Statistique Bayésienne

Propriétés asymptotiques des approches bayésiennes

Anna Simoni[2]

[2]CREST - Ensae and CNRS

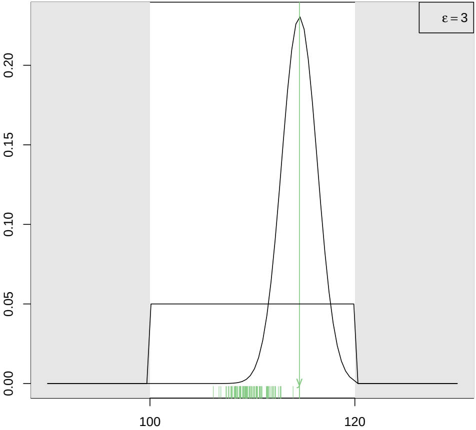
Outline

Setup

- Θ =espace métrique complet separable avec σ-algèbre $\mathrm{B}(\Theta)$;
- P_θ = probabilité sur un espace mesurable (X, B) ;
- $X_1, X_2, \ldots$= suite de variables-aléatoires qui sont, pour tout $\theta \in \Theta$, $i.i.d.$ P_θ ;
- $X^{(n)} := (X_1, \ldots, X_n)$ et P_θ^n est la mesure produit sur $(\mathrm{X}^n, \mathrm{B}^n)$;
- π = prior sur $(\Theta, \mathrm{B}(\Theta))$;
- $\pi(\cdot|\cdot) : \mathrm{B}(\Theta) \times (\mathrm{X}^n, \mathrm{B}^n) \mapsto [0, l]$ est l'a posteriori (i.e. une version de la distribution conditionnelle de θ étant donné $X^{(n)}$).

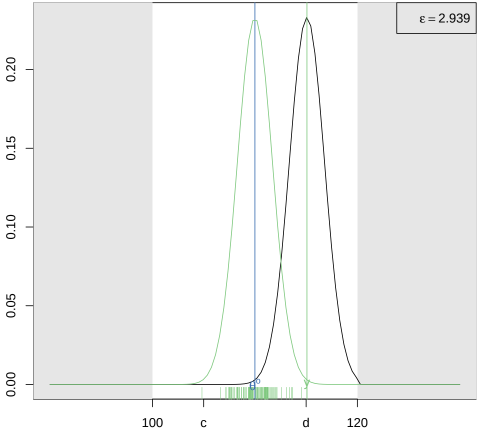
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a,b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



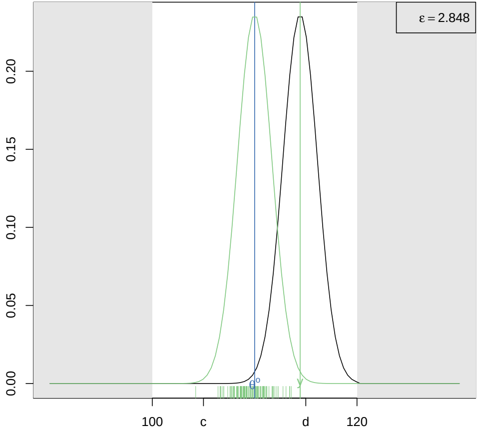
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



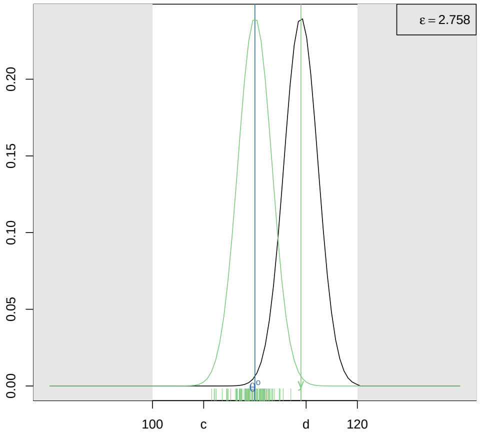
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

