

Statistique Bayésienne

Aziz CHENNOUFI et Auderic Bien Aimée ABESSOLO MINKO

Encadrement par M. Frédéric Proïa

Mai 2024

Table des matières

1	Introduction	3
2	Lois <i>a priori</i> et <i>a posteriori</i>	5
3	Comment choisir l'<i>a priori</i> ?	7
3.1	Lois <i>a priori</i> informatives	7
3.2	Lois <i>a priori</i> conjuguées	8
3.3	Lois <i>a priori</i> non informatives	9
3.4	La loi <i>a priori</i> de Jeffreys	10
4	Esimation	12
4.1	Théorie de la décision	12
4.2	Propriétés asymptotiques	17
5	Simulation de lois <i>a posteriori</i>	19
5.1	Méthode de rejet	19
5.2	Méthode de Monte Carlo	22
6	Etude d'un jeu de données réel	23

1 Introduction

La statistique bayésienne, une approche innovante et de plus en plus prisée dans l'analyse de données, offre des outils puissants pour évaluer les incertitudes et prendre des décisions éclairées dans un monde incertain. Notre Travail d'Etude et de Recherche vise à présenter et explorer les fondements et les avantages de cette méthode qui a révolutionné le paysage traditionnel des sciences empiriques.

Le contexte de cette approche remonte au XVIII^e siècle, avec l'émergence de la **règle de Bayes**, nommée ainsi en l'honneur du mathématicien anglais **Thomas Bayes**. Cette règle représente la pierre angulaire de la statistique bayésienne et constitue le fondement sur lequel repose cette méthode d'analyse.

Exemple 1.1 (L'exemple historique de Bayes). *Publié dans son œuvre *Essay Towards Solving a Problem in the Doctrine of Chances*, 1763, l'exemple de Bayes a donné naissance à la Statistique Bayésienne. Il se formule comme suit :*

Une boule de billard se déplace le long d'une ligne de longueur 1, s'arrêtant à un point avec une probabilité uniforme. Imaginons qu'elle s'arrête à un point nommé p . Une deuxième boule suit le même parcours n fois dans les mêmes conditions, et on observe le nombre de fois, noté X , où elle s'arrête avant la première boule. Bayes se demande alors : étant donné X , quelles conclusions pouvons-nous tirer sur p ? Pour lui, il serait raisonnable de prendre en compte l'observation menée sur la première boule (qui donnera naissance à l'a priori) avant de réaliser des estimations sur la deuxième.

La différence entre l'inférence fréquentiste et l'inférence bayésienne se trouve dans la manière d'estimer une quantité d'intérêt θ .

En approche fréquentiste, on cherche à estimer une valeur ponctuelle inconnue θ à partir des observations grâce à des méthodes telles que l'estimateur du maximum de vraisemblance ou l'estimateur des moments.

En approche bayésienne, θ n'est pas une valeur précise mais une variable aléatoire qui suit une certaine loi.

Exemple 1.2

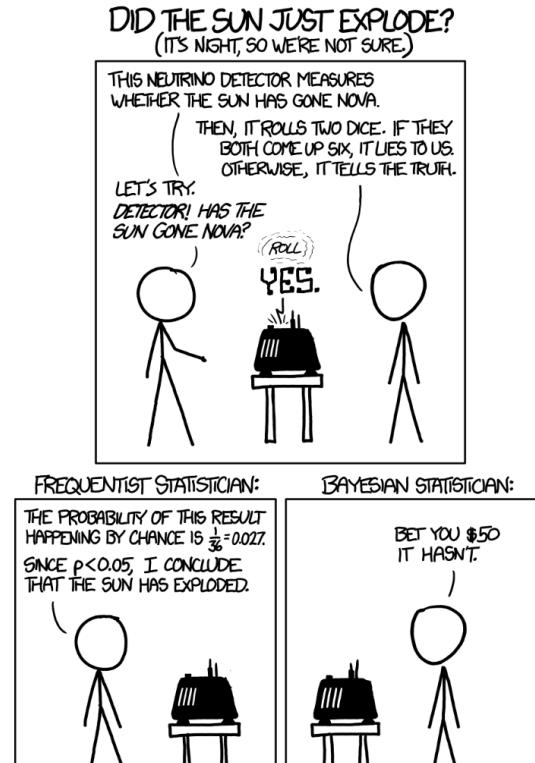


FIGURE 1 – (xkcd)

Dans cet exemple, l'enjeu est d'estimer si "le soleil a explosé" en se basant sur la réponse d'un détecteur de neutrinos. Le détecteur a répondu "oui" à la question "Le soleil a t-il explosé?". La probabilité qu'il mente est $\frac{1}{36}$ et celle qu'il dise la vérité est $\frac{35}{36}$.

En approche fréquentiste, le stasticien aurait raisonné de la façon suivante, comme la probaboilité que le détecteur mente, sachant que la réponse est "oui", est $\frac{1}{36} = 0,027$. Cette probabilité est $< 0,05$, donc le soleil a explosé.

En approche bayésienne, Le statisticien aurait raisonné comme suit, soient, S : "le soleil a explosé", \bar{S} : "le soleil n'a pas explosé" et O : "oui" Par la formule de Bayes, on a :

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

On peut remarquer que pour ϵ très petit $P(O|S)$ est également très petite. On a défini ϵ comme étant $P(S)$ qui est en fait l'*a priori* que l'on définira dans la suite. Logiquement, dans ce cas le statisticien aurait pris comme *a priori* un ϵ très petit. Par conséquent, il aurait conclu que le soleil n'a pas explosé bien que le détecteur ait répondu "oui". Ce qui est plus raisonnable ici.

2 Lois *a priori* et *a posteriori*

En approche bayésienne, on dispose d'une information *a priori* sur le paramètre à estimer θ . Cette information, choisie par le statisticien, prend la forme d'une loi sur l'espace paramétrique Θ et on la note $\pi(\theta)$. On note également \mathcal{P}_θ la loi du modèle qui est en fait la loi conditionnelle sachant θ . On a donc,

$$\underline{X}|\theta \sim \mathcal{P}_\theta \text{ avec } \theta \sim \pi.$$

Il est important de noter pour la suite que les observations $\underline{X} = (X_1, \dots, X_n)$ sont indépendantes conditionnellement à θ . On désigne par f_θ la densité de \mathcal{P}_θ par rapport à la mesure de Lebesgue ou la mesure de comptage. La loi de \underline{X} sachant θ est donc donnée par la vraisemblance du modèle,

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

Définition 2.1 (Loi *a posteriori*). On appelle loi *a posteriori* la loi conditionnelle $\pi(\theta|\underline{X})$.

