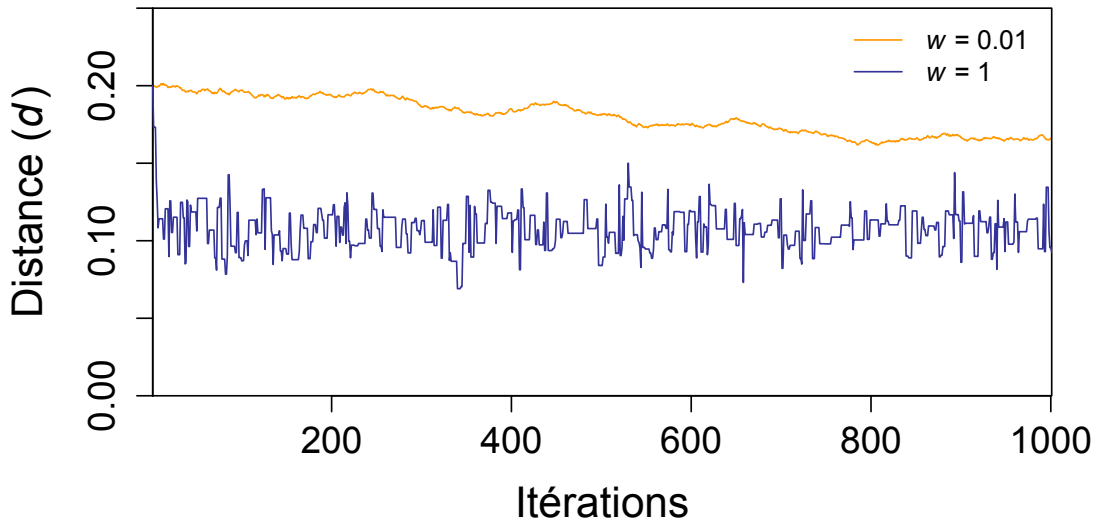
$$r = \min\left[1, \frac{f(d^*)f(p|d^*)}{f(d_i)f(p|d_i)}\right]$$

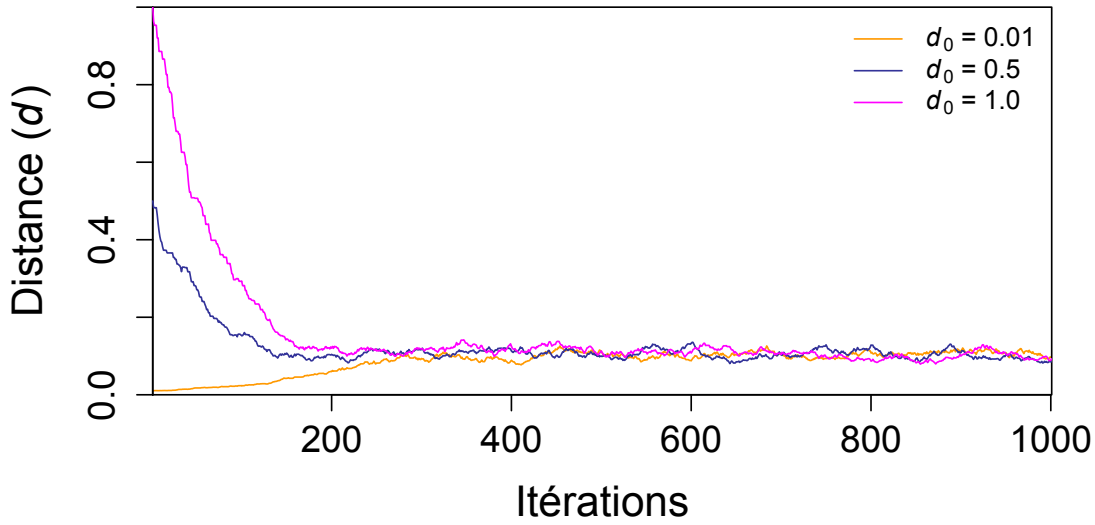
- Choix des propositions pour d :
  - Distribution uniforme, centrée sur la valeur actuelle et ayant une largeur égale à w :

$$d^* = |d_i + u|, \text{ avec } u \sim \mathcal{U}(-w/2, w/2)$$

- Choix de différentes valeurs pour l'amplitude (w) et la distance (d_0) utilisées pour initialiser la chaîne de Markov :
  - Valeurs variables pour w (0.01 et 1) et valeur fixe pour d_0 (0.2).
  - Valeur fixe pour w (0.1) et valeurs variables pour d_0 (0.01, 0.5 et 1).

