

TUTORIEL

L'inférence bayésienne pour l'analyse des données expérimentales

Bruno Lecoutre¹

ERIS, Laboratoire de Mathématiques Raphaël Salem
UMR 6085 C.N.R.S. et Université de Rouen
Avenue de l'Université, BP 12, 76801 Saint-Etienne-du-Rouvray
bruno.lecoutre@univ-rouen.fr
Internet : <http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/Eris>

Résumé Ce tutoriel se situe dans la ligne des articles précédemment publiés dans *la Revue de Modulad* : Lecoutre (1996b, 2005/1997, 2005) ; Lecoutre, Poitevineau & Lecoutre (2005). Il s'appuie sur l'utilisation de programmes informatiques qui ont également fait l'objet d'une présentation dans un numéro précédent (Lecoutre & Poitevineau, 2005). La motivation de ce tutoriel est avant tout méthodologique, et le choix du cadre bayésien ne devrait pas paraître idéologique. Plus précisément, l'objectif est d'apporter aux questions essentielles soulevées par l'analyse des données expérimentales des réponses mieux adaptées que les tests de signification de l'hypothèse nulle. Basées sur des définitions opérationnelles plus utiles que les procédures traditionnelles (tests, intervalles de confiance), les méthodes bayésiennes offrent une souplesse considérable, en rendant tous les choix explicites. De plus, la philosophie bayésienne met en avant la nécessité de réfléchir sur l'information fournie par les données disponibles – “qu'est-ce que les données ont à dire ?” – au lieu d'appliquer des procédures rituelles.

Des procédures bayésiennes *de routine* sont désormais faciles à mettre en œuvre pour toutes les situations courantes. Leurs résultats peuvent être présentés sous une forme intuitivement séduisante et facilement interprétable. Elles ouvrent une nouvelle voie prometteuse dans la méthodologie statistique ¹.

INTRODUCTION

La motivation méthodologique

“The test provides neither the necessary nor the sufficient scope or type of knowledge that basic scientific social research requires.” (Morrison & Henkel, 1969)

La motivation méthodologique résulte de l'examen de l'état actuel de l'utilisation de l'inférence statistique dans la recherche expérimentale. Celle-ci est confrontée à une situation paradoxale.

¹Je remercie Jacques Poitevineau pour son aide dans la tâche ingrate de relecture de ce texte. J'assume l'entière responsabilité des fautes et erreurs qui peuvent subsister.

Les tests ne répondent pas aux bonnes questions

D'une part, les tests de signification de l'hypothèse nulle (en anglais, *Null Hypothesis Significance Tests*, "NHST") sont considérés dans la plupart des publications scientifiques comme une norme incontournable et apparaissent souvent comme une garantie de scientificité. Mais, d'autre part, ces tests conduisent à d'innombrables erreurs d'interprétations et mauvais usages. Leur utilisation a d'ailleurs été explicitement proscrite par les scientifiques les plus éminents et les plus avertis, tant sur des arguments théoriques que sur de considérations méthodologiques.

Pour interpréter leurs données, les utilisateurs doivent recourir à une synthèse "naïve" des résultats des tests de signification avec d'autres informations, d'où un malaise qui va grandissant. Pour répondre à ce malaise, un processus de définition de nouvelles normes de publication pour la recherche expérimentale a été engagé. C'est pourquoi l'époque actuelle est une charnière cruciale.

Vers de nouvelles normes de publications

Des changements dans la façon de rapporter les résultats expérimentaux sont de fait de plus en plus exigés par les revues expérimentales ceci dans tous les domaines. Ces changements concernent tout particulièrement la présentation et l'interprétation de la grandeur des effets (les "*effect sizes*"). Il s'agit de fournir des indicateurs de ces effets ainsi que leurs estimations par intervalle ("*interval estimates*"), en plus ou à la place des tests.

En Psychologie par exemple, cette nécessité de changements a été rendue officielle par l'*American Psychological Association* (APA, 2001). Les recommandations faites à cette occasion (voir Wilkinson et al., 1999) sont révélatrices de ce que pourrait être l'évolution des pratiques. Mais, si on les examine en détail, elles apparaissent en fait à la fois partiellement redondantes et conceptuellement incohérentes. Ces recommandations ne peuvent en effet qu'aboutir à perpétuer des recettes et des rituels – calcul de puissance pour déterminer les effectifs, utilisation des "*p-values*" (on n'abandonne pas les tests), intervalles de confiance (en plus des tests) – qui seraient combinés sans fournir une réelle pensée statistique. On peut craindre dans ces conditions que les utilisateurs de la statistique continuent à se focaliser sur la signification statistique du résultat – notamment en se demandant seulement si l'intervalle de confiance contient la valeur de l'hypothèse nulle – sans considérer réellement l'information supplémentaire apportée par cet intervalle.

"Consequently we automatically ask ourselves "won't the Bayesian choice be unavoidable?" (Lecoutre et al., 2001)

Dans ces conditions, on ne peut que remettre en question le cadre conceptuel des procédures d'inférence traditionnelles – c'est-à-dire l'approche *fréquentiste* – et nous demander si l'approche *bayésienne* ne sera pas tôt ou tard incontournable.

Plan de l'exposé

La présentation est organisée en quatre parties.

Partie I - Les aspects conceptuels : Inférence fréquentiste ou bayésienne ? La première partie de ce tutoriel traitera des aspects conceptuels de base concernant les

différences entre les approches fréquentistes et bayésiennes. Je conclurai que pour l'analyse des données expérimentales une approche bayésienne "objective" est à la fois *désirable* et *faisable*. Dans ce cadre, je n'envisagerai donc qu'avec beaucoup de réserves une approche "subjectiviste" ; j'écarterai également une conception prioritairement décisionnelle de l'inférence statistique.

Partie II - La pratique : Quelques situations de base. L'objectif de la seconde partie sera de démontrer la faisabilité des méthodes bayésiennes, à partir d'un certain nombre de situations de base. La présentation sera essentiellement illustrative, et le moins possible technique. Il s'agira d'introduire des procédures de routine à partir de problèmes relativement simples d'inférence sur des proportions et des moyennes. On traitera à la fois la mise en œuvre pratique de ces procédures – à l'aide de programmes informatiques – et leurs apports conceptuels et méthodologiques, tant pour l'analyse de données que pour la planification et la conduite des expériences. Dans chaque situation, les solutions bayésiennes seront mises en parallèle avec les procédures fréquentistes.

Je développerai plus particulièrement l'inférence sur des proportions. Les problèmes de comparaisons de moyennes – qui se présentent traditionnellement en analyse de variance – ont en effet été largement traités par ailleurs – notamment dans le cadre de ce que nous appelons l'*Analyse Bayésienne des Comparaisons* – et on pourra notamment se rapporter pour compléter le présent exposé aux références suivantes : Lecoutre (1984, 1996a) ; Lecoutre et Poitevineau (2005) ; Lecoutre (2006).

Partie III - Les aspects techniques : Quelques outils de base. La troisième partie sera technique. Il s'agira d'abord d'illustrer les principes des calculs formels nécessaires à l'inférence bayésienne et de fournir quelques outils utiles pour ces calculs. On abordera également brièvement des techniques de base d'intégration numérique – déterministe ou par simulation. Ces dernières techniques revêtent une importance de plus en plus grande dans l'inférence bayésienne (même si elles ne sont pas limitées à celle-ci) ; elles permettent en effet de résoudre relativement facilement les problèmes pratiques de calcul liés aux situations les plus complexes. Une présentation plus approfondie déborderait le cadre du présent exposé et serait plutôt l'objet d'un autre tutoriel.

Partie IV - Retour sur les aspects conceptuels : L'interface de l'inférence fréquentiste et de l'inférence bayésienne. La quatrième partie s'adressera à ceux qui souhaitent approfondir les liens conceptuels et techniques entre les deux approches, qui auront été mis en avant dans les différents exemples d'application.

En conclusion. Je reviendrai brièvement sur les avantages de l'approche bayésienne et aborderai quelques thèmes "pour aller plus loin" et préparer à d'autres lectures.

PARTIE I - LES ASPECTS CONCEPTUELS : INFÉRENCE FRÉQUENTISTE OU BAYÉSIENNE ?

1 La probabilité et l'inférence statistique

De nos jours la probabilité a au moins deux définitions principales (déjà présentes chez Bernoulli, 1713).

1. La probabilité est la fréquence sur le long terme de l'occurrence d'un événement, soit dans une suite d'essais répétés, soit dans un ensemble de systèmes "identiquement" préparés ².

C'est la conception *fréquentiste*, qui semble faire de la probabilité une propriété observable ("objective"), existant dans la nature indépendamment de nous.

2. La probabilité est une mesure du degré de croyance (ou de confiance) dans l'occurrence d'un événement ou dans la véracité d'une proposition.

C'est la conception *bayésienne*. Il n'est souvent pas évident d'attribuer une probabilité fréquentiste à un événement unique, puisque cela nécessite d'imaginer un ensemble de référence d'événements ou une série d'expériences répétées afin d'obtenir des fréquences empiriques. Malheureusement, de tels ensembles sont rarement disponibles pour l'attribution des probabilités dans les problèmes réels. Par contraste la définition bayésienne est plus générale : il n'est pas *conceptuellement* problématique d'attribuer une probabilité à un événement unique.

Les probabilités bayésiennes sont considérées par certains comme le résultat d'une appréciation subjective d'une situation par un observateur (Savage, 1954; de Finetti, 1974). Mais elle peuvent tout aussi bien servir à décrire une connaissance objective, en particulier basée sur des arguments de symétrie ou sur des fréquences.

On remarquera que la définition bayésienne est en accord avec le sens du mot probabilité dans le langage de tous les jours : la conception bayésienne apparaît donc beaucoup plus proche de la façon dont les gens raisonnent intuitivement en présence d'incertitude ³.

1.1 Les termes du débat

"[...] *whether the probabilities should only refer to data and be based on frequency or whether they should also apply to hypotheses and be regarded as measures of beliefs.*" (Lindley, 1993)

²Dans une définition plus précise, la probabilité est la valeur limite de la fréquence dans une suite infinie d'essais, mais c'est ici un point mineur.

³Pour ne prendre qu'un exemple, comment interprétez-vous *l'indice de confiance* (par exemple 4/5) figurant dans les bulletins météorologiques fournis par météo France? Tous ceux que j'ai interrogés l'interprètent comme une probabilité bayésienne : "il y a 4 chances sur 5 que la prévision se réalise", alors que cet indice n'est pas du tout défini comme une telle probabilité : voir http://www.meteofrance.com/FR/glossaire/designation/1036_curieux.view.jsp. On notera que cette définition est réservée aux "curieux", ce qui favorise plus ou moins volontairement l'interprétation bayésienne. Il est intéressant à ce propos de constater que le service météorologique d'Environnement Canada donne au contraire explicitement une définition bayésienne de la "probabilité de précipitations", définie comme "une estimation numérique subjective des risques de précipitations mesurables à tout point du secteur visé. Par exemple, si la probabilité de pluie est de 40 p.100 pour aujourd'hui, il existe quatre chances sur 10 pour qu'il pleuve" (<http://www.smc.ec.gc.ca/cd/brochures/probability.f.cfm>). Cela n'implique évidemment pas que cette estimation soit moins objective que l'indice de météo France.

L'inférence statistique fait typiquement intervenir à la fois des quantités connues – les données observées – et des quantités inconnues – les paramètres et les données qui n'ont pas été observées. Le débat se pose alors en ces termes : “est-ce que les probabilités devraient seulement être relatives aux données et être basées sur des fréquences ou est-ce qu'elles doivent aussi d'appliquer aux paramètres et être considérées comme des mesures de croyance ?”.

1.1.1 L'inférence fréquentiste

Dans l'inférence fréquentiste toutes les probabilités sont conditionnelles aux paramètres qui sont supposés connus. Cela conduit en particulier :

- aux tests de signification, dans lesquels la valeur d'au moins un paramètre est fixée par hypothèse ;
- aux intervalles de confiance.

Mais les paramètres ne peuvent pas et ne doivent pas être probabilisés (des fréquences empiriques ne sont pas disponibles)

1.1.2 L'inférence bayésienne

Dans l'inférence bayésienne au contraire, les paramètres peuvent aussi être probabilisés. Il en résulte des distributions de probabilité qui expriment notre incertitude :

- avant les observations (elle ne dépendent pas des données) : ce sont les probabilités initiales ou (*a priori*) ;
- après les observations (conditionnelles aux données) : ce sont les probabilités révisées, finales ou *a posteriori* ;
- relatives à des données futures : ce sont les probabilités prédictives (avant ou après les observations).

Remarque sur la terminologie. Les appellations *initiales* et *finales* seront préférées ici à *a priori* et *a posteriori* qui renvoient davantage à une conception “subjective”. De même l'appellation *distribution* (de probabilité) sera préférée à *loi* qui évoque une conception particulière du hasard et de la probabilité.

2 Une illustration simple

Comme illustration simple, considérons la situation suivante. On suppose une population finie d'effectif $N = 20$ avec une variable dichotomique “succès/échec” et une proportion φ de succès qui est le paramètre inconnu. Un échantillon de $n = 5$ observations – dont on supposera qu'il a été tiré au sort sans remise – a été observé, d'où les données *connues* :

$$\boxed{0 \mid 0 \mid 0 \mid 1 \mid 0} \quad n = 5 \quad f = \frac{1}{5}$$

Le raisonnement inductif est fondamentalement une généralisation d'une quantité connue – les données – à une quantité *inconnue* – ici le paramètre φ .

2.1 Solution fréquentiste

Dans le cadre fréquentiste nous n'avons pas de probabilités, et par conséquent pas d'inférence inductive possible.

De l'inconnu vers le connu

Aussi l'inférence fréquentiste doit retourner la situation. Mais nous n'avons pas davantage de probabilités... à moins que nous ne fixions une valeur du paramètre. Supposons par exemple $\varphi = 15/20 = 0.75$. Nous obtenons alors des *probabilités d'échantillonnage*, qui peuvent être définies comme des fréquences mettant en jeu un très grand nombre de répétitions *imaginaires* des observations. Nous pouvons obtenir ces fréquences en simulant un grand nombre de tirages sans remise de 5 boules dans une urne contenant 20 boules, dont 15 d'une couleur et 5 d'une autre couleur. Mais il n'est pas nécessaire de recourir à une définition fréquentiste pour obtenir les probabilités d'échantillonnage.

Ici il y a 15 504 échantillons différents possibles (dont un particulier a été observé) et les probabilités d'échantillonnage sont données par une distribution *Hypergéométrique*. Le tableau ci-après donne ces probabilités, ainsi que les fréquences obtenues par deux simulations, l'une de 10 000 tirages et l'autre de 1 000 000 de tirages.

$$\varphi = 15/20 = 0.75$$

Nombre de succès	Nombres d'échantillons	Probabilités d'échantillonnage	Fréquences	
			10 ⁴ tirages	10 ⁶ tirages
$f = 0/5$	1/15 504	= 0.00006	0	0.00006
$f = 1/5$	75/15 504	= 0.0048	0.0054	0.0049
$f = 2/5$	1 050/15 504	= 0.0677	0.066	0.068
$f = 3/5$	4 550/15 504	= 0.2935	0.294	0.293
$f = 4/5$	6 825/15 504	= 0.4402	0.447	0.441
$f = 5/5$	3 003/15 504	= 0.1937	0.187	0.194

2.1.1 Test de signification

Ces probabilités d'échantillonnage sont utilisées pour définir un test de signification. La valeur du paramètre est fixée par l'*hypothèse nulle*, par exemple $H_0 : \varphi = 0.75$. Si $\varphi = 0.75$, on trouve dans seulement 0.49% des répétitions (soit la proportion $0.0049=0.00006+0.0049$) une valeur inférieure ou égale à l'observation ($f \leq 1/5$). Le résultat est dit "significatif" ($p = 0.0049$) : sur la base des données observées, l'hypothèse nulle est rejetée (je n'entre pas ici dans la discussion "test unilatéral/bilatéral", ni dans le choix du "risque" α qui ne sont pas pertinents pour mon propos).

