Statistique Bayésienne

Aziz CHENNOUFI et Auderic Bien-Aimée ABESSOLO MINKO

Mai 2024

Table de matières

1. Introduction
2. Lois *a priori* et *a posteriori*
3. Comment choisir l'*a priori*?
4. Estimation
5. Etude d'un jeu de données réel

Table de matières

1. Introduction
2. Lois *a priori* et *a posteriori*
3. Comment choisir l'*a priori*?
4. Estimation
5. Etude d'un jeu de données réel

Le soleil a-t-il explosé ?

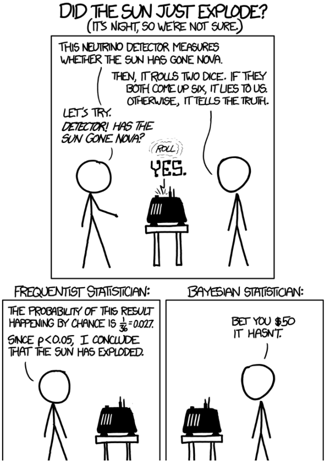

Figure: Exemple introductif (xkcd)

Introduction

S : "le soleil a explosé", $\overline{S}$: "le soleil n'a pas explosé" et O :"oui"
En approche fréquentiste, $\mathbb{P}(O|\overline{S}) = \frac{1}{36} = 0,027$.
Comme $\mathbb{P}(O|\overline{S}) < 0,05$, le soleil a explosé.
En approche bayésienne, Le statisticien aurait raisonné comme suit,

$$\begin{aligned}
P(S|O) &= \frac{P(O|S)P(S)}{P(O)} \text{ et posons } P(S) = \epsilon \\
&= \frac{P(O|S)\epsilon}{P(O|S)P(S) + P(O|\overline{S})P(\overline{S})} \\
&= \frac{\frac{35}{36}\epsilon}{\frac{35}{36}\epsilon + \frac{1}{36}(1-\epsilon)} \\
&= \frac{35\epsilon}{1+34\epsilon}
\end{aligned}$$

Table de matières

1. Introduction
2. Lois *a priori* et *a posteriori*
3. Comment choisir l'*a priori*?
4. Estimation
5. Etude d'un jeu de données réel

Lois *a priori* et *a posteriori*

- Information a priori sur le paramètre : $\theta \sim \pi$
- L'échantillon $\underline{X}|\theta$ est i.i.d suit une loi conditionnelle $\mathcal{P}_\theta$, vraisemblance : $f_X(\underline{X}|\theta)$
- Objectif : déterminer $\pi(\theta|\underline{X}) \rightarrow$ Formule de Bayes
- $P(A|B) = \frac{P(B|A)P(A)}{P(B)} \rightarrow \pi(\theta|\underline{X}) = \frac{f_X(\underline{X}|\theta)\pi(\theta)}{\int_\Theta f_X(\underline{X}|\theta)\,d\pi(\theta)} \propto f_X(\underline{X}|\theta)\pi(\theta)$

Exemple

$$\underline{X} \sim \text{Ber}(\theta) \text{ avec } \theta \sim \text{Beta}(a, b)$$

$$f_X(\underline{X}|\theta) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{n\bar{X}_n}(1-\theta)^{n(1-\bar{X}_n)}$$

$$\pi(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}1_{\{0\leq\theta\leq1\}}$$

$$\begin{aligned}\pi(\theta|\underline{X}) &\propto f_X(\underline{X}|\theta)\pi(\theta) \\ &\propto \theta^{n\bar{X}_n+a-1}(1-\theta)^{n(1-\bar{X}_n)+b-1}1_{\{0\leq\theta\leq1\}}\end{aligned}$$

$$\theta|\underline{X} \sim \text{Beta}(n\bar{X}_n + a, n(1-\bar{X}_n) + b)$$

Definition (Loi a priori impropre)

Une loi a priori impropre est une mesure π sur le paramètre θ qui vérifie simultanément,

$$\int_\Theta \pi(\theta)\, d\theta = +\infty \quad \text{et} \quad \int_\Theta f_X(\underline{X}|\theta)\, d\pi(\theta) < +\infty \text{ p.s.}$$

