

Remerciement

Nous remercions d'abord et avant tout le bon Dieu qui nous donne le courage et la patience pour réaliser ce travail.

Nous remercions notre encadreuse madame ZERARI AMEL , pour ses conseils précieux, ses critiques et ses directions qui nous ont donné le plaisir de travailler assidument.

Nos remerciements s'adressent également à toutes les personnes de l'institut de sciences et à de la technologie surtout les enseignants qui nous ont enseigné durant toutes nos années d'étude.

Nous voudrions dire toute notre reconnaissance à nos parents pour leur dévouement sans limite donné sur tous les plans, et remercier mes familles et mes amis pour leur soutien constant.

Table des matières

Introduction Générale	2
1 L'approche bayésienne	4
1.1 La vraisemblance :	4
1.1.1 Principe de vraisemblance :	4
1.2 Distribution a priori et a posteriori et prédictive :	5
1.3 La loi a priori, a posteriori et prédictive :	6
1.3.1 Loi a priori :	6
1.3.2 Loi a posteriori :	7
1.3.3 Loi prédictive :	7
1.4 Le problème de choix de l'a priori :	7
1.4.1 Lois subjectives :	8
1.4.2 Les Lois a priori conjuguées :	8
1.5 Famille exponentielle :	9
1.6 Lois a priori non informatives :	12
1.6.1 La loi a priori de Jeffreys :	12
1.6.2 L'analyse avec la distribution a priori non informative : . .	13
1.6.3 Facteur de bayes :	15
1.6.4 La fonction d'utilité :	16
2 Les tests	18
2.1 Test dans le cas fréquentiste :	18
2.2 Théorie des tests paramétriques :	19
2.2.1 Introduction : test sur l'espérance d'une loi normale de va- riance connue :	19
2.3 Vocabulaire des tests :	19
2.4 Probabilité d'erreur et risque, puissance de test :	20
2.5 Choix optimal de la statistique de test et de la région de rejet : .	21

2.6	Utilisation de la puissance de test :	23
2.6.1	Résumé :	24
2.6.2	p-value :	25
2.7	Tests sur une population :	25
2.8	Test sur le caractère central d'une population :	25
2.8.1	Cas d'un échantillon grand ou gaussien :	25
2.8.2	Cas d'un petit échantillon non gaussien :	27
2.9	L'approche bayésienne des tests :	27
2.9.1	La fonction de coût :	28
2.9.2	Fonctions de coût usuelles :	28
2.10	La comparaison entre les tests fréquentiste et les testes bayesienne :	29
	Conclusion Générale	30
	Bibliographie	31

Introduction Générale

Dans de nombreuses situations d'expériences aléatoires, il semble raisonnable d'imaginer que le praticien a une certaine idée du phénomène aléatoire qu'il est en train d'observer. Or, la démarche statistique classique repose essentiellement sur un principe de vraisemblance qui consiste à considérer que ce qui a été observé rend compte de manière exhaustive du phénomène. Mais l'observation ne fournit qu'une image et celle-ci peut être mauvaise. Certes cet inconvénient est en général gommé par les considérations asymptotiques et un certain nombre de théorèmes permettent d'évaluer la bonne qualité des estimateurs si le nombre d'observations est suffisant .

L'analyse bayésienne des problèmes statistiques propose d'introduire dans la démarche d'inférence, l'information dans dispose a priori le praticien. Dans le cadre de la statistique paramétrique, ceci se traduira par le choix d'un loi sur le paramètre d'intérêt.

Dans l'approche classique, le modèle statistique est le triplet $(N; A; P_\theta; \theta \in \Theta)$. Ayant un a priori sur le paramètre, modélisé par une densité de probabilité que nous noterons $\pi(\theta)$, on "ré-actualise" cet a priori au vu de l'observation en calculant la densité a posteriori $\pi(\theta/x)$, et c'est à partir de cette loi que l'on mène l'inférence.

On peut alors, par exemple, de manière intuitive pour le moment, retenir l'espérance mathématique ou encore le mode de cette densité a posteriori comme estimateur de le paramètre devient donc en quelque sorte une variable aléatoire, à la quelle on associe une loi de probabilité dite loi a priori.

On sent bien d'emblée que les estimateurs bayésiens sont très dépendants du choix de la priori.

Différentes méthodes existent pour déterminer ces lois a priori. On peut se référer à des techniques bayésiennes empiriques, où l'on construit la loi a priori sur la base d'une expérience passée, usant de méthodes fréquentistes, pour obtenir forme et valeurs de paramètres pour cette loi. Nous verrons que l'on peut aussi modéliser l'absence d'information sur le paramètre au moyen des lois dites lois non informatives.

L'approche bayésienne se différencie dans de l'approche classique dans le sens où le paramètre θ n'est plus considéré comme étant totalement inconnu ; il est devenu une v.a. dont le comportement est supposé connu. On fait intervenir dans l'analyse statistique une distribution associée à ce paramètre.

Comparaison des approches

Statistique Fréquentiste :

- La probabilité est une propriété physique : < Probability lives in the world >.
- Décisions = p-value.
- Incertitude = intervalle de confiance.

Statistique Bayésienne :

- La probabilité est un concept : < Probability lives in the mind >.
- Décisions = distribution a posteriori.
- Incertitude = l'intervalle de crédibilité.

N : l'espace des observations,

Θ : l'espace des états de la nature (l'espace des paramètres dans le cas d'un problème statistique),

A : l'espace des actions ou décisions, dont les éléments sont des images de l'observation par une application δ appelée règle de décisions (une statistique (i.e. fonction de observations) dans le cas d'un problème statistique)

Chapitre 1

L'approche bayésienne

Introduction :

L'approche bayésienne peut être présentée comme une généralisation de l'approche classique. Les paramètres ne sont plus des valeurs inconnues fixées, mais des variables aléatoires. Nous considérons que l'objet principal de la statistique est de tirer, au vu d'observations d'un phénomène aléatoire, une inférence au sujet de la loi générant ces observations; afin, soit d'analyser un phénomène passé, soit de prévoir un événement futur, en nous attachant tout particulièrement aux aspects décisionnels de l'inférence bayésienne.

L'inférence bayésienne s'accompagne d'une modélisation probabiliste du phénomène observé qui est nécessaire, car elle permet de remplacer le phénomène dans une globalité qui autorise des analyses et des généralisations.

1.1 La vraisemblance :

L'inférence statistique cherche à apprendre sur la valeur dans la population d'un paramètre à partir d'un échantillon de données y .

Inférence traditionnelle : une fois y connu $\mathbf{P}(y/\theta)$ (fonction de vraisemblance) quantifie à quel point certaines valeur de sont compatibles avec les valeur observées y .

$P(y/\theta)$ résume toute l'information que les données fournissent sur le paramètre θ .

1.1.1 Principe de vraisemblance :

L'information apportée par une observation de x sur θ est entièrement contenue dans la fonction de vraisemblance $\ell(\theta/x)$. De plus, si x_1 et x_2 sont deux observations qui dépendent du même paramètre θ , et telles qu'il existe une constante c satisfaisant

$$l_1(\theta \setminus x_1) = cl_2(\theta \setminus x_2)$$

Pour tout θ , elles apportent la même information sur θ et doivent conduire à la même inférence.

1.2 Distribution a priori et a posteriori et prédictive :

Supposons une distribution a priori sur θ , $\pi(\theta)$, soit disponible, c'est-à-dire que nous disposons d'un modèle complètement bayésien une fois données ces deux distributions nous pouvons en construire plusieurs autres, à savoir :

a) La distribution jointe de $(\theta; x)$:

$$\varphi(\theta/x) = f(x/\theta) \pi(\theta)$$

b) La distribution marginale de x :

$$m(x) = \int \varphi(\theta/x) d\theta = \int f(x/\theta) \pi(\theta) d\theta$$

c) La distribution a posteriori de θ obtenue par la formule de Bayes :

$$\pi(\theta/x) = \frac{f(x/\theta) \pi(\theta)}{\int f(x/\theta) \pi(\theta) d\theta} = \frac{f(x/\theta) \pi(\theta)}{m(x)}$$

d) La distribution prédictive de y ou $y \sim g(y/\theta; x)$ obtenue par :

$$g(y/x) = \int g(y/\theta; x) \pi(\theta) d\theta$$

Exemple 1.1 : Une boule de billard W roule sur une ligne de longueur un, supposons qu'elle s'arrête en P . Une deuxième boule O roule alors n fois dans les mêmes conditions, on note X le nombre de fois que la boule O s'arrête à gauche de W .