"A hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred." (Jeffreys, 1961)

Mais, comme le soulignait ironiquement Jeffreys, cette conclusion est basée sur la probabilité **des échantillons qui n'ont pas été observés**.

Considérons un autre exemple d'hypothèse nulle, $H_0 : \varphi = 0.50$.

$$\varphi = 10/20 = 0.50$$

Nombre de succès	Nombres d'échantillons	Probabilités d'échantillonnage
$f = 0/5$	252/15 504	= 0.016
$f = 1/5$	2 100/15 504	= 0.135
$f = 2/5$	5 400/15 504	= 0.348
$f = 3/5$	5 400/15 504	= 0.348
$f = 4/5$	2 100/15 504	= 0.135
$f = 5/5$	252/15 504	= 0.016

Dans ce cas, si $\varphi = 0.50$, on trouve dans 15.2% des répétitions (soit la proportion $0.0152 = 0.016+0.135$) une valeur inférieure ou égale à l'observation ($f \leq 1/5$). Le résultat est dit "non significatif" ($p = 0.152$) : sur la base des données observées, l'hypothèse nulle n'est pas rejetée. À l'évidence cela ne prouve pas que $\varphi = 0.50$!

2.1.2 Intervalle de confiance

Un intervalle de confiance peut être construit comme l'ensemble des valeurs possibles du paramètre qui ne sont pas rejetées par les données à un seuil α fixé. Par exemple, pour $\alpha = 0.05$, nous obtenons ici l'intervalle de confiance 95% :

$$[0.05, 0.60]$$

Comment interpréter la confiance ? L'interprétation est basée sur l'énoncé *universel* :

"Quelle que soit la valeur fixée du paramètre, dans 95% (au moins) des répétitions l'intervalle qui serait calculé contiendrait cette valeur."

Cette propriété peut être vérifiée par exemple pour $\varphi = 0.50$:

Répétitions imaginaires des observations

Valeur observée	Intervalle de confiance	Probabilité d'échantillonnage
$f = 0$ (0/5)	[0 , 0.45]	0.016
$f = 0.20$ (1/5)	[0.05 , 0.60]	0.135*
$f = 0.40$ (2/5)	[0.10 , 0.75]	0.348*
$f = 0.60$ (3/5)	[0.15 , 0.90]	0.348*
$f = 0.80$ (4/5)	[0.20 , 0.95]	0.135*
$f = 1$ (5/5)	[0.25 , 1]	0.016*

*contient $\varphi = 0.50$

Ici dans 98.4% des répétitions l'intervalle contient la valeur fixée $\varphi = 0.50$.

Mais cette interprétation est pour le moins étrange puisqu'elle **ne fait pas intervenir les données observées** !

2.2 Solution bayésienne

Du connu vers l'inconnu

Revenons au raisonnement inductif, en partant des données connues, et en adoptant le point de vue bayésien. Nous pouvons utiliser, **en plus** des probabilités d'échantillonnage, des probabilités qui expriment notre incertitude sur toutes les valeurs possibles du paramètre.

Dans l'inférence bayésienne, nous considérons les probabilités fréquentistes – non pas d'échantillons *imaginaires* – mais des *données observées*, ceci **pour toutes les valeurs possibles du paramètre**. C'est la *fonction de vraisemblance* $Pr(f = 1/5 | \varphi)$, que l'on notera encore $v(\varphi | \text{données})$.

$$Pr(f = 1/5 | \varphi) = v(\varphi | \text{données})$$

↪ **Fonction de vraisemblance**

$\varphi = 0/20 \rightarrow 0$	$\varphi = 10/20 \rightarrow 0.135$
$\varphi = 1/20 \rightarrow 0.250$	$\varphi = 11/20 \rightarrow 0.089$
$\varphi = 2/20 \rightarrow 0.395$	$\varphi = 12/20 \rightarrow 0.054$
$\varphi = 3/20 \rightarrow 0.461$	$\varphi = 13/20 \rightarrow 0.029$
$\varphi = 4/20 \rightarrow 0.470$	$\varphi = 14/20 \rightarrow 0.014$
$\varphi = 5/20 \rightarrow 0.440$	$\varphi = 15/20 \rightarrow 0.005$
$\varphi = 6/20 \rightarrow 0.387$	$\varphi = 16/20 \rightarrow 0.001$
$\varphi = 7/20 \rightarrow 0.323$	$\varphi = 17/20 \rightarrow 0$
$\varphi = 8/20 \rightarrow 0.255$	$\varphi = 18/20 \rightarrow 0$
$\varphi = 9/20 \rightarrow 0.192$	$\varphi = 19/20 \rightarrow 0$
	$\varphi = 20/20 \rightarrow 0$

2.2.1 Probabilités initiales

Nous choisissons des probabilités *initiales* avant les observations, par exemple

$$Pr(\varphi)$$

↪ **Probabilités initiales** (avant les observations)

$Pr(\varphi = 0/20) = 0.00000001$	$Pr(\varphi = 10/20) = 0.117$
$Pr(\varphi = 1/20) = 0.0000003$	$Pr(\varphi = 11/20) = 0.160$
$Pr(\varphi = 2/20) = 0.000005$	$Pr(\varphi = 12/20) = 0.180$
$Pr(\varphi = 3/20) = 0.00004$	$Pr(\varphi = 13/20) = 0.166$
$Pr(\varphi = 4/20) = 0.0003$	$Pr(\varphi = 14/20) = 0.124$
$Pr(\varphi = 5/20) = 0.001$	$Pr(\varphi = 15/20) = 0.075$
$Pr(\varphi = 6/20) = 0.005$	$Pr(\varphi = 16/20) = 0.035$
$Pr(\varphi = 7/20) = 0.015$	$Pr(\varphi = 17/20) = 0.012$
$Pr(\varphi = 8/20) = 0.035$	$Pr(\varphi = 18/20) = 0.003$
$Pr(\varphi = 9/20) = 0.071$	$Pr(\varphi = 19/20) = 0.0005$
	$Pr(\varphi = 20/20) = 0.00004$

2.2.2 Probabilités conjointes

Par un simple produit de la vraisemblance et des probabilités initiales, nous obtenons les probabilités *conjointes* des valeurs du paramètre et des données

$Pr(\varphi \text{ et } f = 1/5) = Pr(f = 1/5 | \varphi) \times Pr(\varphi)$
Produit : vraisemblance \times probabilités initiales
 \hookrightarrow **Probabilités conjointes**

$Pr(\varphi = 0/20 \text{ et } f = 1/5) = 0$	$Pr(\varphi = 10/20 \text{ et } f = 1/5) = 0.016$
$Pr(\varphi = 1/20 \text{ et } f = 1/5) = 0.00000008$	$Pr(\varphi = 11/20 \text{ et } f = 1/5) = 0.014$
$Pr(\varphi = 2/20 \text{ et } f = 1/5) = 0.000005$	$Pr(\varphi = 12/20 \text{ et } f = 1/5) = 0.010$
$Pr(\varphi = 3/20 \text{ et } f = 1/5) = 0.00002$	$Pr(\varphi = 13/20 \text{ et } f = 1/5) = 0.005$
$Pr(\varphi = 4/20 \text{ et } f = 1/5) = 0.0001$	$Pr(\varphi = 14/20 \text{ et } f = 1/5) = 0.002$
$Pr(\varphi = 5/20 \text{ et } f = 1/5) = 0.0004$	$Pr(\varphi = 15/20 \text{ et } f = 1/5) = 0.0004$
$Pr(\varphi = 6/20 \text{ et } f = 1/5) = 0.002$	$Pr(\varphi = 16/20 \text{ et } f = 1/5) = 0.00003$
$Pr(\varphi = 7/20 \text{ et } f = 1/5) = 0.005$	$Pr(\varphi = 17/20 \text{ et } f = 1/5) = 0$
$Pr(\varphi = 8/20 \text{ et } f = 1/5) = 0.009$	$Pr(\varphi = 18/20 \text{ et } f = 1/5) = 0$
$Pr(\varphi = 9/20 \text{ et } f = 1/5) = 0.014$	$Pr(\varphi = 19/20 \text{ et } f = 1/5) = 0$
	$Pr(\varphi = 20/20 \text{ et } f = 1/5) = 0$

2.2.3 Probabilités prédictives

La somme des probabilités conjointes donne la probabilité prédictive marginale des données, avant les observations

$$Pr(f = 1/5) = \sum_{\varphi} Pr(\varphi \text{ et } f = 1/5) = 0.078$$

Somme des probabilités conjointes
 \hookrightarrow **Probabilité prédictive**

Ce résultat est très intuitif puisque la probabilité prédictive est une moyenne pondérée de la fonction de vraisemblance, les poids étant les probabilités initiales.

2.2.4 Probabilités finales

“*Bayesian statistics is difficult in the sense that thinking is difficult.*” (Berry, 1997)

Enfin nous calculons les probabilités *finales* après les observations, par une simple application de la définition des probabilités conditionnelles

$$Pr(\varphi | f = 1/5) = \frac{Pr(\varphi \text{ et } f=1/5)}{Pr(f=1/5)}$$

Rapport : probabilité conjointe / probabilité prédictive
 \hookrightarrow **Probabilités finales** (après les observations)

$Pr(\varphi = 0 f = 1/5) = 0$	$Pr(\varphi = 0.50 f = 1/5) = 0.205$
$Pr(\varphi = 0.05 f = 1/5) = 0.000001$	$Pr(\varphi = 0.55 f = 1/5) = 0.180$
$Pr(\varphi = 0.10 f = 1/5) = 0.0006$	$Pr(\varphi = 0.60 f = 1/5) = 0.128$
$Pr(\varphi = 0.15 f = 1/5) = 0.0003$	$Pr(\varphi = 0.65 f = 1/5) = 0.064$
$Pr(\varphi = 0.20 f = 1/5) = 0.001$	$Pr(\varphi = 0.70 f = 1/5) = 0.026$
$Pr(\varphi = 0.25 f = 1/5) = 0.005$	$Pr(\varphi = 0.75 f = 1/5) = 0.005$
$Pr(\varphi = 0.30 f = 1/5) = 0.026$	$Pr(\varphi = 0.80 f = 1/5) = 0.0004$
$Pr(\varphi = 0.35 f = 1/5) = 0.064$	$Pr(\varphi = 0.85 f = 1/5) = 0$
$Pr(\varphi = 0.40 f = 1/5) = 0.115$	$Pr(\varphi = 0.90 f = 1/5) = 0$
$Pr(\varphi = 0.45 f = 1/5) = 0.179$	$Pr(\varphi = 0.95 f = 1/5) = 0$
	$Pr(\varphi = 1 f = 1/5) = 0$

Les probabilités finales sont donc simplement proportionnelles au produit des probabilités initiales et de la vraisemblance :

$$Pr(\varphi | \text{données}) \propto v(\varphi | \text{données}) \times Pr(\varphi)$$

Leur somme devant être égale à 1 (par définition de la probabilité), ce produit est “normalisé” en divisant par la somme :

$$Pr(\text{données}) = \sum_{\varphi} v(\varphi | \text{données}) \times Pr(\varphi)$$

Dans le cas continu, les probabilités discrètes sont remplacées par des densités, par exemple si φ peut prendre toutes les valeurs réelles dans l'intervalle $[0,1]$:

$$p(\varphi | \text{données}) \propto v(\varphi | \text{données}) \times p(\varphi)$$

et la somme par une intégrale :

$$p(\text{données}) = \int_0^1 v(\varphi | \text{données}) \times p(\varphi) d\varphi$$

Ceci se généralise directement au cas de plusieurs paramètres.

3 Nouvelles difficultés avec les intervalles de confiance

Comme résultat d'un processus déjà bien engagé, les intervalles de confiance pourraient rapidement devenir une norme obligatoire dans les publications expérimentales. En pratique, deux probabilités peuvent être associées de manière routinière à un intervalle d'estimation pour un paramètre calculé à partir des données observées.

- La première probabilité est “la proportion des intervalles qui contiennent le paramètre pour un grand nombre de répétitions” ; elle est usuellement appelée la probabilité de *couverture* (ou de *recouvrement*) *fréquentiste*.
- La seconde probabilité est la “probabilité *bayésienne* que cet intervalle contienne le paramètre” (pour une certaine distribution initiale).

Dans l'approche fréquentiste, il est **interdit** d'utiliser la seconde probabilité, tandis que dans l'approche bayésienne, les deux probabilités sont **valides**.

Pour de nombreuses raisons dues à leur conception fréquentiste, les intervalles de confiance peuvent difficilement apparaître comme “LA méthode ultime”. En effet la raison qui les rend attrayants résulte d'une incompréhension fondamentale. Il est si étrange de traiter les données comme aléatoires même *après avoir recueilli les observations* que l'interprétation fréquentiste *orthodoxe* des intervalles de confiance n'a pas de sens pour la plupart des utilisateurs. C'est indiscutablement l'interprétation bayésienne (naturelle) des intervalles de confiance qui les rend attrayants. Ironiquement cette interprétation *hérétique* est encouragée par la duplicité de la plupart des formateurs en statistique qui les tolèrent et même les utilisent. Par exemple, Pagano (1990, page 288), dans un ouvrage dont le titre affiche l'objectif de faire comprendre la statistique (“*understanding statistics...*”), décrit un intervalle de confiance 95% comme un intervalle “*such that the probability is 0.95 that the interval contains the population value*”.

Sous une autre forme, on voit très couramment des interprétations telles que “avec 5% de risque de se tromper, on peut dire que π est dans l'intervalle $[0.27, 0.42]$ ”.

D'autres auteurs affirment que l'interprétation fréquentiste "correcte" qu'ils défendent peut être exprimée comme "nous pouvons être confiants à 95% que le paramètre est contenu dans l'intervalle"; par exemple Kirk (1982, page 43) énonce "*we can be 95% confident that the population mean is between 114.06 and 119.94*". Il est difficile d'imaginer que les lecteurs puissent comprendre que "confiant" renvoie à une conception fréquentiste de la probabilité!

"We [statisticians] will all be Bayesians in 2020, and then we can be a united profession." Lindley (in Smith, 1995, page 317))

La littérature est remplie d'interprétations bayésiennes des intervalles de confiance et des tests de signification. Toutes les tentatives des fréquentistes pour corriger ces "errements" se sont révélées vaines : c'est un combat perdu d'avance. En fait la plupart utilisateurs de l'inférence statistique sont des bayésiens sans le savoir (Lecoutre, 1997/2005)!

4 L'approche bayésienne "objective"

4.1 Où est l'objectivité ?

"A common misconception is that Bayesian analysis is a subjective theory; this is neither true historically nor in practice. The first Bayesians, Bayes (see Bayes (1763)) and Laplace (see Laplace (1812)) performed Bayesian analysis using a constant prior distribution for unknown parameters. . ." (Berger, 2004, page 3)

L'inférence statistique fréquentiste s'auto-proclame "objective" contrairement à l'inférence bayésienne qui serait nécessairement "subjective". Cette affirmation se trouve renforcée par certaines conceptions bayésiennes extrémistes, dans lesquelles les opinions – et non seulement les connaissances *a priori* – pourraient (devraient) être intégrées dans l'inférence scientifique. Il n'est donc pas étonnant que la critique la plus commune adressée par les fréquentistes à l'approche bayésienne soit la nécessité de probabilités initiales. Il est évidemment facile, dans l'exemple précédent, de choisir des probabilités initiales quelconques – ce que j'ai fait à titre d'illustration des calculs. Quelqu'un de mal intentionné pourra alors dire (à juste titre ?) que l'on peut dans ce cas obtenir "ce que l'on veut".