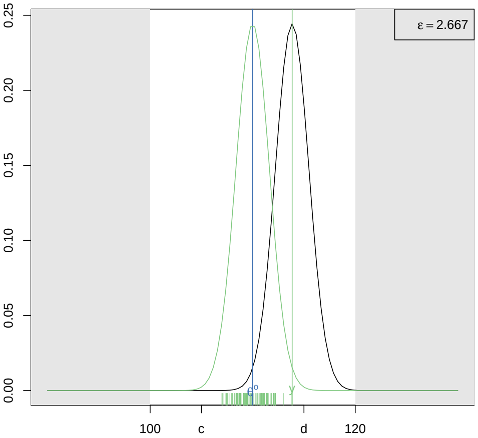


Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

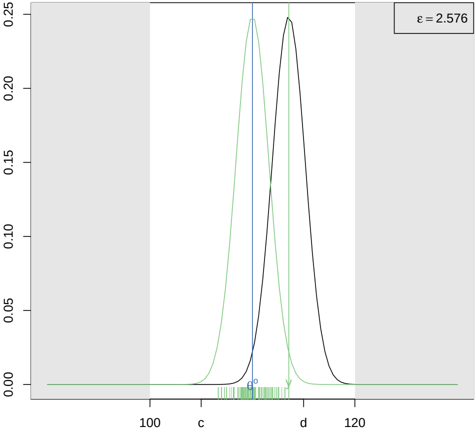


- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



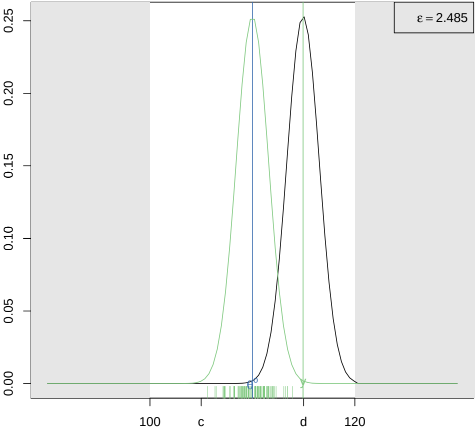
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



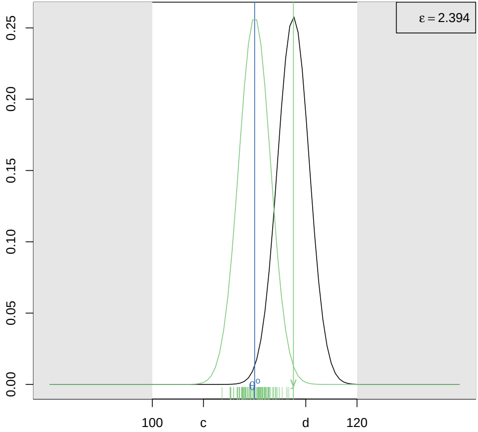
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



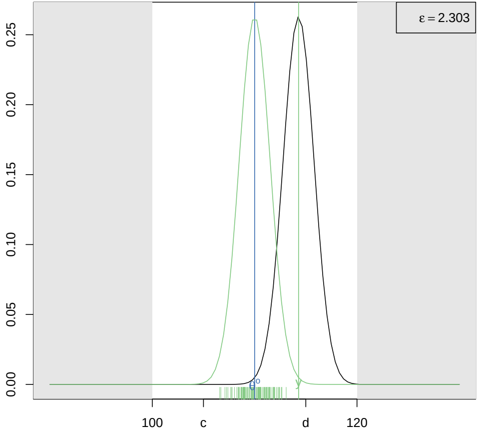
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



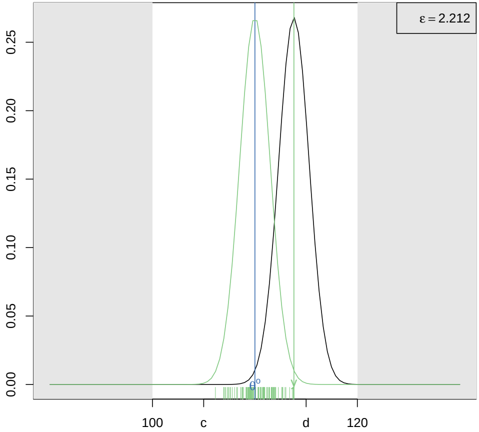
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

