

Raisonnement probabiliste

Approche bayésienne *versus* fréquentiste

Marc Fischer

marc.fischer@emse.fr

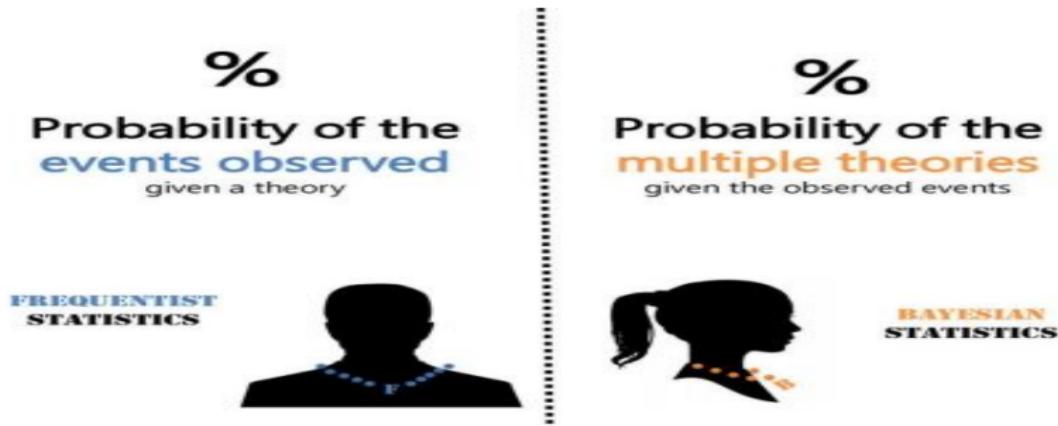
Xavier Bay

August 20, 2025

Contenu et objectifs du cours

Citation de E. T. Jaynes

"The necessity of reasoning as best we can in situations where our information is incomplete is faced by all of us, every waking hour of our lives." — E. T. Jaynes, Bayesian Methods: General Background, An Introductory Tutorial, 1996



Organisation du cours : 1h30 de cours + 1h30 de TP + 3h00 de présentations PowerPoint (**évaluation !**)

Probabilité conditionnelle

(Ω, \mathcal{F}, P) est un espace probabilisé (cadre axiomatique classique, *Foundations of the Theory of Probability*, A. Kolmogorov, 1933).

Définition

Probabilité conditionnelle : soit $B \in \mathcal{F}$ tel que $P(B) > 0$. Lorsque B est réalisé, la mesure P doit être mise à jour et remplacée par la nouvelle mesure de probabilité P_B , définie par :

$$P_B(A) = \frac{P(A \cap B)}{P(B)} = P(A | B), \quad \text{pour tout } A \in \mathcal{F}.$$

Exercice : Trouvez un exemple d'application lorsque deux pièces non biaisées sont lancées indépendamment !



Exemple d'application : Lancer de deux pièces non biaisées

Solution possible

Soit A l'événement "la première pièce montre face" et B l'événement "la seconde pièce montre face". Les résultats possibles lors du lancer de deux pièces non biaisées sont :

$$\{(F, F), (F, P), (P, F), (P, P)\}$$

où F représente "face" et P représente "pile".

Nous avons donc les probabilités suivantes :

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{2}, \quad P(A \cap B) = \frac{1}{4}.$$

La probabilité conditionnelle que A se produise sachant que B s'est produit est donnée par :

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.$$

Formules de Bayes



Thomas BAYES, 1702-1761, pasteur et mathématicien
Membre de la Royal Society en 1742

Formule de Bayes : *Essai sur la manière de résoudre un problème dans la doctrine des risques (Essay Towards Solving a Problem in the Doctrine of Chances - 1763)*

Formule déjà connue et redécouverte par Pierre-Simon de LAPLACE, 1749-1827 : loi de Bayes-Laplace

Aspect “révolutionnaire” de la formule :

$$P(H | O) = \frac{P(O | H) \times P(H)}{P(O)}$$

Probabilité a posteriori Vraisemblance Probabilité a priori

H = hypothèse (ou cause) O = observation « constante » de normalisation

Peut être déduite de la formule des probabilités conditionnelles!