Quelle inférence pouvons-nous mener sur P ?

Dans la terminologie moderne, le problème est de déterminer la distribution a posteriori de P conditionnellement à X , quand la distribution a priori de P est uniforme sur $[0; 1]$ et $X \sim B(n; p)$, variable aléatoire binomiale comme :

$$P(X = \theta / P) = \binom{n}{\theta} P^\theta (1 - P)^{n-\theta}$$

$$P(a < P < b / X = \theta) = \int_a^b \binom{n}{\theta} P^\theta (1 - P)^{n-\theta} dp \text{ et}$$

$$P(X = \theta) = \int_0^1 \binom{n}{\theta} P^\theta (1 - P)^{n-\theta} dp$$

nous trouvons que :

$$P(a < P < b / X = \theta) = \frac{\int_a^b \binom{n}{\theta} P^\theta (1 - P)^{n-\theta} dp}{\int_0^1 \binom{n}{\theta} P^\theta (1 - P)^{n-\theta} dp} = \frac{\int_a^b \binom{n}{\theta} P^\theta (1 - P)^{n-\theta} dp}{B(\theta + 1, n - \theta + 1)}$$

donc la distribution de P conditionnellement à $X = \theta$ est une distribution bêta, $B(\theta + 1, n - \theta + 1)$.

1.3 La loi a priori, a posteriori et prédictive :

1.3.1 Loi a priori :

On entend par information a priori sur le paramètre θ toute information disponible sur en dehors de celle apportée par les observations.

L'information a priori sur est entachée d'incertitude (si ce n'était pas le cas, le paramètre serait connu avec certitude et on n'aurait pas à l'estimer).

Il est naturel de modéliser cette information a priori au travers d'une loi de probabilité appelée *loi a priori*, notée $\pi(\theta)$.

1.3.2 Loi a posteriori :

C'est la loi conditionnelle de θ sachant x , sa densité notée $\pi(\theta/x)$, en vertu de la formule de Bayes de (1,1), et on a aussi :

a/ La loi du couple (θ, x) : sa densité est notée $\varphi(\theta, x)$ on a donc :

$$\varphi(\theta/x) = f(x/\theta) \pi(\theta)$$

b/ La loi marginale de x sa densité est notée $m(x)$, on a donc :

$$m(x) = \int f(x/\theta) \pi(\theta) d\theta$$

c/ La distribution a posteriori de obtenue par la formule de bayes :

$$\pi(\theta/x) = \frac{f(x/\theta) \pi(\theta)}{\int f(x/\theta) \pi(\theta) d\theta} = \frac{f(x/\theta) \pi(\theta)}{m(x)}$$

1.3.3 Loi prédictive :

La probabilité prédictive d'obtenir la conclusion souhaitée est un élément important à prendre en compte dans la décision. Une probabilité très élevée ou très faible est un argument en faveur de l'interruption de l'essai. En outre si on décide de poursuivre l'essai elle fournit un outil pour révaluer le nombre de sujets supplémentaires nécessaire

$$g(y/\theta) = \int g(y/\theta, x) \pi(\theta) d\theta$$

1.4 Le problème de choix de l'a priori :

L'aspect de l'analyse bayésienne le plus critiqué et le plus délicat est certainement le choix de la loi a priori des paramètres ; en particulier, le recours à des lois usuelles comme la loi normale, gamma, bêta, ne peut pas être défendu comme approche systématique plus aisée.

La détermination de la loi a priori donc est basée sur l'information a priori.

1.4.1 Lois subjectives :

Précisons tout d'abord que cette démarche n'est pas forcément facile dans la pratique. L'idée est d'utiliser les données antérieures. Par exemple, dans un cadre paramétrique, cela revient à choisir une valeur particulière du paramètre.

Dans un cas concret, il peut être judicieux de baser son raisonnement sur les dires d'experts, notamment à l'aide de questionnaires. Il est alors nécessaire de veiller à ce que les questions soient compréhensibles, par exemple en prenant comme base les quantiles plutôt que les moments. Pour plusieurs experts, il peut être utile de pondérer leurs réponses et d'utiliser des modèles hiérarchiques.

Ainsi, la difficulté ici n'est pas mathématique mais plus psychométrique pour réduire les biais sur les réponses fournies. Nous allons nous concentrer sur le second aspect de la détermination.

1.4.2 Les Lois a priori conjuguées :

Une des difficultés de l'approche bayésienne est le calcul de la loi a posteriori. ce calcul est facilité lorsque loi a priori et loi a posteriori ont la même forme.

Dans ce cas, on parle de loi a priori conjuguée.

Définition 1.2 : une famille \mathbf{F} de lois, sur Θ est dite conjuguée si pour tout $\pi \in \mathbf{F}$, la loi a posteriori $\pi(\theta/x)$ appartient également à \mathbf{F}

Un exemple trivial de famille conjuguée est bien sûr \mathbf{F}_0 , ensemble de toutes les lois de probabilité sur Θ , qui n'est d'aucune utilité dans le choix d'une loi a priori.

En fait, on essaie de prendre \mathbf{F} aussi petite que possible et paramétrée.

<i>loidesobservation</i> $P(x/\theta)$	loi a priori $P(\theta/T)$	loi a posteriori $P(\theta/T) \propto P(\theta/T) P(x/\theta)$
	variables discrètes	
Binomiale $Bin(x/n, \theta)$	Bêta $Bet(\theta/\alpha, \beta)$	Bêta $Bet(\theta/\alpha + x, \beta + n - x)$
Binomiale $Bin(x/n, \theta)$	$P(\theta) = \theta(1 - \theta)^{-1}$	Bêta $Bet(\theta/x, n - x), x \neq 0, x \neq n$
<i>Binomialenégative</i> $NegBin(x/n, \theta)$	Bêta $Bet(\theta/\alpha, \beta)$	Bêta $Bet(\theta/\alpha + n, \beta + x)$
<i>Multinomiale</i> $M_k(x/\theta_1, \dots, \theta_k)$	Dirichlet $Di_k(\theta/\alpha_1, \dots, \alpha_k)$	<i>Dirichlet</i> $Di_k(\theta/\alpha_1 + x_1, \dots, \alpha_k + x_k)$
Poisson $P_n(x/\theta)$	Gamma $Gam(\theta/\alpha, \beta)$	Gamma $Gam(\theta/\alpha + x, \beta + 1)$
Gamma $Gam(x/v, \theta)$	Gamma $Gam(\theta/\alpha, \beta)$	Gamma $Gam(\theta/\alpha + v, \beta + x)$
Bêta $Bet(x/\alpha, \theta)$	Exponentielle $Ex(\theta/\lambda)$	<i>Exponentielle</i> $Ex(\theta/\lambda - \log(1 - x))$
Normale $N(x/\theta, \sigma^2)$	Normale $N(\theta/\mu, T^2)$	Normale $N\left(\mu/\frac{\mu\sigma^2 + T^2 X}{\sigma^2 + T^2}, \frac{\sigma^2 T^2}{\sigma^2 + T^2}\right)$
	variables continues	
Normale $N(x/\mu, 1/\theta)$	Gamma $Gam(\theta/\alpha, \beta)$	<i>Gamma</i> $Gam(\theta/\alpha + \frac{1}{2}, \beta + \frac{1}{2}(\mu - x)^2)$
Normale $N(x/\theta, \theta^2)$	<i>Normale inversegénéralisée</i> $ING(\theta/\alpha, \mu, \sigma) \propto \theta ^{-\alpha}$ $\exp\left[-\frac{1}{2\sigma^2}\left(\frac{1}{\theta} - \mu\right)^2\right]$	<i>Normaleinversegénéralisée</i> $ING(\theta/\alpha_n, \mu_n, \sigma_n)$

1.5 Famille exponentielle :

Un type particulier de lois de probabilité permet une détermination directe des familles de lois conjuguées. De telles lois sont dites familles exponentielles.