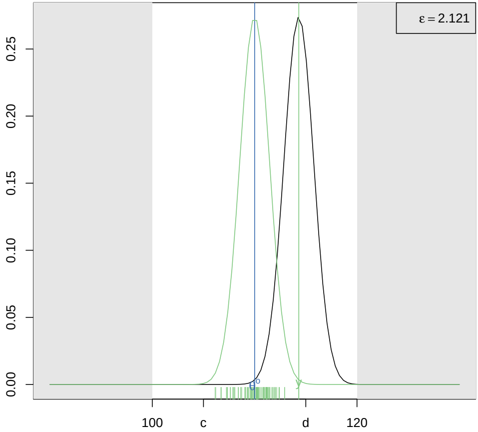


Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

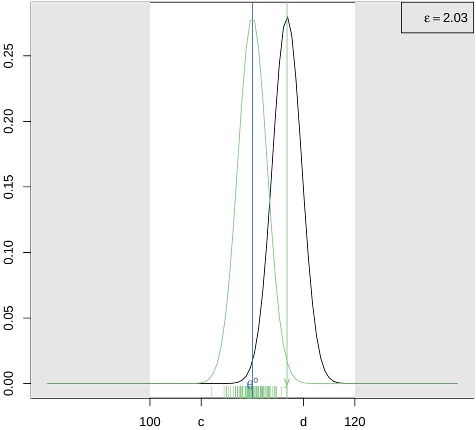


- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

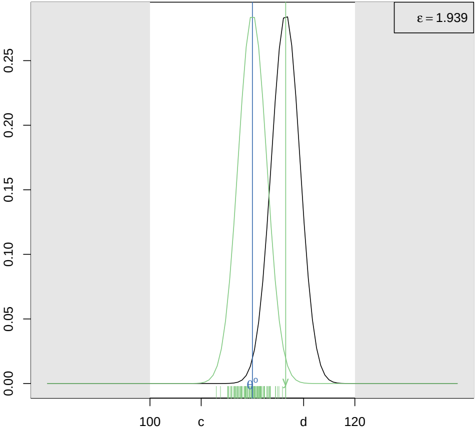


Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

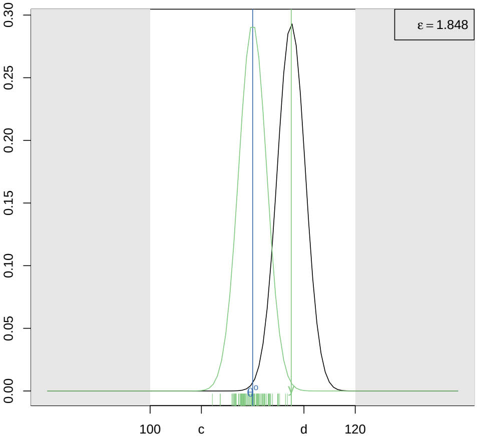


- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a,b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



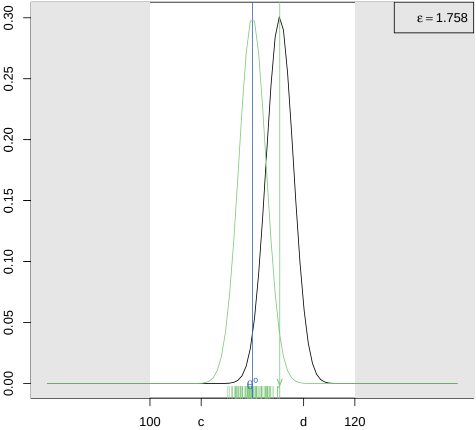
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



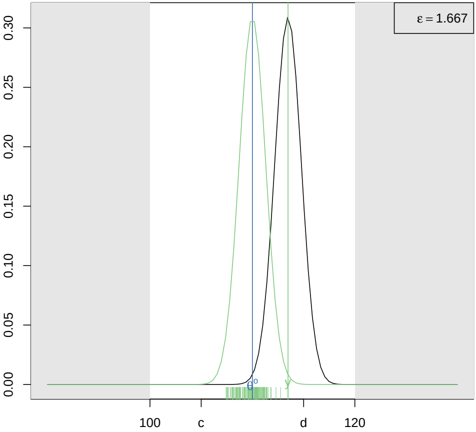
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

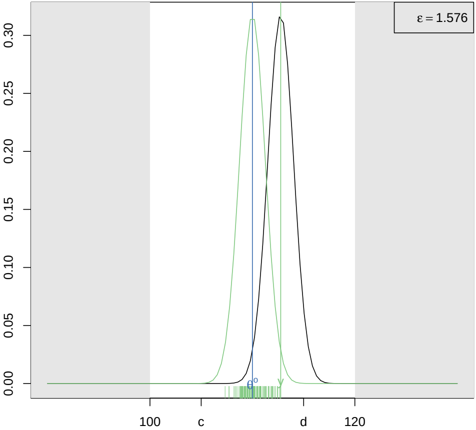


Exemple

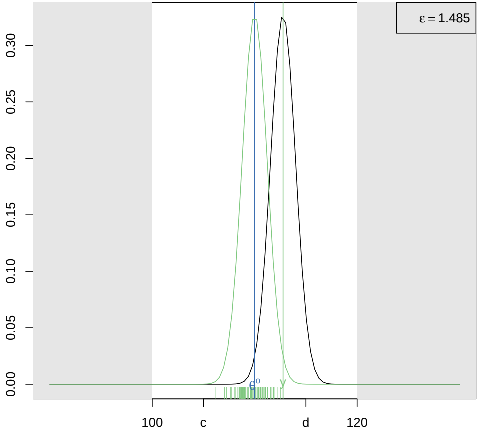
- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



