# La statistique bayésienne(1)

Jean-Louis Piednoir

La statistique enseignée dans les collèges et les lycées est d'abord une statistique descriptive. On cherche quelques indicateurs pour résumer un grand nombre de données. Depuis la rentrée 2000 le programme de seconde aborde une autre statistique, dite inductive, en lien avec le calcul des probabilités. Comme ce dernier n'est pas au programme, on passe par l'observation des fluctuations d'échantillonnage. Cette statistique inductive fait partie des programmes de BTS. Elle est fondée sur une certaine vision de la probabilité, elle est actuellement dominante dans les applications. Mais il existe une autre interprétation de la probabilité qui engendre d'autres techniques de statistique inductive que celles de la statistique classique. L'ensemble porte le nom de statistique bayésienne. Son usage rencontre à la fois enthousiasme et scepticisme selon les acteurs. On présentera d'abord différentes interprétations de la probabilité, puis sur un exemple ce qu'est une procédure bayésienne pour terminer par quelques problèmes et quelques applications.

## I. Qu'est ce que la probabilité ?

La légende, répandue par Leibniz, raconte que Blaise Pascal a forgé le calcul des probabilités à partir d'une remarque faite par un joueur invétéré, le Chevalier de Méré. Ce dernier avait remarqué qu'au jeu de dés une configuration sortait plus souvent qu'une autre et qu'il avait intérêt à parier sur elle. En dénombrant les configurations de base et celles qui donnent un résultat favorable dans l'un ou l'autre cas, on trouve l'explication : les configurations du cas expérimentalement le plus favorable sont les plus nombreuses. Avant lui Galilée s'était intéressé à un problème analogue dit du Grand Duc de Toscane. D'une façon implicite on suppose que les chances des configurations de base sont égales pour employer le langage de l'époque. Il reviendra à Jacques Bernoulli(2) de préciser l'explication donnée en démontrant la loi des grands nombres dont un énoncé intuitif est du type : si je lance un grand nombre de fois un dé, il est très peu probable que la proportion observée d'une face donnée s'écarte beaucoup de 1/6. La probabilité apparaît alors comme une propriété du phénomène étudié et que des répétitions, évidemment indépendantes, du phénomène et nombreuses permettent de lui attribuer une valeur. On a ainsi une conception dite objective de la probabilité car indépendante de l'observateur.

Cette conception de la probabilité sous-tend les programmes de probabilité et de statistique de l'enseignement secondaire. L'introduction de la simulation a, en particulier, pour objectif de mettre en évidence la loi des grands nombres puis de s'en servir pour déterminer expérimentalement la valeur d'une probabilité inconnue. On

(1) Thomas Bayes (1702-1761) mathématicien anglais.
(2) Jacques Bernoulli (1637-1705) mathématicien suisse.

peut citer les travaux pratiques, avec l'exemple fameux de la stratégie des familles pour avoir un garçon parmi leurs enfants, en classe de seconde, les programmes de terminale ES et S pour l'exécution d'un test d'adéquation à une distribution théorique équirépartie (cf. sujet national ou sujet de Pondichéry du baccalauréat ES 2003).

Pour certains cette conception de la probabilité est à la fois mal assurée logiquement et trop étroite. Mal assurée logiquement car pour définir la probabilité à partir de la loi des grands nombres il faut parler de probabilité. En effet, son énoncé exact est le suivant : « soit une épreuve aléatoire donnant 1 avec la probabilité $p$ et 0 avec la probabilité $(1 - p)$, $n$ répétitions indépendantes de cette épreuve et X le nombre de 1 obtenable. Alors si P désigne la probabilité, le théorème de Bernoulli affirme : $\forall \varepsilon > 0, \mathrm{P}\left[\left|\frac{\mathrm{X}_n}{n} - p\right| \geq \varepsilon\right] \underset{n\to\infty}{\to} 0$. On voit bien qu'il y a un cercle vicieux logique. Bernoulli l'avait d'ailleurs flairé et il se lance dans un commentaire pour le lever. Au XX$^{\text{e}}$ siècle des mathématiciens comme Von Mises, Wald ont proposé des axiomatiques reflétant la conception objective de la probabilité. Elles ont été critiquées par Bruno de Finetti dans un article célèbre de 1936 où il fonde un autre point de vue sur la probabilité.

La conception objective paraît trop étroite surtout dans un univers où il faut décider. Soit l'histoire suivante : vous jouez au dé, celui-ci roule sur la table mais son résultat est caché à votre vue par une couverture. Un compère passant par-là n'a pas pu lire exactement le numéro, il vous glisse à l'oreille : « je crois bien que la face qui est sortie porte un numéro inférieur ou égal à 3 ». Vous savez, par ailleurs, que le dé ainsi que le lanceur sont honnêtes. Objectivement toutes les faces ont la même probabilité d'apparaître. Pourtant l'information qui vous a été donnée est à intégrer dans votre pari. Il est tentant d'attribuer aux numéros 1, 2 et 3 une probabilité plus grande qu'aux numéros 4, 5 et 6. On quitte les bases objectives, les modifications que vous êtes tentés de faire aux probabilités 1/6 de départ pour chaque face risque de tenir compte de divers paramètres dont ce que vous savez de votre informateur. Les nouvelles probabilités attribuées sont dites subjectives. La probabilité n'est plus une qualité du phénomène étudié, elle devient une opinion sur les choses, opinion que l'on peut modifier selon des règles rationnelles (cf. ci-dessous). Cette conception de la probabilité permet de donner un sens à des expressions du type : « il est probable que Vercingétorix a été exécuté à Rome dans la prison Mamertime en 45 avant J.-C. ». Comme il n'y a aucune répétition possible, d'un point de vue objectif, cette phrase n'a rigoureusement aucun sens.