Le calcul de cette loi se repose sur le théorème fondamental suivant :

Théorème 2.1 (Formule de Bayes). Soient A et B deux événements tels que $P(B) \neq 0$, alors

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

En notant $f_X(\underline{X}|\theta)$ la vraisemblance de l'échantillon \underline{X} conditionnellement à θ , la **formule de Bayes** nous permet de déduire la loi *a posteriori* de la manière suivante

$$\pi(\theta|\underline{X}) = \frac{f_X(\underline{X}|\theta)\pi(\theta)}{\int_{\Theta} f_X(\underline{X}|\theta) d\pi(\theta)}$$

En effet, le numérateur représente la loi jointe du couple (\underline{X}, θ) et le dénominateur est la loi marginale de \underline{X} . Le dénominateur ne dépend pas de θ , il s'apparente donc à une constante de normalisation qui garantit que $\pi(\theta|\underline{X})$ soit une densité (donc de mesure égale à 1). Il nous sera donc utile de noter simplement,

$$\pi(\theta|\underline{X}) \propto f_X(\underline{X}|\theta)\pi(\theta)$$

Exemple 2.1 (Beta-Binomial). *Dans les modèles où le paramètre joue le rôle d'une probabilité, il est utile de prendre la loi $Beta(a,b)$ avec $a, b > 0$ comme a priori. En effet, elle est portée par $[0,1]$ et ses deux paramètres lui permettent une flexibilité qui nous permettra d'ajuster convenablement notre loi a priori. Supposons donc*

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

La vraisemblance du modèle est

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

La loi a priori s'écrit :

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

On obtient donc comme loi a posteriori :

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

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

Remarque 2.1. *On voit bien que la loi a posteriori a du sens puisque les paramètres obtenus représentent respectivement le nombre de succès obtenus ($n\bar{X}_n$) augmenté de a , et le nombre d'échecs ($n(1-\bar{X}_n)$) augmenté de b . Nous reviendrons par la suite sur cet exemple une fois que d'autres notions utiles seront définies.*

Exemple 2.2 (Modèle et a priori gaussiens). *Supposons que*

$$\underline{X}|\theta \sim \mathcal{N}(\theta, 1) \text{ avec } \theta \sim \mathcal{N}(0, 1)$$

Alors,

$$\begin{aligned}
\pi(\theta|\underline{X}) &\propto \left(\frac{1}{\sqrt{2\pi}}\right) \cdot e^{-\frac{\theta^2}{2}} \cdot \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(X_i-\theta)^2}{2}}\right) \\
&\propto \exp\left(-\sum_{i=1}^n \frac{1}{2}(X_i - \theta)^2 - \frac{\theta^2}{2}\right) \\
&\propto \exp\left(-\frac{n+1}{2}\theta^2 + n\bar{X}_n\theta\right) \\
&\propto \exp\left(-\frac{\left(\theta - \frac{n\bar{X}_n}{n+1}\right)^2}{2(n+1)}\right)
\end{aligned}$$

Par suite, $\theta|\underline{X} \sim \mathcal{N}\left(\frac{n\bar{X}_n}{n+1}, \frac{1}{n+1}\right)$.

3 Comment choisir l'*a priori* ?

Le choix de l'*a priori* est une étape fondamentale en statistique bayésienne car cet *a priori* permet de décrire l'information disponible sur le paramètre avant observation des données. L'enjeu ici est donc de représenter cet information sous la forme d'une distribution sur le paramètre. Il est important de noter qu'un "bon" choix d'*a priori* est d'autant plus important que l'échantillon est de petite taille. En effet, sur des échantillons de grande taille l'*a priori* devient complètement négligeable (cf. le théorème de Bernstein-von Mises).

3.1 Lois *a priori* informatives

Il existe des stratégies pour transformer des connaissances *a priori* en distribution *a priori* à l'aide d'avis d'experts qui peuvent être non-statisticiens. La méthode la plus simple est de demander aux experts de donner des poids ou des probabilités à des intervalles de valeurs (méthode des histogrammes).

1. Dans le cas discret, on a k experts et chacun d'eux propose une valeur μ_j avec $j = 1, \dots, k$ pour le paramètre θ . Pour chaque valeur proposée par un expert, on associe un réel positif π_j qui mesure la fiabilité de l'avis de l'expert j . Si on n'a pas d'information sur la fiabilité de chaque avis, on peut considérer que chaque avis a le même niveau de fiabilité (ou poids) $\pi_j = \frac{1}{k}$. On a ainsi comme loi *a priori* discrète,

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

Au préalable, il est possible de normaliser les π_j afin que π soit une mesure de probabilité, i.e $\sum_{i=1}^k \pi_j = 1$, sans perte de généralité. Ici, la loi *a posteriori* est elle aussi discrète et on a

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

Cependant, si la loi *a posteriori* est une loi discrète, la probabilité qu'elle contienne la "vraie" valeur du paramètre est faible.

2. On peut généraliser ce qui précède au cas où au lieu de donner des valeurs précises au paramètre, les experts propose une distribution représentative τ_j des valeurs probables de θ . Chaque expert j peut proposer une distribution sous forme d'intervalle de confiance auquel on associe une loi uniforme \mathcal{U}_j . Autrement, chaque expert peut proposer une distribution gaussienne avec pour paramètres une valeur moyenne et une incertitude pour θ . Dans ce dernier cas, la loi *a priori* est alors un mélange de lois gaussiennes. La loi *a posteriori* se calcule donc comme suit,

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

3.2 Lois *a priori* conjuguées

Définition 3.1 (Loi *a priori* conjuguée). *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.*

Par conséquent, une famille conjuguée est stable par passage de l'*a priori* à l'*a posteriori*. Et l'avantage des familles conjuguées est avant tout de simplifier les calculs.

Exemple 3.1 (Lois normales). $\underline{X} \sim \mathcal{N}(\theta, 1)$ et $\mathcal{F} = \{\mathcal{N}(\lambda, \mu) \mid \mu > 0, \lambda \in \mathbb{R}\}$.

On a,

$$\begin{aligned} f_X(\underline{X}|\theta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_i - \theta)^2} \\ &= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \sum_{i=1}^n (X_i - \theta)^2} \\ \pi(\theta) &= \frac{1}{\sqrt{2\pi\mu}} e^{\frac{1}{2\mu}(\theta - \lambda)^2} \\ D'où, \pi(\theta|\underline{X}) &\propto e^{\theta^2(-\frac{1}{2\mu} - \frac{n}{2}) + \theta(n\bar{X}_n + \frac{\lambda}{\mu})} \\ \pi(\theta|\underline{X}) &\sim \mathcal{N}\left(\frac{\lambda + n\bar{X}_n\mu}{1 + n\mu}, \frac{\mu}{1 + n\mu}\right) \text{ avec } \frac{\mu}{1 + n\mu} > 0. \end{aligned}$$

Donc \mathcal{F} est une famille conjuguée.

Exemple 3.2 (Lois gamma). $\underline{X} \sim \mathcal{E}(\theta)$ et $\mathcal{F} = \{\Gamma_{a,b}(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \mathbb{1}_{\{\theta > 0\}} \mid a, b > 0\}$.

On a,

$$\begin{aligned} f_X(\underline{X}|\theta) &= \prod_{i=1}^n \theta e^{\theta X_i} \mathbb{1}_{\{X_i \geq 0\}} \\ &= \theta^n e^{-n\theta \sum_{i=1}^n X_i} \mathbb{1}_{\{\min X_i \geq 0\}} \\ D'où, \pi(\theta|\underline{X}) &\propto e^{-\theta(b + n^2 \bar{X}_n)} \theta^{n+a-1} \mathbb{1}_{\{\theta > 0\}} \\ \pi(\theta|\underline{X}) &\sim \Gamma_{n+a, b+n^2 \bar{X}_n} \end{aligned}$$

Avec $n + a > 0$ et $b + n^2 \bar{X}_n > 0$ (car les X_i sont > 0 p.s.)

Donc \mathcal{F} est une famille conjuguée.

3.3 Lois *a priori* non informatives

Dans le cas où on dispose de très peu d'informations, voire pas du tout, sur θ , on peut choisir des lois *a priori* dites non informatives. Dans ce cas, on souhaite que l'*a priori* intervienne de façon minimale dans la loi *a posteriori*, i.e que les données parlent d'elles même. On peut donc dire qu'une loi non informative est une distribution de probabilité qui ne favorise pas spécifiquement une valeur particulière ou un ensemble de valeurs pour les paramètres du modèle. Parmi les lois *a priori* non informatives, certaines peuvent être impropres.