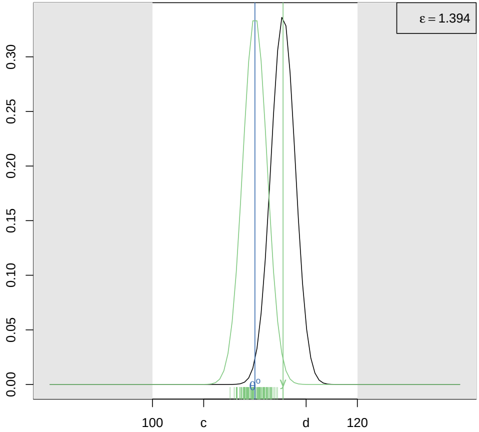
- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

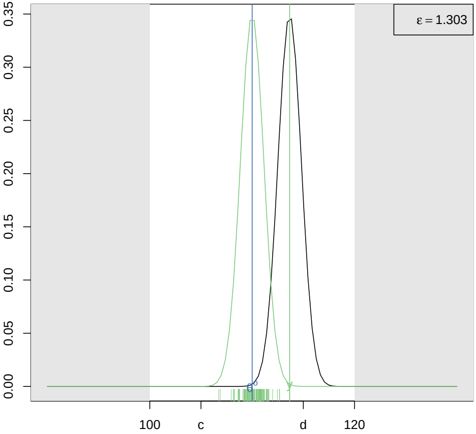


- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \frac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



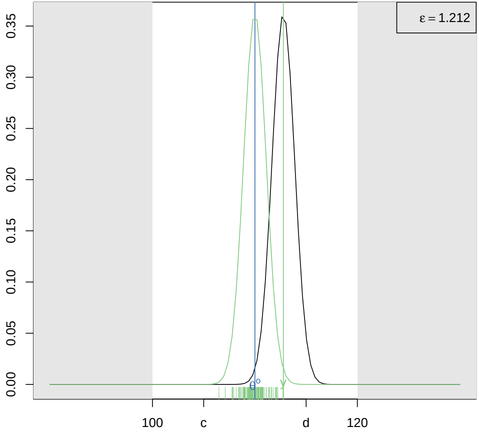
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a,b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