Définition 1.3 : Soient μ , mesure σ finie sur X , et Θ l'espace des paramètres. On définit C et H , respectivement fonction de X et dans \mathbb{R}_+ , et R et T , fonction de et X dans \mathbb{R}^k . La famille de distributions de densité (par rapport à) :

$$f(x/\theta) = C(\theta) H(x) \exp\{R(\theta) T(x)\}$$

est dite famille exponentielle de dimension K . Dans le cas particulier où $\theta \in \mathbb{R}^k, x \in \mathbb{R}^k$ et :

$$f(x/\theta) = C(\theta) H(x) \exp\{\theta x\}$$

La famille est dite naturelle.

Exemple 1.4 : Le calcul de la loi a posteriori dans le modèle gaussien lorsque σ^2 est connue. On suppose qu'on observe l'observation $x \sim N(\theta, \sigma^2)$ avec

σ^2 connue, et que la vraisemblance est

$$f(x/\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right), x \in \mathbb{R}, \theta \in \mathbb{R} \text{ et } \sigma > 0$$

Si la distribution a priori est spécifique comme $\pi(\theta) = N(\theta/\mu, \tau^2)$ avec (μ, τ^2) fixé, donc on obtient

$$\begin{aligned} \pi(\theta/x) &= \frac{f(x/\theta)\pi(\theta)}{\int_{\Theta} f(x/\theta)\pi(\theta)d\theta} \\ &= \frac{\frac{1}{\tau\sqrt{2\pi}} \exp\left(-\frac{(\theta-\mu)^2}{2\tau^2}\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)}{\int_{-\infty}^{+\infty} \frac{1}{\tau\sqrt{2\pi}} \exp\left(-\frac{(\theta-\mu)^2}{2\tau^2}\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right) d\theta} \\ &= \frac{\frac{1}{\tau\sigma 2\pi} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2}\right]}{\int_{-\infty}^{+\infty} \frac{1}{\tau\sigma 2\pi} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2}\right] d\theta} \end{aligned}$$

on pose

$$m(x) = \int_{-\infty}^{+\infty} \frac{1}{\tau\sigma 2\pi} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2}\right] d\theta$$

et

$$\begin{aligned} \frac{(x-\theta)^2}{\sigma^2} - \frac{(\theta-\mu)^2}{\tau^2} &= \frac{x^2 - 2\theta x + \theta^2}{\sigma^2} - \frac{\theta^2 - 2\theta\mu + \mu^2}{\tau^2} \\ &= \theta^2 \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right) - 2\theta \left(\frac{x}{\sigma^2} - \frac{\mu}{\tau^2}\right) + \frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2} \\ &= \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right) \left[\theta - \frac{\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right]^2 + \frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2} - \frac{\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)^2}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}} \end{aligned}$$

d'où

$$m(x) = \frac{1}{2\tau\sigma\pi} \int_{-\infty}^{+\infty} \exp\left[-\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right) \left[\theta - \frac{\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right]^2 \exp\left[\frac{\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)^2}{2\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)} - \frac{x^2}{\sigma^2} - \frac{\mu^2}{\tau^2}\right]\right] d\theta$$

posons

$$\begin{aligned} t &= \left(\theta - \frac{\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right) \sqrt{\frac{1}{\sigma^2} + \frac{1}{\tau^2}} \\ &= \sqrt{\frac{1}{\sigma^2} + \frac{1}{\tau^2}} d\theta \end{aligned}$$

$$m(x) = \frac{\exp \left[\frac{\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2} \right)^2}{2 \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2} \right)} - \frac{x^2}{\sigma^2} - \frac{\mu^2}{\tau^2} \right]}{2\tau\sigma\pi\sqrt{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}} \int_{-\infty}^{+\infty} \exp -\frac{t^2}{2} dt$$

comme

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp -\frac{t^2}{2} dt = 1$$

on en déduit

$$\begin{aligned} m(x) &= \frac{\exp \left[\frac{\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2} \right)^2}{2 \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2} \right)} - \frac{x^2}{\sigma^2} - \frac{\mu^2}{\tau^2} \right]}{\tau\sigma\sqrt{2\pi}\sqrt{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{\exp \left[\frac{\tau^2\sigma^2 \left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2} \right)^2}{2 \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2} \right)} - \frac{x^2}{\sigma^2} - \frac{\mu^2}{\tau^2} \right]}{\sqrt{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \exp \frac{\tau^2\sigma^2}{2(\tau^2 + \sigma^2)} \left(\frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2} + 2\frac{x\mu}{\tau^2\sigma^2} \right) - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \exp \left[\frac{1}{2(\tau^2 + \sigma^2)} \left(\frac{x^2\tau^2}{\sigma^2} + \frac{\mu^2\sigma^2}{\tau^2} + 2x\mu - \frac{x^2}{\sigma^2}(\tau^2 + \sigma^2) - \frac{\theta^2}{\sigma^2}(\tau^2 + \sigma^2) \right) \right] \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \exp \frac{1}{2(\tau^2 + \sigma^2)} [2x\mu - x^2 - \mu^2] \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \exp \left[-\frac{1}{2(\tau^2 + \sigma^2)} (x - \mu)^2 \right] \end{aligned}$$

$$x \sim (\mu, \sigma^2 + \tau^2)$$

On substitue $m(x)$ dans la formule de $\pi(\theta/x)$ et on trouve que :

$$\begin{aligned} \pi(\theta/x) &= \frac{\frac{1}{\tau\sigma 2\pi} \exp \left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2} \right]}{\frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \exp \left[-\frac{1}{2(\tau^2 + \sigma^2)} (x - \mu)^2 \right]} \\ &= \frac{\sqrt{\tau^2 + \sigma^2}}{\tau\sigma\sqrt{2\pi}} \exp -\frac{1}{2} \left[\frac{(\theta - \mu)^2}{\tau^2} + \frac{(x - \theta)^2}{\sigma^2} - \frac{(x - \mu)^2}{\tau^2 + \sigma^2} \right] \\ &= \frac{\sqrt{\tau^2 + \sigma^2}}{\tau\sigma\sqrt{2\pi}} \exp -\frac{(\tau^2 + \sigma^2)}{2\tau^2\sigma^2} \left(\theta - \frac{\sigma^2\mu + \tau^2x}{\tau^2 + \sigma^2} \right)^2 \end{aligned}$$

donc

$$\theta/x \sim N \left(\frac{\sigma^2\mu + \tau^2x}{\tau^2 + \sigma^2}, \frac{\tau^2\sigma^2}{\tau^2 + \sigma^2} \right)$$

1.6 Lois a priori non informatives :

Dans de telles situations, il est impossible de bâtir une distribution a priori sur des considérations subjectives. On peut alors chercher à utiliser malgré tout des techniques bayésiennes qui intègrent notre ignorance sur les paramètres du modèle, de telles méthodes sont appelées de manière évidente, **non informative**.