Définition 3.2 (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 3.3. Considérons le modèle gaussien $G = \{\mathcal{N}(\theta, 1), \theta \in \mathbb{R}\}$, si l'on prend comme loi a priori π la mesure de Lebesgue sur \mathbb{R} , on a :

$$\begin{aligned} \int_{\Theta} \pi(\theta) d\theta &= \int_{\mathbb{R}} d\theta = +\infty \\ \text{Et, } \int_{\Theta} f_X(\underline{X}|\theta) d\pi(\theta) &= \frac{1}{\sqrt{2\pi}^{n-1}} e^{-\frac{n}{2}\bar{X}^2} \int_{\theta \in \mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{n}{2}[(\theta - \bar{X})^2 - \bar{X}^2]} d\theta \\ &= \frac{1}{\sqrt{2\pi}^{n-1}} e^{-\frac{n}{2}(\bar{X}^2 - \bar{X}^2)} \int_{\theta \in \mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{n}{2}(\theta - \bar{X})^2} d\theta < +\infty \text{ p.s.} \end{aligned}$$

3.4 La loi a priori de Jeffreys

L'approche la plus aboutie des *a priori* faiblement informatifs est peut-être celle de Jeffreys. Ce dernier a proposé une procédure pour trouver une loi *a priori* avec une propriété d'invariance par rapport à la paramétrisation.

a) Cas unidimensionnel.

Définition 3.3. L'a priori de Jeffreys est défini par

$$\pi(\theta) \propto \sqrt{I(\theta)}$$

$I(\theta)$ est la matrice d'information de Fisher, définie comme :

$$I(\theta) = -\mathbb{E} \left(\frac{\partial^2 \log(f_X(\underline{X}|\theta))}{\partial \theta^2} \right)$$

Exemple 3.4. Soit $\underline{X} \sim \mathcal{E}(\lambda)$ avec $\lambda > 0$. On a $f_X(\underline{X}|\lambda) = \lambda^n \exp(-n\lambda\bar{X}_n) \mathbf{1}_{[0, +\infty[}(\min_{i=1}^n X_i) \propto \lambda^n \exp(-n\lambda\bar{X}_n)$.

Calculons l'information de Fisher :

$$\begin{aligned} \frac{\partial}{\partial \lambda} \log f_X(\underline{X}|\lambda) &= \frac{n}{\lambda} - n\bar{X}_n \\ \frac{\partial^2}{\partial \lambda^2} \log f_X(\underline{X}|\lambda) &= -\frac{n}{\lambda^2} \end{aligned}$$

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

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

b) Cas multi-dimensionnel.

Définition 3.4. La loi *a priori* de Jeffreys est donnée par

$$\pi(\theta) \propto \sqrt{\det(I(\theta))}$$

Si $\theta \in \mathbb{R}^k$, alors $I(\theta)$ est une matrice dont les coefficients sont donnés par :

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

Exemple 3.5. Prenons l'exemple de la loi gaussienne de paramètres μ et σ^2 i.e $\underline{X} \sim \mathcal{N}(\mu, \sigma^2)$. Posons $\theta = (\mu, \sigma^2)$.

$$f_X(\underline{X}|\theta) \propto \frac{1}{\sigma^{2/n}} \exp \left(-\frac{1}{2\sigma^2} (n\bar{X}_n^2 - 2n\bar{X}_n\mu + n\mu^2) \right)$$

On a :

$$\begin{aligned} \log f_X(\underline{X}|\theta) &\propto -\frac{1}{2} \log \sigma^2 - \frac{(n\bar{X}_n^2 - 2n\bar{X}_n\mu + n\mu^2)}{2\sigma^2} \\ \frac{\partial}{\partial \mu} \log f_X(\underline{X}|\theta) &= \frac{n(\bar{X}_n - \mu)}{\sigma^2} \\ \frac{\partial^2}{\partial \mu^2} \log f_X(\underline{X}|\theta) &= -\frac{n}{\sigma^2} \\ \frac{\partial^2}{\partial \mu \partial \sigma^2} \log f_X(\underline{X}|\theta) &= -\frac{n(\bar{X}_n - \mu)}{(\sigma^2)^2} \\ \frac{\partial}{\partial \sigma^2} \log f_X(\underline{X}|\theta) &= -\frac{1}{2\sigma^2} + \frac{(n\bar{X}_n^2 - 2n\bar{X}_n\mu + n\mu^2)}{2(\sigma^2)^2} \\ \frac{\partial^2}{\partial (\sigma^2)^2} \log f_X(\underline{X}|\theta) &= \frac{1}{2(\sigma^2)^2} - \frac{(n\bar{X}_n^2 - 2n\bar{X}_n\mu + n\mu^2)}{(\sigma^2)^3} \end{aligned}$$

Or, $\mathbb{E}(\bar{X}_n - \mu) = 0$ et $\mathbb{E}[(n\bar{X}_n^2 - 2n\bar{X}_n\mu + n\mu^2)] = n\sigma^2$.

On a donc :

$$I(\theta) = I(\mu, \sigma^2) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n-1}{2\sigma^2} \end{pmatrix}$$

D'où la loi de Jeffreys : $\pi(\theta) \propto \frac{1}{\sigma^2}$.

De plus, la loi de Jeffreys est invariante par reparamétrisation. En effet, soit $\varphi = h(\theta)$ avec h un C^1 -difféomorphisme. Notons π la loi *a priori* de θ , donc φ est de loi $\tilde{\pi}$ avec $\tilde{\pi}(\varphi) = \pi(\varphi) |((h^{-1})'(\varphi))|$. De plus, on a $\tilde{I}(\varphi) = I(\varphi) |((h^{-1})'(\varphi))|^2$ donc $\tilde{\pi}(\varphi) \propto \sqrt{\tilde{I}(\varphi)}$. Ce calcul justifie en particulier la présence de la racine carrée.

4 Estimation

La loi *a posteriori* nous permet de déduire plusieurs estimateurs potentiels du paramètre, notamment l'espérance *a posteriori*, la médiane *a posteriori*, le mode *a posteriori* etc. On se demande donc quel est l'estimateur optimal parmi ces candidats ? Pour répondre à cette problématique, nous allons introduire un nombre de règles de décisions basées sur une fonction de perte $L(\theta, \hat{\theta})$ qui quantifie le coût de l'erreur d'estimation.

4.1 Théorie de la décision

La théorie de la décision repose sur deux éléments :

- Une fonction de coût ou fonction de perte $L : \Theta \times \mathcal{D} \rightarrow \mathbb{R}$, qui associe à une valeur de paramètre $\theta \in \Theta$ et une décision $T \in \mathcal{D}$, un coût $L(\theta, T)$.
- Une règle de décision qui est une fonction mesurable T telle que l'image par l'échantillon $T(\underline{X})$ soit un estimateur.

Parmi les fonctions de coût les plus classiques, on peut citer :

- La perte \mathbb{L}^2 ou quadratique $L(\theta, T(\underline{X})) = \|\theta - T(\underline{X})\|^2$
- La perte \mathbb{L}^1 ou perte absolue $L(\theta, T(\underline{X})) = |\theta - T(\underline{X})|$

Une des méthodes les plus fréquentes pour préférer une règle de décision T_1 à une autre T_2 est la notion du risque fréquentiste.