Convergence des chaînes





Estimations de la distance

- Paramètres choisis : $\mu = 0.2$, $w = 0.1$ et $d_0 = 0.5$.
- Élimination de la zone d'approche (400 premières itérations).
- Échantillonnage de 1000 itérations prélevées à intervalles réguliers dans une chaîne :
  - Utilisation de la moyenne des valeurs pour l'estimation.
- Estimations obtenues après :
  - 1400 itérations : $\mathbb{E}(d|p) = 0.1073 \pm 5.18 \times 10^{-4}$
  - 10000 itérations : $\mathbb{E}(d|p) = 0.1072 \pm 7.03 \times 10^{-4}$
  - 100000 itérations : $\mathbb{E}(d|p) = 0.1071 \pm 7.23 \times 10^{-4}$

  avec, dans chaque cas, un intervalle de crédibilité à 95%.
- Variations stochastiques autour de la valeur obtenue par calcul direct.

Notations pour la phylogénie

- En phylogénie moléculaire, les données sont représentées par un ensemble de séquences alignées S.
- Par ailleurs, le vecteur des paramètres est $\boldsymbol{\theta} = (\tau, \mathbf{b}, \boldsymbol{\vartheta}, \alpha)$, avec :
  - τ la topologie de l'arbre.
  - $\mathbf{b}$ le vecteur des longueurs de branches.
  - $\boldsymbol{\vartheta}$ le vecteur des paramètres du modèle d'évolution utilisé.
  - α le paramètre de forme de la loi Gamma, le cas échéant.
- Le formule permettant de déterminer la probabilité postérieure est donc égale à :

$$f(\tau, \mathbf{b}, \boldsymbol{\vartheta}, \alpha | S) = \frac{f(\tau, \mathbf{b}, \boldsymbol{\vartheta}, \alpha) f(S | \tau, \mathbf{b}, \boldsymbol{\vartheta}, \alpha)}{f(S)}$$