## II. Le modèle probabiliste et ses applications

Quelle que soit l'interprétation épistémologique de la probabilité, le modèle mathématique reste le même, il ignore la polémique. Il a été mis au point par Kolmogorov dans les années trente et est caractérisé par les éléments suivants :

– un ensemble Ω de tous les cas possibles appelé ensemble fondamental ;

– un sous-ensemble $\mathcal{A}$ de parties de $\Omega$ ($\mathcal{A} \subset \mathcal{P}(\Omega)$) stable par passage au complémentaire et union dénombrable appelé tribu des événements et tel que $\varnothing$ et $\Omega$ appartiennent à $\mathcal{A}$ ;
– une application $P : \mathcal{A} \to [0 ; 1]$ telle que :
  (i) $P(\Omega) = 1$ ;
  (ii) $P(A \cup B) = P(A) + P(B)$ si $A \cap B = \varnothing$ ;
  (iii) Si $(A_n)_{n\in\mathbf{N}}$ est une suite d'événements emboîtés $(\forall n, A_n \subset A_{n+1})$,

$$P\left(\bigcup_n A_n\right) = \lim_{n\to\infty} P(A_n).$$

– une définition de la probabilité conditionnelle : si A et B sont deux événements avec $P(B) \neq 0$,on définit la probabilité de A sachant B notée $P_B(A)$ ou bien $P(A/B)$ comme étant égale à $\dfrac{P(A\cap B)}{P(B)}$.

Ainsi définie la probabilité est une mesure finie et le calcul des probabilités devient un chapitre de la théorie de la mesure et tenants de l'interprétation objective ou tenants de l'interprétation en terme d'opinion développent le même modèle et reconnaissent les mêmes théorèmes. Parmi ceux-ci deux d'entre eux auront une importance particulière. Le premier a déjà été mentionné, il s'agit de la loi des grands nombres. On en donne un énoncé plus complet :

Si $X_1, \ldots, X_n$ sont $n$ variables aléatoires indépendantes et de même loi telle que $E(X_i)$ existe, alors : $\forall \varepsilon > 0, P\left[\left|\dfrac{1}{n}\sum_{i=1}^{n} X_i - E(X)\right| \geq \varepsilon\right] \underset{n\to\infty}{\to} 0$. Il s'agit de la loi faible des grands nombres.

Pour introduire le second théorème, on utilisera l'exemple qu'Henri Poincaré proposait et qui fut repris par Émile Borel(3) et Georges Darmois. Considérons le jeu de l'écarté : on tire une carte dans un jeu de 32 cartes, vous avez gagné si vous tirez le roi de cœur. Un individu arrive, il tire le roi de cœur et vous vous posez la question « est-il un tricheur ? ». Il vous faut modéliser la situation : $t$ symbolise le tricheur, $\bar{t}$ le joueur honnête, $r$ tirer le roi de cœur, $\bar{r}$ tirer une autre carte.

L'espace fondamental est : $\Omega = \{t, \bar{t}\} \times \{r, \bar{r}\}$, $\mathcal{A} = \mathcal{P}(\Omega)$.

T est l'événement « être un tricheur » : $T = \{t\} \times \{r, \bar{r}\}$ ;

$\overline{T}$ est l'événement « être un joueur honnête » : $\overline{T} = \{\bar{t}\} \times \{r, \bar{r}\}$ ;

R est l'événement « tirer le roi de cœur » : $R = \{t, \bar{t}\} \times \{r\}$.

On a évidemment $P_{\bar{T}}(R) = \dfrac{1}{32}$. Supposons que le tricheur réussisse son coup deux fois sur trois, alors $P_T(R) = \dfrac{2}{3}$. Posons $p = P(T)$. Que voulons-nous connaître : la

(3) Émile Borel (1871-1956) mathématicien français.

probabilité d'être un tricheur sachant que l'on a tiré le roi de cœur, donc la quantité P(T/R).

Par définition : $P_R(T) = \dfrac{P(T \cap R)}{P(R)}$ et $P(T \cap R) = P_T(R) \times P(T)$.

De même $P\left(\overline{T} \cap R\right) = P_{\overline{T}}(R) \times P\left(\overline{T}\right)$.

On remarque que : $R = \left(T \times R\right) \cup \left(\overline{T} \times R\right)$ et que $\left(T \times R\right)$ et $\left(\overline{T} \times R\right)$ sont incompatibles. Il vient $P_R(T) = \dfrac{\dfrac{2}{3}p}{\dfrac{2}{3}p + \dfrac{1}{32}(1-p)}$ soit $P_R(T) = \dfrac{64p}{61p+3}$.

On introduira un autre langage, on peut dire que le fait de tirer le roi de cœur avait deux causes : joueur honnête ou tricheur et on cherche la probabilité des causes quand on sait l'événement réalisé.

Plus généralement, on a le théorème suivant appelé **Théorème de Bayes** :
Soit $(A_1, A_2, \ldots, A_k)$ une partition de $\Omega$ en $k$ événements incompatibles, E un événement. Alors :

$$P_E\left(A_j\right) = \frac{P_{A_j}(E) \times P\left(A_j\right)}{\sum_{j=1}^{k} P_{A_j}(E) \times P\left(A_j\right)}.$$

L'interprétation possible est claire : E étant dû aux « causes » $A_1, \ldots, A_k$, quelle est la probabilité de chaque cause quand E est réalisé ? Il s'agit d'un problème inverse déjà exploité par Simon Laplace au début du XIX$^e$ siècle et qui est le paradigme des subjectivistes. $P\left(A_j\right)$ quantifie mon opinion a priori sur $A_j$, j'observe E, je tiens compte de cette information pour modifier cette opinion et j'obtiens une opinion a posteriori $P\left(A_j / E\right)$. Il semble toutefois dans ce schéma que les $P\left(E / A_j\right)$ n'aient pas tout à fait le même statut d'opinion, ils ont un relent de probabilités objectives !

Si l'étude mathématique du modèle probabiliste peut ignorer les interprétations épistémologiques, celles-ci resurgissent dès lors qu'il faut savoir quels sont les phénomènes susceptibles d'être représentés par un modèle probabiliste et surtout quand il faut attribuer des valeurs aux paramètres inconnus c'est-à-dire quand on fait de la statistique inductive.