Définition 4.1 (Risque fréquentiste). *Le risque fréquentiste d'une règle de décision T pour une fonction de perte L est défini par*

$$\begin{aligned} R_T : \Theta &\rightarrow \mathbb{R}_+ \\ \theta &\mapsto \mathbb{E}[L(\theta, T(\underline{X})) | \theta] \end{aligned}$$

Le risque au point θ est donc l'erreur moyenne de T en θ . On dira donc que T_1 est préférable à T_2 si $R_{T_1}(\theta) \leq R_{T_2}(\theta)$ pour tout $\theta \in \Theta$

Définition 4.2 (Règle de décision admissible). *Une règle de décision T^* est dite admissible si c'est la meilleure règle de décision au sens du risque fréquentiste, c'est à dire,*

$$\forall T \in \mathcal{D}, R_{T^*}(\theta) \leq R_T(\theta) \text{ pour tout } \theta \in \Theta$$

Un des critères de sélection d'une règle de décision est le critère **minimax**, qui consiste à choisir la règle de décision qui a, pour T sur l'ensemble des $\theta \in \Theta$, le plus petit risque maximal. Donc On dit que T_0 est minimax si :

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

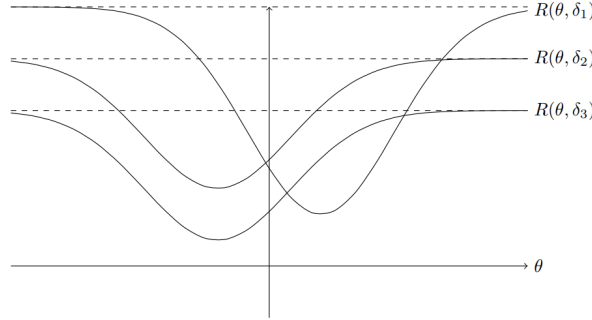


FIGURE 2 – (ISAE-SUPAERO). Un exemple de trois risques fréquentistes. δ_3 est préférable à δ_2 donc δ_2 n'est pas admissible.

Sur l'exemple de la Figure 2, le critère minimax va favoriser la règle de décision δ_3 .

Bien que ce critère soit considéré comme assez "prudent", son inconvénient est qu'il peut favoriser des règles de décision avec des risques fréquentistes bas, sans pour autant donner une meilleure prédiction du paramètre qu'une autre règle de décision, dont le risque fréquentiste est plus élevé.

Pour palier à ce problème, on introduit le **critère Bayésien**, basé sur le risque fréquentiste intégré, aussi appelé risque Bayésien. En fait, ce risque prend en compte tous les risques possibles, mais en pondérant les valeurs du paramètre par les probabilités qu'elles ont d'arriver. On le définit comme suit,

Définition 4.3 (Risque Bayésien). *Le risque Bayésien d'une règle de décision T pour un a priori π est défini par*

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

où $R_T(\theta)$ est le risque fréquentiste associé à T .

Ce risque bayésien donne naissance au principe de **règle de décision Bayésienne**, qui est la meilleure règle de décision au sens du risque Bayésien. Une règle de décision T_π est donc Bayésienne si :

$$r(T_\pi) \leq r(T) \text{ pour tout } T \in \mathcal{D}$$

Notons qu'il est souvent possible de proposer une construction d'un estimateur Bayésien pour une fonction de coût L et un a priori π donnés.

Théorème 4.1 (Construction d'un estimateur Bayésien). *Soient L une fonction de coût et π un a priori, alors la règle de décision T_π vérifiant*

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

est une règle de décision Bayésienne.

Démonstration. Supposons que $r(T) < \infty$. Pour toute règle de décision T , on a

$$r(T) = \mathbb{E}[L(\theta, T(\underline{X})) | \underline{X}] = \mathbb{E}[\mathbb{E}[L(\theta, T(\underline{X})) | \underline{X}]]$$

Donc,

$$r(T) \geq \mathbb{E}[\min_{T \in \mathcal{D}} \mathbb{E}[L(\theta, T) | \underline{X}]] = \mathbb{E}[\mathbb{E}[L(\theta, T_\pi(\underline{X})) | \underline{X}]] = \mathbb{E}[L(\theta, T_\pi(\underline{X}))] = r(T_\pi).$$

□

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

Proposition 4.1 (Estimateur de Bayes pour la fonction de perte quadratique). *Soit L la fonction de perte quadratique et π un a priori. On suppose que $\mathbb{E}(\theta | \underline{X}) < \infty$. Un estimateur Bayésien pour L et la loi π est donné par :*

$$T_\pi(\underline{X}) = \mathbb{E}[\theta | \underline{X}]$$

Démonstration. On applique le Théorème 4.1 pour la fonction de perte quadratique, on obtient

$$T_\pi(\underline{X}) = \operatorname{argmin}_{T \in \mathcal{D}} \mathbb{E}[L(\theta, T(\underline{X})) | \underline{X}] = \operatorname{argmin}_{T \in \mathcal{D}} \mathbb{E}[\|\theta - T(\underline{X})\|^2 | \underline{X}]$$

Or, la fonction $g : a \mapsto \mathbb{E}[(Z - a)^2]$ est minimale pour $a = \mathbb{E}[Z]$ car

$$g(a) = \mathbb{E}[(Z - \mathbb{E}[Z])^2] + (\mathbb{E}[Z] - a)^2 \geq g(\mathbb{E}[Z]).$$

Par suite, $T_\pi(\underline{X}) = \mathbb{E}[\theta | \underline{X}]$

□

Exemple 4.1. *Dans le modèle*

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

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

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

Pour illustrer l'avantage de l'approche bayésienne, nous avons simulé l'échantillon précédent en prenant $\theta = 0,6$ pour différentes valeurs de a et b comme *a priori*. On obtient la figure suivante,

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import beta

# Définition des valeurs de a et b pour les lois beta
a_values = [1, 3, 1]
b_values = [1, 1, 3]

# Données observées
X = np.random.binomial(n=1, p=0.6, size=10)
echecs = np.sum(X == 0)
succes = np.sum(X == 1)
EMV = np.mean(X)

# Plot des distributions a posteriori pour différentes valeurs de a et b
for a, b in zip(a_values, b_values):
    # Mise à jour des paramètres a et b en fonction des données observées
    a_posterior = a + succes
    b_posterior = b + echecs

    # Création de la distribution a posteriori beta
    posterior_beta = beta(a_posterior, b_posterior)

    # Plot de la distribution a posteriori
    x_values = np.linspace(0, 1, 100)
    plt.plot(x_values, posterior_beta.pdf(x_values), label=f'a priori a={a}, b={b}')
    EAP = a_posterior / (a_posterior + b_posterior)
    plt.scatter(EAP, 0, label="EAP")

# Affichage des autres éléments du graphique
plt.scatter(EMV, 0, color='red', label='EMV')
plt.scatter(0.6, 0, color='blue', label='vraie valeur de theta')
```