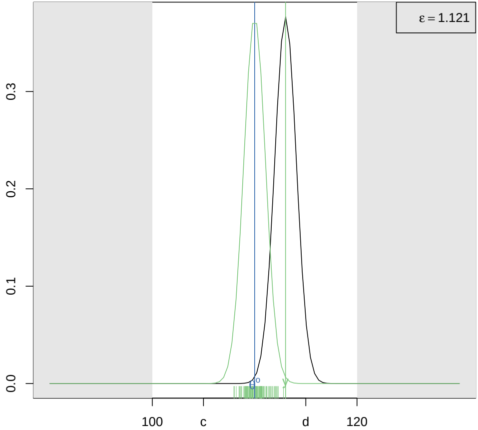


Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

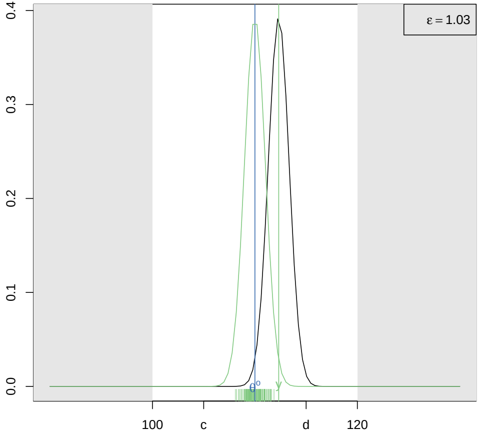


- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

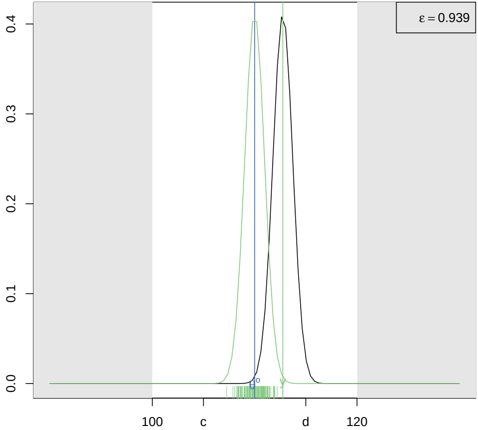


Exemple

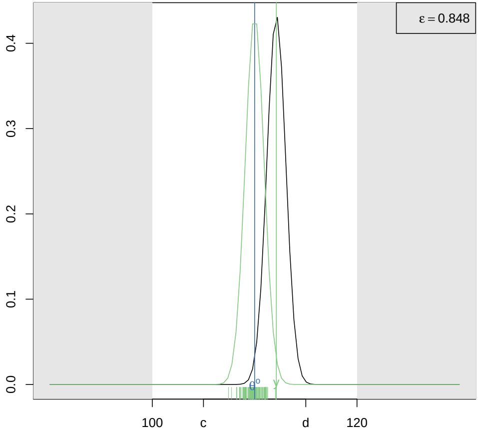
- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



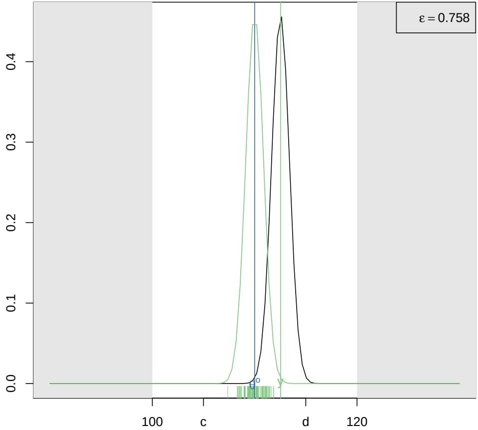
- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

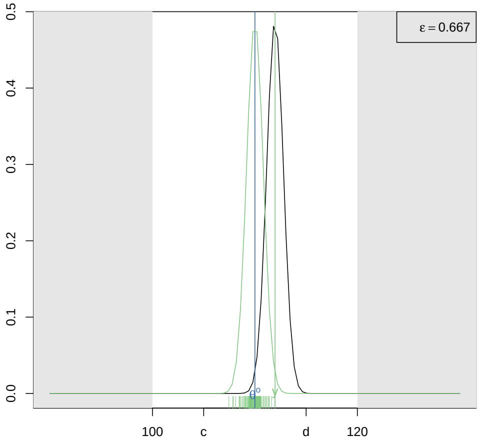


- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



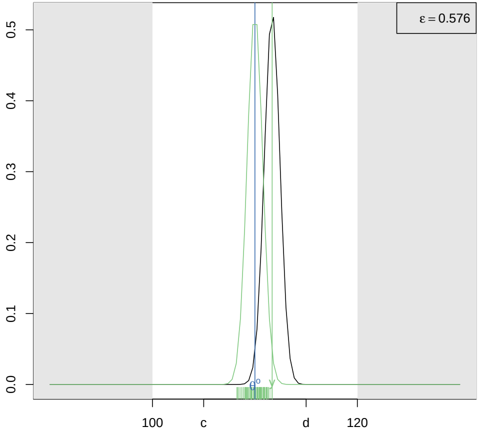
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



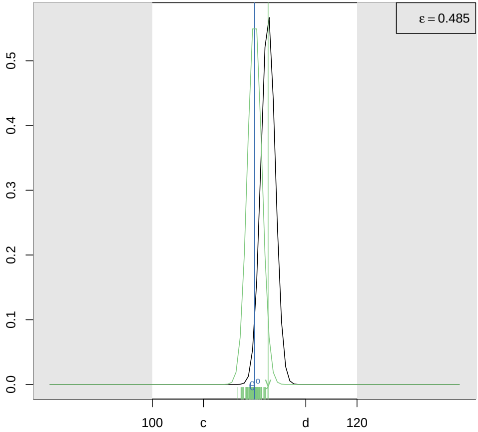
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a,b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



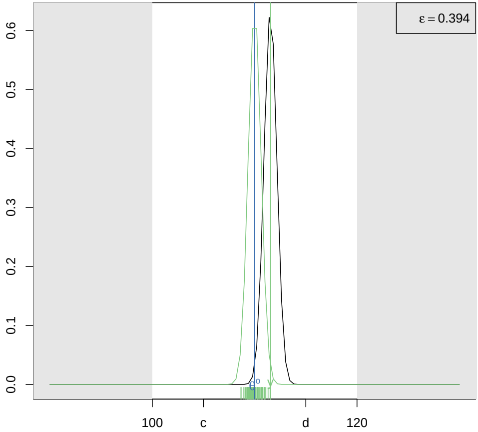
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