Illustration : cas d'un test de dépistage d'une maladie

Illustration : cas d'un test de dépistage d'une maladie

Test de dépistage $\Rightarrow T^+ =$ "test positif" ou (exclusif) $T^- = \overline{T^+} =$ "test négatif"

⊗ Une personne nommée Julian fait le test et le résultat est « positif » : ici $\mathbf{O} = T^+$ et $\mathbf{H} = \mathbf{M}$ où $M =$ "Julian est atteint de la maladie "

Table d'analyse bayésienne				
	Loi a priori	Vraisemblance	Loi a posteriori non normalisée	Loi a posteriori
M	p	$P(T^+ M)$	$P(T^+ M) \times P(M)$	$\frac{P(T^+ M) \times P(M)}{P(T^+)}$
\bar{M}	$1 - p$	$P(T^+ \bar{M})$	$P(T^+ \bar{M}) \times P(\bar{M})$	$\frac{P(T^+ \bar{M}) \times P(\bar{M})}{P(T^+)}$

Petite application numérique

$$P(T^+ | M) = 90\% = P(T^- | \neg M) \quad \text{et} \quad p = 1\% \Rightarrow$$

Illustration : cas d'un test de dépistage d'une maladie

Solution

$$P(T^+ | M) = 90\% = P(T^- | \neg M) \quad \text{et} \quad p = 1\% \Rightarrow$$

$$P(T^+) = P(T^+ | M)P(M) + P(T^+ | \neg M)P(\neg M)$$

$$P(M | T^+) = \frac{1}{12} \quad (\text{donc moins de 1 chance sur 10 !})$$



Fréquentisme versus Bayésianisme

1. Soit A un événement.
2. Pour un fréquentiste, $p(A) = \lim_{n \rightarrow +\infty} f_n(A)$ est la **fréquence** de l'événement A sur un très grand nombre d'essais.
3. Pour un Bayésien, $p(A) \in [0, 1]$ représente le **degré de croyance** d'une personne en A .
4. A peut aussi être une proposition, par exemple : "*La mécanique des solides continus contredit la mécanique quantique*".
5. $p(A) = 0.99$: je suis presque sûr de A . $p(\bar{A}) = 0.01$: je suis presque sûr de \bar{A} .
6. $p(A) = 0.5$: il y a autant de raisons de penser que A est vrai que faux.
7. $p(A) \in [0.01, 0.99]$: je ne sais presque rien sur A .

L'importance des priors (distributions a priori)

1. Le calcul de $p(A)$ est souvent influencé par notre connaissance initiale, représentée par un **prior**.
2. Un prior est une distribution de probabilité qui reflète nos croyances avant l'observation des données.
3. Lorsque nous observons un nouvel événement B , la probabilité de A est mise à jour selon le **théorème de Bayes** :

$$p(A | B) = \frac{p(B | A) p(A)}{p(B)} = \frac{p(B | A) p(A)}{\sum_i p(B | A_i) p(A_i)}.$$

4. Cette mise à jour permet d'incorporer les nouvelles informations tout en tenant compte de notre croyance a priori sur A .
5. Par exemple, si $p(A) = 0.3$ (probabilité a priori de A) et que la probabilité d'observer B sachant A est $p(B | A) = 0.8$, alors après observation de B , la probabilité de A devient :

$$p(A | B) = \frac{0.8 \times 0.3}{p(B)}.$$

6. L'importance des priors réside dans leur capacité à influencer la probabilité a posteriori, en particulier lorsque les données sont peu informatives.

Le choix des priors peut être problématique

1. Choisir $p(A) = 0.5$ peut sembler raisonnable, mais c'est en réalité une hypothèse forte, suggérant une **symétrie parfaite** dans notre connaissance de l'événement A . Cela suppose que A est aussi probable que son complément \bar{A} , ce qui n'est pas toujours justifié, en particulier lorsqu'on manque d'informations.
2. Par exemple, lorsqu'on ne sait rien sur A , il est plus prudent de choisir un prior qui reflète notre incertitude, comme $p(A) \in [0.1, 0.9]$, plutôt que de fixer une probabilité exacte de 0.5. Cela permet d'exprimer le fait que nous avons peu de certitudes et que A pourrait être à la fois improbable et probable.
3. Le prior $p(A) = 0.5$ suppose une **connaissance parfaite ou une symétrie** que nous ne possédons pas nécessairement, ce qui pourrait fausser l'analyse, en particulier si les données sont limitées ou non informatives.
4. En utilisant un intervalle plus large comme $[0.1, 0.9]$, on exprime une plus grande incertitude, ce qui est souvent plus approprié dans des situations où l'on ne dispose pas de connaissances approfondies sur le phénomène étudié.
5. Le choix du prior a un impact direct sur les résultats de l'inférence bayésienne, surtout lorsque les données sont limitées. Il est donc crucial de réfléchir soigneusement au prior choisi.