```
plt.title('Densité a posteriori')
plt.legend()
```

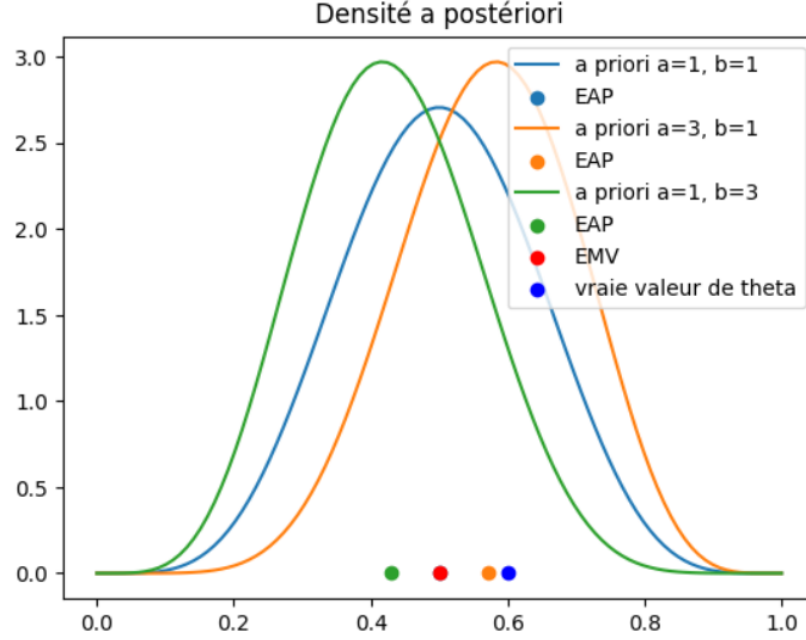


FIGURE 3 – Densité *a posteriori* et estimateurs en résultant pour différents *a priori*.

Le cas $a=1$ et $b=1$ (courbe en bleu) n'apporte pas d'information supplémentaire, ce qui explique que l'esperance *a posteriori* (EAP) est confondu avec l'EMV (obtenu en calculant la moyenne empirique). Le cas $a=3$ et $b=1$ (courbe en orange) vise à donner plus de poids aux succès qu'aux échecs, ce qui explique l'obtention d'une valeur plus proche de la vraie valeur de θ comparé à l'EMV. Enfin, l'*a priori* $a=1$ et $b=3$ est un mauvais *a priori* puisqu'il apporte de la fausse information sur le vrai paramètre. Ceci nous permet de conclure que lorsque l'*a priori* est bien choisi, l'approche bayésienne nous fournit des estimateurs plus pertinents que l'approche fréquentiste.

Exemple 4.2. Dans le modèle

$$\underline{X}|\theta \sim \mathcal{N}(\theta, 1) \text{ avec } \theta \sim \mathcal{N}(0, 1)$$

On a montré que $\theta|\underline{X} \sim \mathcal{N}(\frac{n\bar{X}_n}{n+1}, \frac{1}{n+1})$

Une estimateur Bayésien est simplement $\mathbb{E}[\theta|\underline{X}] = \frac{n\bar{X}_n}{n+1}$

Proposition 4.2 (Estimateur Bayésien pour la perte absolue). Soit L la fonction de perte absolue et π un *a priori*. Un estimateur Bayésien pour l et la loi π est donné par :

$$T_\pi(\underline{X}) = \text{med}(\theta|\underline{X})$$

Démonstration. D'après le Théorème 4.1, on cherche

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

Montrons que pour une variable aléatoire réelle Z intégrable de fonction de répartition F et de médiane $m = F^{-1}(1/2)$, on a

$$m \in \operatorname{argmin}_{a \in \mathbb{R}} E[|Z - a|].$$

Pour tout $a \in \mathbb{R}$, on a

$$E[|Z - a|] = \int_{-\infty}^{\infty} P(|Z - a| > t) dt = \int_a^{\infty} P(Z > t) dt + \int_{-\infty}^a P(Z < t) dt$$

Supposons que $a < m$, alors

$$E[|Z - a|] - E[|Z - m|] = \int_m^a \{P(Z > t) - P(Z < t)\} dt.$$

Si $t < m$, on a $P(Z > t) = 1 - P(Z \leq t) > 1/2$, donc l'intégrale ci-dessus est positive. Et si $a > m$, on a

$$E[|Z - a|] - E[|Z - m|] = \int_a^m \{P(Z < t) - P(Z > t)\} dt.$$

Si $t > m$, on a $P(Z < t) \geq P(Z \leq m) \geq 1/2$, donc dans ce cas aussi, l'intégrale est positive. En appliquant ce résultat à la loi *a posteriori*, on obtient bien

$$\operatorname{med}(\theta | \underline{X}) \in \operatorname{argmin}_{T \in \mathcal{D}} E[|\theta - T(X)| | \underline{X}].$$

□

4.2 Propriétés asymptotiques

Le théorème de Bernstein-von Mises est l'un des résultats clés de la statistique bayésienne. Il permet d'établir, sous certaines conditions de régularité¹, que 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

$$\sqrt{n}(\hat{\theta}_{\text{bay}} - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, I(\theta_0)^{-1})$$

1. https://fr.wikipedia.org/wiki/Théorème_de_Bernstein-von_Mises

De plus, un des résultats les plus importants de la statistique inférentielle est,

$$\sqrt{n}(\hat{\theta}_{\text{MV}} - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, I(\theta_0)^{-1})$$

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 $\mathcal{N}(0, I(\theta_0)^{-1})$.

Revenons sur l'exemple où $\underline{X} \sim \text{Ber}(\theta)$ avec $\theta \sim \text{Beta}(a, b)$. On sait que $\hat{\theta}_{\text{MV}} = \bar{X}_n$.

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

Donc un estimateur bayésien est l'espérance conditionnelle $\hat{\theta}_{\text{bay}} = \mathbb{E}[\theta|\underline{X}] = \frac{n\bar{X}_n + a}{n + a + b}$

On a, $\hat{\theta}_{\text{bay}} - \hat{\theta}_{\text{MV}} = \bar{X}_n \left[\frac{-\frac{a+b}{n}}{1 + \frac{a+b}{n}} \right] + \frac{\frac{a}{n}}{1 + \frac{a+b}{n}}$

D'où, $\sqrt{n}(\hat{\theta}_{\text{bay}} - \hat{\theta}_{\text{MV}}) = \bar{X}_n \left[\frac{-\frac{a+b}{\sqrt{n}}}{1 + \frac{a+b}{n}} \right] + \frac{\frac{a}{\sqrt{n}}}{1 + \frac{a+b}{n}} \xrightarrow[n \rightarrow \infty]{} 0$ p.s. (car $\bar{X}_n \xrightarrow{\text{p.s.}} \theta$ 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, $\sqrt{n}(\hat{\theta}_{\text{MV}} - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \theta(1 - \theta))$

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

On retrouve bien les résultats du théorème.

Pour mettre en évidence le théorème de Bernstein, nous avons repris l'exemple Beta-Binomial en agrandissant la taille de l'échantillon simulé, on obtient :

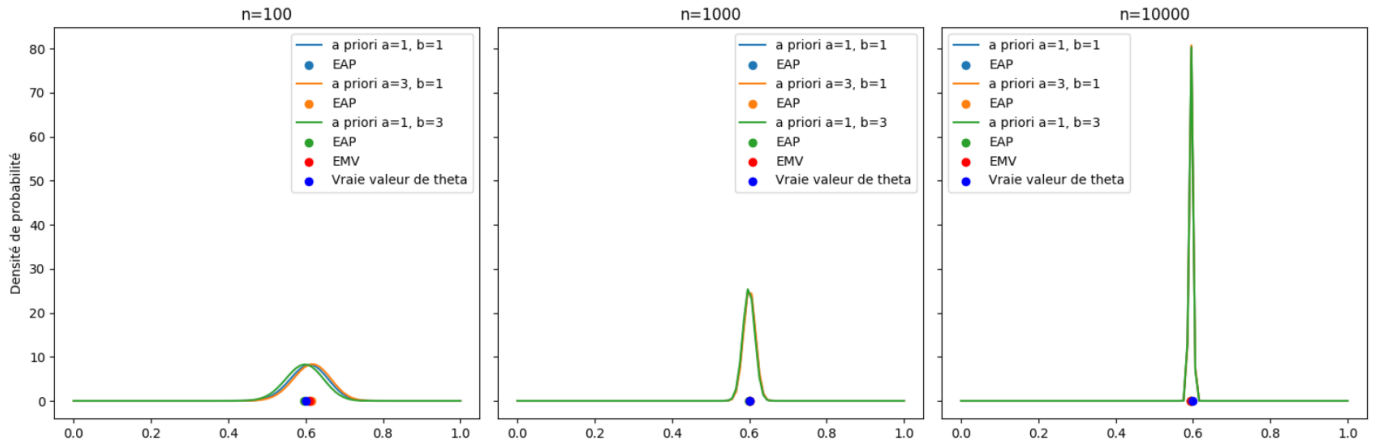


FIGURE 4 – Evolution du comportement asymptotique de la densité à postériori

On remarque que plus l'échantillon est grand, plus l'influence de l'*a priori* est non significative. De plus, la densité *a posteriori* se concentre sur la vraie valeur du paramètre θ .

5 Simulation de lois *a posteriori*

En pratique, la loi *a posteriori* n'est souvent pas une loi usuelle dont on sait facilement calculer des quantités comme la moyenne *a posteriori*, la médiane *a posteriori*, etc. Cependant, il existe des méthodes de simulation, nous permettant soit

- D'échantillonner des valeurs selon cette loi, c'est-à-dire de générer des données qui suivent cette distribution, même si celle-ci est difficile à caractériser directement. Nous verrons ici la **méthode de rejet**.
- D'approcher des intégrales nous permettant d'obtenir des valeurs approximatives des estimateurs. Nous verrons la **méthode de Monte Carlo**.

Ces deux approches présentent l'avantage de ne nécessiter que la connaissance de la densité *a posteriori*.

5.1 Méthode de rejet

L'idée principale de cette méthode est de dominer la densité de la loi que l'on cherche à simuler par la densité d'une autre loi que l'on sait simuler. On se place dans le cadre général où μ est une loi de densité f qu'on cherche à simuler. On se donne

- g une densité d'une loi que l'on sait simuler. On note ν cette loi.
- $(U_n)_{n \geq 1}$ une suite i.i.d de loi $\mathcal{U}([0, 1])$.
- $(Y_n)_{n \geq 1}$ une suite i.i.d de loi ν et indépendante de $(U_n)_{n \geq 1}$.

On suppose qu'il existe $c \geq 1$ tel que $f \leq cg$, et on définit

$$T := \{\inf k \geq 1 \mid f(Y_k)/g(U_k) > cU_k\}$$

Théorème 5.1 (Méthode de rejet). *La variable aléatoire Y_T suit la loi μ et admet donc pour densité f .*

Algorithme L'algorithme de rejet se résume donc comme suit :

1. Générer une observation Y_k selon la loi $g(y)$ (loi instrumentale).
2. Générer un nombre aléatoire U_k selon la loi uniforme sur l'intervalle $[0, 1]$.
3. Calculer le rapport d'acceptation $\frac{f(Y_k)}{cg(Y_k)}$.

4. Comparer $\frac{f(Y_k)}{cg(Y_k)}$ à U_k :
 - Si $\frac{f(Y_k)}{cg(Y_k)} > U_k$, accepter Y_k comme une observation de Y_T .
 - Sinon, rejeter Y_k et retourner à l'étape 1.
5. Répéter les étapes 1 à 4 jusqu'à obtenir le nombre souhaité d'observations.

Cet algorithme permet de générer des observations selon la densité cible $f(y)$, en utilisant une densité instrumentale $g(y)$ et en calculant le rapport d'acceptation $\frac{f(y)}{cg(y)}$ pour décider si une observation est acceptée ou rejetée. L'application de cet algorithme dans le cadre Bayésien peut être particulièrement simple.

Application de la méthode de rejet au cadre Bayésien

Dans le cadre bayésien, il est utile de considérer la loi *a priori* π comme loi instrumentale. En effet, pour trouver la constante de majoration c , on cherche à majorer le rapport $\frac{\pi(\theta|\underline{X})}{\pi(\theta)}$ pour tout θ . Or

$$\frac{\pi(\theta|\underline{X})}{\pi(\theta)} = \frac{f_X(\underline{X}|\theta)}{\int_{\Theta} f_X(\underline{X}|\theta) d\pi(\theta)} \leq c := \frac{f_X(\underline{X}|\theta = \theta^{MV})}{\int_{\Theta} f_X(\underline{X}|\theta) d\pi(\theta)}$$

. Pour appliquer la méthode de rejet, il suffit donc de simuler selon la loi *a priori* π et de calculer le rapport d'acceptation

$$\frac{\pi(\theta|\underline{X})}{c\pi(\theta)} = \frac{f_X(\underline{X}|\theta)}{f_X(\underline{X}|\theta = \hat{\theta}^{MV})}$$

L'intérêt de cette méthode est de ne pas faire appel à la densité marginale $f(\underline{X}) = \int_{\Theta} f_X(\underline{X}|\theta) d\pi(\theta)$, dont on ne connaît pas l'expression en pratique. L'inconvénient revient au fait qu'il faut pouvoir calculer facilement l'EMV.

Exemple 5.1. *Exemple Numérique Soit le modèle*

$$\underline{X} \sim \mathcal{N}(\theta, 1) \text{ avec } \theta \sim \mathcal{C}(\theta)$$

où \mathcal{C} est la loi de Cauchy. La loi *a posteriori* est

$$\pi(\theta|\underline{X}) \propto \frac{1}{\pi(1+\theta^2)} \exp\left(-\frac{n}{2}\theta^2 + \theta \sum_{i=1}^n X_i\right)$$

qui n'est pas une loi usuelle, et que l'on sait donc pas facilement simuler. Comme l'EMV de cet échantillon est $\hat{\theta}^{MV} = \bar{X}_n$, le rapport d'acceptation s'écrit

$$\frac{\pi(\theta|\underline{X})}{c\pi(\theta)} = \frac{f_X(\underline{X}|\theta)}{f_X(\underline{X}|\theta = \hat{\theta}^{MV})} = \frac{\exp\left\{-\frac{1}{2} \sum_{i=1}^n (X_i - \theta)^2\right\}}{\exp\left\{-\frac{1}{2} \sum_{i=1}^n (X_i - \bar{X}_n)^2\right\}} = \exp\left\{-\frac{n}{2}(\bar{X}_n - \theta)^2\right\}.$$

Pour obtenir une simulation selon la loi *a posteriori*, il suffit donc de simuler des couples *i.i.d.* $(\theta_k, U_k)_{k \geq 1}$, où θ_k est simulé selon la loi *a priori* (Cauchy) indépendamment de U_k (uniforme), uniquement si $U_k \leq \exp\left\{-\frac{n^2}{2}(\bar{X}_n - \theta_k)^2\right\}$.