"But the primary aim of a scientific experiment is not to precipitate decisions, but to make an appropriate adjustment in the degree to which one accepts, or believes, the hypothesis or hypotheses being tested." (Rozeboom, 1960, in Morrison & Henkel, 1970, page 221)

Une autre difficulté vient de l'insistance de certains théoriciens bayésiens à vouloir faire dépendre l'inférence statistique de la "Théorie de la Décision". S'il peut être difficile et sujet à critique d'assigner les probabilités initiales, il est encore bien plus problématique de choisir une fonction de coût.

La conséquence de cette mise en avant des aspects subjectifs de l'inférence bayésienne et d'une approche décisionnelle, a été d'obscurcir sa contribution à l'analyse des données expérimentales. En fait la définition bayésienne peut parfaitement être utilisée pour décrire une "connaissance objective", en particulier basée sur des arguments de symétrie ou sur des données de fréquences. Il ne s'agit pas d'affirmer qu'une analyse statistique peut être

entièrement objective : elle met nécessairement en jeu des éléments subjectifs – notamment le choix du modèle – et comporte en fait une part de *conventions*. Mais l’inférence statistique bayésienne n’est pas moins objective que l’inférence fréquentiste. C’est même le contraire dans de nombreux contextes.

4.2 L’inférence fiducio-bayésienne

Il existe une voie, de plus en plus reconnue, qui vise à concilier la théorie bayésienne avec la conception fréquentiste. Dans cette perspective, l’approche développée par Jeffreys dans les années trente (Jeffreys, 1961/1939) a un statut privilégié. Dans la ligne de Laplace (1986/1825), la philosophie de cette approche est de choisir les probabilités initiales en écartant toute connaissance éventuelle sur la valeur du paramètre. En pratique, ces probabilités initiales “non informatives” sont des distributions vagues qui, *a priori* ne favorisent aucune valeur, aucune hypothèse, particulière. En conséquence elles laissent les données “parler pour elles-mêmes”.

“1. A major goal of statistics (indeed science) is to find a completely coherent objective Bayesian methodology for learning from data. This is exemplified by the attitudes of Jeffreys (1961) and Jaynes (1999 [2003]).

2. Objective Bayesian analysis is the best method for objectively synthesizing and communicating the uncertainties that arise in a specific scenario, but is not necessarily coherent in a more general sense.

My general view is that 1) is not attainable; 2) is often attainable and should be done if possible.” (Berger, 2004, page 2)

Sous cette forme le paradigme bayésien fournit, sinon des méthodes objectives et complètement cohérentes, au moins des méthodes de référence, pleinement justifiées et appropriées pour la communication scientifique. En outre, dans les situations courantes où l’on utilise traditionnellement les tests t , F ou χ^2 , ces méthodes sont très faciles à mettre en œuvre. Elles peuvent maintenant être utilisées aussi aisément que ces tests, tout en procurant des avantages conceptuels et méthodologiques considérables.

“A widely accepted objective Bayes theory, which fiducial inference was intended to be, would be of immense theoretical and practical importance.” (Efron, 1998)

Dans le but de les promouvoir, il nous a semblé important de leur donner un nom plus explicite que standard, non informatives, de référence, conventionnelles, etc. (Lecoutre, Lecoutre & Poitevineau, 2001). Nous les appelons

fiducio-bayésiennes.

“The statistics profession, in general, hurts itself by not using attractive names for its methodologies, and we should start systematically accepting the ‘objective Bayes’ name before it is co-opted by others.” (Berger, 2004, page 3)

Avec une motivation similaire, Berger, 2004 défend l’appellation

bayésiennes objectives.

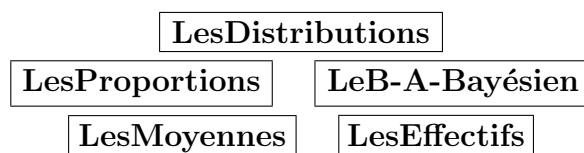
Dans ce qui suit j'utiliserai ces deux appellations, en évitant toutefois l'appellation fiducio-bayésiennes dans le cas de l'inférence sur les proportions, pour lequel elle nécessite une discussion particulière (qui est abordée dans la partie IV).

PARTIE II - LA PRATIQUE : QUELQUES SITUATIONS DE BASE

Programmes informatiques. Les programmes informatiques utilisés pour la mise en œuvre des méthodes sont regroupés dans le logiciel **LePAC**, qui peut être téléchargé à l'adresse

<http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/Eris>

On en trouvera une présentation dans un numéro précédent de *la Revue de Modulad* (Lecoutre & Poitevineau, 2005). Dans ce qui suit, sont fournis des écrans (ou plus souvent des extraits d'écrans) des programmes utilisés. Ceux-ci sont précédés du nom du programme utilisé



ainsi que d'un mode d'emploi pour obtenir les résultats présentés.

5 Inférence sur une proportion

L'objectif de ce premier exemple est de présenter une "transition en douceur" des procédures fréquentistes traditionnelles vers les procédures bayésiennes. Techniquement simple, il permettra d'illustrer en détail la méthodologie bayésienne.

5.1 Le problème : Planification et conduite d'une expérimentation

Supposons la situation suivante. Un collègue enseignant "fréquentiste" m'affirme avoir développé une méthode d'enseignement individuel de l'intervalle de confiance conduisant à un taux élevé d'interprétations correctes. Nous convenons que sa méthode est efficace si le taux d'interprétations correctes ("succès") φ après l'enseignement est supérieur à 0.85, et qu'elle est inefficace si ce taux est inférieur à 0.70. Ma probabilité *a priori* que la méthode soit inefficace est si élevée que je pense qu'il n'est pas nécessaire de recueillir des données. Bien entendu mon collègue ne peut pas être convaincu par ce subjectivisme flagrant et il propose de planifier une expérimentation dans le cadre fréquentiste traditionnel de Neyman-Pearson.

5.2 Une solution fréquentiste : Test binomial et test d'interruption stochastique

5.2.1 La planification de l'expérience

Considérant l'hypothèse nulle $H_0 : \varphi = 0.70$, il décide d'utiliser un test binomial unilatéral pour un échantillon de taille fixé avec des probabilités d'erreurs de Types I et II respectivement égales à $\alpha = 0.05$ et $\beta = 0.20$, soit une puissance $1 - \beta = 0.80$ pour l'hypothèse alternative $H_a : \varphi = 0.85$ (celle qu'il souhaite accepter!). Il correspond à ces conditions un effectif de l'échantillon $n = 59$, pour lequel le test binomial rejette H_0 au seuil 0.05 si le nombre de succès observé a est plus grand que 47.

En effet, pour un échantillon de taille n , la probabilité d'observer a succès est donnée par la distribution *Binomiale*

$$a|\varphi \sim \text{Bin}(\varphi, n)$$

$$\text{Pr}(a|\varphi) = \binom{n}{a} \varphi^a (1 - \varphi)^{n-a}$$

soit la fonction de vraisemblance

$$v(\varphi | \text{données}) \propto \varphi^a (1 - \varphi)^{n-a}.$$

Pour $n = 59$ (que l'on trouve par itérations successives), on obtient :

$$\text{Pr}(a > 47 | H_0 : \varphi = 0.70) = 0.035 < 0.05 (\alpha)$$

$$\text{Pr}(a > 47 | H_a : \varphi = 0.85) = 0.834 > 0.80 (1 - \beta)$$

LesDistributions

Dans **LePAC** activez le menu **LesDistributions** et le sous-menu **Binomiale - Standard**, ce qui affiche la fenêtre pour la distribution *Binomiale*. Entrez dans les champs appropriés les **paramètres** de la distribution :

- pour φ : **0.70** (ou simplement **.7**)
- pour n : **59**

Sélectionnez les boutons d'option

- **Pr(X>x)**
- **probabilité**

Entrez dans le champ **probabilité** : 0.05 et appuyez sur touche "**Entrée**" du clavier (ou cliquez sur le bouton **Calculer**). Vous obtenez la première figure ci-après.

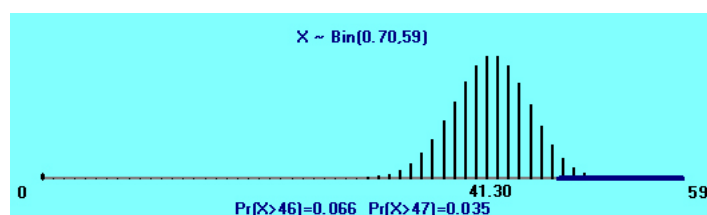
Pour la seconde figure, recommencez avec **0.85** pour φ et **0.80** pour la probabilité.

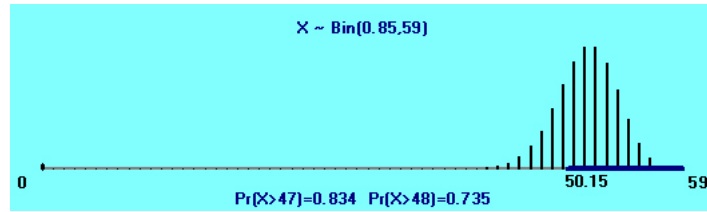
Vous pouvez sélectionner le nombre de décimales voulues.

Pour obtenir une probabilité associée à une valeur donnée, sélectionnez le bouton d'option

- **Limite**

et entrez dans le champ **limite** la valeur dont vous voulez la probabilité





Il convient de noter qu'en raison du caractère discret de la distribution, le taux d'erreur et la puissance réels ne peuvent pas être égaux à α et $1 - \beta$ mais leur sont respectivement inférieur et supérieur.

Je réussis à convaincre mon collègue qu'il serait préférable d'arrêter l'expérimentation avant son terme si sa méthode se révèle inefficace (compte tenu du fait que celle-ci nécessite un enseignement individuel et est lourde à mettre en œuvre). En conséquence il planifie une analyse intermédiaire après l'inclusion de 20 sujets. Les notations sont résumées dans le tableau suivant

	effectif	nombre de	
		succès	erreurs
Données intermédiaires	$n_1 = 20$	a_1	$b_1 = n_1 - a_1$
Données futures	$n_2 = 39$	a_2	$b_2 = n_2 - a_2$
Données complètes	$n = 59$	$a = a_1 + a_2$	$b = n - a$

5.2.2 Le test d'interruption stochastique et la puissance conditionnelle

Le cadre traditionnel de Neyman-Pearson nécessitant la spécification de toutes les possibilités avant le recueil des données, il prévoit d'effectuer un "test d'interruption stochastique" (*stochastically curtailed test*). L'interruption stochastique suggère de stopper une expérience à une étape intermédiaire quand l'information disponible détermine le résultat de l'expérience avec une probabilité élevée, soit sous H_0 soit sous H_a .

La *puissance conditionnelle* à l'analyse intermédiaire est définie comme la probabilité, étant donné φ et les données disponibles, que le test rejette H_0 au terme prévu de l'expérience.

(1) A l'analyse intermédiaire, l'expérience est interrompue et on *rejette* H_0 si la puissance conditionnelle à la valeur spécifiée par l'hypothèse nulle est élevée, disons supérieure à 0.80. Dans notre exemple, même si après 20 observations on n'a observé que des succès ($a_1 = 20$) nous n'arrêtons pas l'expérience, car la probabilité de rejeter H_0 au terme prévu – donnée par la distribution $Bin(0.70, 39)$ – est inférieure à 0.80

$$Pr(a > 47 | H_0 : \varphi = 0.70 \text{ and } a_1 = 20) = Pr(a_2 > 27 | H_0 : \varphi = 0.70) = 0.482 < 0.80$$

(2) De manière similaire, à l'analyse intermédiaire, l'expérience est interrompue et on *accepte* H_0 si la puissance conditionnelle à la valeur spécifiée par l'hypothèse alternative – donnée par la distribution $Bin(0.85, 39)$ – est faible, disons inférieure à 0.20. Par exemple, si 12 succès sont observés après 20 observations, cette règle suggère d'arrêter l'expérience et d'accepter l'hypothèse nulle

$$Pr(a > 47 | H_a : \varphi = 0.85 \text{ and } a_1 = 12) = Pr(a_2 > 35 | H_a : \varphi = 0.85) = 0.143 < 0.20$$

Une critique à l'adresse de cette procédure est qu'il ne semble guère pertinent de considérer une prédiction qui est basée sur des hypothèses qui peuvent ne plus être plausibles étant donné les données disponibles. En fait la procédure ignore purement et simplement la connaissance sur le paramètre obtenue au moment de l'analyse intermédiaire.

5.3 Une solution hybride : la puissance prédictive

Pour répondre à cette critique, de nombreux auteurs ont défendu l'idée de calculer la *puissance prédictive*, c'est-à-dire de moyenniser la puissance conditionnelle sur toutes les valeurs du paramètre par un calcul bayésien. Cela nous conduit à une approche bayésienne, mais en gardant la procédure de test fréquentiste pour l'analyse des données.

Formellement, on utilise les données disponibles à l'analyse intermédiaire et la distribution finale de φ qui en résulte. La solution la plus simple est de choisir comme distribution initiale une distribution *conjuguée*

$$\varphi \sim \text{Bêta}(a_0, b_0)$$

L'avantage est que la distribution finale est également une distribution *Bêta* (ce qui explique l'appellation "conjuguée"). Les "poids" initiaux a_0 et b_0 s'ajoutent aux effectifs observés a_1 et b_1 , soit

$$\varphi | a_1 \sim \text{Bêta}(a_1 + a_0, b_1 + b_0)$$

et la distribution prédictive, qui est un mélange de distributions *Binomiales*, s'appelle tout naturellement une distribution *Bêta-Binomiale*

$$a_2 | a_1 \sim \text{Bêta-Bin}(a_1 + a_0, b_1 + b_0; n_2)$$

Pour obtenir la puissance prédictive, on choisit une distribution initiale *non informative*. Celle-ci correspond à des poids a_0 et b_0 faibles, compris entre 0 et 1⁴. Ici, je retiendrai la distribution initiale

$$\varphi \sim \text{Bêta}(0, 1)$$

un choix qui est cohérent avec la procédure de test utilisée : j'y reviendrai plus loin.

Reprenons l'exemple traité ci-dessus en (1), soit $n_1 = 20$ et $a_1 = 20$. La probabilité prédictive de rejeter H_0 au terme prévu ($n = 59$) prend explicitement en compte les données disponibles (aucune erreur n'a été observée). Elle est donnée par la distribution *Bêta-Bin*(20, 1; 39) et est sans surprise largement supérieure à la probabilité conditionnelle à la valeur spécifiée par l'hypothèse nulle ($\varphi = 0.70$)

$$Pr(a > 47 | a_1 = 20) = Pr(a_2 > 27 | a_1 = 20) = 0.997 > 0.80$$

d'où la décision d'interrompre l'expérience et de rejeter H_0 .