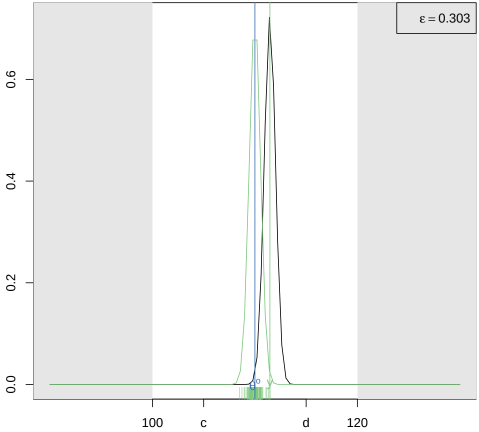
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta-y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right)-\Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

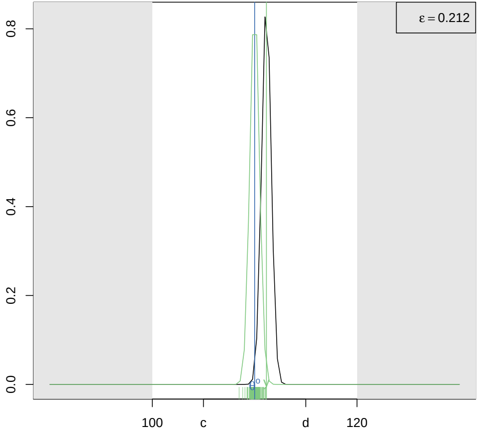


Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$

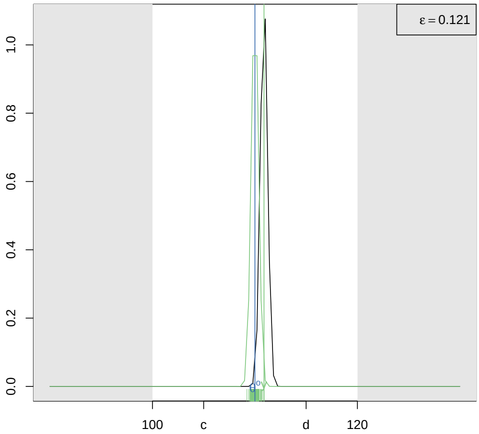


- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



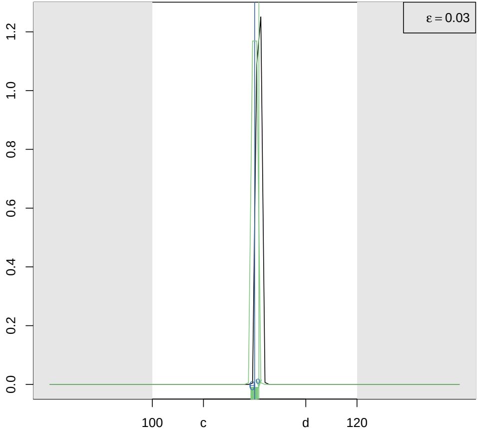
Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \dfrac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



Exemple

- likelihood : $Y|\theta \sim \mathrm{N}(\theta, \varepsilon)$, prior : $\theta \sim \mathrm{U}[a, b]$
- posterior density : $\pi_{\theta|Y}(\theta|y) = \frac{\varepsilon^{-1/2}\phi\left(\frac{\theta - y}{\sqrt{\varepsilon}}\right)}{\Phi\left(\frac{b-y}{\sqrt{\varepsilon}}\right) - \Phi\left(\frac{a-y}{\sqrt{\varepsilon}}\right)} 1_{[a,b]}(\theta)$



Consistance de la loi a posteriori. I

Définition

Pour tout n, soit $\pi(\cdot|X^{(n)})$ l'a posteriori étant donné $X^{(n)}$. La suite $\{\pi(\cdot|X^{(n)})\}$ est consistante en θ_0 si pour tout voisinages U *de θ_0,*

$$\lim_{n\to\infty} \pi(\mathrm{U}|X^{(n)}) \to 1 \qquad P^n_{\theta_0} - p.s. \tag{1}$$

- Soit d une métrique définiè sur Θ, alors on peut reécrire (1) comme :

  $$\lim_{n\to\infty} \pi(\theta; d(\theta,\theta_0) \leq \epsilon|X^{(n)}) \to 1 \qquad P^n_{\theta_0} - p.s.$$

  pour tout $\epsilon > 0$.

- Si Θ est un espace métrique separable alors la consistance de l'a posteriori en θ_0 est équivalente à :

  $$\pi(\cdot|X^{(n)}) \to \delta_{\theta_0} \qquad P^n_{\theta_0} - p.s.$$

  où la convergence est dans le sens de la convergence faible.