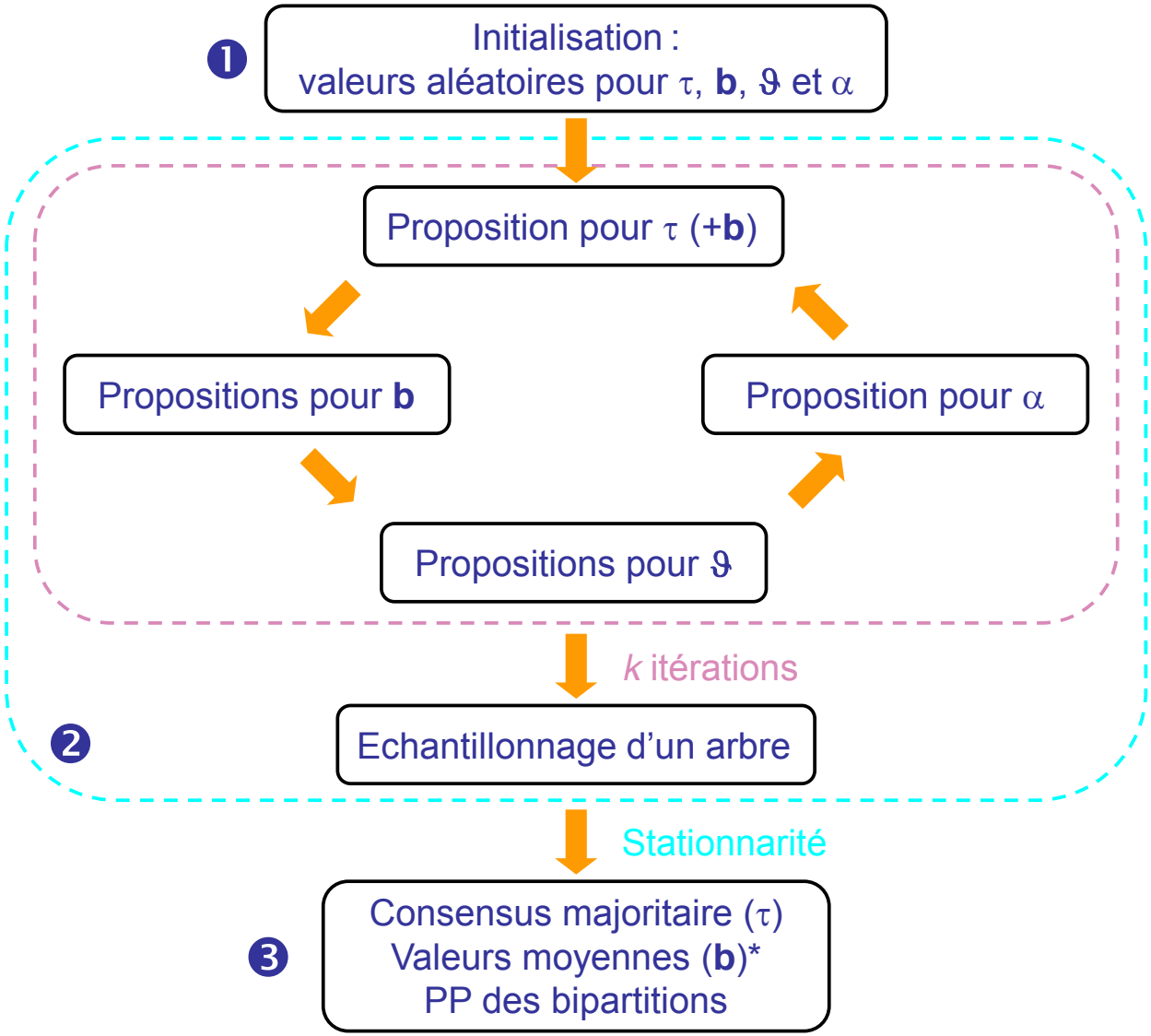
avec :

$$f(S) = \sum_{\tau} \int_{\mathbf{b}} \int_{\boldsymbol{\vartheta}} \int_{\alpha} f(S | \tau, \mathbf{b}, \boldsymbol{\vartheta}, \alpha) f(\mathbf{b}) f(\boldsymbol{\vartheta}) f(\alpha) \mathrm{d}\mathbf{b} \mathrm{d}\boldsymbol{\vartheta} \mathrm{d}\alpha$$

Choix possibles pour les *a priori*

- Topologies :
  - Distribution uniforme $\mathcal{U}(N)$.
- Longueurs des branches :
  - Distribution uniforme $\mathcal{U}(0, 10)$.
  - Distribution exponentielle $\mathcal{E}(0.1)$.
- Paramètres du modèle d'évolution :
  - Distributions de Dirichlet plates $\mathcal{D}(1, 1, 1, 1)$ pour les échangeabilités et les fréquences à l'équilibre.
- Paramètre α de la loi Gamma :
  - Distribution exponentielle $\mathcal{E}(1)$.