LesDistributions

Dans **LePAC** activez le menu **LesDistributions** et le sous-menu **Bêta - Binomiale**, ce qui affiche la fenêtre pour la distribution *Bêta-Binomiale* (ou si la fenêtre est déjà ouverte pour une autre distribution, cliquez sur le bouton correspondant à cette distribution pour la changer). Entrez dans les champs appropriés les **paramètres** de la distribution :

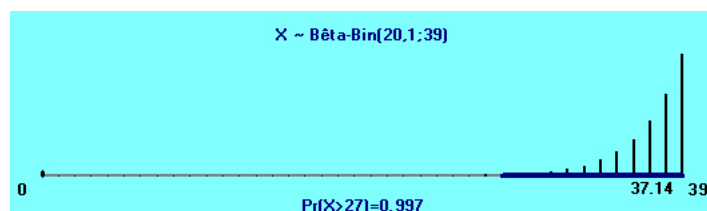
⁴La distribution *Bêta*(a_0, b_0) n'est en toute rigueur définie que si a_0 et b_0 sont strictement positifs. Lorsque a_0 ou b_0 est nul, la densité est impropre ($\int_0^1 p(\varphi)d\varphi = +\infty$), mais cela n'empêche pas le calcul de la densité finale qui sera, sauf cas particuliers, une densité propre. En tout état de cause, il ne s'agit que d'un problème théorique, car en pratique, une distribution "propre" obtenue en remplaçant 0 par une valeur ϵ très petite ne fera aucune différence.

- pour α : **20**
- pour β : **1**
- pour n : **59**

Sélectionnez les boutons d'option

- **Pr(X>x)**
- **limite**

Entrez dans le champ **limite** : **27** et appuyez sur touche “**Entrée**” du clavier (ou cliquez sur le bouton **Calculer**). Vous obtenez la figure ci-après.



Cette probabilité prédictive est une moyenne pondérée des probabilités conditionnelles à φ (les poids étant donnés par la distribution finale)

$$Pr(a > 47 \mid a_1 = 20 \text{ et } \varphi) = Pr(a_2 > 27 \mid a_1 = 20 \text{ et } \varphi)$$

dont quelques exemples sont

φ	$Pr(a > 47 \mid a_1 = 20 \text{ et } \varphi)$
1	1
0.95	0.9999997
0.85	0.990
0.70	0.482

Cette approche de la puissance prédictive étant une approche *hybride* n'est pas très satisfaisante. En particulier elle ne nous donne pas une information bayésienne directe sur φ : la distribution finale n'est utilisée que comme intermédiaire de calcul. Ce qui est troublant est qu'une *décision* (accepter H_0 ou accepter H_a) est prise à l'analyse terminale – ou éventuellement à l'analyse intermédiaire – *même si la proportion observée est située dans la région de non conclusion* $[0.70, 0.85]$, auquel cas on n'a évidemment rien prouvé pour ces hypothèses.

Ce dont on a réellement besoin, c'est de pouvoir évaluer à toute étape de l'expérience la probabilité des régions spécifiées auxquelles on s'intéresse et l'aptitude d'un échantillon futur à corroborer les résultats déjà obtenus. L'analyse bayésienne traite directement ces questions.

5.4 Solution bayésienne

La méthodologie bayésienne nous permet d'obtenir les probabilités des régions spécifiées, ce qui apporte des réponses directes aux questions sur la grandeur des effets et n'a pas de contrepartie fréquentiste. Considérons un nouvel exemple d'analyse intermédiaire, avec 10 succès observés ($n_1 = 20$ et $a_1 = 10$).

5.4.1 Evaluer la probabilité des régions spécifiées

Choisissons la distribution initiale noninformative $B\acute{e}ta(1/2, 1/2)$, qui donnera la “méthode bayésienne objective” (je reviendrai également sur ce choix plus loin) ⁵. Dans ce cas la distribution finale est $B\acute{e}ta(10.5, 10.5)$ et on obtient une probabilité élevée (0.971) que la méthode d’enseignement soit inefficace ($\varphi < 0.70$).

$Pr(\varphi < 0.70 a_1 = 10)$	$Pr(0.70 < \varphi < 0.85 a_1 = 10)$	$Pr(\varphi > 0.85 a_1 = 10)$
0.971	0.029	0.0001

L’utilisation de l’ordinateur résoud les calculs nécessaires pour l’utilisation des distributions bayésiennes. Cela donne à l’utilisateur un moyen à la fois intuitif et attrayant pour comprendre les rôles de la taille de l’échantillon, des données et de la distribution initiale. En particulier la distribution finale peut être visualisée et explorée interactivement.

LesProportions

Dans **LePAC** activez le menu **LesBayésiens** et le sous-menu **LesProportions**, ce qui affiche la fenêtre pour l’inférence sur des proportions (par défaut, c’est l’option **1 groupe** qui est active). Entrez dans les champs appropriés les **effectifs** :

- pour **1** : **10**
- pour **0** : **10**

Entrez s’il y a lieu les paramètres de la distribution **initiale bêta** (mais c’est en principe inutile car c’est l’option par défaut) :

- pour **1** : **0.5** (que vous pouvez entrer comme **1/2**)
- pour **0** : **0.5**

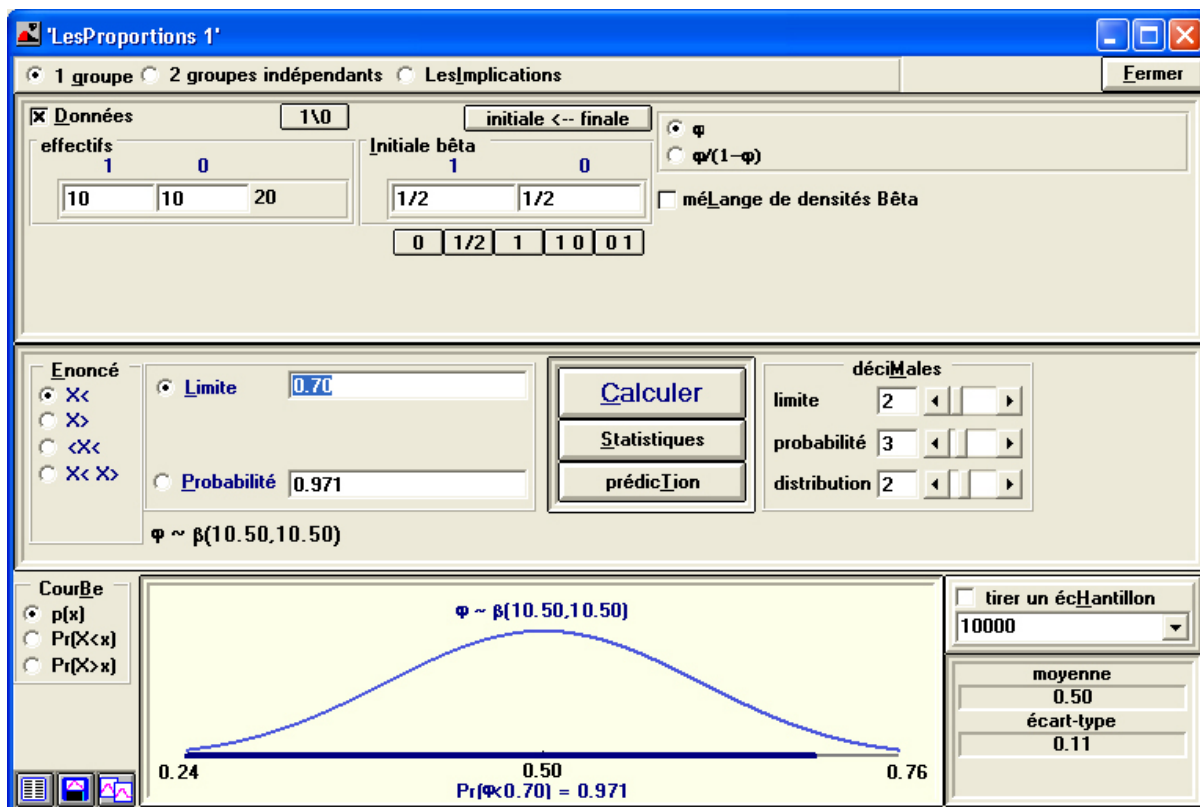
Remarque : ces valeurs peuvent être affichées en cliquant sur le bouton **1/2**.

Sélectionnez les boutons d’option

- **X<**
- **limite**

Entrez dans le champ **limite** : **0.70** et sélectionnez les décimales appropriées : **2** pour la limite et **3** pour la probabilité. Cliquez sur le bouton **Calculer** et vous obtenez la figure ci-après.

⁵Dans ce cas, l’inférence bayésienne sur φ effectuée à l’analyse intermédiaire ne prend pas explicitement en compte la règle d’arrêt (qui est cependant prise en compte dans le calcul de la probabilité prédictive). Dans le cadre fréquentiste, il est habituel que les inférences intermédiaires soient modifiées pour tenir compte de la règle d’arrêt. Ce point – qui pourrait apparaître comme un sujet de discordance entre les deux approches – est abordé dans la partie IV.



La probabilité associée à une limite fixée, ou inversement les limites associées à une probabilité (ou garantie) donnée (sélectionnez le bouton d'option **probabilité**), peuvent être calculées.

5.4.2 Evaluer l'aptitude d'un échantillon futur à corroborer les résultats déjà obtenus

Pour résumer l'information obtenue quant à la décision d'interrompre ou non l'expérience, nous pouvons calculer la probabilité prédictive de confirmer la conclusion d'inefficacité. Si on souhaite une garantie d'au moins 0.95 pour la conclusion finale, c'est-à-dire $Pr(\varphi < 0.70 | a) > 0.95$, il faut que le nombre total de succès a soit inférieur à 36 sur 59. Il s'agit donc, puisque l'on a déjà observé $a_1 = 10$ succès, de calculer la probabilité prédictive d'observer dans les données futures un nombre de succès a_2 compris entre 0 et 25 sur 39.

Ici, sur la base des données disponibles, il y a 87.3% de chances que cette conclusion soit obtenue si l'expérience est menée à son terme.

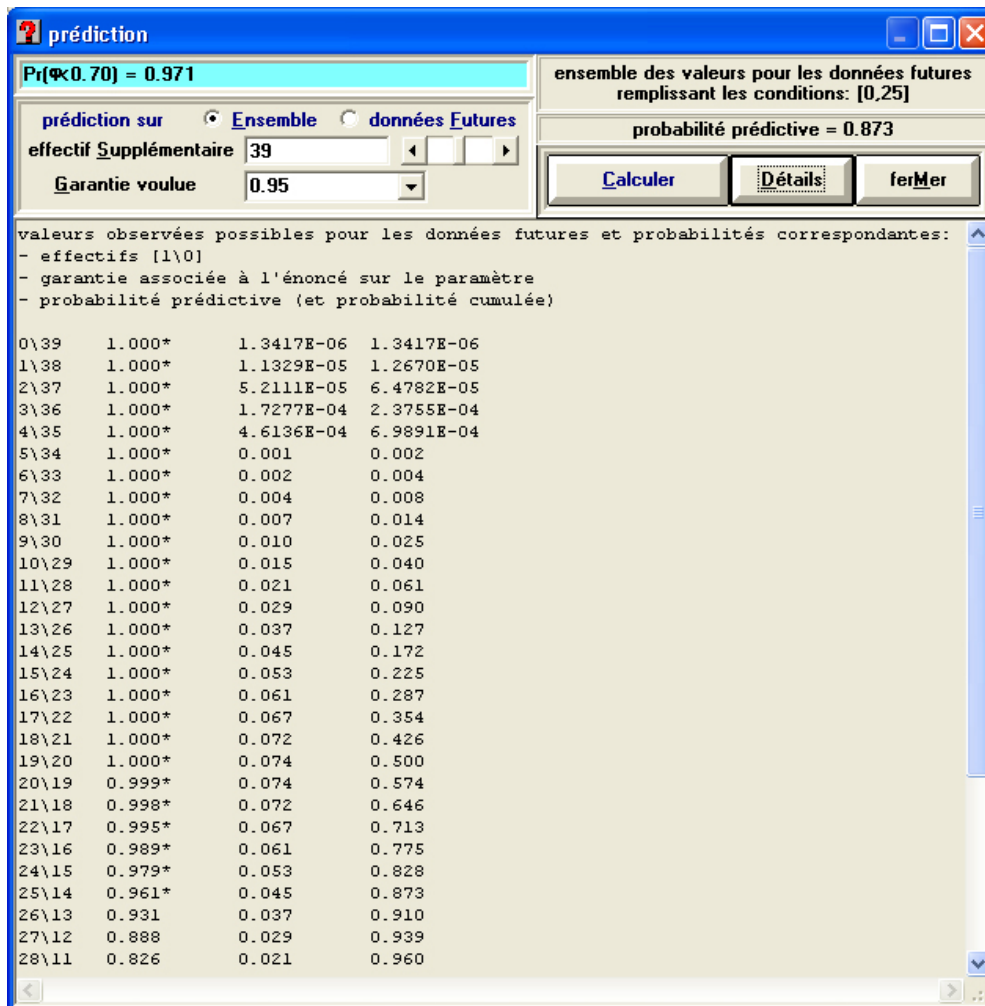
LesProportions

Effectuez s'il y a lieu les opérations décrites pour la figure précédente, puis cliquez sur le bouton **prediction**.

Sélectionnez le bouton d'option

- **ensemble**

Entrez dans le champ **effectif supplémentaire** : **39** et sélectionnez la **garantie voulue** dans la liste déroulante : **0.95**. Cliquez sur le bouton **Calculer** et vous obtenez la probabilité prédictive : **0.873**. Cliquez sur le bouton **détail** pour obtenir la figure ci-après.



Le tableau ci-après donne un résumé des analyses pour l'exemple précédent, ainsi que pour un autre exemple plus favorable à la méthode d'enseignement.

Exemple 1 : $n_1 = 20$ et $a_1 = 10$

Inférence sur φ	Probabilité prédictive ($n = 59$)
Probabilité finale	Conclusion avec une garantie ≥ 0.95
$Pr(\varphi < 0.70 a_1 = 10)$	$\varphi < 0.70$
0.971	0.873 ($a < 36$)
$Pr(\varphi < 0.85 a_1 = 10)$	$\varphi < 0.85$
0.9999	0.9998 ($a < 46$)

Exemple 2 : $n_1 = 20$ et $a_1 = 18$

Inférence sur φ	Probabilité prédictive ($n = 59$)
Probabilité finale	Conclusion avec une garantie ≥ 0.95
$Pr(\varphi > 0.70 a_1 = 10)$	$\varphi > 0.70$
0.982	0.939 ($a > 46$)
$Pr(\varphi > 0.85 a_1 = 10)$	$\varphi > 0.85$
0.717	0.301 ($a > 54$)

5.4.3 Choisir l'effectif de l'échantillon

Les procédures prédictives sont aussi des outils adaptés pour aider au choix de l'effectif de l'échantillon. Supposons qu'en vue de planifier une expérience pour démontrer l'efficacité de la méthode d'enseignement nous ayons réalisé une expérience préliminaire, par exemple avec 10 sujets et que n'ayons observé que des succès. Dans ce cas la distribution finale pour l'expérience préliminaire – en commençant avec la distribution non informative $Bêta(1/2, 1/2)$ – est utilisée comme distribution initiale. On obtient ici la probabilité finale $Pr(\varphi > 0.85) = 0.932$.

Les Proportions

Procédez comme précédemment pour l'inférence sur une proportion et entrez dans les champs appropriés les **effectifs** :

- pour **1** : **10**
- pour **0** : **0**

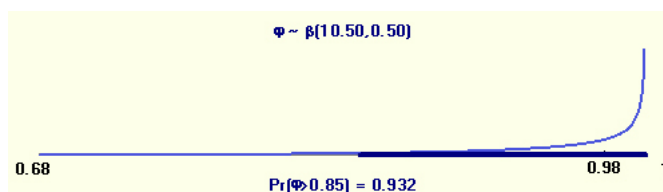
Entrez la distribution **initiale bêta**

- pour **1** : **1/2**
- pour **0** : **1/2**

Sélectionnez les boutons d'option

- **X>**
- **limite**

Entrez dans le champ **limite** : **0.85** pour obtenir la figure ci-après (avec les décimales appropriées).