1.6.1 La loi a priori de Jeffreys :

Jeffreys (1946; 1961) propose une approche intrinsèque qui évite effectivement le besoin de prendre en compte une structure d'invariance potentielle, tout en étant souvent compatible lorsque cette structure existe, les lois a priori non informatives de **Jeffrey** consiste à assigner à un modèle d'échantillonnage caractérisé par sa vraisemblance $\ell(\varphi_1, \varphi_2 / \text{Données})$ la loi de densité : $P(\varphi_t) = [\det I(\varphi_t)]^{\frac{1}{2}}$ où $I(\varphi_t)$ est la matrice d'information de **Fisher** :

$$I(\varphi_t) = -E \left[\frac{\partial^2}{\partial \varphi_1 \partial \varphi_2} \ln \ell(\varphi_1, \varphi_2 / \text{Données}) \right]$$

La vraisemblance du modèle est :

$$\ell(\varphi_1, \varphi_2 / n_{11}, n_{10}, n_{21}, n_{20}) = \frac{1}{2} \varphi_1^{n_{11}} (1 - \varphi_1)^{n_{10}} \varphi_2 (1 - \varphi_2)^{n_{20}}$$

La matrice de Fisher est :

$$I_n(\varphi_1, \varphi_2) = -E \left(\begin{array}{cc} \frac{\partial^2 \log \ell(\varphi_1, \varphi_2 / n_{11}, n_{10}, n_{21}, n_{20})}{\partial^2 \varphi_1} & \frac{\partial^2 \log \ell(\varphi_1, \varphi_2 / n_{11}, n_{10}, n_{21}, n_{20})}{\partial \varphi_1 \partial \varphi_2} \\ \frac{\partial^2 \log \ell(\varphi_1, \varphi_2 / n_{11}, n_{10}, n_{21}, n_{20})}{\partial \varphi_2 \partial \varphi_1} & \frac{\partial^2 \log \ell(\varphi_1, \varphi_2 / n_{11}, n_{10}, n_{21}, n_{20})}{\partial^2 \varphi_2} \end{array} \right)$$

$$I_n(\varphi_1, \varphi_2) = \left(\begin{array}{cc} \frac{E(T_n^1)}{\varphi_1(1-\varphi_1)} & 0 \\ 0 & \frac{E(T_n^2)}{\varphi_2(1-\varphi_2)} \end{array} \right)$$

$$\text{Avec } E(T_n^1) = \left(n\psi_1 + (z_0 - \psi_1) \frac{1 - h^n}{1 - h} \right)$$

$$E(T_n^2) = \left(n\psi_2 + (z_0 - \psi_2) \frac{1 - h^n}{1 - h} \right)$$

Où, rappelons-le , $\psi_1 = \frac{1 - \varphi_2}{2 - \varphi_1 - \varphi_2}$ et $h = \varphi_1 + \varphi_2 - 1$

La loi a priori de Jeffreys est de densité

$$P(\varphi_1, \varphi_2) = [\det I(\varphi_1, \varphi_2)]^{\frac{1}{2}}$$

$$= (E(T_n^1) E(T_n^2))^{\frac{1}{2}} \varphi_1^{-\frac{1}{2}} (1 - \varphi_1)^{-\frac{1}{2}} \varphi_2^{-\frac{1}{2}} (1 - \varphi_2)^{-\frac{1}{2}}$$

On remarque que cette densité est beaucoup plus complexe que dans le cas du modèle d'échantillonnage avec deux distributions binomiales indépendantes.

La densité de la loi a priori de Jeffreys est simplement le terme :

$$\varphi_1^{-\frac{1}{2}} (1 - \varphi_1)^{-\frac{1}{2}} \varphi_2^{-\frac{1}{2}} (1 - \varphi_2)^{-\frac{1}{2}}$$

Ceci correspond à des lois a priori indépendantes pour $\varphi_1 \sim \text{Beta}(\frac{1}{2}, \frac{1}{2})$ et $\varphi_2 \sim \text{Beta}(\frac{1}{2}, \frac{1}{2})$.

Exemple 1.5 : si $x \sim \mathbf{B}(n, p)$

$$f(x/p) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Donc :

$$I(p) = n \left[\frac{1}{p} + \frac{1}{1-p} \right] = \frac{1}{P(1-P)}$$

Donc la loi de Jeffreys pour ce modèle est :

$$\pi^*(p) \propto [p(1-p)]^{-\frac{1}{2}}$$

et est alors propre, car il s'agit de la distribution $\text{Beta}(\frac{1}{2}, \frac{1}{2})$

Dans le cas où θ est un paramètre multidimensionnel, on définit la matrice d'information de Fisher, pour $\theta \in \mathbb{R}^k$; et $I(\theta)$ aux éléments suivants :

$$I_{ij}(\theta) = -E_0 \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x/\theta) \right] \quad i, j = 1, \dots, k$$

et la loi non informative de Jeffreys est alors définie par :

$$\pi^*(p) \propto [\det(I(\theta))]^{\frac{1}{2}}$$

1.6.2 L'analyse avec la distribution a priori non informative :

La distribution a priori pour le modèle de moyenne et de variance inconnues pour les données de distribution normale est proportionnelle à :

$$\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$$

La fonction de densité de probabilité pour la distribution normale de variable aléatoire est :

$$f(x/\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right] \quad (1,2)$$

où μ est la moyenne et σ^2 est la variance.

avec :

$$f(x_1, \dots, x_n/\mu, \sigma^2) = \prod_{i=1}^{i=n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i - \mu)^2\right] \quad (1,3)$$

La fonction de vraisemblance de l'équation (1,3) et la distribution a priori donnée dans l'équation (1,2) conduisent à la distribution a posteriori jointe de $(\mu; \sigma^2)$ qui est proportionnelle à :

$$\pi(\mu, \sigma^2/X) \propto (\sigma^2)^{-\frac{n}{2}-1} \exp\left[-\sum_{i=1}^{i=n} \frac{1}{2\sigma^2}(x_i - \mu)^2\right]$$

Avec $:X = (x_1, \dots, x_n)$, qui est une fonction de densité propre.

On suppose que μ soit le paramètre d'intérêt, pour obtenir la distribution a posteriori marginale de μ , on doit intégrer par rapport à σ^2 donc :

$$\pi(\mu/X) \propto \int_0^{+\infty} (\sigma^2)^{-\frac{n}{2}-1} \exp\left[-\sum_{i=1}^{i=n} \frac{1}{2\sigma^2}(x_i - \mu)^2\right]$$

donc on peut écrire l'équation comme suit :

$$\begin{aligned} \pi(\mu/X) &\propto \frac{\Gamma\left(\frac{n}{2}\right)}{\left(\sum_{i=1}^{i=n} -\frac{(x_i - \mu)^2}{2}\right)^{\frac{n}{2}}} \int_0^{+\infty} \frac{\left(\sum_{i=1}^{i=n} -\frac{(x_i - \mu)^2}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{n}{2}\right)} (\sigma^2)^{-\frac{n}{2}-1} \exp\left[-\sum_{i=1}^{i=n} \frac{1}{2\sigma^2}(x_i - \mu)^2\right] d\sigma \\ &= \frac{\Gamma\left(\frac{n}{2}\right)}{\left(\sum_{i=1}^{i=n} -\frac{(x_i - \mu)^2}{2}\right)^{\frac{n}{2}}} \end{aligned}$$

on note que :

$$\sum_{i=1}^{i=n} (x_i - \mu)^2 = n(\bar{x} - \mu)^2 + \sum_{i=1}^{i=n} (x_i - \bar{x})^2$$

donc :

$$\pi(\mu/X) \propto \left[1 - \frac{(\bar{x} - \mu)^2}{(n-1) \frac{S^2}{n}} \right]$$

$$\text{où : } \bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i \text{ et } S^2 = \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{(n-1)},$$

Donc la distribution a posteriori marginale de μ est une distribution de Student de $n - 1$ degrés de liberté et de moyenne \bar{x} ($n > 1$), et le paramètre de balance $\frac{S^2}{n}$,

Supposons maintenant que le paramètre d'intérêt est σ^2 , et que μ est le paramètre de nuisance. Pour obtenir la distribution a priori marginale de σ^2 .

nous devons intégrer en dehors de μ : Nous pouvons écrire :

$$\begin{aligned} \pi(\mu/X) &\propto \int_{-\infty}^{+\infty} (\sigma^2)^{-\frac{n}{2}-1} \exp \left[\sum_{i=1}^{i=n} -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right] d\mu \\ &= \int_{-\infty}^{+\infty} (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{(\bar{x} - \mu)^2}{2\frac{\sigma^2}{n}} - \frac{(n-1)S^2}{2\sigma^2} \right] d\mu \\ &\propto (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{(n-1)S^2}{2\sigma^2} \right] \sqrt{\frac{2\pi\sigma^2}{n}} \times \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\frac{2\pi\sigma^2}{n}}} \exp \left[-\frac{(\bar{x} - \mu)^2}{2\frac{\sigma^2}{n}} \right] d\mu \\ &\propto (\sigma^2)^{-\frac{(n-1)}{2}-1} \exp \left[-\frac{(n-1)S^2}{2\sigma^2} \right] \end{aligned}$$

qui est proportionnel à une densité de probabilité de Gamma inverse des paramètres : $\frac{(n-1)}{2}$ et $\frac{(n-1)S^2}{2}$

1.6.3 Facteur de bayes :

L'idée à l'origine de cette notion est de limiter l'importance du choix a priori de a_0 et a_1 introduits ci-dessus.