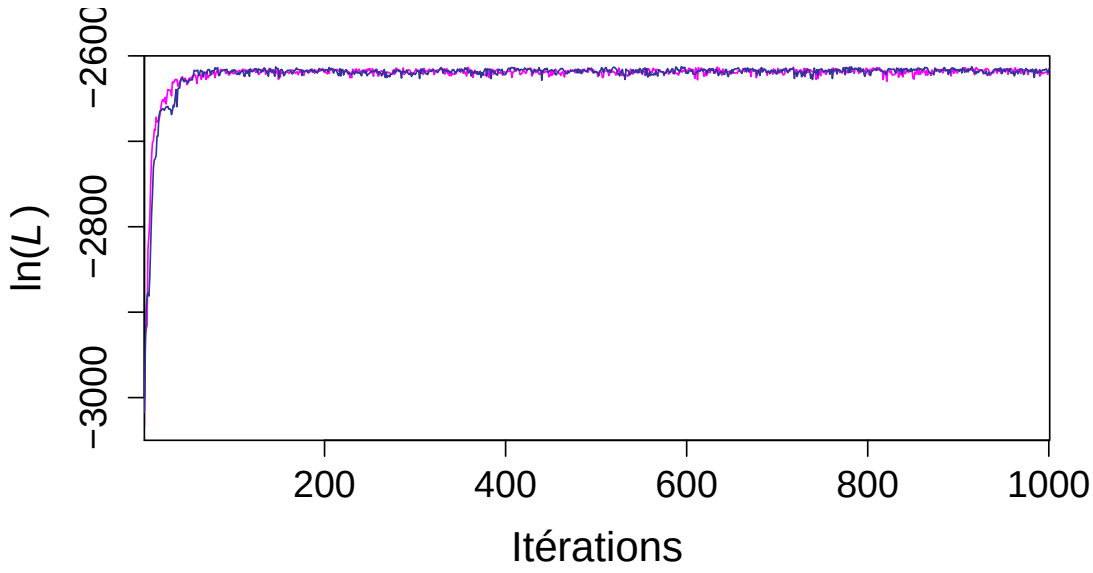
Procédure générale

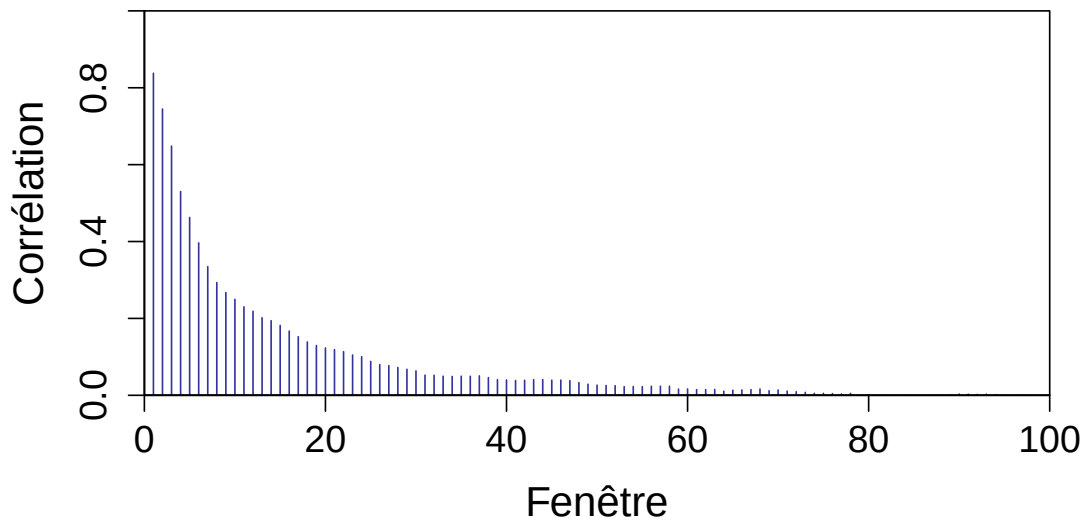


Phylogénie des Hominoïdes

- Sélection du modèle HKY+Γ après un test BIC.
- Utilisation de MrBayes pour reconstruire la phylogénie :
  - Valeurs par défaut des probabilités *a priori*.
  - Deux chaînes froides partant de points de départ différents.
  - Trois chaînes chaudes lancées en parallèle de chaque chaîne froide.
  - Test de Gelman et Rubin pour déterminer si convergence.
  - Arrêt après 10000 itérations et fréquence d'échantillonnage de 1/10 :
    - Jeu de données de petite taille.