Si on intègre les données préliminaires dans l'analyse de l'expérience, la procédure est exactement la même que dans le cas de l'analyse intermédiaire ; on parle dans ce cas d'approche bayésienne complète ("*full Bayesian*"). Mais la plupart temps, dans l'expérimentation, on ne souhaite pas prendre en compte les données préliminaires dans l'analyse de l'expérience, et celle-ci sera donc effectuée elle aussi avec la distribution initiale *Bêta* ($1/2, 1/2$).

La procédure reste analogue : nous calculons la probabilité prédictive que dans l'échantillon futur d'effectif n (et non pour l'ensemble des données) la conclusion d'efficacité ($\varphi > 0.85$) soit atteinte avec une garantie donnée γ , d'où par exemple les probabilités prédictives suivantes pour $\gamma = 0.95$ (avec entre parenthèses les valeurs de a qui remplissent la condition) :

$n = 20 \mapsto 0.582$	($a > 19$)
$n = 30 \mapsto 0.696$	($a > 28$)
$n = 40 \mapsto 0.744$	($a > 37$)
$n = 50 \mapsto 0.770$	($a > 46$)
$n = 60 \mapsto 0.787$	($a > 55$)
$n = 70 \mapsto 0.798$	($a > 64$)
$n = 71 \mapsto 0.795$	($a > 65$)
$n = 72 \mapsto 0.829$	($a > 65$)
Etc.	

Les Proportions

Effectuez s'il y a lieu les opérations décrites pour la figure précédente, puis cliquez sur le bouton **prediction**.

Sélectionnez le bouton d'option

- **données futures**

Entrez dans le champ **effectif supplémentaire** : **39** et sélectionnez la **garantie voulue** dans la liste déroulante : **0.95**. Cliquez sur le bouton **Calculer** et vous obtenez la probabilité prédictive : **0.798** et la figure ci-après.



Ainsi, sur la base des résultats préliminaires, il faut donc un effectif de l'ordre de 70 pour avoir à peu près 80% de chances de démontrer l'efficacité. On ne s'étonnera pas du fait que les probabilités puissent ne pas être croissantes ; cela résulte du caractère discret de la variable (il en est de même pour la puissance).

5.5 Commentaires sur le choix de la distribution initiale non informative

Baucoup d'utilisateurs potentiels des méthodes bayésiennes continuent de penser qu'elles sont trop "subjectives" pour être scientifiquement acceptables. Pourtant les méthodes fréquentistes sont pleines de conventions plus ou moins *ad hoc*. Ainsi le traditionnel seuil

observé p (p value) du test est basé sur les échantillons qui sont “plus extrêmes” que les données observées (sous l’hypothèse nulle). Mais, pour des données discrètes, cela dépend du fait que l’on inclut ou non les données observées.

Supposons qu’à l’analyse finale on observe 47 succès ($n = 59$ et $a = 47$), soit la valeur au delà de laquelle le test binomial rejette l’hypothèse nulle $H_0 : \varphi = 0.70$. On peut alors calculer le seuil observé p suivant une des trois possibilités suivantes :

$p_{inc} = Pr(a \geq 47 \mid H_0 : \varphi = 0.70) = 0.066$ [solution usuelle “incluante”] $\Rightarrow H_0$ n’est pas rejetée au seuil $\alpha = 0.05$ (test “conservateur”)
$p_{exc} = Pr(a > 47 \mid H_0 : \varphi = 0.70) = 0.035$ [solution “excluante”] $\Rightarrow H_0$ est rejetée au seuil $\alpha = 0.05$ (test “libéral”)
$p_{moyen} = 1/2(p_{inc} + p_{exc}) = 0.051$ [“ p moyen”]

Dans la solutions usuelle la valeur observée est incluse et le test est “conservateur” : la probabilité de rejeter H_0 , si elle est vraie, est inférieure à α . Mais si la valeur observée est exclue, le test devient “libéral” : la probabilité de rejeter H_0 , si elle est vraie, est supérieure à α . Une solution typique pour surmonter ce problème consiste à considérer un “ p moyen” ($mid-p$ value) (Routledge, 1994 ; Berry & Armitage, 1995), mais dans le cadre fréquentiste cette pratique a seulement des justifications *ad hoc*.

À l’évidence dans ce cas le choix d’une distribution initiale non informative implique aussi des *conventions*. Mais le choix particulier d’une telle distribution initiale n’est ni plus ni moins qu’une contrepartie exacte de l’arbitraire en jeu dans l’approche fréquentiste (Bernard, 1996). Ainsi, dans notre situation, différentes distributions ont été proposées pour une analyse bayésienne objective (pour une discussion, voir par exemple Lee, 2004, pages 79-81). La plupart appartiennent à la famille des distributions *Bêta* (a_0, b_0) avec des valeurs de a_0 et b_0 comprises entre 0 et 1, en particulier

- *Bêta*(1, 1) – qui est la distribution uniforme sur $[0,1]$ (Laplace, 1774) ;
- *Bêta*(0, 0) qui est une distribution impropre – mais qui correspond à la distribution uniforme pour $\log(\frac{\varphi}{1-\varphi})$ (Lhoste, 1923 ; Haldane, 1948) ;
- *Bêta*($1/2, 1/2$) – qui correspond à la distribution uniforme pour $\arcsinus(\sqrt{\pi})$ (Jeffreys, 1961 ; Perks, 1947) et qui est dérivée en utilisant la *règle de Jeffreys* dont on trouvera la justification dans la partie IV.

5.5.1 Interprétation bayésienne du seuil observé

Pour notre problème, il existe dans cette famille deux distributions initiales extrêmes – *Bêta*(0, 1) et *Bêta*(1, 0) – qui sont respectivement la plus défavorable et la plus favorable par rapport à l’hypothèse nulle. Les seuils de signification observés associés aux solutions incluante et excluante sont exactement les probabilités bayésiennes finales que φ soit plus petit que 0.70 (la valeur spécifiée par H_0) respectivement associées à ces deux distributions initiales extrêmes. La distribution initiale de Jeffreys *Bêta*($1/2, 1/2$) avec des poids égaux à $1/2$ est intermédiaire et donne une probabilité proche du “ p moyen”.

$Pr(\varphi < 0.70 \mid a = 47) = 0.066 = p_{inc}$ pour la distribution initiale $\varphi \sim B\hat{e}ta(0, 1)$ (la plus d\`efavorable \`a H_0) soit la distribution finale $\varphi \mid \sim B\hat{e}ta(47, 13)$
$Pr(\varphi < 0.70 \mid a = 47) = 0.035 = p_{exc}$ pour la distribution initiale $\varphi \sim B\hat{e}ta(1, 0)$ (la plus favorable \`a H_0) soit la distribution finale $\varphi \mid \sim B\hat{e}ta(48, 12)$
$Pr(\varphi >< 0.70 \mid a = 47) = 0.049 \approx p_{moyen}$ pour la distribution initiale $\varphi \sim B\hat{e}ta(1/2, 1/2)$ soit la distribution finale $\varphi \mid \sim B\hat{e}ta(47.5, 12.5)$

LesProportions

Proc\`edez comme pr\`ec\`edemment pour l'inf\`erence sur une proportion et entrez dans les champs appropri\`es les **effectifs** :

- pour **1** : **47**
- pour **0** : **12**

Entrez la distribution **initiale b\`eta**

- pour **1** : **0**
- pour **0** : **1**

Remarque : ces valeurs peuvent \^etre affich\`ees en cliquant sur le bouton **0 1**.

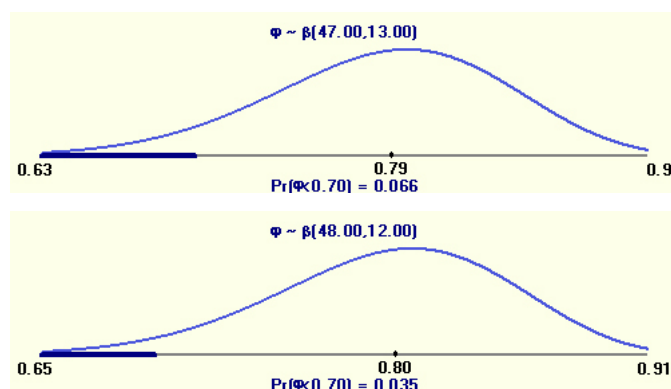
S\`electionnez les boutons d'option

- **X<**
- **limite**

Entrez dans le champ **limite** : **0.70** pour obtenir la premi\`ere figure ci-apr\`es (avec les d\`ecimales appropri\`ees).

La seconde figure est obtenue pour la distribution **initiale b\`eta**

- pour **1** : **1**
- pour **0** : **0**



La critique facile \`a l'\`egard de l'approche bay\`esienne concernant l'existence de divergences quant au choix de la distribution non informative se retourne donc \`a l'encontre de l'approche fr\`equentiste. La r\`eponse des bay\`esiens ne devrait donc pas \^etre de m\`esestimer l'impact du choix d'une distribution non informative particuli\`ere, comme cela est souvent le cas

“In fact, the [different non informative priors] do not differ enough to make much difference with even a fairly small amount of data.” (Lee, 2004, page 81)

mais au contraire de l’accepter.

5.5.2 Intervalles de crédibilité bayésien et taux de couverture fréquentistes

Dans d’autres situations où nous ne nous intéressons pas à des valeurs particulières, nous pouvons considérer un intervalle (ou plus généralement une *région*) d’estimation pour φ . Dans le cadre bayésien un tel intervalle est habituellement appelé un *intervalle de crédibilité* pour souligner la différence d’interprétation avec l’intervalle de confiance fréquentiste.

Intervalles “à queues égales”

LesProportions

Procédez comme précédemment pour l’inférence sur une proportion et entrez dans les champs appropriés les **effectifs** :

- pour **1** : **19**
- pour **0** : **1**

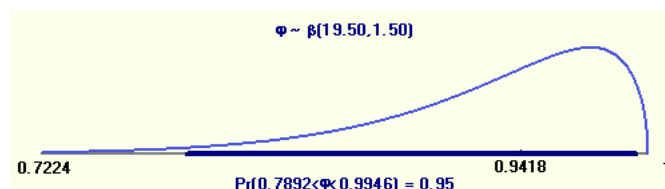
Entrez la distribution **initiale bêta**

- pour **1** : **1/2**
- pour **0** : **1/2**

Sélectionnez les boutons d’option

- **<X<**
- **probabilité**

Entrez dans le champ **probabilité** : **0.95** pour obtenir la figure ci-après (avec **4** décimales pour la limite).



On trouve ici, pour les différentes distributions initiales considérées les intervalles à 95% pour les deux exemples suivants

$$n_1 = 20, a_1 = 19$$

$Bêta(0, 1)$	$Bêta(1, 1)$	$Bêta(1/2, 1/2)$	$Bêta(0, 0)$	$Bêta(1, 0)$
[0.7513, 0.9877]	[0.7618, 0.9883]	[0.7892, 0.9946]	[0.8235, 0.9987]	[0.8316, 0.9987]

$$n_1 = 59, a_1 = 32$$

$Bêta(0, 1)$	$Bêta(1, 1)$	$Bêta(1/2, 1/2)$	$Bêta(0, 0)$	$Bêta(1, 0)$
[0.4075, 0.6570]	[0.4161, 0.6633]	[0.4158, 0.6649]	[0.4240, 0.6728]	[0.4240, 0.6728]

Parmi ces distributions initiales, $Bêta(1, 0)$ donne les plus grandes limites et a les propriétés fréquentistes suivantes : la proportion des échantillons pour lesquels la limite supérieure est inférieure à φ est plus petite que $\alpha/2$ et la proportion des échantillons pour

pour pour lesquels la limite inférieure est supérieure à φ est plus grande que $\alpha/2$. La distribution $B\acute{e}ta(0, 1)$ donne les plus petites limites et les propriétés inverses. Par conséquent, considérer simultanément les limites de ces deux intervalles protège l'utilisateur à la fois d'une acceptation et d'un rejet erronés des hypothèses sur φ au seuil unilatéral $\alpha/2$.

Si l'on souhaite un seul intervalle pour résumer les résultats, ces propriétés conduisent à privilégier celui associé à la distribution "intermédiaire" $B\acute{e}ta(1/2, 1/2)$ (qui est la distribution de Jeffreys). Effectivement cet intervalle a de très bonnes propriétés fréquentistes qui justifient pleinement le nom de "méthode bayésienne objective". La probabilité de couverture de cet intervalle est très proche du seuil nominal, même pour des échantillons d'effectifs très faibles. Il se compare favorablement à la plupart des intervalles fréquentistes disponibles et est recommandable même d'un point de vue fréquentiste, comme l'ont montré Brown, Cai et DasGupta (2001) (voir aussi Agresti & Min, 2005)

"We revisit the problem of interval estimation of a binomial proportion. . . We begin by showing that the chaotic coverage properties of the Wald interval are far more persistent than is appreciated. . . We recommend the Wilson interval or the equal tailed Jeffreys prior interval for small n." (Brown, Cai & DasGupta, 2001, page 101).

Intervalles "de plus haute densité finale"

Une approche qui a été souvent recommandée par les bayésiens est de considérer l'intervalle "de plus haute densité finale" (*highest posterior density*, en bref HPD). Pour un tel intervalle (qui peut en fait, si la distribution n'est pas unimodale, être une réunion d'intervalles disjoints), la densité de probabilité est plus élevée pour toute valeur à l'intérieur de l'intervalle que pour toute valeur extérieure. Ceci répond à l'objectif d'obtenir l'intervalle le plus court possible. Mais, sauf pour une distribution symétrique, chacune des probabilités latérales sera différente de $\alpha/2$. Cette propriété apparaît indésirable pour l'analyse de données expérimentales où la plupart des questions sont, comme dans le présent exemple "unilatérales".

De plus un tel intervalle n'est pas invariant par transformation (excepté pour une transformation linéaire), ce qui peut être considéré avec Agresti et Min (2005, page 3) comme "*a fatal disadvantage*". Ainsi, pour les données $n = 59, a = 32$ et la distribution initiale $B\acute{e}ta(1/2, 1/2)$, nous obtenons les intervalles de plus haute densité finale

$$[0.4167, 0.6658] \text{ pour } \varphi \text{ et } [0.7481, 2.1594] \text{ pour } \frac{\varphi}{1-\varphi}$$

avec les probabilités correspondantes :

$$Pr(\varphi < 0.4167) = 0.026 \text{ et } Pr\left(\frac{\varphi}{1-\varphi} < 0.7481\right) = 0.039$$

$$Pr(\varphi > 0.6658) = 0.024 \text{ et } Pr\left(\frac{\varphi}{1-\varphi} > 2.1594\right) = 0.011$$

Pour ces raisons, nous n'avons pas retenu ces intervalles de plus haute densité finale dans nos programmes.

On remarquera en passant, à propos de cet exemple qu'il est tout aussi facile d'obtenir la distribution finale de $\frac{\varphi}{1-\varphi}$, qui est une distribution F de Fisher Snedecor. On trouve l'intervalle de crédibilité 95% à queues égales [0.712, 1.984].

Procédez comme précédemment pour l'inférence sur une proportion, mais activez le bouton d'option $\varphi/(1 - \varphi)$. Entrez dans les champs appropriés les **effectifs** :

- pour **1** : **32**
- pour **0** : **27**

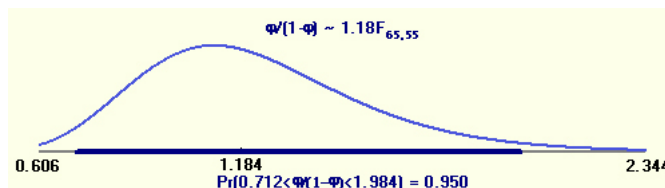
Entrez la distribution **initiale bêta**

- pour **1** : **1/2**
- pour **0** : **1/2**

Sélectionnez les boutons d'option

- **<X<**
- **probabilité**

Entrez dans le champ **probabilité** : **0.95** pour obtenir la figure ci-après (avec **3** décimales pour la limite).