Intervalle de confiance (statistique fréquentiste)

Considérons θ_v une grandeur physique « déterministe » inconnue
(par exemple une perméabilité k)

Observations (mesures) : $z_i = \theta_v + \varepsilon_i$; $1 \leq i \leq n$ avec les ε_i i.i.d. $N(0, \sigma)$, σ connu

Inférence classique : estimation ponctuelle par maximum de vraisemblance et intervalle de confiance à 95% donné par

$$IC_{95\%} = [\bar{z} - 1.96 \times \frac{\sigma}{\sqrt{n}} ; \bar{z} + 1.96 \times \frac{\sigma}{\sqrt{n}}]$$

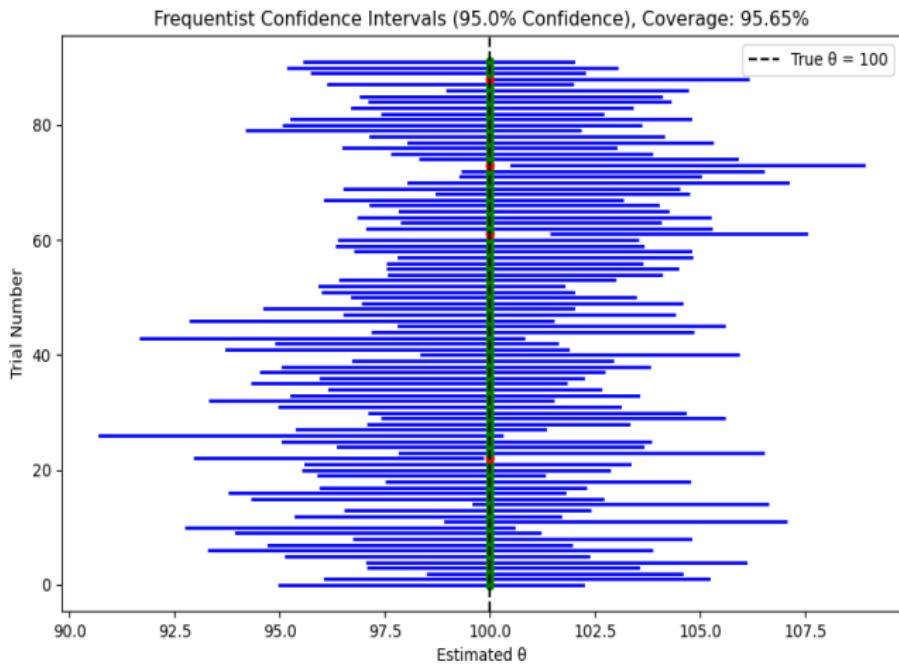
$$P(\theta_v \in IC_{95\%}) = 95\%$$

Interprétation « fréquentiste » : $P(\theta_v \in [\bar{Z} - 1.96 \times \frac{\sigma}{\sqrt{n}} ; \bar{Z} + 1.96 \times \frac{\sigma}{\sqrt{n}}]) = 95\%$

où \bar{Z} = estimateur usuel de la moyenne (et non pas l'estimation \bar{z} calculée sur la série de mesures).

Intervalle de confiance (Exemple)

La vraie valeur n'est pas toujours dans l'intervalle ALEATOIRE



Inférence Bayésienne

Inférence bayésienne : on dispose d'informations a priori sur le paramètre θ que l'on traduit sous la forme d'une loi de probabilité, par exemple une loi gaussienne :

$$P(\theta \in [a, b]) = \int_a^b \pi(t) dt$$

avec $\pi(t)$ densité de la loi $N(\theta_0, \tau_0)$ et θ_0 et τ_0 connus

\Rightarrow Loi a priori = loi normale $N(\theta_0, \tau_0)$

Vraisemblance : $L(t) = f_{Z|\theta=t}(z)$ densité jointe des $Z_i = \theta + \varepsilon_i$ si $\theta = t$

Densité de la loi a posteriori (« petite » extension de la formule de Bayes) :

Densité a posteriori
 $\pi(t | z)$

$$f_{\theta|Z=z}(t) \propto f_{Z|\theta=t}(z) \times \pi(t)$$

Vraisemblance $\pi(z | t)$

Densité a priori ou prior

Choix du prior

Résultat :