## III. Le problème statistique

### 1. Un exemple

On illustrera le problème statistique et les diverses manières de l'aborder par un exemple simple. À la veille de l'ouverture de la chasse un commerçant se fait livrer

une caisse de cartouches pour ensuite les débiter à ses clients. Pour son calcul de rentabilité, pour savoir s'il doit accepter la caisse livrée ou au contraire la retourner à son fournisseur, il a besoin de connaître la proportion $\theta$ de mauvaises cartouches dans la caisse. Il y a évidemment un moyen de connaître exactement $\theta$ : c'est d'essayer toutes les cartouches. Évidemment, le procédé est inapplicable quand il s'agit comme ici d'essais destructifs. S'il enquêtait sur le mode de stockage de la caisse, peut-être aurait-il des informations sur la géométrie des mauvaises cartouches dans la caisse : sur le dessus du fait d'un arrosage intempestif ? au fond car séjour prolongé sur un sol humide ? Impossible de la savoir. On va rendre aléatoire les paramètres inconnus afin de les neutraliser. Cette opération s'appelle randomisation. On va considérer la caisse comme une urne et on va prélever $n$ cartouches parmi les N de la caisse, $n$ très petit devant N, sur le mode aléatoire, chaque cartouche ayant la même probabilité de figurer dans l'échantillon. On essaie successivement les $n$ cartouches choisies au hasard, on note 1 quand elle est mauvaise, 0 quand elle est bonne. On a donc le schéma suivant : $\Omega = \{0, 1\}^n$, $\mathcal{A} = \mathcal{P}(\Omega)$.
Si $n \ll$ N, on peut considérer les tirages comme indépendants et trouver en fonction de $\theta$, proportion inconnue de mauvaises cartouches, la loi de probabilité sur $\Omega$. On définit les variables $X_1,\ldots, X_n$ telles que :

$X_i = 1$ si la $i$-ème est mauvaise
$X_i = 0$ sinon et

$$K = \sum_{i=1}^{n} X_i.$$

$K(\omega)$ sera donc le nombre de mauvaises cartouches dans l'échantillon $\omega$ et, du fait des hypothèses précédentes, $P(\{\omega\}) = \theta^{K(\omega)}(1-\theta)^{n-K(\omega)}$.
P sera noté $P_\theta$ car il dépend du paramètre inconnu $\theta$. L'objet de la statistique est de répondre à des questions du genre :

- Quelle valeur attribuer à $\theta$ ? (estimation ponctuelle),
- Dans quel intervalle se trouve $\theta$ ? (estimation par intervalle),
- A-t-on $\theta \leq 0{,}1$ ou $\theta > 0{,}1$ ? (problème de test) si on suppose que 0,1 est le seuil de rentabilité du détaillant.

On va voir que les tenants de la probabilité objective appelés objectivistes ou fréquentistes ne mettent pas en action les mêmes méthodes que les tenants de la probabilité subjective appelés subjectivistes ou bayésiens.

**2. Les méthodes fréquentistes**

Un fréquentiste va définir pour répondre à la première question un estimateur de $\theta$ noté $\hat{\theta}$ qui sera une application de $\Omega$ dans [0,1] ( $\hat{\theta} : \Omega \to [0,1]$). À tout constat expérimental on affecte une valeur de $\theta$. Si le dit constat est $\omega$, $\hat{\theta}(\omega)$ sera appelé estimation de $\theta$. Bien entendu $\hat{\theta}$ doit avoir certaines qualités. Dans le cas qui nous occupe, on prendra comme estimateur la fréquence observée : $\hat{\theta} = \frac{1}{n}K = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Si $\theta$ est l'état de la nature, les théorèmes du calcul des probabilités nous garantissent que :

$\mathrm{E}_\theta(\hat{\theta}) = \theta$ : l'estimateur est dit sans biais.

$\mathrm{Var}_\theta(\hat{\theta}) = \dfrac{\theta(1-\theta)}{n}$ : c'est la plus petite valeur possible pour les estimateur sans biais.

$\forall\, \varepsilon > 0, \mathrm{P}_\theta\left[\left|\hat{\theta} - \theta\right| > \varepsilon\right] \underset{n\to\infty}{\to} 0$ d'après la loi des grands nombres : l'estimateur est dit convergent.

De plus, si on pose $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x$ fonction de répartition de la loi gaussienne on a si $n$ n'est pas trop petit $\mathrm{P}_\theta\left[\sqrt{n}\,\dfrac{\hat{\theta}-\theta}{\sqrt{\theta(1-\theta)}} \leq x\right] \cong \Phi(x)$.

Et comme $\theta(1-\theta) \leq \dfrac{1}{4}$, on peut écrire $\mathrm{P}_\theta\left[\hat{\theta} - \dfrac{t_p}{2\sqrt{n}} \leq \theta \leq \hat{\theta} + \dfrac{t_p}{2\sqrt{n}}\right] \geq 1 - 2p$

avec $1 - \Phi\left(t_p\right) = \alpha$ et $p \in ]0,1[$.

Les interprétations en langage courant sont les suivantes :

(i) Si $n$ est assez grand, il y a de fortes chances pour que l'estimation ne soit pas trop loin de la quantité à estimer.

(ii) C'est le meilleur estimateur parmi les sans biais.

(iii) On peut construire un intervalle de confiance. Avec une confiance supérieure à $1 - 2p$ où $p$ est fixé (0,01 ou 0,025 ou 0,05) on dit que :

$$\theta \in \left[\hat{\theta}(\omega) - \frac{t_p}{2\sqrt{n}}, \hat{\theta}(\omega) + \frac{t_p}{2\sqrt{n}}\right].$$

Les calculs précédents supposent que $n$ est très petit devant N mais assez grand pour que les approximations précédentes soient valides. Si $n \geq 100$ et $\mathrm{N} \geq 1\,000$, les approximations précédentes sont très satisfaisantes.

Attention, pour un fréquentiste, confiance n'est pas probabilité. On ne peut pas dire que $\theta$ est dans l'intervalle, par exemple [0,09 ; 0,21] (avec $n$ = 100 ; $\mathrm{K}(\omega) = 15$ ; $p = 0{,}2$). Cela n'a pas de sens, $\theta$ n'est pas aléatoire, il est inconnu ce qui n'est pas la même chose. Le mot confiance est là pour rappeler que la procédure utilisée pour trouver l'intervalle avait une probabilité de $1 - 2\alpha$ de recouvrir la vraie valeur de $\theta$. En effet la probabilité n'est pas dans le résultat, mais dans l'action : quand un dé roule, il y a une probabilité $\dfrac{1}{6}$ d'avoir l'as, quand il s'arrête il y a un constat qui est sûr, il n'y a donc plus de probabilité.