```

import numpy as np
import matplotlib.pyplot as plt

# Génération de l'échantillon

theta0 = np.random.standard_cauchy()
ptheta = np.random.normal(loc=theta0, scale=1)
n = 100
X = np.random.normal(loc=theta0, scale=1, size=n)
EMV = np.mean(X)

# Algorithme de rejet
N = 500
echantillon = []
while len(echantillon) < N:
    # Générer un échantillon de theta selon la distribution prior
    theta = np.random.standard_cauchy()
    # Générer un seuil
    U = np.random.uniform()
    # Calculer le critère de rejet
    crit = np.exp(-n/2 * (EMV - theta)**2)
    if U <= crit:
        echantillon.append(theta)

echantillon = np.asarray(echantillon)

# Affichage des résultats
plt.hist(echantillon, label="Loi a posteriori")
plt.legend()
plt.show()

```

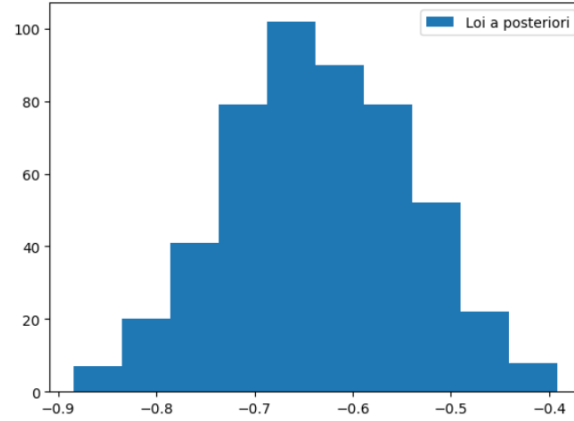


FIGURE 5 – Densité *a posteriori*

5.2 Méthode de Monte Carlo

Une fois la densité *a posteriori* simulée, on cherche à calculer les estimateurs en découlant. Le calcul d'estimateurs bayésiens revient souvent à des calculs d'intégrale. Par exemple, l'estimateur Bayésien classique (pour la perte quadratique) est donné par

$$\hat{\theta}_n^{bay} = \mathbb{E}[\theta|\underline{X}] = \int_{\Theta} \theta d\pi(\theta|\underline{X})$$

qui est souvent une intégrale que l'on ne sait pas calculer. La méthode de Monte Carlo nous fournit une méthode simple pour calculer cette intégrale en se basant seulement sur l'échantillon *a posteriori*.

Théorème 5.2. *Loi forte des Grands Nombres*

Soit X_1, \dots, X_n une suite de variables aléatoires i.i.d intégrables, alors,

$$\bar{X}_n \xrightarrow{p.s.} \mathbb{E}[X] \quad \text{quand } n \rightarrow \infty$$

Ainsi, pour simuler la moyenne *a posteriori*, il suffit de calculer la moyenne empirique de la distribution *a posteriori* obtenue après simulation. Dans l'exemple 5.1, on ajoute la ligne suivante