5.5.3 Distributions initiales “informatives”

L'utilisation de distributions initiales non informatives a manifestement un statut privilégié pour obtenir des énoncés “à usage public”. Cependant d'autres techniques bayésiennes peuvent aussi avoir un rôle important à jouer dans la recherche expérimentale. En particulier, elles sont idéalement adaptées pour combiner des informations provenant de plusieurs sources, et par suite pour planifier une série d'expériences. Des utilisations plus ou moins réalistes et plus ou moins convaincantes de ces techniques ont été proposées. Quand une analyse bayésienne objective suggère une conclusion donnée, différentes distributions *a priori* traduisant les résultats d'autres expériences ou des opinions plus ou moins subjectives d’“experts”, soit *sceptiques* soit *enthousiastes*, peuvent être utilisées pour éprouver la robustesse des conclusions (voir en particulier Spiegelhalter, Freedman & Parmar, 1994).

On peut aussi considérer que la construction d'une distribution initiale à partir d'opinions d’“experts” du domaine peut être utile dans certaines études, mais cela nécessite des techniques appropriée (voir pour un exemple dans les essais cliniques Tan *et al.*, 2003).

Il convient encore de mentionner le fait que les distributions prédictives sont également un outil utile pour construire une distribution initiale subjective, car il est souvent plus facile d'exprimer une opinion relativement à des données attendues plutôt qu'à des paramètres.

L'utilisation de ces techniques pour l'analyse des données expérimentales n'a cependant pas assez été explorée pour pouvoir réellement apprécier leur apport, notamment dans le cas où l'information *a priori* reflète d'avantage une opinion qu'une connaissance réelle.

Les exemples donnés ci-après doivent être vus comme des exercices pour mieux comprendre comment l'inférence bayésienne permet de combiner des informations. Je laisse au lecteur le soin d'exercer son esprit critique quant à l'apport de ces méthodes pour l'analyse de données expérimentales.

Distributions “sceptiques” et “enthousiastes”

Reprenons l'exemple des données $n = 59, a = 32$, pour laquelle la procédure bayésienne objective conduisait à conclure que la méthode d'enseignement était inefficace ($\varphi < 0.70$). Considérons à titre d'illustration les deux distributions initiales, respectivement *a priori* très sceptique et très enthousiaste envers la méthode

$$\begin{aligned} \varphi &\sim \beta(20, 80) && \text{de moyenne } 0.200 && \text{pour laquelle } Pr(\varphi < 0.70) \approx 1 \\ \varphi &\sim \beta(98, 2) && \text{de moyenne } 0.980 && \text{pour laquelle } Pr(\varphi > 0.85) = 0.999998 \end{aligned}$$

Les distributions finales correspondantes sont respectivement

$$\begin{aligned} \varphi &\sim \beta(52, 107) && \text{de moyenne } 0.327 && \text{pour laquelle } Pr(\varphi < 0.70) \approx 1 \\ \varphi &\sim \beta(130, 29) && \text{de moyenne } 0.818 && \text{pour laquelle } Pr(\varphi > 0.85) = 0.143 \end{aligned}$$

La première “renforce” bien entendu la conclusion d'inefficacité. On peut le voir sur la figure ci-après, qui montre la distribution initiale (en noir) et la distribution finale (en rouge); celle-ci peut être comparée à la distribution finale objective – pour l'initiale *Bêta*(1/2, 1/2) (en bleu).

LeB-A-Bayésien

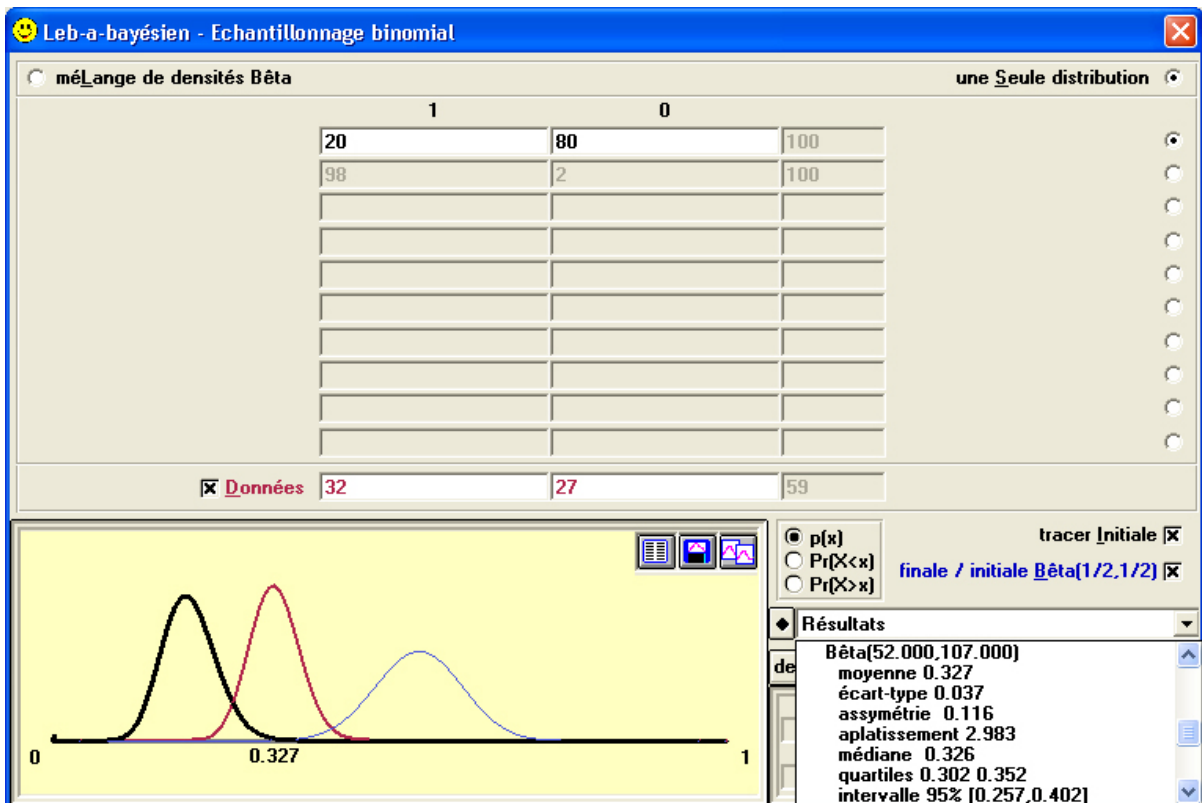
Dans **LePAC** activez le menu **LeB-A-Bayésien** et le sous-menu **échantillonnage Binomial**, ce qui affiche la fenêtre correspondante. Activez la case **données** et entrez les valeurs :

- pour **1** : **32**
- pour **0** : **27**

Entrez les paramètres de la distribution **initiale** (par défaut c'est l'option **une seule distribution** qui est active) :

- pour **1** : **20**
- pour **0** : **80**

Activez la case **finale / initiale Bêta(1/2,1/2)** pour obtenir la figure ci-après. Les courbes sont mises à jour automatiquement à chaque nouvelle entrée; au besoin, cliquez sur le bouton **calculer**.



Mais cette distribution initiale ne laissait “aucune chance” aux données d’infirmier l’opinion *a priori*, puisque même si on observait 59 succès et 0 erreur, on aurait quand même

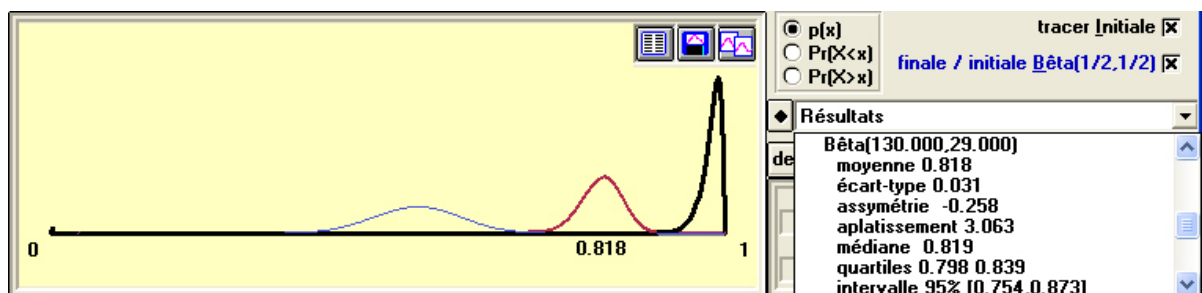
$$Pr(\varphi < 0.70) | a = 59) = 0.99999995$$

La seconde permet une conclusion beaucoup moins défavorable à la méthode d’enseignement, comme le montre la figure suivante

LeB-A-Bayésien

Dans la configuration de la figure précédente, activez le deuxième bouton d’option de la colonne **une seule distribution** et entrez les paramètres de la distribution **initiale** (voir la figure ci-dessus) :

- pour 1 : 98
- pour 0 : 2



mais les données ne permettent cependant pas de conclure à son efficacité

$$Pr(\varphi > 0.70 | a = 32) = 0.997 \text{ mais } Pr(\varphi > 0.85 | a = 32) = 0.143$$

Il est éclairant d'examiner l'effet de la distribution initiale $B\hat{e}ta(a_0, b_0)$ sur la moyenne de la distribution finale. En posant $n_0 = a_0 + b_0$, les rapports $n_0/(n_0 + n)$ et $n/(n_0 + n)$ représentent les poids relatifs de la distribution initiale et des données. La moyenne finale peut être écrite sous la forme

$$\frac{a_0 + a}{n_0 + n} = \frac{n_0}{n_0 + n} \times \frac{a_0}{n_0} + \frac{n}{n_0 + n} \times \frac{a}{n}$$

Elle est donc égale à

“poids relatif de l'initiale \times moyenne initiale + poids relatif des données \times moyenne observée”

soit ici les moyenne finales

$$100/159 \times 0.200 + 59/159 \times 0.542 = 0.327 \text{ pour la distribution initiale } \varphi \sim \beta(20, 80)$$

$$100/159 \times 0.980 + 59/159 \times 0.542 = 0.818 \text{ pour la distribution initiale } \varphi \sim \beta(98, 2)$$

Mélanges de densités $B\hat{e}ta$

Une technique qui reste simple de mise en œuvre est d'utiliser une distribution initiale dont la densité est un mélange de densités de distributions $B\hat{e}ta$, la distribution finale étant encore un tel mélange.

Ceci peut avoir deux intérêts, d'une part approcher une distribution initiale complexe quelconque qui nécessiterait le recours à des méthodes d'intégration numérique, d'autre part combiner différentes sources d'informations (ou différentes opinions). A titre d'illustration, considérons pour les mêmes données un mélange des deux distributions précédentes avec des poids égaux, soit

$$\varphi \sim 1/2 B\hat{e}ta(20, 80) \oplus 1/2 B\hat{e}ta(98, 2)$$

où \oplus signifie que c'est la densité de φ qui est un mélange, soit symboliquement

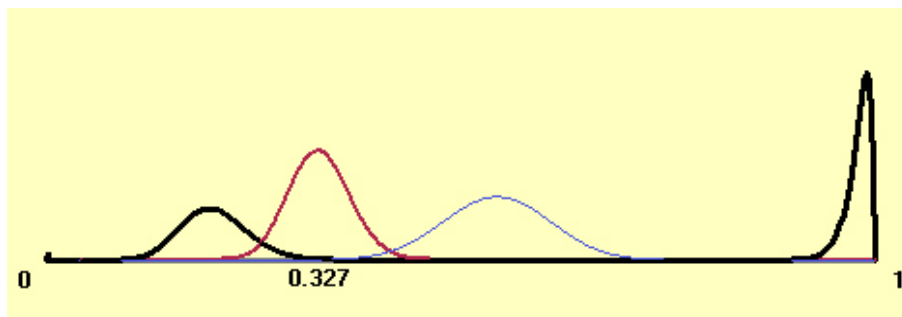
$$p(\varphi) = 1/2 p(B\hat{e}ta(20, 80)) + 1/2 p(B\hat{e}ta(98, 2))$$

Cette distribution ne doit pas être confondue avec celle d'une combinaison linéaire de deux variables ayant des distributions $B\hat{e}ta$ indépendantes (qui aurait une densité beaucoup plus complexe).

La figure ci-après montre la distribution initiale correspondante (en noir), qui est bimodale, et la distribution finale (en rouge); celle-ci peut être comparée à la distribution finale objective (en bleu).

LeB-A-Bayésien

Dans la configuration de la figure précédente, activez maintenant le bouton d'option **mélange de densités $B\hat{e}ta$** pour obtenir la figure ci-après.



En fait, dans ce cas, les données $n = 59, a = 32$ permettent en quelque sorte de trancher entre les deux distributions du mélange puisque la distribution finale est

$$0.999999903B\hat{e}ta(52, 107) \oplus 0.000000097B\hat{e}ta(130, 29)$$

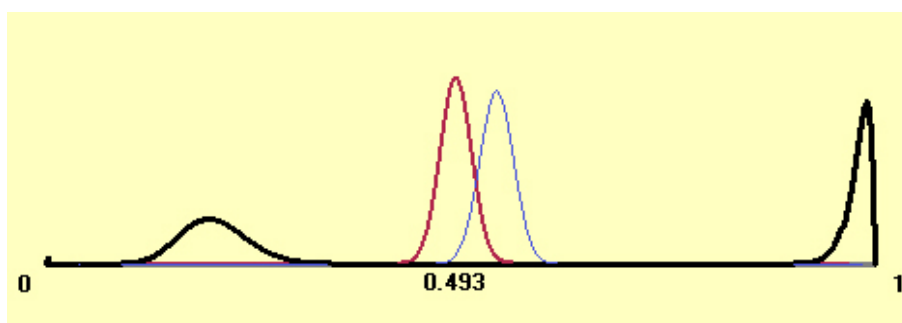
de sorte qu'elle est virtuellement confondue avec la distribution $B\hat{e}ta(52, 107)$ associée à la distribution initiale $B\hat{e}ta(20, 80)$ considérée précédemment.

Avec 10 fois plus de données, et la même proportion observée de succès ($n = 590, a = 320$), la distribution finale, que l'on peut voir sur la figure suivante serait pratiquement indiscernable de la distribution $B\hat{e}ta(340, 350)$ associée à la distribution initiale $B\hat{e}ta(20, 80)$, mais elle se rapprocherait bien entendu de la solution objective.