Le facteur de bayes est défini par :

$$B_{0/1}^\pi = \frac{p^\pi(\Theta_0/X)}{p^\pi(\Theta_1/X)} \times \frac{\pi(\Theta_1)}{\pi(\Theta_0)}$$

Cas spécial : $\Theta_0 = \{\theta_0\}, \Theta_1 = \{\theta_1\}$

$$\begin{aligned}
 B_{01} &= \left\{ \frac{P(\theta_0/x)}{P(\theta_1/x)} \right\} \Big/ \left\{ \frac{P(\theta_0)}{P(\theta_1)} \right\} \\
 &= \left\{ \frac{\frac{\pi(\theta_0/x)}{m(x)}}{\frac{\pi(\theta_1/x)}{m(x)}} \right\} \Big/ \left\{ \frac{P(\theta_0)}{P(\theta_1)} \right\} \\
 &= \left\{ \frac{f(\theta_0/x)\pi(\theta_0)}{f(\theta_1/x)\pi(\theta_1)} \right\} \Big/ \left\{ \frac{\pi(\theta_0)}{\pi(\theta_1)} \right\} \\
 &= \frac{P(x/\theta_0)}{P(x/\theta_1)}
 \end{aligned}$$

A la suite de Jeffrey (1939) et de Good (1952), le facteur de Bayes est désormais un outil à part entière, voir (Kass et Raftery, 1995). En particulier, Jeffrey (1939) a développé une échelle “absolue” pour évaluer le degré de certitude en faveur ou au détriment de H_0 apporté par les données, en l’absence d’un cadre décisionnel véritable, l’échelle de Jeffrey est la suivante :

- (i) Si $0 \prec \log(B_{10}^\pi) \prec \frac{1}{2}$, la certitude que H_0 est faible.
- (ii) Si $\frac{1}{2} \prec \log(B_{10}^\pi) \prec 1$, cette certitude est substantielle.
- (iii) Si $1 \prec \log(B_{10}^\pi) \prec 2$, elle est forte.
- (iiii) Si $\log(B_{10}^\pi) \succ 2$, elle est décisive.

Où $B_{10}^\pi = \frac{1}{B_{01}^\pi}$.

Donc le facteur de Baye permet l’utilisation de test bayésien sans avoir recours à une fonction de perte spécifique.

1.6.4 La fonction d’utilité :

D’après (Robert (2006)), la notion d’utilité (définie comme l’opposé d’une fonction de coût) est utilisée non seulement en statistique, mais aussi en économie, et dans d’autres domaines comme la théorie des jeux où il est nécessaire d’ordonner les conséquences d’action ou de décisions. Conséquences (ou récompenses) sont des notions génériques qui résument l’ensemble des résultats émanant de l’action du décideur. Dans les cas les plus simples, il peut s’agir d’un gain ou d’un coût financier dû à cette décision.

Dans le cas de l’estimation, l’utilité peut être une mesure de la distance entre l’évaluation et la vraie valeur du paramètre. Les bases axiomatiques de l’utilité ont été attribuées

à **Von Neumann et Morgenstern (1947)** et ont mené à des nombreuses extensions, particulièrement en théorie des jeux. Dans un cadre statistique, cette approche a été considérée par **Wald (1950)** et **Ferguson (1967)**; voir aussi **Chamberlain (2000)** pour une connexion avec l'économétrie.

Définition 1.6 : *la fonction d'utilité est utilisée en statistique et en particulier dans le domaine économique. Son but principal est d'ordonner les conséquences d'action d'un décideur. Pour cela on considère que \mathfrak{R} l'espace des récompenses est supposé complètement connu, et d'autre part on suppose qu'il est possible d'ordonner les récompenses. On suppose qu'il existe une relation d'ordre total tel que : $r_1 \leq r_2$ ou : $r_2 \leq r_1$; si $r_1 \leq r_2$ et $r_2 \leq r_3$ alors $r_1 \leq r_3$ permet de choisir la meilleure.*

Nous considérons d'abord dans la section suivante l'approche bayésienne standard des tests, qui repose sur une évaluation des décisions par des coûts 0-1 et comparerons les procédures bayésiennes avec leurs homologues fréquentistes. Nous proposerons ensuite une alternative à l'approche décisionnelle fondée sur des coûts plus adaptés qui mettent en avant l'évaluation ex post pour des procédures de tests, par opposition aux procédures de **Neymann-Pearson** pour lesquelles l'évaluation fonctionne dans un esprit ex ante.

Chapitre 2

Les tests

Introduction :

Nous allons présenter les tests bayésiens et pour plus de détails, voir (Robert 1922, Lehmann 1986, Berger 1985), dont nous reprenons ici quelques notions de base sur les tests bayésiens, ainsi que les développements récents de cette théorie qui possède de nombreux avantages dans la statistique appliquée.

Les divergences exhibées entre les diverses approches possibles du problème des tests ont pour but de mettre en évidence l'aspect incomplet et peu naturel de la modélisation classique qui, au contraire de l'approche bayésienne (pour laquelle la notion de probabilité d'une hypothèse a véritablement un sens, a priori comme a posteriori), doit recourir à des concepts artificiels pour fonder ses procédures. De plus, à l'opposé du cadre de l'estimation ponctuelle où les estimateurs optimaux au sens classique sont minimax et admissible et sont des limites des estimateurs de Bayes. Les réponses bayésiennes et fréquentistes sont différentes ; non seulement en nature, mais aussi numériquement, donc peuvent conduire à des conclusions opposées en pratique.

L'approche des tests d'hypothèses est généralement traitée comme un problème des tests, plutôt que l'estimation plus détaillée.

2.1 Test dans le cas fréquentiste :

L'hypothèse des tests classiques est construite selon le lemme de **Neyman- Pearson** [**Lehmann (1986)**]. Et les résultats dans la règle de décision sont les règles 0 – 1. Ces formes de test bien qu'elles soient optimales dans le sens des fréquentistes strictement mais elles ont des critères pour des directions différentes. Peut être la critique de **Neyman-Pearson** devient active en pratique, car les formules de la théorie de Neyman-Pearson ne sont pas largement utilisées en pratique.

D'autre part, la p-valeur est implicitement utilisée comme une mesure de la preuve pour les hypothèses, donc Kiefer considère la p-valeur comme une évaluation de la vraisemblance de l'hypothèse nulle.

2.2 Théorie des tests paramétriques :

2.2.1 Introduction : test sur l'espérance d'une loi normale de variance connue :

Soit un échantillon $(X_1; \dots; X_n)$ de loi $N(\mu, \sigma^2)$, avec μ inconnue et σ^2 connue. On cherche à tester si l'espérance est égale ou non à une valeur de référence μ_0 :

$$\mu = \mu_0 \quad \text{contre} \quad H_1 : \mu \neq \mu_0$$

Sous l'hypothèse H_0 , la statistique suivante suit une loi $N(0, 1)$:

$$T = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Ainsi, si H_0 est vraie, la valeur de cette statistique pour l'échantillon observé devrait appartenir à l'intervalle $[u_{\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}}]$ avec la probabilité $1 - \alpha$. Ce qui revient à dire que la réalisation de \bar{X} appartient à l'intervalle : $\left[\mu_0 + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \mu_0 + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$

avec une probabilité de $1 - \alpha$

Ainsi, si l'observation \bar{x} de \bar{X} n'est pas dans cet intervalle on peut décider de rejeter l'hypothèse H_0 . Le risque de se tromper en rejetant H_0 est α .