### 3. La méthode bayésienne

Un bayésien voudra donner un sens à l'expression $\mathrm{P}(\theta \in [0{,}09\,;0{,}21])$, ce sera une opinion a posteriori sur la proportion de mauvaises cartouches dans la caisse qu'il faudra construire. Il s'interrogera sur les connaissances que l'on peut avoir sur le fournisseur : qualité des fournitures précédentes, de celles faites à d'autres détaillants, réputation d'honnêteté, etc. À partir de ces informations on va se forger une opinion a priori sur $\theta \in [0, 1]$ : il y a peu de chances que $\theta \in 0{,}5$ par exemple. Le bayésien va tenter de traduire cette opinion a priori par une distribution de probabilité sur [0,1]. Le paramètre $\theta$ considéré comme inconnu par le fréquentiste devient un élément d'un espace probabilisé. Reste à choisir la loi de probabilité censée représenter l'opinion a priori. Pour des raisons de faisabilité de traitement mathématique, on prendra comme loi de probabilité sur [0,1], une loi bêta dont la densité de probabilité notée $\Pi$ dépend de deux paramètres $\alpha$ et $\beta$ et s'écrit :

$$\Pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},\ \alpha>0,\ \beta>0 \text{ avec } \Gamma(t) = \int_0^{\infty} \mathrm{e}^{-x}\, x^{t-1}\,\mathrm{d}\,x.$$

On rappelle que, si $n$ est entier, $\Gamma(n) = (n-1)!$ et $\Gamma(t+1) = t\Gamma(t)$.

Des calculs simples permettent de calculer la moyenne $\mu$ et la variance $\sigma^2$ de cette loi : $\mu = \dfrac{\alpha}{\alpha+\beta}$ ; $\sigma^2 = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Si $\alpha > 1$ et $\beta > 1$, la loi bêta est unimodale et le mode en est $m = \dfrac{\alpha-1}{\alpha+\beta-2}$.

Ces divers indicateurs permettent de choisir $\alpha$ et $\beta$ selon l'opinion que l'on a sur la valeur de $\theta$. Par exemple, si on suppute que $\theta$ s'écarte peu d'une valeur $a$ on choisira $\alpha$ et $\beta$ grands avec $a = \dfrac{\alpha-1}{\alpha+\beta-2}$. En effet, on voit que $\sigma^2$ est faible donc la distribution de probabilité sera concentrée autour de son mode $a$.

Par exemple :
Si $a = 0{,}1$, $\alpha = 3$, $\beta = 19$, alors il y a une probabilité a priori de 0,82 pour que $\theta$ soit plus petit que 0,20.
Si $a = 0{,}1$, $\alpha = 2$, $\beta = 10$, alors il y a une probabilité a priori de 0,68 pour que $\theta$ soit plus petit que 0,20.

De même $\alpha < 1$ et $\beta$ grand traduit une excellente opinion sur la qualité de la marchandise livrée. Si $\alpha = \beta = 1$, la loi a priori est la loi uniforme qui traduit une indifférence mais pas forcément une absence d'information (cf. ci-dessous).

Une fois choisie la loi a priori, il faut la combiner avec l'observation faite $\mathrm{K}(\omega) = k$ selon la règle de Bayes pour obtenir la loi a posteriori tenant compte de l'observation et traduisant la modification d'opinion induite par l'observation. Quand la loi a priori est une loi bêta de paramètres $\alpha$ et $\beta$, alors la loi a posteriori est encore une loi $\beta$ de paramètres $\alpha + k$, $\beta + n - k$ (voir démonstration en Annexe).

Ses caractéristiques sont :

$$\text{moyenne } \mu^* = \frac{k+\alpha}{n+\alpha+\beta} \ ;$$

$$\text{variance } \sigma^{*2} = \frac{(k+\alpha)(n-k+\beta)}{(n+\alpha+\beta)^2(n+\alpha+\beta+1)}.$$

On voit qu'elle est plus concentrée que la loi a priori ( $\sigma^{*2} < \sigma^2$ ) et que $\sigma^{*2} = \mathrm{O}\left(\frac{1}{x}\right)$ quand $n \to \infty$.

Le bayésien prendra comme estimateur de $\theta$ la moyenne $\mu^*$. En appelant $\hat{\theta}_b$ l'estimateur bayésien de $\theta$ et en posant $\mathrm{K}(\omega) = k$, on a : $\hat{\theta}_b(\omega) = \frac{k+\alpha}{n+\alpha+\beta}$.

C'est un estimateur convergent : $\forall \varepsilon > 0, \mathrm{P}_\theta\left[\left|\frac{\mathrm{K}+\alpha}{n+\alpha+\beta} - \theta\right| > \varepsilon\right] \underset{n\to\infty}{\to} 0.$

Il n'est pas sans biais : $\mathrm{E}_\theta\left(\hat{\theta}_b\right) = \frac{n\theta+\alpha}{n+\alpha+\beta}$ ; sa variance est $\mathrm{Var}_\theta\left(\hat{\theta}_b\right) = \frac{n\theta(1-\theta)}{(n+\alpha+\beta)^2}$.

Pour trouver un intervalle de confiance il faut trouver, en consultant une table de la loi bêta ou par des méthodes d'analyse numérique, deux valeurs $\theta_1$ et $\theta_2$ telles que :

$$\frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha)\Gamma(n+k+\beta)} \int_{\theta_1}^{\theta_2} \theta^{k+\alpha+1}(1-\theta)^{n-k+\beta-1}\,\mathrm{d}\theta = 1-p \text{ , } (\theta_2 - \theta_1 \text{ minimum})$$

où $p$ est le risque pris. Il est alors licite de dire : $\mathrm{P}\left[\theta_1 \le \theta \le \theta_2\right] = 1-p$ . L'intervalle $\left[\theta_1, \theta_2\right]$ est appelé intervalle de confiance bayésien.

L'examen de la formule donnant l'estimateur bayésien montre que, si $n$ est grand et donc probablement $k$ et si $\theta$ n'est pas trop proche de 0, l'influence de la loi a priori est faible : cela rassure ceux dont les convictions bayésiennes sont chancelantes. De plus, la loi a priori intervient d'autant moins que $\alpha$ et $\beta$ sont petits.