LeB-A-Bayésien

Dans la configuration de la figure précédente, modifiez les données :

- pour **1** : **320**
- pour **0** : **270**



Un autre exemple de mélanges

Lee (2004, page 59) donne un exemple sans doute plus réaliste d'utilisation de mélanges. Cet exemple est emprunté à Diaconis et Ylvisaker (1985). On part de la constatation suivante : quand on lance en l'air un grand nombre de fois une pièce de monnaie, on obtient approximativement une proportion 1/2 de *Face*, mais lorsqu'on la fait tourner sur elle même on obtient des proportions telles que 1/3 ou 2/3. Admettant ce résultat, on peut considérer qu'un mélange de deux densités de distributions $B\hat{e}ta$ de moyennes respectives 1/3 et 2/3 est un choix approprié pour la distribution initiale de la proportion φ de *Face*, soit par exemple

$$\varphi \sim \frac{1}{2}B\hat{e}ta(10, 20) \oplus \frac{1}{2}B\hat{e}ta(20, 10)$$

Si pour une pièce donnée on fait $n = 10$ essais et si on observe $a = 3$ *Face*, la distribution finale est

$$\frac{115}{129}B\hat{e}ta(13, 27) \oplus \frac{14}{129}B\hat{e}ta(23, 17)$$

Elle a pour moyenne 0.352 et donne un intervalle de crédibilité 95%

$$[0.194, 0.633]$$

Pour comparaison, la distribution finale objective a pour moyenne 0.318 avec un intervalle de crédibilité 95%

$$[0.093, 0.606]$$

Si on fait $n = 100$ essais et si on observe $a = 30$ *Face*, la distribution finale (de moyenne 0.308) est

$$\varphi \sim 0.9984\text{Bêta}(40, 90) \oplus 0.0016\text{Bêta}(50, 80)$$

L'intervalle de crédibilité 95% est

$$[0.232, 0.390]$$

qui est bien entendu dans ce cas beaucoup plus proche de celui obtenu pour la solution objective

$$[0.217, 0.395]$$

On peut choisir des poids différents ou encore faire intervenir dans le mélange une troisième distribution *Bêta* de moyenne $1/2$, etc.

5.5.4 Remarques

Interprétation des poids du mélange à partir des probabilités prédictives

Le poids associé à une distribution *Bêta* du mélange final est proportionnel au produit du poids initial par la probabilité prédictive associée à la distribution initiale *Bêta* correspondante. Reprenons l'exemple précédent d'une distribution initiale $\frac{1}{2}\text{Bêta}(10, 20) \oplus \frac{1}{2}\text{Bêta}(20, 10)$ avec les observations $n = 10$ et $a = 3$. Les probabilités prédictives d'observer $a = 3$ sont respectivement

$$Pr(a = 3) = 0.227632 \text{ pour la distribution initiale } \text{Bêta}(10, 20)$$

$$Pr(a = 3) = 0.027712 \text{ pour la distribution initiale } \text{Bêta}(20, 10)$$

Les poids du mélange finale sont proportionnels à $\frac{1}{2} \times 0.227632$ et $\frac{1}{2} \times 0.027712$ et on vérifie que

$$\frac{0.227632}{0.227632 + 0.027712} = \frac{115}{129} \quad \text{et} \quad \frac{0.027712}{0.227632 + 0.027712} = \frac{14}{129}$$

Modèle binomial négatif

Les mêmes procédures pour les paramètres – faisant intervenir des distributions *Bêta* ou des mélanges – s'appliquent au modèle binomial négatif dans lequel on arrête l'expérience quand on a observé un nombre fixé à l'avance de succès ou d'échecs (échantillonnage de Pascal), l'effectif n étant une variable aléatoire. On notera toutefois que la distribution initiale de Jeffreys est différente (voir l'annexe A). La distribution d'échantillonnage étant différente, la distribution prédictive est donc également différente; il s'agit dans ce cas d'une distribution *Bêta-Pascal*.

5.6 Le facteur de Bayes

Pour compléter la présentation de cet exemple, j'introduirai le *facteur de Bayes*, même si l'utilisation de celui-ci relève davantage d'une approche décisionnelle et pourra donc apparaître plus appropriée dans d'autres contextes.

Reprenons encore, pour l'expérience sur la méthode d'enseignement, l'exemple des données $n = 59, a = 32$, avec la distribution initiale enthousiaste $\varphi \sim \beta(98, 2)$ et les probabilités *a priori* $Pr(\varphi > 0.85) = 0.99999810$ (qu'on notera π_1) et donc $Pr(\varphi < 0.85) = 0.00000190$ (π_0). Les notations π_0 et π_1 sont usuelles car le facteur de Bayes est généralement présenté comme une approche bayésienne au tests d'hypothèses classiques;

dans ce cadre, π_0 et π_1 sont les probabilités de l'hypothèse nulle H_0 et de l'hypothèse alternative H_1 .

Il apparaît alors assez naturel de considérer :

- le rapport de ces deux probabilités *a priori*, soit

$$\frac{\pi_0}{\pi_1} = \frac{Pr(\varphi < 0.85)}{Pr(\varphi > 0.85)} = 0.0000019$$

qui est évidemment ici très faible ;

- et leur rapport *a posteriori*, soit

$$\frac{p_0}{p_1} = \frac{Pr(\varphi < 0.85 | a = 32)}{Pr(\varphi > 0.85 | a = 32)} = \frac{0.8570}{0.1430} = 5.99$$

qui est maintenant nettement supérieur à 1.

On définira alors le facteur de Bayes (associé à l'observation a) comme le rapport de ces deux rapports

$$B(a) = \frac{p_0/p_1}{\pi_0/\pi_1} = \frac{p_0\pi_1}{p_1\pi_0} = 3\,154\,986$$

ce qui “évalue donc la modification de la vraisemblance relative de l'hypothèse nulle qui est due aux observations” (Robert, 1992, page 166). Mais ceci n'est évidemment qu'un résumé incomplet, qui ne peut remplacer l'information fournie par les probabilités finales.

Le facteur de Bayes s'applique de la même manière à des hypothèses H_0 et H_1 “non complémentaires”, par exemple ici $\varphi < 0.70$ et $\varphi > 0.85$; mais l'interprétation est encore plus problématique puisqu'on ignore la zone de “non conclusion” $0.70 < \varphi < 0.85$.

Dans le cas particulier où H_0 et H_1 sont des hypothèses simples $\varphi = \varphi_0$ et $\varphi = \varphi_1$, le facteur de Bayes est simplement le classique *rapport de vraisemblance*

$$B(a) = \frac{p(\varphi_0 | a) p(\varphi_1)}{p(\varphi_1 | a) p(\varphi_0)} = \frac{p(a | \varphi_0)}{p(a | \varphi_1)}$$

puisque $p(\varphi_0 | a) \propto p(a | \varphi_0) p(\varphi_0)$ et $p(\varphi_1 | a) \propto p(a | \varphi_1) p(\varphi_1)$.

On notera encore que dans le cas où H_0 et H_1 sont des hypothèses “complémentaires” (donc $p_1 = 1 - p_0$), comme dans l'exemple précédent, on peut retrouver leurs probabilités finales à partir des probabilités initiales ($\pi_1 = 1 - \pi_0$) et du rapport de Bayes, puisqu'on vérifie facilement que

$$\frac{1}{p_0} = 1 + \frac{1 - \pi_0}{\pi_0} \frac{1}{B(a)}$$

Pour une discussion de l'utilisation du facteur de Bayes en liaison avec la problématique des tests d'hypothèses, on pourra consulter Robert (1992, pages 166-168).

6 Comparaison de deux proportions

Conceptuellement, les solutions précédentes pour une proportion s'étendent immédiatement à deux échantillons binomiaux indépendants, pourvu que les distributions initiales soient également indépendantes (Lecoutre, Grouin & Derzko, 1995). Nous illustrerons ces solutions à partir d'un modèle plus complexe, qui conduit aux mêmes procédures pour les paramètres et permet d'illustrer la souplesse de l'approche bayésienne.

6.1 Le problème : Règle d'expérimentation "rejouez le gagnant"

Pour comparer deux traitements, dans le plan d'expérience usuel "en groupes indépendants" les sujets sont affectés par tirage au sort à chacun des deux traitements. Ceci conduit à deux groupes d'effectifs égaux (en contraignant un peu le tirage au sort). Pour des considérations éthiques, on peut préférer un plan permettant d'attribuer le "meilleur" traitement un plus grand nombre de fois. Comme on ne sait pas *a priori* de quel traitement il s'agit ou qu'on ne veut pas faire intervenir une éventuelle connaissance dans la planification de l'expérience, on peut utiliser une règle appropriée basée sur un processus séquentiel. Ainsi la règle "rejouez le gagnant" (*play-the-winner*), décrite ci-après, répond à cet objectif (Zelen, 1969).

Le premier sujet reçoit l'un des deux traitements, t^1 ou t^2 , avec des probabilités égales. Ensuite, si le sujet $k - 1$ a reçu un traitement t dont le résultat est un "succès", le sujet suivant k reçoit le même traitement ; si au contraire le résultat est "un échec", le sujet k reçoit l'autre traitement.

La règle présuppose que le résultat du sujet $k - 1$ est connu quand le sujet k est inclus, mais elle pourrait être étendue pour tenir compte du cas où la réponse est différée. En dépit de son déterminisme apparent, cette règle est un processus stochastique, puisqu'elle dépend des probabilités de succès φ_1 et φ_2 de chacun des traitements.

Pour un total, supposé fixé à l'avance, de n sujets, la séquence des traitements attribués $(t_1, t_2 \dots t_k, t_{k+1} \dots t_{n+1})$ contient toute l'information des données. En effet $t_{k+1} = t_k$ implique qu'on a observé un succès à t_k , tandis que $t_{k+1} \neq t_k$ implique qu'on a observé un échec à t_k . La fonction de vraisemblance est donc simplement

$$v(\varphi_1, \varphi_2) | (t_1, \dots, t_{n+1}) = 1/2 \varphi_1^{n_{11}} (1 - \varphi_1)^{n_{10}} \varphi_2^{n_{21}} (1 - \varphi_2)^{n_{20}}$$

où n_{i1} ($i = 1, 2$) est le nombre de paires (t_k, t_{k+1}) égales à (t^i, t^i) , de sorte que n_{11} et n_{21} sont les nombres respectifs de succès aux traitements t^1 et t^2 ; n_{i0} est le nombre de paires (t_k, t_{k+1}) égales à (t^i, t^j) , avec $j \neq i$, de sorte que n_{10} et n_{20} sont les nombres d'échecs ; $1/2$ est la probabilité de t_1 .

On peut voir que la fonction de vraisemblance est proportionnelle au produit des fonctions de vraisemblance associées à chacun de deux échantillons binomiaux, d'effectifs respectifs $n_{11} + n_{10}$ et $n_{21} + n_{20}$ et de paramètres φ_1 et φ_2 . Elle est donc identique – à une constante multiplicative près – à celle associée à la comparaison de deux proportions binomiales indépendantes. Cependant, ici seul $n = n_{11} + n_{10} + n_{21} + n_{20}$ est fixé, alors que pour deux échantillons binomiaux $n_{11} + n_{10}$ et $n_{21} + n_{20}$ sont tous les deux fixés.

6.2 Solutions fréquentistes

Alors que, comme nous le verrons, les procédures bayésiennes sont les mêmes que pour deux échantillons binomiaux indépendantes, les procédures fréquentistes sont différentes. En effet, même si les fonctions de vraisemblance sont proportionnelles, les probabilités d'obtenir n_{11} succès au traitement t_1 et n_{21} succès au traitement t_2 avec la règle "rejouez le gagnant" sont données par des distributions beaucoup plus complexes que la distribution *Binomiale*, qui en outre ne sont bien entendu pas indépendantes. En raison de cette complexification, seules des procédures asymptotiques ont été proposées ; elles ne sont guère satisfaisantes en dehors d'échantillons d'effectif élevé.

6.3 Solution bayésienne

Dans le cadre bayésien au contraire, puisque leurs fonctions de vraisemblance sont proportionnelles, les procédures sur les paramètres sont les mêmes pour les deux modèles, bien que les probabilités d'échantillonnage soient très différentes. Une solution simple et usuelle suppose deux distributions initiales *Bêta* indépendantes pour φ_1 et φ_2 , respectivement :

$$\varphi_1 \sim \text{Bêta}(\nu_{11}, \nu_{10}) \text{ et } \varphi_2 \sim \text{Bêta}(\nu_{21}, \nu_{20})$$

Les distributions finales marginales sont encore deux distributions *Bêta* indépendantes :

$$\varphi_1 | (t_1, \dots, t_{n+1}) \sim \text{Bêta}(\nu_{11} + n_{11}, \nu_{10} + n_{10}) \text{ et } \varphi_2 | (t_1, \dots, t_{n+1}) \sim \text{Bêta}(\nu_{21} + n_{21}, \nu_{20} + n_{20})$$

6.3.1 Exemple numérique

Considérons à titre d'illustration les résultats suivants pour une expérience avec $n = 150$ sujets. On remarquera que, par définition de la règle, les nombres d'échecs (ici 20 et 21) ne peuvent différer que d'au plus une unité.

	succès	échecs	
traitement t_1	$n_{11} = 74$	$n_{10} = 20$	94
traitement t_2	$n_{21} = 35$	$n_{20} = 21$	56
	109	41	150

6.3.2 La simplicité conceptuelle des méthodes bayésiennes objectives

Un énoncé de probabilité conjoint est, dans un sens, le meilleur résumé de la distribution finale. Si nous adoptons la distribution initiale de Jeffreys ($\nu_{11} = \nu_{10} = \nu_{21} = \nu_{20} = 1/2$), nous obtenons par exemple la probabilité finale conjointe (en omettant le conditionnement par les données pour simplifier l'écriture) :

$$Pr(\varphi_1 > 0.697 \text{ et } \varphi_2 < 0.743) = 0.95$$

qui se déduit de

$$Pr(\varphi_1 > 0.697) = Pr(\varphi_2 > 0.743) = \sqrt{0.95} = 0.974679$$

(obtenues comme dans le cas de l'inférence sur une proportion) en utilisant l'indépendance des deux distributions finales

$$\varphi_1 \sim \text{Bêta}(74.5, 20.5) \text{ et } \varphi_2 \sim \text{Bêta}(35.5, 21.5)$$

Cependant nous pouvons préférer un énoncé qui permet de comparer directement les deux traitements. Ainsi nous avons

$$Pr(\varphi_1 > \varphi_2) = 0.984$$

De plus il est facile de considérer les principaux critères classiques pour comparer deux proportions (différence, rapport, etc.). Dans le cadre fréquentiste, chacun de ces critères nécessite des procédures différentes. On sait en outre que dans le cas de deux échantillons binomiaux indépendants il existe pour chacun d'eux une pléthore de procédures⁶.

⁶De nouvelles procédures sont en outre régulièrement proposées. Certaines d'entre elles réservent d'ailleurs des surprises. C'est par exemple le cas des "new confidence intervals" développés par Zhou et Qin (2004, 2005). Pour l'exemple d'effectifs observés pour deux groupes indépendants (2,8) et (1,35), les auteurs (Zhou & Qin, 2004, pages 108-109) donnent pour la différence $\varphi_1 - \varphi_2$ (avec les notations

De même qu'il existe une correspondance entre le seuil observé du test binomial et la probabilité bayésienne finale pour une distribution initiale appropriée, on trouve également dans ce cas un lien avec les tests conditionnels de Fisher. Ceci sera illustré plus loin à propos d'une autre situation. Mais bien entendu ces procédures fréquentistes ne seraient pas applicables ici puisque la distribution d'échantillonnage n'est pas la même.

Au contraire la solution bayésienne est conceptuellement immédiate. En effet la distribution de n'importe quel paramètre dérivé peut être obtenue à partir de la distribution finale conjointe⁷. Le problème est seulement technique : il faut en général recourir à des méthodes d'intégration numérique. Par exemple nous trouvons les intervalles de crédibilité 95% (à "queues égales")

[+0.013, +0.312]	pour	$\varphi_1 - \varphi_2$
[1.02, 1.62]	pour	$\frac{\varphi_1}{\varphi_2}$
[1.07, 4.64]	pour	$\frac{\varphi_1/(1-\varphi_1)}{\varphi_2/(1-\varphi_2)}$

LesProportions

Dans **LePAC** activez le menu **LesBayésiens** et le sous-menu **LesProportions**, ce qui affiche la fenêtre pour l'inférence sur des proportions. Activez le bouton d'option **2 groupes indépendants**. Entrez dans les champs appropriés les **effectifs** :

- pour **g1** : **74** et **20**
- pour **g2** : **35** et **21**

Entrez s'il y a lieu les paramètres de la distribution **initiale bêta** – soit dans chaque cas **1/2** (mais c'est en principe inutile car c'est l'option par défaut).