2.3 Vocabulaire des tests :

Un test est un procédé qui permet de trancher entre deux hypothèses, au vu des résultats d'un échantillon : on teste une **hypothèse nulle** contre une **hypothèse alternative**. L'hypothèse nulle H_0 est l'hypothèse que l'on veut contrôler. Elle est toujours de forme **simple**

$$H_0 : \theta = \theta_0$$

où θ_0 est une valeur donnée du paramètre. Le choix de cette hypothèse est fait de manière **conservative** : si on test un médicament, on prendra H_0 l'hypothèse où le médicament n'a pas d'effet. C'est également souvent la plus importante des deux hypothèses

puisque c'est celle dont on contrôle le risque. L'hypothèse alternative H_1 est quant à elle généralement composite :

$$H_1 : \theta \in \Theta_1$$

où Θ_1 est une partie de \mathbb{R} non nécessairement réduite à un élément. Cette hypothèse se ramène souvent à un des cas suivants : $\theta < \theta_0$, $\theta > \theta_0$ (test unilatéraux) ou $\theta \neq \theta_0$ (test bilatéral)

Suivant la justesse de la décision prise à l'issue du test, on est en présence de 4 cas de figure.

<i>Vérité</i> <i>Décision</i>	H_0	H_1
H_0	conclusion correcte	erreur de deuxième espèce
H_1	erreur de première espèce	conclusion correcte

TABLE – Erreurs associés à un test

Exemple 2.1 : (*Importance du choix des hypothèses*). Considérons le test des hypothèses suivantes :

- hypothèse H_0 : le patient doit être hospitalisé,
- hypothèse alternative H_1 : le patient ne doit pas être hospitalisé

L'erreur de première espèce consiste à ne pas hospitaliser un patient qui en avait besoin. Cette erreur est très grave, puisqu'elle peut conduire au décès du patient. Le risque de deuxième espèce, qui consiste à hospitaliser un patient qui n'en avait pas besoin peut s'avérer moins grave.

Pour l'exemple du médicament, l'erreur de première espèce consiste à mettre sur le marché un médicament qui n'a pas d'effet.

2.4 Probabilité d'erreur et risque, puissance de test :

On associe aux erreurs de première et deuxième espèces les probabilités (**risques**) associées (tableau). Le niveau de confiance du test est la probabilité $1 - \alpha$ de ne pas rejeter à raison H_0 . Le risque de première espèce α est le risque de rejeter H_0 à tort. Le risque de deuxième espèce β est le risque de conserver H_0 à tort.

<i>Vérité</i> <i>Décision</i>	H_0	H_1
H_0	niveau de confiance $1 - \alpha$	risque β
H_1	risque α	$1 - \beta$

TABLE – Risques associés à un test

En pratique il est d'usage de **fixer le risque** α : 5%, 1%, 10%. Ainsi, on contrôle le risque associé à l'erreur de première espèce, qui nous l'avons vu est l'erreur la plus grave.

Choisir un risque α trop petit va conduire à ne rejeter que très rarement H_0 (si on ne la rejette pas on ne risque pas de la rejeter à tort!). Au contraire, choisir un risque trop grand va conduire à n'accepter que très rarement α .

Le risque β se déduit alors par le calcul, si la loi sous H_1 est connue. Il varie en sens contraire de α . Ainsi, en diminuant le risque α , on augmente le risque β . On définit alors la puissance du test par $1 - \beta$, qui correspond à la probabilité de rejeter H_0 à raison.

Le choix d'un test sera donc le résultat d'un compromis entre risque de première espèce et puissance du test.

Une fois que l'on a fixé raisonnablement α , il faut choisir une variable de décision, qui doit apporter le maximum d'information sur le problème posé, et dont la loi sera différente selon que H_0 ou H_1 est vraie. La loi sous H_0 doit être connue. On définit alors **la région critique** W qui est l'ensemble des valeurs de la variable de décision qui conduisent à rejeter H_0 au profit de H_1 . Sa forme est déterminée par la nature de H_1 , et sa détermination exacte est donnée par $p(W/H_0) = \alpha$. **La région d'acceptation** est son complémentaire \overline{W} .

2.5 Choix optimal de la statistique de test et de la région de rejet :

Le choix de la statistique de test et de la région de rejet est fait de sorte à maximiser la puissance du test $1-\beta$ pour un risque de première espèce α fixé.

Plaçons nous dans le cadre d'un test entre hypothèses simples :

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1$$

Neyman et Pearson (1933) ont montré que **le test du rapport de vraisemblance** est le test le plus puissant au niveau de confiance α .

Théorème 2.2 : (Neyman et Pearson). *La région critique optimale est définie par les points $x = (x_1, \dots, x_n)$ vérifiant*

$$W = \left\{ x : \frac{L(x, \theta_1)}{L(x, \theta_0)} \succ c_\alpha \right\}$$

La constante c_α , qui dépend de α , est déterminée par $\alpha = P_{\theta_0}(x \in W)$.

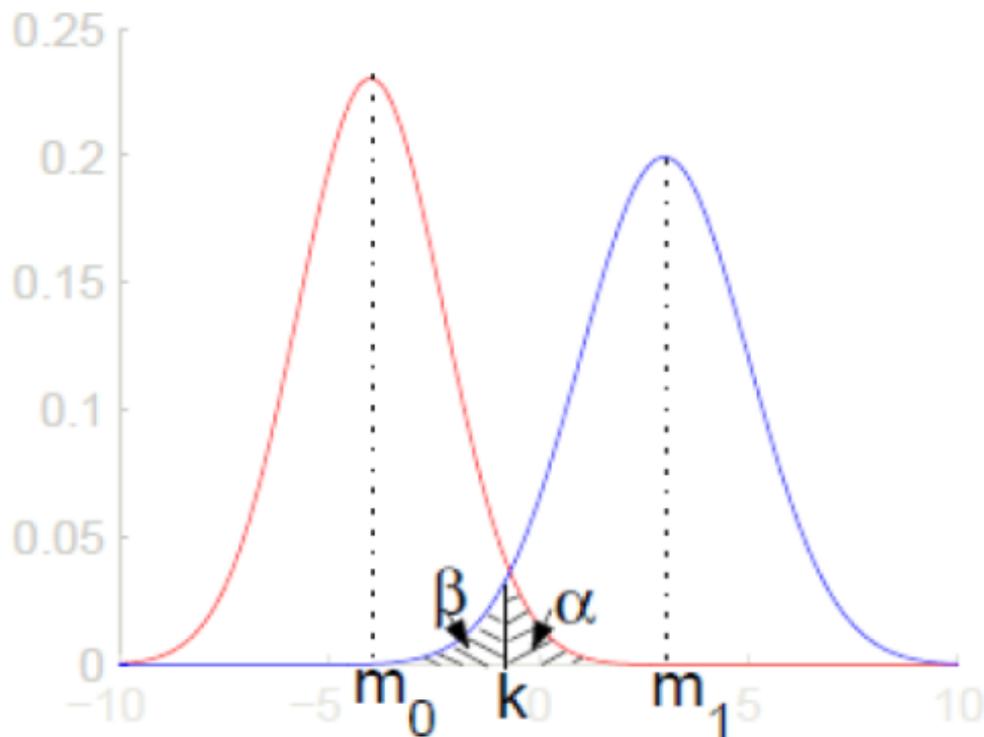


Figure 2.1 - Courbe de puissance d'un test

Exemple 2.3 : Reprenons le test d'introduction, où (X_1, \dots, X_n) est de loi normale de variance σ^2 connue et d'espérance μ inconnue, avec cette fois une hypothèse alternative simple :

$$H_0 : \mu = \mu_0 \text{ contre } H_1 : \mu = \mu_1.$$

On suppose $\mu_0 < \mu_1$. La vraisemblance de l'échantillon gaussien s'écrit

$$L(x, \mu) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

d'où le rapport de vraisemblance

$$\frac{L(x, \theta_1)}{L(x, \theta_0)} = \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n 2(\mu_1 - \mu_0)x_i - \frac{n}{2\sigma^2}(\mu_1^2 - \mu_0^2)\right)$$

Ainsi, $\frac{L(x, \theta_1)}{L(x, \theta_0)} > c_\alpha$ est équivalent à $\bar{x} > \log(c_\alpha) \frac{\sigma^2}{n(\mu_1 - \mu_0)} + \frac{\mu_1 + \mu_0}{2} = C$, où la constante C est déterminée

$$P_{\mu_0}(x \in W) = P_{\mu_0}(\bar{x} > C) = \alpha.$$

La région critique optimale du test de Neyman-Pearson est donc .

$$W = \left\{ x : \bar{x} > \mu_0 + u_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

et on retombe bien sur le test « intuitif » de l'introduction.