Exemple: Soit $\underline{X} \sim \mathcal{N}(\theta, 1)$, si l'on prend comme loi a priori π la mesure de Lebesgue sur $\mathbb{R}$, on a:

$$\int_\Theta \pi(\theta)\, d\theta = \int_{\mathbb{R}} d\theta = +\infty$$

Et, $$\int_\Theta f_X(\underline{X}|\theta)\, d\pi(\theta) = \frac{1}{\sqrt{2\pi}^{n-1}} e^{-\frac{n}{2}\overline{X^2}} \int_{\theta\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{n}{2}[(\theta-\overline{X})^2-\overline{X}^2]}\, d\theta$$
$$= \frac{1}{\sqrt{2\pi}^{n-1}} e^{-\frac{n}{2}(\overline{X^2}-\overline{X}^2)} \int_{\theta\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{n}{2}(\theta-\overline{X})^2}\, d\theta < +\infty$$

Table de matières

1. Introduction
2. Lois *a priori* et *a posteriori*
3. Comment choisir l'*a priori*?
4. Estimation
5. Etude d'un jeu de données réel

Lois a priori informatives

- Cas discret : k experts, k valeurs μ_j ,avec $j = 1, ..., k$, pour le paramètre θ.
  $\pi_j > 0$ mesure la fiabilité de l'avis de l'expert j (par défaut $\pi_j = \frac{1}{k}$).
  loi *a priori* (discrète),

$$\pi(\theta) = \sum_{i=1}^{k} \pi_j \delta_{\mu_j}(\theta)$$

  Au préalable, $\sum_{i=1}^{k} \pi_j = 1$
  Loi *a posteriori* (discrète),

$$\begin{aligned} P(\theta = \mu_j | \underline{X}) = \pi(\mu_j | \underline{X}) \\ &= \frac{f_X(\underline{X}|\mu_j)\pi(\mu_j)}{\sum_{i=1}^{k} f_X(\underline{X}|\mu_j)\pi(\mu_j)} \\ &\propto f_X(\underline{X}|\mu_j)\pi(\mu_j) \end{aligned}$$

Lois a priori informatives

- Cas continu : k experts, k distributions représentatives τ_j avec $j = 1, ..., k$ pour le paramètre. A priori : mélange de lois $\pi = \sum_{j=1}^{k} \tau_j \pi_j$
- Lois uniformes $\tau_j = \mathcal{U}_{I_j}$ où I_j est l'intervalle de confiance pour θ associé à l'expert j.
- Lois gaussiennes $\tau_j = \mathcal{N}(\mu_j, \sigma_j^2)$, où μ_j est la valeur moyenne du paramètre et σ_j^2 son incertitude.

$$\begin{aligned}
\pi(\theta|\underline{X}) &\propto f_X(\underline{X}|\theta)\pi(\theta) \\
&\propto f_X(\underline{X}|\theta) \sum_{j=1}^{k} \pi_j \tau_j(\theta) \\
&\propto \sum_{j=1}^{k} \pi_j f_X(\underline{X}|\theta) \tau_j
\end{aligned}$$

Lois a priori conjuguées

Définition

On se donne une famille $\mathcal{F}$ de lois de probabilité sur Θ. On suppose que la loi *a priori* appartient à $\mathcal{F}$. Si dans ces conditions, la loi *a posteriori* appartient encore à $\mathcal{F}$, on dira que la loi *a priori* est conjuguée.

Exemple : $\underline{X} \sim \mathcal{N}(\theta, 1)$ et $\mathcal{F} = \{\mathcal{N}(\lambda, \mu) \mid \mu > 0, \lambda \in \mathbb{R}\}$.

$$f_X(\underline{X}|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_i-\theta)^2} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2}$$

$$\pi(\theta) = \frac{1}{\sqrt{2\pi\mu}} e^{\frac{1}{2\mu}(\theta-\lambda)^2}$$

D'où,

$$\pi(\theta|\underline{X}) \propto e^{\theta^2\left(-\frac{1}{2\mu}-\frac{n}{2}\right)+\theta\left(n\bar{X}_n+\frac{\lambda}{\mu}\right)}$$

$$\pi(\theta|\underline{X}) \sim \mathcal{N}\left(\frac{\lambda + n\bar{X}_n\mu}{1+n\mu}, \frac{\mu}{1+n\mu}\right) \text{ avec } \frac{\mu}{1+n\mu} > 0.$$