```
plt.plot(echantillon.mean(), 0, 'o', label="Moyenne a posteriori")
plt.plot(EMV,0,'o',color = 'red', label='EMV')
```

Et on obtient

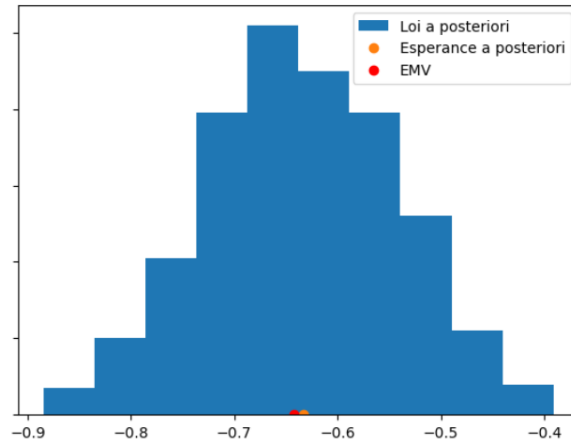


FIGURE 6 – Densité *a posteriori*

On voit bien que la simulation précédente est légitime puisque la moyenne *a posteriori* et l'EMV sont proches. Ceci illustre encore une fois le Théorème de Bernstein Von Mises.

6 Etude d'un jeu de données réel

Le jeu de données 'duree-de-vie.txt'² représente la durée de vie d'un échantillon de 1000 ampoules.

On commence par représenter la densité de l'échantillon comme suit

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from scipy.stats import expon

fig, axs = plt.subplots(1, 2, figsize=(15, 5), sharey=True)

# Chargement des données
df = pd.read_csv('duree-de-vie.txt', header=None)
data = df[0].tolist()

x = np.linspace(min(data), max(data), 1000)

axs[0].hist(data, bins=30, density=True, alpha=0.5, label='Données')
```

2. <https://www.math.sciences.univ-nantes.fr/~philippe/data/duree-de-vie.txt>

```

axs[0].legend()
axs[0].grid(True)

```

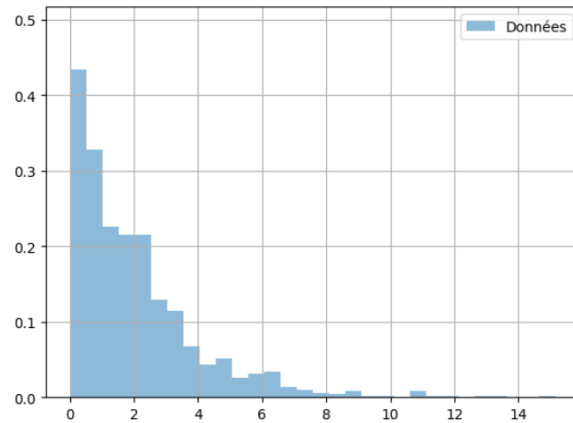


FIGURE 7 – Histogramme de la distribution des données (durées de vie des ampoules)

L'histogramme obtenu nous fait penser à une distribution exponentielle. On représente la densité de la loi exponentielle de paramètre adapté à notre distribution pour s'en convaincre.

```

loc, scale = expon.fit(data)
y = expon.pdf(x, loc=loc, scale=scale)

axs[1].hist(data, bins=30, density=True, alpha=0.5, label='Données')
plt.plot(x, y, label='Distribution exponentielle', color='red')
axs[1].legend()
axs[1].grid(True)

```

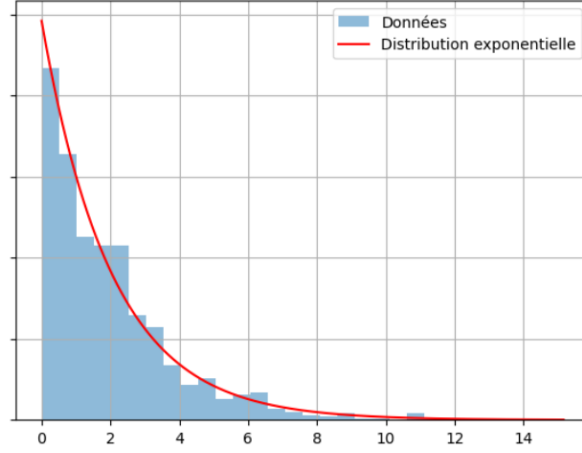



FIGURE 8 – Approximation de la distribution à une loi exponentielle

Ceci nous suggère le modèle suivant

$$\underline{X}|\theta \sim \mathcal{E}(\theta) \text{ avec } \theta \sim \Gamma(\alpha, \beta)$$

Le choix de l'*a priori* Gamma revient à sa nature de famille conjuguée. La vraisemblance est donc

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

La densité de $\pi(\theta)$ est donnée par :

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

Par suite,

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

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

On se propose d'étudier le comportement de cette loi *a posteriori* en prenant différentes valeurs de α et β pour l'*a priori*. L'EMV de ce modèle étant $\frac{1}{\bar{X}_n} = 0,49$. On choisit α et β tels que $\mathbb{E}[\theta] = \frac{\alpha}{\beta} = \frac{1}{2}$.

```
from scipy.stats import gamma

alphas = [0.5, 1, 2]
betas = [1, 2, 4]
n = 25
EMV = 1/xn_barre

fig, axs = plt.subplots(1, 3, figsize=(15, 5))

for i, (alpha, beta) in enumerate(zip(alphas, betas)):
    x = np.linspace(0, 5, 1000)

    a_priori = gamma(alpha, scale=1/beta)
    axs[i].plot(x, a_priori.pdf(x), 'r-', lw=2, label='Densité de la loi a priori')

    a_posteriori = gamma(alpha + n, scale=1/(beta + n * xn_barre))
    axs[i].plot(x, a_posteriori.pdf(x), color='green', lw=2, label='Densité de la loi
a posteriori')

    EAP = (alpha + n) / (beta + n * xn_barre)
    MedAP = gamma.ppf(0.5, alpha + n, scale=1/(beta + n * xn_barre))
    axs[i].set_title(f'A priori et a posteriori avec alpha={alpha} et beta={beta}')
    axs[i].scatter(EMV, 0, marker='o', color='blue', label=f'EMV={round(EMV,4)}')
    axs[i].scatter(EAP, 0, marker='^', color='black', label=f'EAP={round(EAP,4)}')
    axs[i].scatter(MedAP, 0, color='purple', marker='s', label=f'MedAP =
{round(MedAP,4)}')

    axs[i].legend()

plt.show()
```