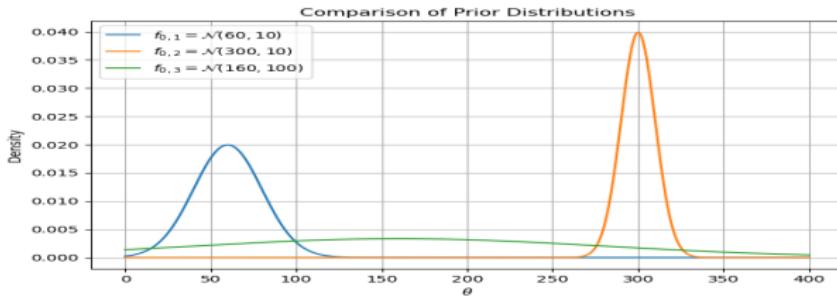
$$\text{Loi a posteriori} = \text{loi normale } N\left(\frac{\tau_0^2}{\sigma^2/n + \tau_0^2} \bar{z} + \frac{\sigma^2/n}{\sigma^2/n + \tau_0^2} \theta_0, \frac{(\sigma^2/n) \times \tau_0^2}{\sigma^2/n + \tau_0^2}\right)$$

Pour établir ce résultat, on utilise la **formule de Bayes**

$$f_{\theta|Z=z}(t) \propto \exp\left(-\frac{1}{2} q(t)\right)$$

avec $q(t)$ forme quadratique qu'il suffit de réduire...

- ▶ Souvent, on n'est pas capable d'exprimer notre ignorance initiale à travers un seul prior.
- ▶ Il est souvent préférable de considérer **plusieurs** priors afin d'obtenir plusieurs postérieurs.



Actualisation d'un prior uniforme

1. Le prior uniforme est supposé représenter notre ignorance en ne favorisant aucune valeur.
2. Pour en obtenir un, il suffit de considérer $f_0 = \mathcal{N}(\theta_0, \tau_0 \rightarrow \infty)$.
3. Il s'agit d'un prior **diffus** (improper prior) dont l'intégrale n'est pas égale à 1.
4. Si $\tau_0 \rightarrow +\infty$, alors la loi a posteriori est donnée par :

$$P(\theta|z) = \mathcal{N}(z, \sigma^2).$$

5. L'**intervalle crédible** est alors le même que l'intervalle de confiance:

$$IC_{95\%}(\theta) = \left[z - 1.96 \times \frac{\sigma}{\sqrt{n}} ; z + 1.96 \times \frac{\sigma}{\sqrt{n}} \right].$$

6. Notre **degré de croyance** que θ appartient à $IC_{95\%}(\theta)$ est égal à 0.95!

Théorème de convergence des priors

1. Rappel: expression du posterior pour tout prior normal.

$$\text{Loi a posteriori} = \text{loi normale } N\left(\frac{\tau_0^2}{\sigma^2/n + \tau_0^2} \bar{z} + \frac{\sigma^2/n}{\sigma^2/n + \tau_0^2} \theta_0, \frac{(\sigma^2/n) \times \tau_0^2}{\sigma^2/n + \tau_0^2}\right)$$

2. Que se passe t-il quand la taille de l'échantillon n devient de plus en plus grande ?

Théorème de convergence des priors

$$\text{Loi a posteriori} = \text{loi normale } N\left(\frac{\tau_0^2}{\sigma^2/n + \tau_0^2} \bar{z} + \frac{\sigma^2/n}{\sigma^2/n + \tau_0^2} \theta_0, \frac{(\sigma^2/n) \times \tau_0^2}{\sigma^2/n + \tau_0^2}\right)$$

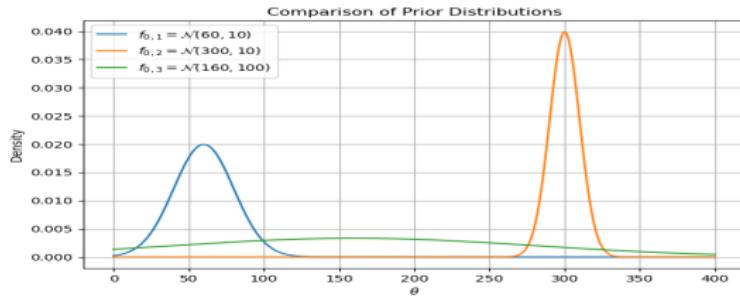
- Quand $n \rightarrow \infty$, le posterior est donné par

$$f_{n \gg 1} = \mathcal{N}\left(\bar{z}, \frac{\sigma^2}{n}\right).$$

- C'est la même expression que pour le posterior du prior uniforme!
- Quand $n \rightarrow +\infty$, les posteriors convergent vers une même distribution et finalement vers δ_{θ_V} .
- Qu'est-ce qu'est δ_{θ_V} ?