Lois a priori non informatives

- Pas d'information disponible sur $\theta \to$ a priori non informatif
- Par exemple, $\pi \sim \text{Leb}_{\mathbb{R}}$ (a priori impropre)
- A priori de Jeffreys : $\pi(\theta) \propto \sqrt{det(I(\theta))}$ où

$$I_{ij}(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j} \log f_X(\underline{X}|\theta)\right]$$

- En dimension 1, $I(\theta) = -\mathbb{E}\left(\frac{\partial^2 \log(f_X(\underline{X}|\theta))}{\partial\theta^2}\right)$ et $\pi(\theta) \propto \sqrt{I(\theta)}$

Lois a priori non informatives

Exemple : Soit $\underline{X} \sim \mathcal{E}(\lambda)$ avec $\lambda > 0$.

$$f_X(\underline{X}|\lambda) = \lambda^n \exp(-n\lambda\bar{X}_n)\mathbf{1}_{[0,+\infty[}(\min_{i=1}^{n} X_i) \propto \lambda^n \exp(-n\lambda\bar{X}_n)$$

$$\frac{\partial}{\partial\lambda}\log f_X(\underline{X}|\lambda) = \frac{n}{\lambda} - n\bar{X}_n$$

$$\frac{\partial^2}{\partial\lambda^2}\log f_X(\underline{X}|\lambda) = -\frac{n}{\lambda^2}$$

$$\text{Donc, } I(\lambda) = \frac{n}{\lambda^2}$$

D'où la loi de Jeffreys : $\pi(\lambda) \propto \frac{1}{\lambda}$.

Table de matières

1. Introduction
2. Lois *a priori* et *a posteriori*
3. Comment choisir l'*a priori*?
4. Estimation
5. Etude d'un jeu de données réel

Théorie de la décision

- Règle de décision T $\rightarrow$ estimateur T($\underline{X}$)
- Fonction de coût L : mesure l'erreur $L(\theta, T(\underline{X}))$
- Choisir la règle qui minimise le **risque fréquentiste** $R_T(\theta) = \mathbb{E}[L(\theta, T(\underline{X}))|\theta]$ uniformément en θ.
- Utiliser le **critère minimax**, qui consiste à choisir la règle de décision T_0 telle que :

$$\forall T \in \mathcal{D}, \max_{\theta \in \Theta} R_{T_0}(\theta) \leq \max_{\theta \in \Theta} R_T(\theta)$$

- Choisir la règle qui minimise le **risque bayésien** $r(T)$ défini par :

$$r(T) = \int_{\Theta} R_T(\theta)\, d\pi(\theta)$$

Une telle règle de décision est appelée **règle de décision Bayésienne**.

Construction d'estimateurs Bayésiens

Théorème

Soient L une fonction de coût et π un a priori, alors la règle de décision T_π vérifiant

$$T_\pi(\underline{X}) \in \mathrm{argmin}_{T\in\mathcal{D}}\mathbb{E}[L(\theta, T(\underline{X}))|\underline{X})]$$

est une règle de décision Bayésienne.

Ce théorème fondamental donne naissance à des formes usuelles simples d'estimateurs Bayésiens, selon la fonction de perte choisie.

- **Pour la perte** $\mathbb{L}^2$: $L(\theta, T(\underline{X})) = (\theta - T(\underline{X}))^2 \Rightarrow T_\pi(\underline{X}) = \mathbb{E}[\theta|\underline{X}]$
- **Pour la perte** $\mathbb{L}^1$: $L(\theta, T(\underline{X})) = |\theta - T(\underline{X})| \Rightarrow T_\pi(\underline{X}) = \mathrm{med}(\theta|\underline{X})$

Exemple

Dans le modèle

$$\underline{X}|\theta \sim \text{Ber}(\theta) \text{ avec } \theta \sim \text{Beta}(a, b)$$

On a montré que $\theta|\underline{X} \sim \text{Beta}(n\bar{X}_n + a, n(1 - \bar{X}_n) + b)$