Consistance de la loi a posteriori. II

Théorème (Doob 1949)

Soit $\mathrm{X} \subset \mathbb{R}^p$ *pour quelque* $p < \infty$*. Soit* $\theta \in \Theta$ *à valeurs reéls, soit* $\theta \mapsto P_\theta$ $1-1$*, et soit* π *la distribution a priori. Alors, il* $\exists A \subseteq \Theta$ *tel que* $\pi(A) = 1$ *et* $\forall \theta_0 \in A$*,* $\forall \epsilon > 0$

$$\lim_{n\to\infty} \pi(\theta; \|\theta - \theta_0\| \leq \epsilon | X^{(n)}) \to 1 \qquad P^n_{\theta_0} - p.s.$$

Consistance de la loi a posteriori. III

- Le théorème de Doob ne repond pas à la question de la consistance de l'a posteriori à un θ_0 specifique.

- Pour avoir la consistance de l'a posteriori à tous θ_0 il faut imposer des conditions.

- De plus, dans le cas ∞-dimensionnelle, l'ensemble de valeurs θ_0 pour lesquelles la consistance n'est pas satisfaite peux être très grand (voir, Freedman 1963).

- Si $\theta_0 \notin Supp(\pi)$, alors $\exists$ un voisinage U de θ_0 tel que $\pi(\mathrm{U}) = 0$. Ceci implique que $\pi(\mathrm{U}|X^{(n)}) = 0$ p.s.. Il n'est donc pas raisonnable de s'attendre à avoir la consistance en dehors du support de l'a priori.

Théorème (1)

Suppose que K soit un sousensemble compact d'un espace métrique separable. Soit $T(\cdot,\cdot) : \theta \times \mathbb{R} \to \mathbb{R}$ telle que

(i) *pour tout x, $T(\cdot, x)$ est continue en θ et*

(ii) *pour tout θ, $T(\theta, \cdot)$ est mesurable.*

Soient $X_1, X_2, \ldots$ variables aléatoires i.i.d. definie sur $(\mathrm{X}, \mathrm{B}, P)$ avec $\mathbf{E}(T(\theta, X_1)) = \mu(\theta)$ et suppose en plus que

$$\mathbf{E}\left(\sup_{\theta\in K} |T(\theta, X_i)|\right) < \infty.$$

Alors, pour $n \to \infty$

$$\sup_{\theta\in K}\left|\frac{1}{n}\sum_{i=1}^{n} T(\theta, X_i) - \mu(\theta)\right| \to 0 \qquad p.s.P.$$

Consistance de la loi a posteriori. V

On suppose P_θ dominée par la mesure de Lebesgue et on note $p_\theta = dP_\theta/d\theta$.

Théorème (2)

Soit Θ un espace métrique compact. Pour un θ_0 fixé, soit

$$T(\theta, x) = \log\left(\frac{p_\theta(x)}{p_{\theta_0}(x)}\right).$$

Si $T(\theta, X_i)$ satisfait les hypothèses du Théorème (1) avec $P = P_{\theta_0}$, alors

- *l'estimateur de maximum de vraisemblance est consistant pour θ_0 ;*
- *si π est une a priori sur Θ et si θ_0 est dans le support de π alors l'a posteriori defini par la densité (par rapport à Lebesgue)*

$$\frac{\prod_{i=1}^n p_\theta(X_i)\pi(\theta)}{\int_\theta \prod_{i=1}^n p_\theta(X_i)\pi(\theta)}d\theta$$

est consistent en θ_0.

Outline

Vitesses de convergence

L'a posteriori converge à la vitesse $\varepsilon_n \to 0$ pour la distance d en θ_0 si

$$\mathbf{E}_{\theta_0} \pi(\theta; d(\theta, \theta_0) \leq \varepsilon_n | X^{(n)}) \to 1.$$

C'est une borne supérieure : on cherche ε_n le plus petit possible.

On dit que ζ_n est une borne inférieure pour la vitesse pour d en θ_0 si

$$\mathbf{E}_{\theta_0} \pi(\theta; d(\theta, \theta_0) \geq \zeta_n | X^{(n)}) \to 1.$$

Exemple : Modèles paramétriques *réguliers*. Pour tout $M_n \to +\infty$,

$$\mathbf{E}_{\theta_0} \pi \left(\theta; \frac{1}{M_n \sqrt{n}} \leq \|\theta - \theta_0\| \leq \frac{M_n}{\sqrt{n}} | X^{(n)} \right) \to 1.$$

Forme de l'a posteriori I

Encore plus précisément, on peut s'intéresser à la forme de la loi a posteriori.