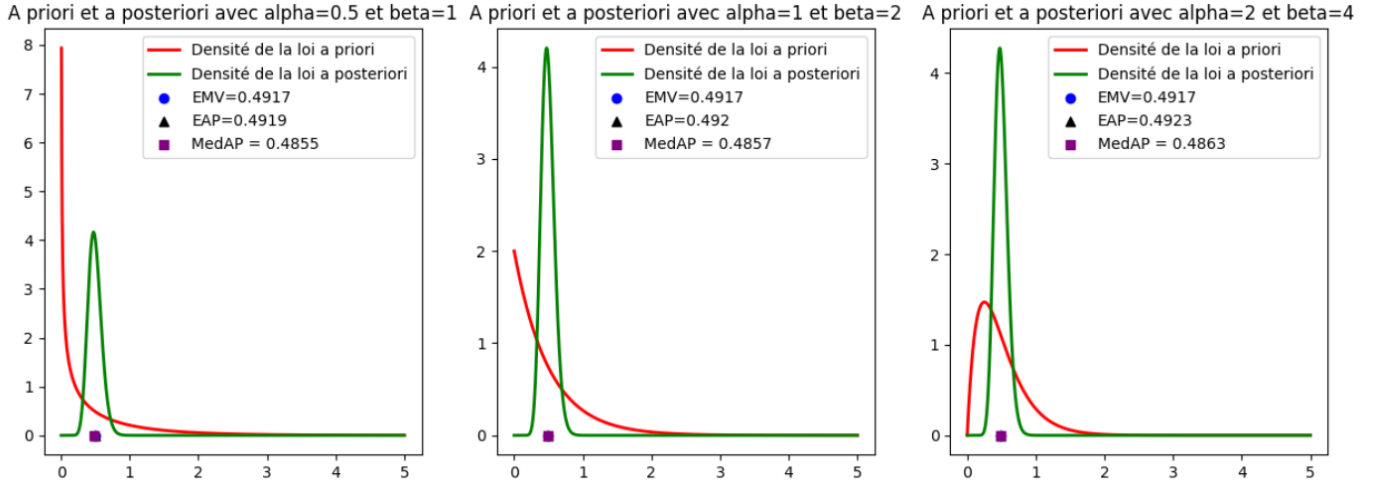


FIGURE 9 – Distributions *a priori* et *a posteriori* pour différentes valeurs de α et β , $n = 25$

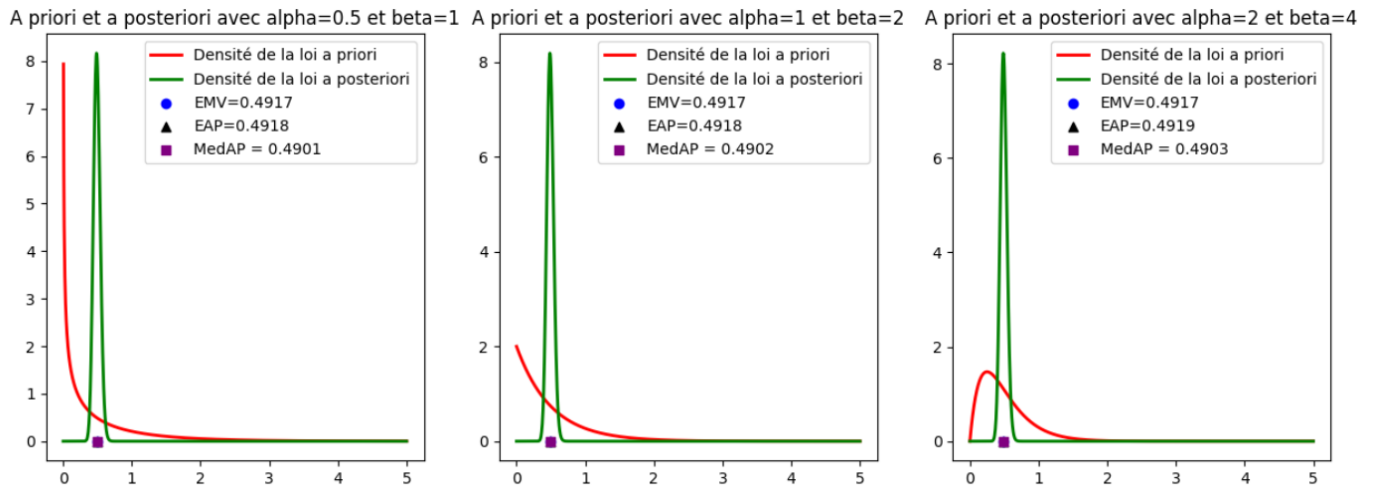


FIGURE 10 – Distributions *a priori* et *a posteriori* pour différentes valeurs de α et β , $n=100$

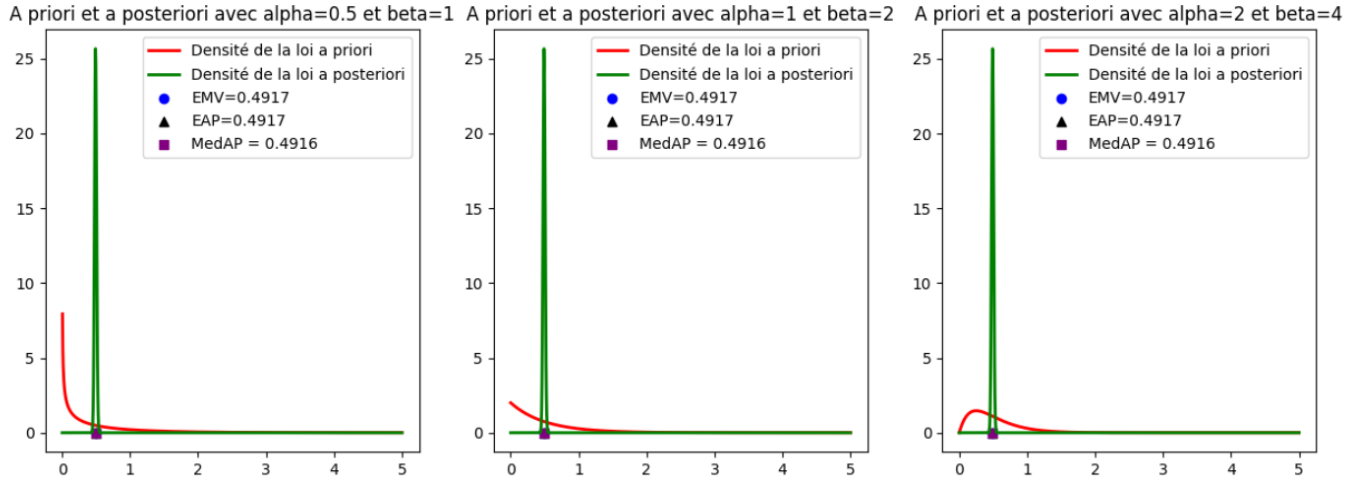


FIGURE 11 – Distributions *a priori* et *a posteriori* pour différentes valeurs de α et β , $n=1000$

En faisant varier la taille de l'échantillon n , on remarque que l'influence de l'*a priori* est de moins en moins significative (cf. Bernstein-von Mises). En effet, plus on augmente le n , plus les estimateurs bayésiens pour les risques \mathbb{L}^2 et \mathbb{L}^1 sont très proches de l'EMV.

Références

- [1] https://www.math.sciences.univ-nantes.fr/~rochet/enseignement/cours_bayes.pdf
- [2] https://pagespro.isae-supaero.fr/IMG/pdf/notes_de_cours-sb.pdf
- [3] <https://www.math.u-bordeaux.fr/~arichou/Bayes/StatBayesiennes.pdf>
- [4] https://borishejblum.science/files/STA305cours_etud.pdf
- [5] <https://www.math.univ-toulouse.fr/~dupuis/bayesmim08.pdf>
- [6] <http://web.univ-ubs.fr/lmba/gouno/BAYES/COURS/Cours3.pdf>
- [7] <https://perso.lpsm.paris/~aguyader/files/teaching/M1/PolycopiePartie2.pdf>