Un estimateur Bayésien est simplement
$\mathbb{E}[\theta|\underline{X}] = \frac{n\bar{X}_n+a}{n\bar{X}_n+a+n(1-\bar{X}_n)+b} = \frac{n\bar{X}_n+a}{n+a+b}$

Exemple

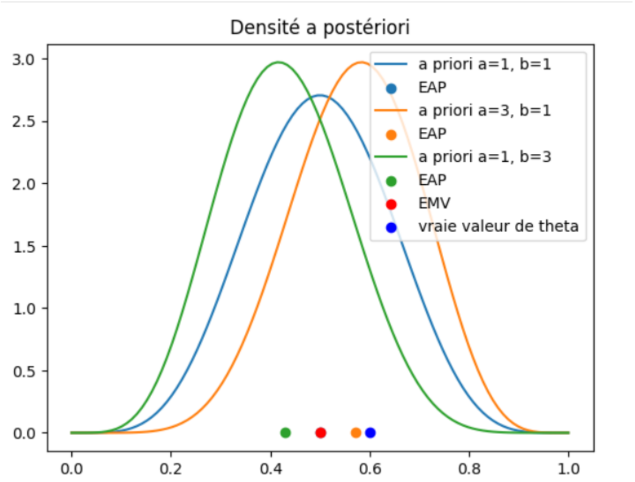

Figure: Densité a posteriori et estimateurs en résultant pour différents a priori.

Propriétés asymptotiques

- **Théorème de Bernstein Von-Mises:** sous certaines conditions de régularité, la loi a posteriori converge étroitement vers la masse de Dirac centrée sur la vraie valeur θ_0 du paramètre lorsque la taille de l'échantillon tend vers $+\infty$.
- Une conséquence de ce théorème est que $\sqrt{n}(\hat{\theta}_{\text{bay}} - \theta_0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, I(\theta_0)^{-1})$, avec $\hat{\theta}_{\text{bay}} = \mathbb{E}[\theta|\underline{X}]$
- De plus,sous certaines conditions de régularité, un des résultats les plus importants de la statistique inférentielle est $\sqrt{n}(\hat{\theta}_{\text{MV}} - \theta_0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, I(\theta_0)^{-1})$
- Sous certaines conditions de régularité, on s'attend à ce que $\sqrt{n}(\hat{\theta}_{\text{bay}} - \hat{\theta}_{\text{MV}})$ converge vers 0 l'espérance de la loi normale $\mathcal{N}(0, I(\theta_0)^{-1})$.

Exemple

Illustration calculatoire du théorème de Bernstein Von-Mises

Le modèle

$\underline{X} \sim \text{Ber}(\theta)$ avec $\theta \sim \text{Beta}(a, b) \rightarrow \pi(\theta|\underline{X}) \propto \text{Beta}(n\overline{X}_n + a, n(1 - \overline{X}_n) + b)$

Un estimateur bayésien est $\hat{\theta}_{\text{bay}} = \mathbb{E}[\theta|\underline{X}] = \dfrac{n\overline{X}_n + a}{n + a + b}$

On a, $\hat{\theta}_{\text{bay}} - \hat{\theta}_{\text{MV}} = \overline{X_n}[\dfrac{-\frac{a+b}{n}}{1 + \frac{a+b}{n}}] + \dfrac{\frac{a}{n}}{1 + \frac{a+b}{n}}$

D'où, $\sqrt{n}(\hat{\theta}_{\text{bay}} - \hat{\theta}_{\text{MV}}) = \overline{X_n}[\dfrac{-\frac{a+b}{\sqrt{n}}}{1 + \frac{a+b}{n}}] + \dfrac{\frac{a}{\sqrt{n}}}{1 + \frac{a+b}{n}} \xrightarrow[n \to \infty]{} 0$ p.s.

car $\overline{X_n} \xrightarrow{\text{p.s}} \theta$ par la LFGN.