Dans le cas où l'hypothèse alternative est composite ($\theta \in \Theta_1$), la puissance du test est fonction de θ : $1 - \beta(\theta)$ est appelée la fonction puissance du test.

Un test est dit uniformément le plus puissant (UPP) si quelque soit la valeur de θ appartenant à l'hypothèse alternative, sa puissance est supérieure à celle de tout autre test.

Exemple 2.4 : On a vu précédemment pour le test $H_0 : \mu = \mu_0$ contre

$H_1 : \mu = \mu_1 > \mu_0$ que la région critique ne dépend pas de μ_1 , et qu'elle est donc la même pour tout $\mu_1 > \mu_0$.

Le test est donc UPP pour $H_0 : \mu = \mu_0$ contre $H_1 : \mu > \mu_0$.

Si cette fois $\mu_1 < \mu_0$, on obtient encore un test UPP $H_0 : \mu = \mu_0$ contre $H_1 : \mu < \mu_0$, mais différent du précédent. Il n'existe donc pas de test UPP pour

$$H_0 : \mu = \mu_0 \text{ contre } H_1 : \mu \neq \mu_0$$

2.6 Utilisation de la puissance de test :

Dans le cas d'un test entre deux hypothèses simples avec variance σ^2 connue $H_0 : \mu = \mu_0$ contre $H_1 : \mu = \mu_0 + \delta$

nous avons vu que la région critique avait la forme

$$W = \left\{ x : \bar{x} > \mu_0 + u_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

On peut calculer le risque de second espèce :

$$\beta = p(\text{décider } H_0 / H_1) = \Phi(u_{1-\alpha} \frac{\delta\sqrt{n}}{\sigma}) :$$

La puissance du test, $1 - \beta$, est donc fonction de α , n et δ . En considérant α et n fixés,

on peut représenter la courbe de puissance du test par la Figure (2.2).

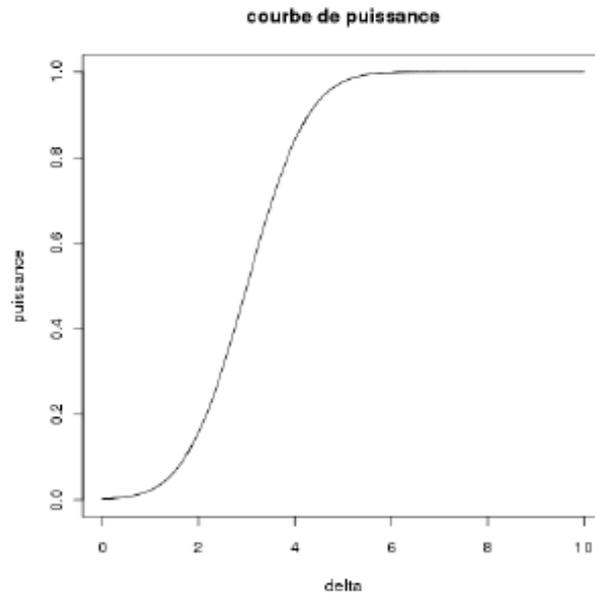


Figure 2.2- illustration de la règle de décision

La courbe de puissance peut donc permettre

- de choisir entre plusieurs tests en fonction de leur courbes de puissance (que l'on veut la plus forte possible, i.e. proche de la droite d'ordonnée 1) ,
- pour un problème donné, dans lequel α et δ sont fixés, on pourra choisir le nombre de sujets nécessaire n pour atteindre une puissance donnée à l'aide de l'équation (2.1).

2.6.1 Résumé :

La démarche de construction d'un test est la suivante :

- choix de H_0 et H_1 ,
- détermination de la variable de décision,
- allure de la région critique en fonction de H_1 ,
- calcul de la région critique en fonction de α ,
- calcul de la valeur expérimentale de la variable de décision,
- conclusion : rejet ou acceptation de H_0 .

2.6.2 p-value :

En pratique, plutôt que de calculer la région critique en fonction de α , on préfère donner un seuil critique α^* , appelée p-value, qui est la plus grande valeur de α conduisant à ne pas rejeter H_0 . Cette information permet au lecteur de conclure à l'acceptation de H_0 pour tout risque de première espèce $\alpha \leq \alpha^*$, et à son rejet pour tout $\alpha > \alpha^*$.

2.7 Tests sur une population :

Nous pouvons maintenant présenter les différents tests statistiques classiques, obtenus par la méthode de Neyman-Pearson lorsque les échantillons sont gaussiens (voir de grandes tailles). Dans le cas de petits échantillons non gaussiens, des alternatives non paramétriques seront présentées.

2.8 Test sur le caractère central d'une population :

2.8.1 Cas d'un échantillon grand ou gaussien :

Soit un n-échantillon (X_1, \dots, X_n) issu d'une population de moyenne μ et de variance σ^2 . Nous supposons que au moins l'une des deux conditions suivantes est satisfaite :

- la population est de loi normale,
- l'échantillon est de taille n suffisamment grande ($n \geq 30$).

Test $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$ lorsque σ^2 est connue La statistique de test est :

$$U = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Sous H_0 , cette statistique suit une loi normale centrée réduite d'après les conditions précédentes (via le théorème centrale limite si seule la seconde condition est satisfaite).

La région critique, définie par $|U| > k$, se traduit par $|\bar{X} - \mu_0| > u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, où $u_{\frac{\alpha}{2}}$ est le quantile de la loi normale centrée réduite d'ordre $\frac{\alpha}{2}$.

Ainsi,

$$\text{on rejette } H_0 \text{ si } |\bar{x} - \mu_0| > u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Remarque 2.5 : (Calcul de la p-value). Pour ce test, on rejette H_0 dès que $\frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} > u_{\frac{\alpha}{2}}$.

La p-value est la valeur critique α^* de α telle que $\frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} > u_{\frac{\alpha^*}{2}}$, d'où $\alpha^* = 2\Phi\left(-\frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}}\right)$

avec Φ la fonction de répartition de la loi normale centrée réduite. Ainsi, dès que l'on choisit un risque α plus grand que α^* , on a $-u_{\frac{\alpha^*}{2}}$ et donc on rejette H_0 . Au contraire, si le risque est plus petit, on aura cette fois $\frac{|\bar{x}-\mu_0|}{\frac{\sigma}{\sqrt{n}}} = -u_{\frac{\alpha^*}{2}} \prec -u_{\frac{\alpha}{2}}$ et on conserve H_0 .

Remarque 2.6 : (Tests unilatéraux). Si le test est unilatéral, $H_0 : \mu = \mu_0$ contre $H_1 : \mu < \mu_0$, on rejette H_0 si la vraie valeur de μ est trop éloignée inférieurement de μ_0 , ce qui se traduit par
$$\bar{x} < \mu_0 - u_{\alpha} \frac{\sigma}{\sqrt{n}}.$$

Si le test est $H_0 : \mu = \mu_0$ contre $H_1 : \mu > \mu_0$, on rejette H_0 si $\bar{x} > \mu_0 + u_{\alpha} \frac{\sigma}{\sqrt{n}}$.

Test $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$ lorsque σ^2 est inconnue Ce test est généralement connu sous le nom de **test de Student**.

Dans ce cas la variance σ^2 est estimée par son estimateur S^2 . La statistique de test est :

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

qui suit une loi de Student à $n - 1$ degré de liberté.