On remarque que si $\alpha = \beta = 0$, l'estimateur bayésien et celui des fréquentistes est le même. La loi a priori serait alors $\Pi(\theta) = \frac{1}{\theta(1-\theta)}$ . Malheureusement, il ne s'agit pas d'une loi de probabilité car $\int_0^1 \frac{d\theta}{\theta(1-\theta)}$ n'existe pas. Les bayésiens appellent une telle loi une loi impropre car si par le théorème de Bayes on la combine avec une loi de probabilité, ici la loi binomiale, on obtient une vraie loi de probabilité de densité :

$$\Gamma\left(\frac{\theta}{k}\right) = \frac{\Gamma(n)}{\Gamma(k)\Gamma(n-k)}\theta^{k-1}(1-\theta)^{n-k-1} \quad \text{pourvu que } k \neq 0.$$

Une interprétation possible est la suivante : la loi impropre donne le même estimateur que celui des fréquentistes, donc on n' a aucune information a priori sur le paramètre $\theta$, la loi impropre caractérise donc cette absence d'information, elle est dite non informative.

**4. Interprétation d'une distribution a priori**

Le phénomène observé « nombre de mauvaises cartouches dans un lot tiré au hasard » obéit à une loi binomiale : $\mathrm{P}_\theta(\mathrm{K}=k)=\mathrm{C}_n^k\,\theta^k(1-\theta)^{n-k}$. Le bayésien a pris comme loi a priori sur [0,1] une loi bêta de paramètres $\alpha$ et $\beta$. Son estimateur est : $\hat{\theta}_b(\omega)=\dfrac{k+\alpha}{n+\alpha+\beta}$ et le calcul donne $\mathrm{E}_\theta\left(\hat{\theta}_b\right)=\dfrac{n\theta+\alpha}{n+\alpha+\beta}$ et $\mathrm{Var}_\theta\left(\hat{\theta}_b\right)=\dfrac{n\theta(1-\theta)}{(n+\alpha+\beta)^2}$.

Le fréquentiste veut interpréter ce résultat. La loi a priori bêta est dite distribution (ou loi) conjuguée de la loi binomiale. Supposons qu'avant l'expérience une autre a eu lieu avec un lot de $n_1$ cartouches et $\mathrm{K}_1$ mauvaises. Pour lui l'estimateur optimal de $\theta$ sera obtenu en réunissant les deux échantillons et on aura $\hat{\theta}=\dfrac{\mathrm{K}+\mathrm{K}_1}{n+n_1}$. Mais la première expérience a eu lieu et on a observé $k_1$ mauvaises cartouches. Cherchons les deux premiers moments de $\hat{\theta}$ conditionnellement à $\{\mathrm{K}_1=k_1\}$. On a :

$$\mathrm{E}_\theta\left[\hat{\theta}/\mathrm{K}_1=k_1\right]=\frac{n\theta+k_1}{n+n_1},\quad \mathrm{Var}_\theta\left[\hat{\theta}/\mathrm{K}_1=k_1\right]=\frac{n\theta(1-\theta)}{(n+n_1)^2}.$$

Ils sont de la même forme que $\hat{\theta}$ avec $\alpha=k_1$, $\alpha+\beta=n_1$. D'où l'interprétation de l'estimateur bayésien obtenu avec la distribution conjuguée bêta de paramètres $\alpha$ et $\beta$ : c'est comme si on prenait en compte une expérience antérieure dans laquelle il y aurait eu $\alpha$ mauvaises cartouches dans un lot de $\alpha+\beta$. L'information sur $\theta$ véhiculée par la loi des distributions a priori est assimilable à celle donnée par une expérience aléatoire. La loi a priori, impropre, précédente dite non informative est telle que $\alpha=\beta=0$.On peut l'interpréter comme celle correspondant à une absence d'expérience antérieure, ce qui justifie l'appellation non-informative.

## IV. Les points de vue des décideurs

**1. Le décideur bayésien**

Reprenons l'exemple des cartouches. On procède à une expérience : tirage d'un échantillon au hasard et examen de celui-ci, on la modélise par une loi de probabilité comportant un paramètre inconnu $\theta$ figurant la proportion de cartouches défectueuses dans la caisse. Il s'interroge ensuite sur les informations qu'il peut avoir sur le fournisseur. Appelons H son univers de connaissance. Dans H des valeurs de $\theta$ lui paraîtront plus crédibles que d'autres. Il affecte donc aux différentes valeurs de $\theta$ un indice de crédibilité sous forme d'une loi a priori sur $\theta$. Cette loi synthétise ses connaissances dans l'univers H. Il peut modifier cette loi si H devient H′ après la prise en compte d'autres informations. Ensuite on combine les connaissances a priori

et le résultat à l'aide du théorème de Bayes et on obtient une connaissance a posteriori sur $\theta$ intégrant l'information a priori et le résultat de l'expérience vu comme une information supplémentaire.

Pour la justification épistémologique de la procédure deux problèmes se posent.

1. Représentation de la crédibilité d'un état de la nature (ici le paramètre $\theta$) par une loi de probabilité a priori.
2. Combinaison des probabilités définies par le modèle et de la probabilité a priori pour aboutir à une probabilité a posteriori sur $\Theta$.

Pour justifier le point 1. le bayésien fait appel à la théorie de la décision. Cette dernière se fonde sur une observation ; le décideur doit choisir entre plusieurs actions dans un contexte d'incertitude, par exemple faut-il accepter la caisse de cartouches livrée ou faut-il la renvoyer et rechercher un autre fournisseur lui garantissant, moyennant un prix plus élevé, une caisse de bonne qualité ?

Ensuite la théorie explicite des axiomes du comportement rationnel. De la nécessité de décider et d'agir rationnellement, on peut déduire les résultats suivants :
Soit
A l'ensemble des décisions possibles ;
$\Theta$ l'ensemble des états de la nature possibles, l'incertitude est le $\theta \in \Theta$ qui régit le phénomène ;
C l'ensemble des conséquences, si je mène l'action $a$ et que l'état de la nature est $\theta$, la conséquence est un $c \in$ C dépendant de $a$ et de $\theta$.