De plus, $\sqrt{n}(\hat{\theta}_{\text{bay}} - \theta) = \sqrt{n}(\hat{\theta}_{\text{bay}} - \hat{\theta}_{\text{MV}}) + \sqrt{n}(\hat{\theta}_{\text{MV}} - \theta)$

Or, $\sqrt{n}(\hat{\theta}_{\text{MV}} - \theta) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, \theta(1 - \theta))$

Par le lemme Slutsky, $\sqrt{n}(\hat{\theta}_{\text{bay}} - \theta) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, \theta(1 - \theta))$

Exemple

Illustration graphique du théorème de Bernstein-Von Mises

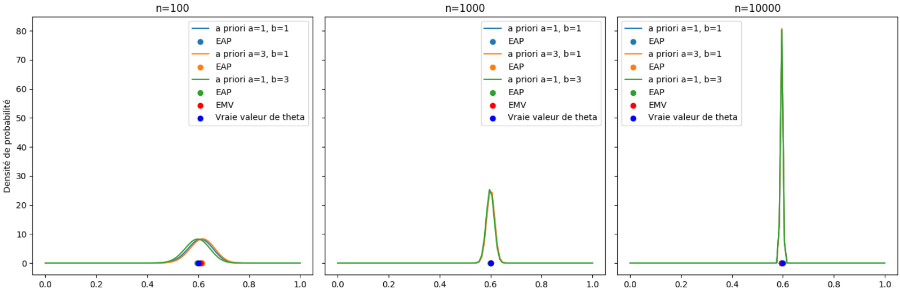

Figure: Evolution du comportement asymptotique de la densité a posteriori

Table de matières

1. Introduction
2. Lois *a priori* et *a posteriori*
3. Comment choisir l'*a priori*?
4. Estimation
5. Etude d'un jeu de données réel

Etude d'un jeu de données réel

	duree de vie
0	0.779152
1	1.661016
2	1.682125
3	3.509372
4	0.054527
...	...
995	2.238471
996	1.530724
997	1.423945
998	2.695116
999	0.125841

1000 rows × 1 columns

Etude d'un jeu de données réel

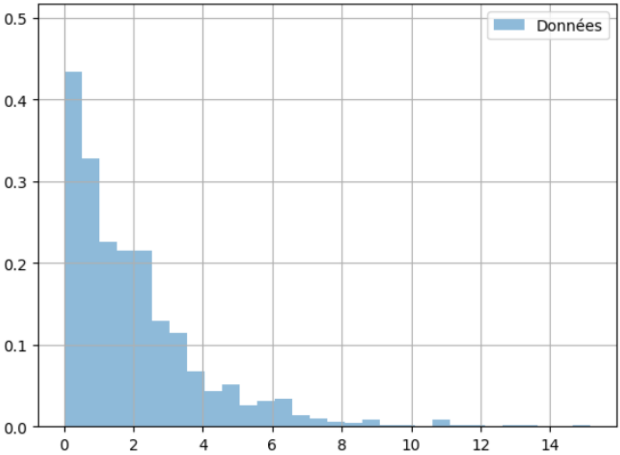

Figure: Histogramme de la distribution des données

Etude d'un jeu de données réel

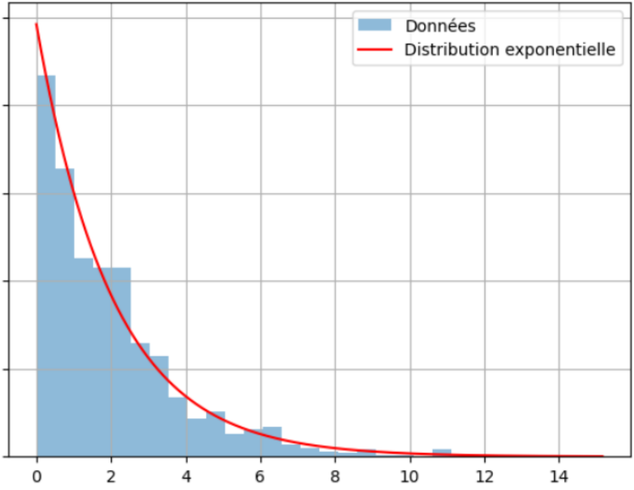

Figure: Approximation des données à une loi exponentielle

Etude d'un jeu de données réel

On pose le modèle $\underline{X}|\theta \sim \mathcal{E}(\theta)$ avec $\theta \sim \Gamma(\alpha, \beta)$

$$
\begin{aligned}
f_X(\underline{X}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta X_i} \mathbb{1}_{X_i>0} \\
&= \theta^n e^{-\theta n \bar{X}_n} \mathbb{1}_{\min X_i>0}
\end{aligned}
$$

$$
\pi(\theta) = \frac{\beta^\alpha \theta^{\alpha-1} e^{-\theta\beta}}{\Gamma(\alpha)} \mathbb{1}_{\theta>0}
$$