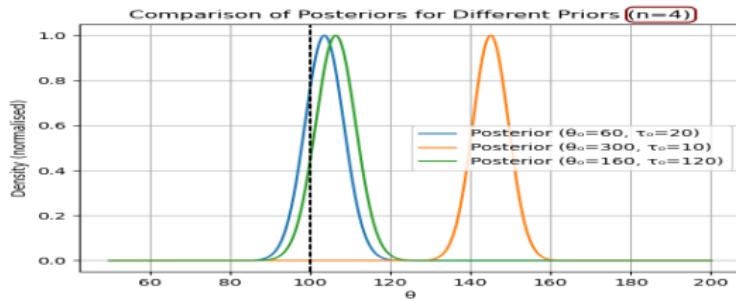
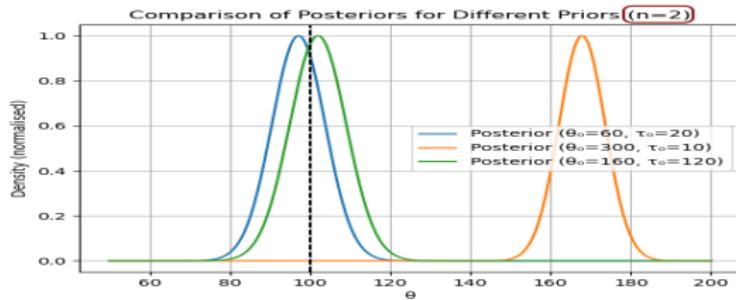
Théorème de convergence des priors

- ▶ Il s'agit de la distribution de Dirac de θ singulière en $\theta = \theta_V$.
- ▶ Nous allons maintenant comparer la convergence des priors que nous avons définis avant.



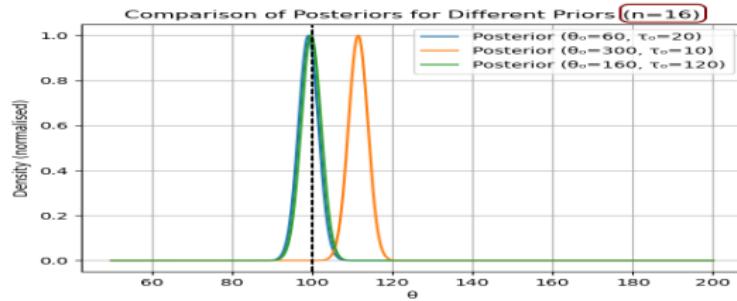
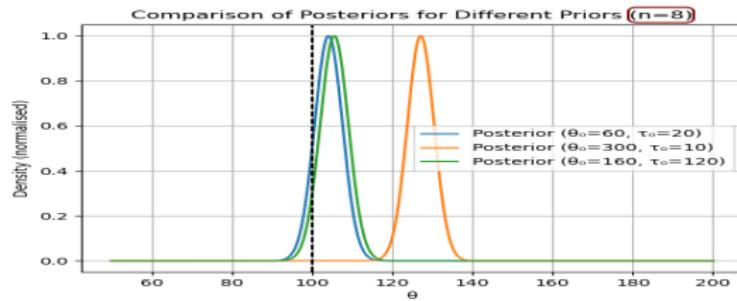
Théorème de convergence des priors

- Voici les résultats pour différentes tailles d'échantillons n .



Théorème de convergence des priors

- On observe une convergence vers la vraie valeur $\theta_V = 100$.



Conclusion

- ▶ La méthode bayésienne repose sur une philosophie des probabilités différente du fréquentisme.
- ▶ Elle permet d'incorporer facilement ses connaissances a priori sous la forme d'un prior.
- ▶ Dans de nombreuses situations, le choix du prior n'est pas évident.
- ▶ Dans ce cas, il est préférable de considérer plusieurs priors mis à jour avec nos données.
- ▶ Si les postérieurs sont similaires, cela signifie que l'on dispose d'assez de données.
- ▶ Si les postérieurs sont trop différents, cela signifie que l'on a besoin de plus de données.

Évaluation

- ▶ TD/TP non noté.
- ▶ Préparation pour l'examen!