Décider c'est classer l'ensemble des décisions possibles ; deux décisions étant données, c'est dire quelle est la meilleure ou bien dire qu'elles sont équivalentes, donc mathématiquement, mettre un préordre complet sur A ; ce préordre dépend du décideur. Moyennant ce formalisme, on démontre que :

1. Chaque conséquence peut être affectée d'un nombre réel $u$ appelé utilité ;
2. Il existe une distribution de probabilité sur $\Theta$ ;
3. Le préordre sur A est induit par les $E[u(c(a, \theta))]$ où l'opérateur E est l'espérance mathématique selon la distribution de probabilité sur $\Theta$.

Autrement dit si vous devez décider et être rationnel, vous êtes bayésien.

Pour justifier le point 2, le bayésien examine la théorie de la décision statistique telle qu'elle a été établie par les fréquentistes. Ceux-ci recherchent une stratégie S qui à tout constat expérimental $\omega$ associe une action. Si l'état de la nature est $\theta$, il aurait fallu décider $h(\theta)$, il y a donc une erreur formalisée par un coût qui est un réel positif $L(S(\omega), h(\theta))$. Celui-ci est comme une loterie. La valeur d'une loterie, depuis Pascal, est son espérance mathématique. En appliquant la stratégie S on prend donc un risque $R(\theta, S) = E_\theta[L(S, h(\theta))]$, l'espérance est prise pour l'état de la nature $\theta$ auquel est associée la loi $P_\theta$ « régissant l'arrivée » du constat $\omega$. Une stratégie $S_1$ est dite meilleure qu'une stratégie $S_2$ si :

$$\forall \theta \in \Theta, R(\theta, S_1) \leq R(\theta, S_2).$$

Une stratégie est dite admissible si aucune autre n'est meilleure qu'elle.

Un théorème de la théorie de la décision énonce que les stratégies bayésiennes, comme celle développée dans l'exemple de la caisse de cartouches, sont admissibles. Le statisticien bayésien vous dira : prenez ma stratégie, elle vaut ce qu'elle vaut mais vous n'en trouverez pas de meilleure qu'elle dans tous les cas.

**2. Le décideur fréquentiste**

Le décideur fréquentiste, lui, pense que le modèle probabiliste n'est applicable que si, au moins conceptuellement, on peut appliquer au phénomène la loi des grands nombres. Certes, un indice de crédibilité peut être rendu par une loi de probabilité, mais mélanger les deux, c'est mélanger les torchons et les serviettes. Pour lui, parler de probabilité subjective n'a pas de sens : une probabilité traduit un phénomène naturel qui existe en l'absence d'observateur.

En outre, l'introduction de croyance a priori sur l'état de la nature comporte un risque : faire une étude scientifique pour justifier un préjugé, pour défendre des intérêts. Pour cela, il suffit de bien choisir la probabilité a priori. Plus généralement, on dénonce l'arbitraire de son choix.

De plus, une fois choisie cette loi a priori sur l'état de la nature, le calcul de la loi a posteriori se révèle extrêmement difficile. Trop souvent il n'y a pas d'expression analytique exprimable à l'aide des fonctions classiques ce qui limite les applications concrètes.

## V. Pour une utilisation raisonnable des procédures bayésiennes

Dans la pratique, les choses sont moins tranchées que le développement précédent pourrait le faire penser. Des distributions a priori peuvent parfois être interprétées en termes fréquentistes. Surtout, il paraît absurde d'ignorer des informations existant sur les valeurs possibles du paramètre même si leur codification par une loi de probabilité pose problème. Les exemples suivants illustrent des utilisations des procédures bayésiennes.

Par exemple, lorsque l'on cherche les pièces de rechange à embarquer dans un sous-marin où il y a peu de place. En regardant les statistiques, on observe que les pannes sont peu nombreuses (heureusement !) et les informations disponibles par les statistiques insuffisantes. Les auteurs de l'étude ont fait appel à d'autres sources d'information. Des pièces n'étaient pas tombées en panne car dès qu'elles présentaient une faiblesse un réparateur sur place était intervenu. Ce sont les fourriers (gestionnaires de stocks) qui pouvaient avoir la mémoire de ces incidents mineurs. On lança l'enquête parmi ceux qui avaient une certaine ancienneté en leur demandant leur avis sur les pièces à embarquer. C'est à partir de ces opinions qu'à été construite la loi a priori. Ici, elle a un relent d'objectivité, ce n'est pas qu'une opinion mais le résumé d'expériences, un fréquentiste pourrait à la limite valider cette procédure !

**Remarque**

Quand on parle de probabilité subjective, il n'est pas en général possible de l'établir à partir des réponses à un questionnaire de type psychologique. En effet de nombreuses expériences ont montré que la perception intuitive de l'incertitude ne se

fait pas selon les axiomes du calcul des probabilités. Par exemple, pour une maladie grave avec un traitement difficile, on annonce 68 % de chances de survie, alors 44 % des personnes sont prêtes à subir le traitement mais si on annonce 32 % d'échec, alors seulement 18 % des personnes sont prêtes au traitement. On voit qu'avec la même information sous deux formes différentes la probabilité subjective varie dans des proportions importantes.

**Recherche de lois a priori**

Pour tirer parti d'une information subjective et la traduire en une loi de probabilité sur le paramètre, il existe plusieurs procédures utilisant les moments ou les quantiles. Elles sont aussi utilisées pour déduire, par exemple, la loi a priori de l'histoire du phénomène. Soit à estimer la probabilité $\theta_k$ qu'un individu d'un lycée donné soit reçu au baccalauréat l'année $k$. Dans les années antérieures, cette probabilité a été estimée à l'aide de la proportion de reçus. Si $\theta_i$ est cette probabilité l'année $i$, $n_i$ le nombre de reçus et $\mathrm{N}_i$ le nombre de présentés, $\theta_i$ a pu être estimée par $\dfrac{n_i}{\mathrm{N}_i} = \hat{\theta}_i$.

Des valeurs $\hat{\theta}_1$, $\hat{\theta}_2$, …, $\hat{\theta}_{k-1}$, on peut déduire une estimation des paramètres $\alpha$ et $\beta$ de la distribution a priori de $\theta$ prise dans la famille bêta (Cf. II.3.). Elle est de densité $\Pi(\theta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$.