Le théorème de Bernstein-von Mises (BvM) est un exemple de tel résultat.

Hypothèses :

1. $X_1, X_2, \ldots, X_n$ sont i.i.d. p_θ et $p_\theta d\theta = dP_\theta$.
2. Θ sousensemble ouvert de $\mathbb{R}$.
3. $\{x;\, p_\theta(x) > 0\}$ est le même $\forall \theta \in \Theta$.
4. $L(\theta, x) = \log p_\theta(x)$ est 3 fois differentiable par rapport à θ en un voisinage $(\theta_0 - \delta, \theta_0 + \delta)$, $\delta > 0$. Dénote : $\dot{L}$, $\ddot{L}$ et $\dddot{L}$ la première, seconde et troisième derivée. Alors $\mathbf{E}_{\theta_0}\dot{L}(\theta_0) < \infty$ et $\mathbf{E}_{\theta_0}\ddot{L}(\theta_0) < \infty$ et
$$\sup_{\theta \in (\theta_0 - \delta, \theta_0 + \delta)} |\dddot{L}(\theta_0, x)| < M(x) \qquad \text{et} \qquad \mathbf{E}_{\theta_0} M < \infty.$$
5. Échange de l'ordre de l'$\mathbf{E}_{\theta_0}$ et de la differentiation en θ_0 est justifié, et donc
$$\mathbf{E}_{\theta_0}\dot{L}(\theta_0) = 0, \qquad \mathbf{E}_{\theta_0}\ddot{L}(\theta_0) = -\mathbf{E}_{\theta_0}\left(\dot{L}(\theta_0)\right)^2.$$

Forme de l'a posteriori II

6. $I(\theta_0) := \mathbf{E}_{\theta_0} \left(\dot{L}(\theta_0)\right)^2$.

Soit $L_n(\theta) = \sum_{i=1}^n L(\theta, X_i)$.

Théorème (BvM)

Soupposons que $\{p_\theta;\ \theta \in \Theta\}$ *satisfait les hypothèses 1-6 et que* $\widehat{\theta}_n$ *est une solution consistante de l'équation de vraisemblance. De plus, supposons que*

(i). $\forall \delta > 0$, $\exists$ *un* $\epsilon > 0$ *tel que*

$$P_{\theta_0}\left\{ \sup_{|\theta-\theta_0|>\delta} \frac{1}{n}(L_n(\theta) - L_n(\theta_0)) \leq -\epsilon \right\} \to 1;$$

(ii). *l'a priori a une densité* $\pi(\theta)$ *par rapport à la mesure de Lebesgue qui est continue et positive en* θ_0.

Dénote $\pi^*(s|X^{(n)})$ *la densité a posteriori de* $s := \sqrt{n}(\theta - \widehat{\theta}_n(X^{(n)}))$. *Alors, pour* $n \to \infty$,

$$\int_{\mathbb{R}} \left| \pi^*(s|X^{(n)}) - \frac{\sqrt{I(\theta_0)}}{\sqrt{2\pi}} e^{-s^2 I(\theta_0)/2} \right| ds \to 0 \qquad en\ P^n_{\theta_0} - proba.$$

Remarques :

- La distance en variation totale entre deux probabilités P et Q est donnée par

$$\begin{aligned} \|P-Q\|_1 &= 2\sup_B |P(B)-Q(B)| \\ &= \int |p-q|d\mu. \end{aligned}$$

- Alors, le théorème de BvM dit que :

$$\left\| \pi(\cdot|X^{(n)}) - \mathrm{N}\left(\widehat{\theta}_n, \frac{I_{\theta_0}^{-1}}{n}\right)(\cdot) \right\|_1 \to 0 \qquad \text{en } P_{\theta_0}^n - \text{proba}$$

  où I_{θ_0} est l'information de Fisher.

- On remarque que l'a priori est asymptotiquement "effacée" de la loi a posteriori.

Normalité asymptotique de l'estimateur Bayésien

Théorème (3)

Supposons que les hypothèses du théorème de BvM sont satisfaites et que $\int |\theta|\pi(\theta)d\theta < \infty$. *Soit* $\theta_n^* := \int_{\mathbb{R}} \theta\pi(\theta|X^{(n)})d\theta$ *l'estimateur de Bayes par rapport à la fonction de coût quadratique. Alors :*

(i) $\sqrt{n}(\widehat{\theta}_n - \theta_n^*) \to 0$ *en* $P_{\theta_0}^n$*- proba.*

(ii) $\sqrt{n}(\theta_n^* - \theta_0)$ *converge en distribution vers une* $\mathrm{N}(0, I_{\theta_0}^{-1})$.