La conclusion du test devient alors :

$$\text{on rejette } H_0 \text{ si } |\bar{x} - \mu_0| > t_{n-1, \frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

où $t_{n-1, \frac{\alpha}{2}}$ est le quantile d'ordre $\frac{\alpha}{2}$ de la loi de Student à $n - 1$ degrés de liberté, et :

$$S^2 = \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{(n - 1)}$$

Logiciel **R** : les tests sur la moyenne s'effectuent à l'aide de la fonction *t.test*.

Logiciel **SAS** : *proc ttest* ou *proc univariate*

Attention : seuls des tests bilatéraux sont possibles sous **SAS**. Dans le cas d'un test unilatéral, il conviendra donc d'ajuster la p-value (en la divisant par deux), et de s'assurer avant de rejeter H_0 que la statistique de test est bien du bon côté de l'hypothèse nulle.

2.8.2 Cas d'un petit échantillon non gaussien :

Le caractère central de la population sera testé cette fois, non plus en travaillant sur l'espérance de la loi comme précédemment, mais en testant la symétrie de la distribution par rapport à une valeur μ_0 d'intérêt. Nous supposons, sans perte de généralité, que $\mu_0 = 0$.

Les hypothèses que nous testons sont donc :

- $H_0 : F(x) = 1 - F(-x)$ la distribution est symétrique par rapport à 0
- contre $H_1 : F(x + \delta) = 1 - F(\delta - x)$ la distribution est symétrique par rapport à δ .

où F est la fonction de répartition de la variable aléatoire testée.

Les tests que nous allons présenter dans cette section seront basés sur les rangs des observations et nécessitent quelques notions introduites dans le paragraphe suivant.

2.9 L'approche bayésienne des tests :

Dans ce chapitre nous nous limiterons à la situation suivante. On suppose que l'espace Θ des paramètres est partitionné en Θ_0 et Θ_1 , et que $P(\theta \in \Theta_k) > 0$ où $k = 0, 1$ on souhaite tester : $H_0 : \theta \in \Theta_0$ contre l'alternative $H_1 : \theta \in \Theta_1$. En statistique bayésienne, la réponse à un tel test repose sur les probabilités a posteriori des hypothèses H_0 et H_1 :

$$P(H_0/x) = P(\theta \in \Theta_0/x) = \int_{\Theta_0} \pi(\theta/x) d\theta$$

A noter que $P(H_1/x) = 1 - P(H_0/x)$. En pratique, une hypothèse (H_0 ou H_1) est acceptée dès que sa probabilité a posteriori est jugée suffisamment forte (typiquement supérieur à 0.9 ou à 0.95). Il se peut qu'aucune des probabilités a posteriori excède 0.9 ; deux attitudes sont alors possibles :

- soit on ne prend pas de décision, et on décide par exemple de recueillir davantage d'observations ;

- soit on choisit l'hypothèse dont la probabilité a posteriori est la plus grande ;

A noter que cela a aussi un sens de définir les probabilités a priori de chacune des deux hypothèses H_0 et H_1 . Ainsi on a :

$$P(H_0) = P(\theta \in \Theta_0) = \int_{\Theta_0} \pi(\theta) d\theta$$

Il convient en n de noter que les écritures $P(H_0)$ et $P(H_0/x)$ n'ont pas de sens en statistique classique.

2.9.1 La fonction de coût :

On appelle fonction de coût, toute fonction L de $\Theta \times A$ dans \mathbb{R} .

$L(\theta, a)$ évalue le coût d'une décision a quand le paramètre vaut θ . Elle permet donc, en quelque sorte, de quantifier la perte encourue par une mauvaise décision, une mauvaise évaluation de .

2.9.2 Fonctions de coût usuelles :

Coût quadratique :

La fonction de coût quadratique est la fonction définie par :

$$L(\theta, \delta(x)) = (\theta - \delta(x))^2 = \omega(\theta) (\theta - \delta(x))$$

Une variante de cette fonction de coût est une fonction de coût quadratique pondérée de la forme : $L(\theta, \delta(x))$

Coût 0-1 :

Définition 2.7 : On appelle coût 0 -1, l'application L définie par .:

$$L(\theta, \delta(x)) = \begin{cases} 0 & \text{si la décision est bonne} \\ 1 & \text{sinon} \end{cases}$$

on retrouve en utilisant cette fonction de coût, les résultats de la théorie des tests d'hypothèses.

Un problème de test est un problème de choix (de prise de décision) entre $H_0 : \theta \in \Theta_0$ et $H_1 : \theta \in \Theta_1$

On définit donc la décision de la manière suivante :

$L(X) = 1$ on accepte H_0 ;

$L(X) = 0$ on rejette H_0 . (nb : ceci ne dépend pas de θ).

On a un espace d'actions de la forme : $A = \{0; 1\}$

Soit ω la région de rejet i.e. le sous ensemble de \mathfrak{N} qui conduit à rejeter H_0 : On peut construire une fonction de coût de la manière suivant :

supposons : $\theta \in \Theta_0$

Si $x \in \omega$ on prend la décision de rejeter i.e. $\delta(x) = 0$ mais la décision n'est pas bonne, on va pénaliser et $L(\theta, \delta(x)) = 1$

Si $x \notin \omega$ on ne rejette pas, on prend la décision $\delta(x) = 1$, la décision est bonne $L(\theta, \delta(x)) = 0$.

Le coût s'écrit donc :

$$L(\theta, \delta(x)) = \begin{cases} 1 - \delta(x) & \text{si } \theta \in \Theta_0 \\ \delta(x) & \text{sinon} \end{cases}$$

Ce qu'on peut écrire $L(\theta, \delta(x)) = 1(x \in \omega)$ et on calcule on fonction de risque :

$$R(\theta, \delta) = E[L(\theta, \delta(x))] = \int_{\mathfrak{X}} L(\theta, \delta(x)) dp_{\theta}(x \in \omega), \theta \in \Theta_0$$

2.10 La comparaison entre les tests fréquentiste et les testes bayesienne :

On distingue différentes catégories de tests :

— les tests dans le cas fréquentiste ont pour objet de tester une certaine hypothèse relative à un ou plusieurs paramètres d'une variable aléatoire de loi spécifiée (généralement supposée normale). Lorsque le test est toujours valide pour des variables non gaussiennes, on dit que le test est robuste (à la loi).

— les tests dans le cas bayesienne qui portent généralement sur la fonction de répartition de la variable aléatoire, sa densité...

Dans ce cours, nous classons les tests en fonction de leur fonctionnalité :

- Tests sur une population :
- test sur le caractère centrale d'une population,
- test sur la variance,
- test sur une proportion,
- test de l'aléatoire d'un échantillon,
- test d'ajustement à une loi spécifiée,
- test de liaison entre variables (quantitatives, qualitatives, mixtes)
- Tests de comparaison de deux populations

Conclusion Générale :

Dans ce travail , nous avons étudié l'approche bayésienne . Au premier lieu, nous avons abordé une modélisation bayésienne parce qu'elle est plus souple et objective dans le traitement statistique des données.

nous avons étudié, au second lieu la théorie mathématique des tests bayésiens, et enfin fait la comparaison entre les tests dans le cas fréquentiste et le cas bayésien.

Bibliographie

[1]

[1] Christian P Robert (2006). Le choix bayésien Principe et pratique. 19-17, 49-48, 130, 139-140.

[2] Jean Michel Marin, Christian P Robert. Paris. Les bases de la statistique bayésienne. 89.

[3] Jerome Dupuis. (2007). Statistique bayésienne et algorithmes. 10-16

[4] Judith ROUSSEAU. (2010). STATISTIQUE BAYÉSIENNE. 22, 39.

[5] Lionel RIOUFRANÇA. (2009). Statistique Bayésienne Elément de Culture Générale. 5, 18, 109.

[6]Julien JACQUES.Statistiques inférentielles.33-38

[7]Judith Rousseau(2009-2010).Statistique Bayésienne.21-25 ;35-39.

[8]Jean-Jacques BoreuxEricParent.Jacques Bernier.Pratique du calcul bayesien7,9,12

[9]Lionel RIOU FRANCA.(2009).Statistique Bayesienne El ements de Culture Générale.4-9,103-107,118,129.