On sait que la moyenne de cette loi est $\dfrac{\alpha}{\alpha+\beta}$ et la variance $\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

On prendra comme valeur pour $\alpha$ et $\beta$ la solution du système d'équation :

$$\begin{cases} \dfrac{\alpha}{\alpha+\beta} = \dfrac{1}{k-1}\displaystyle\sum_{i=1}^{k-1}\hat{\theta}_i = \overline{\theta} \\ \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \dfrac{1}{k-2}\displaystyle\sum_{i=1}^{k-1}\left(\hat{\theta}_i - \overline{\theta}\right)^2 \end{cases}.$$

Ce système est équivalent au suivant :

$$\begin{cases} \dfrac{\alpha}{\alpha+\beta} = \dfrac{1}{k-1}\displaystyle\sum_{i=1}^{k-1}\hat{\theta}_i = \overline{\theta} \\ \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \dfrac{1}{k-1}\displaystyle\sum_{i=1}^{k-1}\left(\hat{\theta}_i\right)^2 \end{cases}.$$

L'utilisateur pratique des procédures bayésiennes a longtemps été freiné par les difficultés de calcul. Une fois choisie une distribution a priori, le calcul analytique de la distribution a posteriori est très souvent inextricable. Mais des méthodes d'analyse numérique ou par simulation ont fait récemment leur apparition avec des logiciels permettant une application pratique, ce qui a permis à beaucoup de statisticiens appliqués de faire appel aux méthodes bayésiennes.

On prendra, comme exemple, l'étude sur les événements hydrologiques extrêmes. Il s'agit de prévoir les crues exceptionnelles de la Garonne ou au moins leur probabilité d'apparition. On connaît la fréquence et la hauteur des crues depuis 1913 mais ces données sont insuffisantes pour estimer valablement les probabilités des valeurs extrêmes possibles indispensables à connaître pour prévoir avec un coût raisonnable la hauteur des digues à construire. À côté de ces données expérimentales on dispose de l'expertise des hydrologues et surtout de données historiques : les grandes crues ont laissé un souvenir dans la mémoire des hommes et sont décrites plus ou moins précisément dans les chroniques. Les méthodes bayésiennes ont permis d'intégrer ces données historiques dans les estimations de quantités du type : quelle est la hauteur de la crue $Q_{1000}$ qui a la probabilité $10^{-3}$ d'apparaître une année donnée. Le modèle utilisé met en œuvre des formulations complexes dont l'exposé dépasse le cadre de cet article. Les méthodes de calculs par simulation ont permis d'estimer $Q_{1000}$ et de fournir un intervalle de confiance.

Un autre exemple d'utilisation des méthodes bayésiennes est fourni par l'étude de la transmission du virus du SIDA par la mère à son enfant. Il s'agit de déterminer le mécanisme de transmission. Trois possibilités existent :
– transmission in utero précoce ;
– transmission in utero tardive ;
– transmission à l'accouchement.
Un échantillon de 700 enfants à risque est étudié, 135 d'entre eux infectés sont décédés avant 18 mois. On veut estimer les probabilités des trois modes possibles d'infection. Les méthodes bayésiennes utilisées ont permis à partir du développement de la maladie après la naissance de tenir compte de circonstances particulières et d'obtenir des résultats validés par les spécialistes.

## VI. En conclusion

On peut être sceptique sur la validité épistémologique des méthodes bayésiennes mais les récents développements, l'introduction de lois a priori dites non informatives, la mise au point de méthodes de calcul performantes permettent un usage raisonnable de ces méthodes. Il ne faut pas rejeter l'enfant avec l'eau du bain mais il ne faut pas non plus regarder le paradigme bayésien comme le seul possible. Il a son efficacité mais aussi ses limites.

## VII. Une étude récente: la salmonellose au Danemark

La salmonellose est une maladie humaine causée par une bactérie, la salmonelle, transmise par des aliments infectés d'origine animale. On distingue 9 variétés de salmonelle. Parmi elles la variété notée « E » comporte 6 sous-variétés différentes et celle notée « T » 11. Ce qui fait 24 types de bactéries différents. L'indice « $i$ » notera le type de bactérie, on a donc: $1 \leq i \leq 24$ ; $i$ varie de 1 à 6 pour la variété « E », de 7 à 17 pour la variété « T » et de 18 à 24 pour les autres variétés. On a repéré 9 sources possibles d'infection d'origine animale. L'indice « $j$ » notera la source, $1 \leq j \leq 9$.

L'étude a été faîte pour les services sanitaires du Danemark. Ils veulent quantifier la contribution à la maladie de chaque source possible. On dispose, pour l'année 1999, des données suivantes :
– $M_j$ quantité de nourriture en provenance de la source « $j$ » consommée dans le pays ;
– $p_{ij}$ pourcentage d'aliments de la source « $j$ » contenant le type « $i$ » de salmonelle ;
– $o_i$ nombre de malades infectés par le type « $i$ » de salmonelle, on compte pour 1 tous les malades d'un même groupe (famille, cantine, etc.) quand ce dernier implique des repas en commun ;
– le nombre de malades dont on connaît seulement la variété mais pas la sous-variété responsable de l'infection ainsi que celui pour lesquels la variété est elle-même inconnue ;
– le nombre de malades infectés par le type « $i$ » de bactérie ou pour lesquels ce type est partiellement ou complètement inconnu et qui ont voyagé à l'étranger dans un passé proche ou bien pour lesquels on ignore si un voyage à l'étranger a eu lieu.
La présence des données incomplètes précédentes complique l'étude, il va falloir corriger les donnée du type « $o_i$ ».

Les services sanitaires ne s'intéressent qu'à la contamination dans le pays, ils ne peuvent agir à l'étranger, il faut donc éliminer les individus ayant voyagé à l'étranger. Quand la donnée voyage est inconnue il va falloir estimer le nombre moyen de voyageurs parmi eux. De même il faut estimer le nombre moyen de cas dont la maladie est due au type « $i$ » de salmonelle quand ce dernier est partiellement ou totalement inconnu. On peut le faire en reportant les cas où des données sont manquantes dans les diverses catégories en utilisant les proportions des diverses catégories quand elles sont connues. L'étude danoise procède d'une façon différente en utilisant une méthode probabiliste. Elle ne sera pas explicitée.