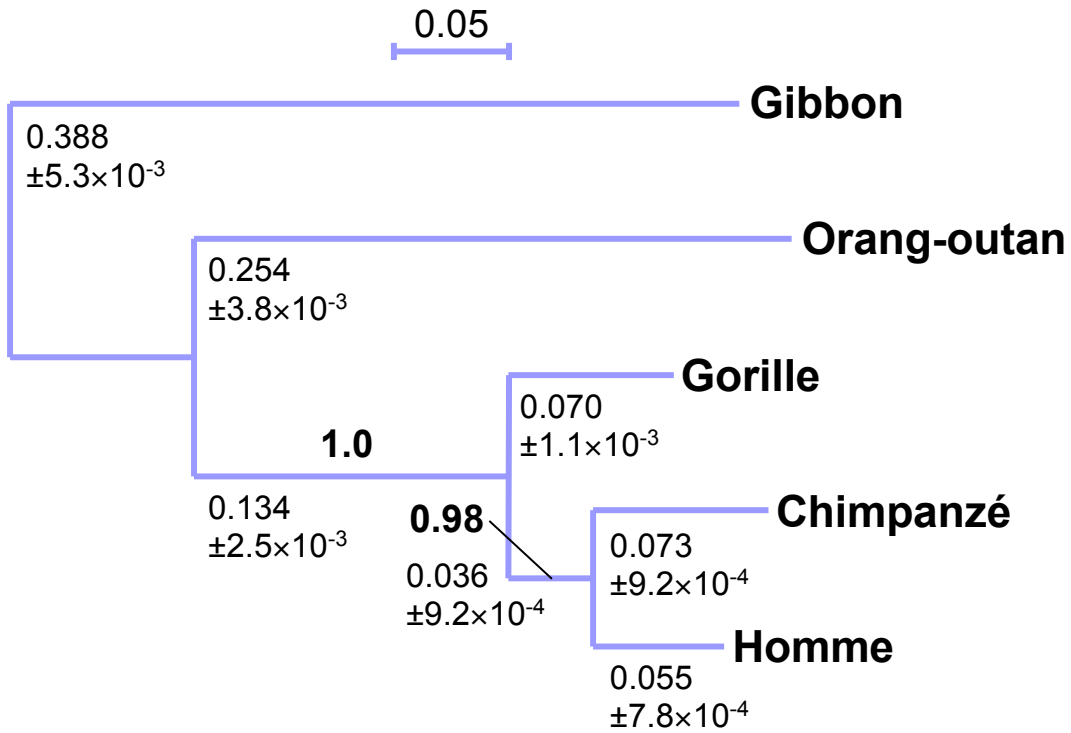
Convergence des chaînes





Arbre obtenu

- Construction par consensus majoritaire à 50% sur les itérations échantillonnées hors *burn-in*.
- Racinement avec la séquence du Gibbon.
- Longueurs des branches avec intervalles de crédibilité à 95%.



Avantages et limitations

- Meilleur comportement que le maximum de vraisemblance avec des modèles comprenant de nombreux paramètres :
  - Intégration des paramètres de nuisance.
- Temps de calcul biens plus longs :
  - Avec les MC^3, de nombreuses chaînes sont lancées en parallèle.
  - Nécessité d'atteindre la distribution stationnaire pour les chaînes froides :
    - Diminution du nombre d'itérations pour raccourcir les temps de calcul.
- Pas de nécessité d'effectuer du rééchantillonnage de type *bootstrap* :
  - Utilisation des valeurs de probabilités postérieures des clades :
    - Valeurs directement interprétables en termes de probabilités.

Bootstrap et probabilités postérieures

- Construction de six phylogénies (Douady *et al.*, 2003) :
  - Vraisemblance et bayésien.
- Comparaison entre valeurs de *bootstrap* (BP) et :
  - Probabilités postérieures (PP) des clades.
  - *Bootstrap* des probabilités postérieures (BPP).
- Valeurs des PP systématiquement plus élevées que celles des BP.