Par suite,

$$
\pi(\theta|\underline{X}) \propto e^{-\theta(n\bar{X}_n+\beta)} \theta^{n+\alpha-1} \mathbb{1}_{\theta>0}
$$

Donc $\theta|\underline{X} \sim \Gamma(n+\alpha, n\bar{X}_n+\beta)$.

Etude d'un jeu de données réel

- EMV $= \frac{1}{\overline{X_n}} = 0,49 \to$ on choisit α et β tels que $\mathbb{E}[\theta] = \frac{\alpha}{\beta} = \frac{1}{2}$

Etude d'un jeu de données réel

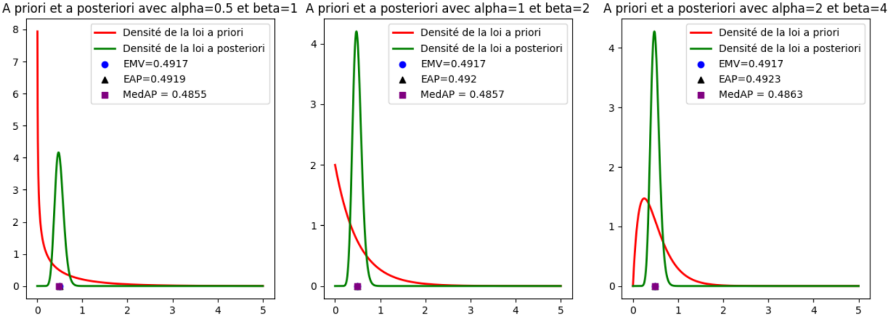

Figure: Distributions *a priori* et *a posteriori* pour différentes valeurs de α et β, n=25

Etude d'un jeu de données réel

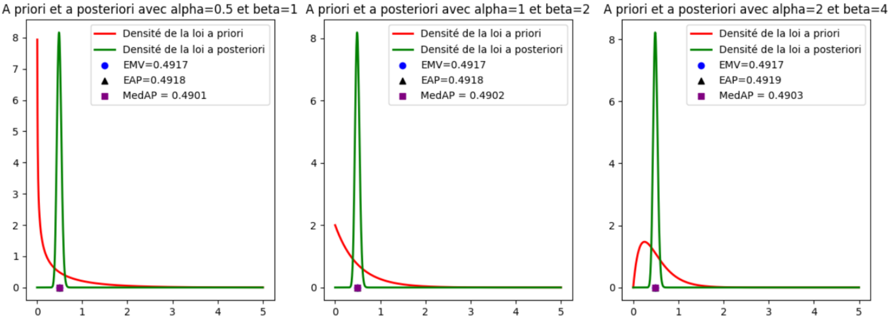

Figure: Distributions *a priori* et *a posteriori* pour différentes valeurs de α et β, n=100

Etude d'un jeu de données réel

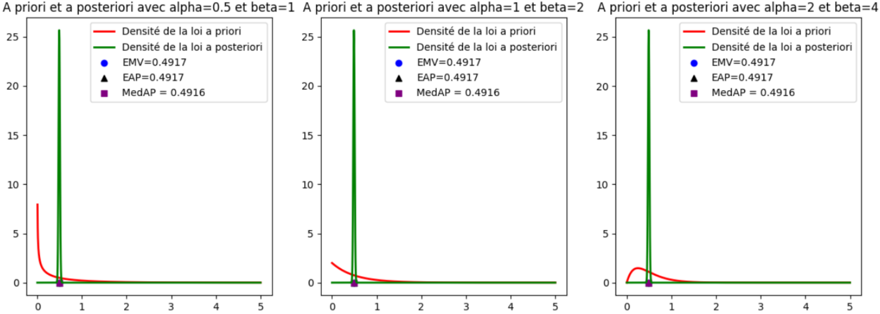

Figure: Distributions *a priori* et *a posteriori* pour différentes valeurs de α et β, n=1000