Soit $X_i$ la variable aléatoire nombre d'individus malades en une année par le type « $i$ » de bactérie et ayant contracté la maladie au Danemark. On pose $x_i$ la valeur prise par $X_i$ en 1999, c'est la valeur observée et corrigée pour tenir compte des données manquantes comme cela a été explicité au paragraphe précédent. On pose $Y_{ij}$ la variable aléatoire : nombre d'individus malades par le type « $i$ » de salmonelle et ayant été infectés par la source « $j$ ». On a :

$$X_i = Y_{i1} + Y_{i2} + \ldots + Y_{i9}.$$

Le modèle adopté est le suivant :
– les variables $Y_{ij}$ sont indépendantes ;
– $Y_{ij}$ suit une loi de Poisson de paramètre $\lambda_{ij}$ ( $P(Y_{ij} = k) = e^{-\lambda} \dfrac{\lambda_{ij}^k}{k!}$ ) rappelons que quand une variable aléatoire suit une loi de Poisson de paramètre $\lambda$, son espérance mathématique est égale à $\lambda$ ;
– on a $\lambda_{ij} = M_j\, p_{ij}\, q_i\, a_j$ ;
– $q_i$ ne dépend que de la variété et non de la sous-variété et mesure sa virulence pathogène, il y a donc 9 paramètres $q_i$ différents ;
– $a_j$ mesure l'intensité avec laquelle la source « $j$ » transmet à l'homme la bactérie, il y a 7 paramètre $a_j$ différents car il y a deux fois deux paramètres égaux.

Comme on s'intéresse aux rapports entre les virulences, le paramètre relatif à la variété « E » sera pris égale à 1, cela revient à choisir des unités pour la mesure de la virulence et de l'intensité de transmission de la source à l'homme. On a donc : 8 + 7 = 15 paramètres à estimer et on dispose des 24 observations $x_i$.

On démontre en calcul des probabilités qu'une somme de variables aléatoires indépendantes dont chacune suit une loi de Poisson suit aussi une loi de Poisson de paramètre la somme des paramètres. Donc $X_i$ suit une loi de Poisson de paramètre $\lambda_i = \sum \lambda_{ij}$.

Pour faire de la statistique bayésienne il faut choisir une loi a priori pour chaque paramètre inconnu. Les $q_i$ seront supposés suivre une loi uniforme sur l'intervalle $[0,b]$ et les $a_j$ suivre une loi uniforme sur l'intervalle $[0,a]$, $a$ et $b$ sont choisis suffisamment grands. Le théorème de Bayes permet, en théorie, de calculer la loi a posteriori des paramètres inconnus. Le calcul analytique est évidemment impossible. Par contre il est possible, par simulation, d'obtenir :
– l'espérance mathématique a posteriori des paramètres ;
– leur médiane a posteriori ;
– pour chacun d'eux un intervalle contenant 95% de la probabilité a posteriori, appelé intervalle de crédibilité.
La médiane et l'espérance a posteriori sont deux estimateurs ponctuels du paramètre et l'intervalle de crédibilité est un estimateur par intervalle du dit paramètre.

À partir de ces estimateurs, un calcul simple permet d'avoir un estimateur des $\lambda_{ij}$. Or $\lambda_{ij}$ est le nombre moyen de malades infectés, sur place, par le type « $i$ » de salmonelle et dont l'origine est la nourriture « $j$ ». Il est intéressant de connaître le nombre moyen $\mu_j$ de malades pour la source « $j$ » ( $\mu_j = \sum \lambda_{ij}$ ). On trouve ainsi, par exemple que les œufs sont responsables, en moyenne, de 47% des cas, avec un intervalle de crédibilité crédibilité à 95% variant de 43% à 51%.

L'adéquation du modèle aux données est bon. Pour s'en assurer on a comparé les $\lambda_i$ estimés et les observations corrigées $x_i$ : l'écart est faible.

## ANNEXE

### Recherche de la distribution a posteriori dans le cas d'une loi binomiale avec une loi bêta comme loi a priori.

En raisonnant sur des infiniment petits on peut écrire :

$$\mathrm{P}\left([\theta;\theta+\mathrm{d}\theta]\right) = \Pi(\theta)\,\mathrm{d}\theta = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\,\mathrm{d}\theta,$$

$$\mathrm{P}\left(\mathrm{K}=k/\theta\right) = \mathrm{C}_n^k\,\theta^k(1-\theta)^{n-k},$$

$$\mathrm{P}\left(\{\mathrm{K}=k\}\cap[\theta;\theta+\mathrm{d}\theta]\right) = \mathrm{C}_n^k\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}\,\mathrm{d}\theta,$$

$$\mathrm{P}\left(\mathrm{K}=k\right)=\mathrm{C}_n^k\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1\theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}\,\mathrm{d}\,\theta,$$

$$\mathrm{P}\left(\mathrm{K}=k\right)=\mathrm{C}_n^k\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\times\frac{\Gamma(k+\alpha)\Gamma(n-k+\beta)}{\Gamma(n+\alpha+\beta)}.$$

L'application du théorème de Bayes donne :

$$\mathrm{P}\left(\left[\theta;\theta+\mathrm{d}\,\theta\right]/\,\mathrm{K}=k\right)=\frac{\Gamma(n+\alpha+\beta)}{\Gamma(k+\alpha)\Gamma(n-k+\beta)}\theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}\,\mathrm{d}\,\theta.$$

La loi a posteriori est donc une loi bêta.

## Bibliographie

Bernier Jacques, Parent Éric, Boreux Jean-Jacques (2000) *Statistique pour l'environnement, traitement bayésien de l'incertitude*. TEC & DOC.

De Finetti Bruno (1937) *La prévision : ses lois logiques, ses sources subjectives*. Annales de l'IHP, Tome 7, nº 1, pages 1 à 68.
Accessible à l'adresse http://www.numdam.org/item ?id=AIHP_1937_7_1_1_0

Robert Christian (1992) *L'analyse Statistique Bayésienne.* Economica.

Savage L.(1962) *The Foundations of Stastistical Inference*. Methuen London.

Droesbeke Jean-Jacques, Fine Jeanne, Saporta Gilbert (éditeurs) (2002) *Méthodes bayésiennes en statistique*. Technip.

N.B. le premier ouvrage est d'accès facile, les trois derniers sont des ouvrages spécialisés.