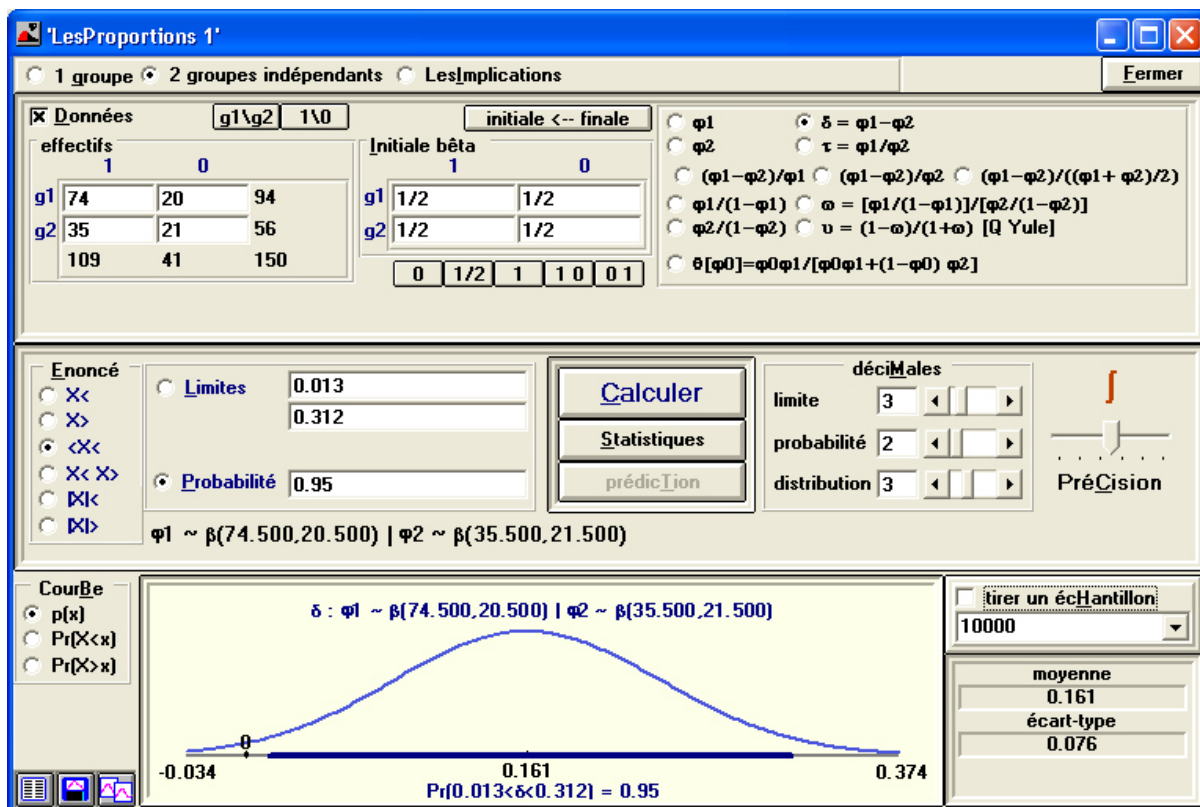
Sélectionnez les boutons d'option

- $\delta = \varphi_1 - \varphi_2$
- **<X<**
- **probabilité**

Entrez dans le champ **probabilité** : **0.95** pour obtenir la figure ci-après (avec **3** décimales pour la limite).

utilisées ici) les intervalles de confiance 95% : [+0.005, +0.516] (pour leur "direct Edgeworth expansion method") et [-0.024, +0.544] (pour leur "transformation method"). Si on a la curiosité de calculer ces mêmes intervalles pour la différence opposée $\varphi_2 - \varphi_1$, on a la surprise de ne pas trouver des intervalles symétriques, mais au contraire des intervalles largement différents : respectivement [-0.373, +0.138] et [-0.401, +0.171] (Lecoutre & Faure, 2005). Les auteurs répondent que leur procédure peut être modifiée pour résoudre ce problème (Zhou & Qin, 2007), mais cela ne fait que renforcer l'impression d'"*ad hocquerie*" des méthodes fréquentistes

⁷ "**Reference priors**". Bernardo (1979) et Berger et Bernardo (1992) ont développé sous cette appellation des distributions initiales qui visent à fournir la distribution objective *optimale* pour un paramètre dérivé θ , quand celui-ci est le *paramètre d'intérêt*. Elles peuvent être regardées comme un raffinement de la règle de Jeffreys. Cette solution présente cependant un certain nombre d'inconvénients (outre l'introduction de calculs plus complexes) : elle nécessite d'utiliser des distributions initiales différentes pour chaque paramètre auquel on peut s'intéresser et, surtout, en cherchant à retrouver certaines propriétés des procédures fréquentistes, elles peuvent conférer aux procédures bayésiennes des propriétés indésirables de ces procédures fréquentistes. On notera d'ailleurs que, bien que préconisant ces "reference priors", Berger (2004) dans une illustration pratique des méthodes bayésiennes subjectives (*diagnostic médical*, qui est traité dans la section suivante) s'en tient à la solution de Jeffreys



On a bien entendu

$$Pr(\varphi_1 - \varphi_2 > 0) = Pr\left(\frac{\varphi_1}{\varphi_2} > 1\right) = Pr\left(\frac{\varphi_1/(1-\varphi_1)}{\varphi_2/(1-\varphi_2)} > 1\right) = Pr(\varphi_1 > \varphi_2) = 0.984$$

Par sa souplesse, l'approche bayésienne est aussi bien appropriée pour un objectif clairement décisionnel (notamment sélectionner le meilleur traitement) que pour l'estimation (par exemple apprécier la différence d'efficacité entre les deux traitements).

Comme dans le cas d'une proportion, pour la distribution initiale de Jeffreys les méthodes bayésiennes ont de très bonnes propriétés fréquentistes de couverture (Lecoutre & ElQasyr, 2005).

Comme pour une proportion, les procédures prédictives fournissent des solutions pour choisir les effectifs ou prendre une décision d'arrêter l'expérience avant son terme. Dans le cas de deux échantillons binomiaux indépendants, les solutions sont une extension immédiate, les distributions prédictives pour deux échantillons futurs étant des distributions *Bêta-Binomiales* indépendantes; elles sont illustrée dans Lecoutre, Derzko et Grouin (1995). Dans le cas de la règle "rejouez le gagnant", les techniques sont les mêmes, mais les distributions d'échantillonnage étant différentes, les distributions prédictives sont bien entendu différentes (ElQasyr, 2006).

7 Une généralisation du modèle binomial avec trois proportions

7.1 Le problème : Diagnostic médical

Mossman & Berger (2001) donnent l'exemple du diagnostic d'une maladie M .

On considère une population pour laquelle la probabilité de la maladie est φ_0 . On utilise un test de diagnostic qui est positif $[+]$ avec une probabilité φ_1 si le patient a la maladie et avec une probabilité φ_2 si le patient n'a pas la maladie

$$Pr(M) = \varphi_0 \quad Pr(+|M) = \varphi_1 \quad Pr(+|\neg M) = \varphi_2$$

On en déduit par le théorème de Bayes la probabilité θ que le patient ait la maladie sachant que le test est positif

$$Pr(M|+) = \frac{Pr(+|M)Pr(M)}{Pr(+|M)Pr(M) + Pr(+|\neg M)Pr(\neg M)} = \frac{\varphi_1\varphi_0}{\varphi_1\varphi_0 + \varphi_2(1 - \varphi_0)} = \theta$$

Généralisant encore la situation de l'inférence sur une proportion binomiale, on suppose que l'on dispose de données indépendantes a_i ($i=0,1,2$) ayant chacune une distribution *Binomiale*

$$a_i|\varphi_i \sim Bin(\varphi_i, n_i)$$

7.2 Solutions fréquentistes

Une des motivations des auteurs était que les approches fréquentistes précédemment développées étaient, soit difficiles à appliquer, soit relativement peu performantes.

7.3 Solution bayésienne

Au contraire la solution bayésienne utilisant trois distributions initiales *Bêta* indépendantes pour les φ_i est conceptuellement très simple. C'est une généralisation immédiate de l'inférence sur deux proportions binomiales. En particulier, pour la solution non informative de Jeffreys (trois distributions $\varphi_i \sim Bêta(1/2, 1/2)$ indépendantes), la procédure est très performante d'un point de vue fréquentiste.

“A great frequentist confidence procedure” (Berger, 2004)

On peut ainsi conclure ironiquement avec Berger que cette procédure est une “grande procédure de confiance fréquentiste”. De la distribution finale conjointe (encore trois distributions *Bêta* indépendantes), on déduit facilement par une méthode numérique un intervalle de crédibilité pour θ .

Ce problème a également été traité, dans un contexte différent, par Zaykin, Meng et Ghosh (2004). Ceux-ci considèrent, outre la solution précédente, le cas le plus simple où φ_0 est une valeur donnée, ainsi que le cas où la distribution de φ_0 est une distribution uniforme sur un intervalle fixé *a priori* (non révisée par les données).

7.3.1 Exemple numérique

Considérons les données suivantes (Mosmann et Berger, 2001, page 505)

$n_0 = 30$	$a_0 = 7$	$n_1 = 20$	$a_1 = 17$	$n_2 = 40$	$a_2 = 1$
------------	-----------	------------	------------	------------	-----------

Pour la méthode bayésienne objective (Jeffreys), on obtient les distributions finales indépendantes

φ_0	φ_1	φ_2
$Bêta(7.5, 23.5)$	$Bêta(17.5, 3.5)$	$Bêta(1.5, 39.5)$

Conditionnellement à φ_0 , en remarquant que

$$\theta = \frac{1}{1 + \frac{1-\varphi_0}{\varphi_0} \frac{\varphi_2}{\varphi_1}}$$

on voit que le problème se ramène à l'inférence traitée sur le rapport de deux proportions binomiales indépendantes traitée dans la section précédente. On a par exemple

$$Pr(\theta < u | \varphi_0) = Pr\left(\frac{\varphi_2}{\varphi_1} > \frac{1-\varphi_0}{\varphi_0} \frac{1-u}{u}\right)$$

On obtient ainsi les intervalles pour différentes valeurs fixées de φ_0 , par exemple

$$\begin{array}{cccc} \varphi_0 = 0.15 & \varphi_0 = 0.20 & \varphi_0 = 0.25 & \varphi_0 = 0.30 \\ [0.569, 0.982] & [0.647, 0.987] & [0.710, 0.990] & [0.759, 0.992] \end{array}$$

LesProportions

Dans **LePAC** activez le menu **LesBayésiens** et le sous-menu **LesProportions**, ce qui affiche la fenêtre pour l'inférence sur des proportions. Activez le bouton d'option **2 groupes indépendants**. Entrez dans les champs appropriés les **effectifs** :

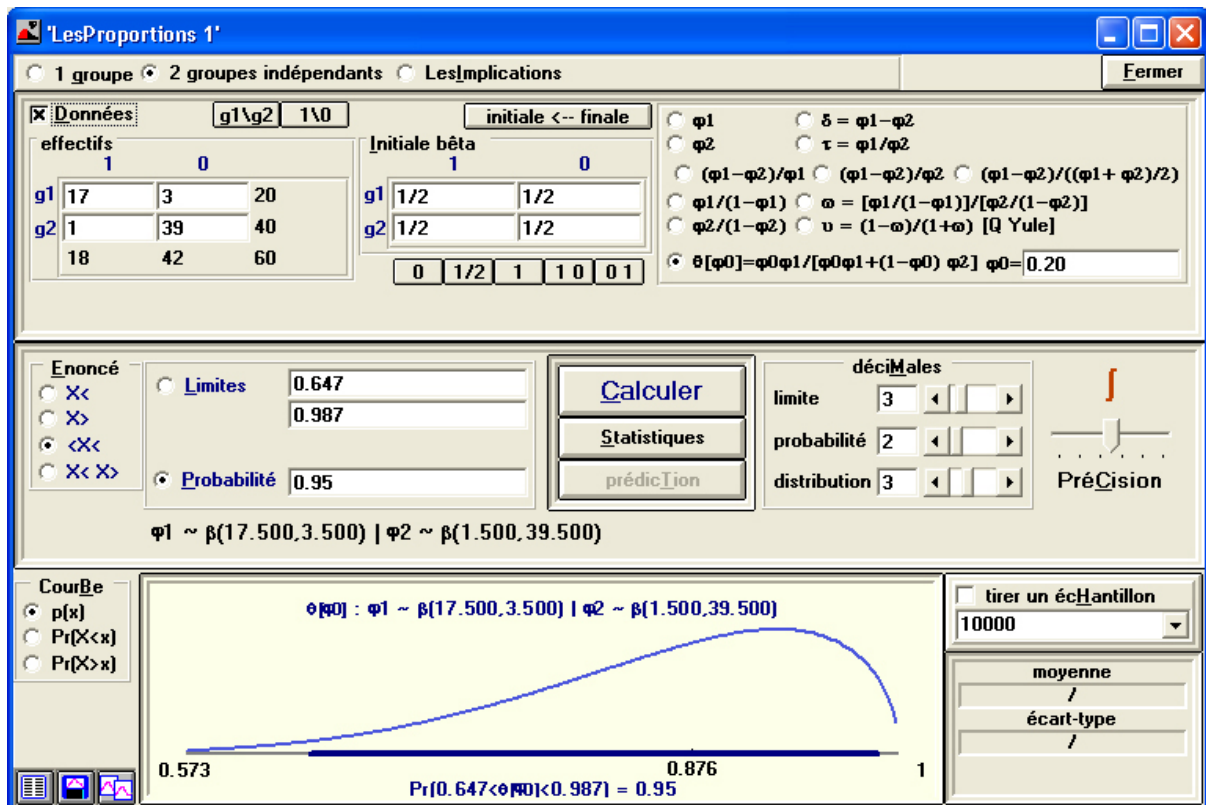
- pour **g1** : **17** et **3**
- pour **g2** : **1** et **39**

Entrez s'il y a lieu les paramètres de la distribution **initiale bêta** – soit dans chaque cas **1/2** (mais c'est en principe inutile car c'est l'option par défaut).

Sélectionnez les boutons d'option

- $\vartheta[\varphi_0] = \varphi_0\varphi_1/[\varphi_0\varphi_1 + (1-\varphi_0)\varphi_2]$
et entrez la valeur pour φ_0 : 0.20
- **<X<**
- **probabilité**

Entrez dans le champ **probabilité** : **0.95** pour obtenir la figure ci-après (avec **3** décimales pour la limite).



La distribution finale de θ est un mélange des distributions conditionnelles à φ_0 . On peut l'obtenir par simulation. En simulant un échantillon de 100 000 valeurs du triplet $(\varphi_0, \varphi_1, \varphi_2)$ et par suite de θ , j'ai obtenu l'intervalle de crédibilité 95% (Mossman et Berger avaient trouvé $[0.633, 0.990]$)

$$[0.632, 0.990]$$

Remplaçons maintenant la distribution finale de φ_0 par la distribution uniforme sur l'intervalle $[0.111, 0.404]$, qui est l'intervalle de crédibilité 95% pour φ_0 obtenu pour la distribution finale précédente $Bêta(7.5, 23.5)$. On obtient alors, toujours pour 100 000 simulations, l'intervalle réellement très proche

$$[0.632, 0.991]$$

On peut également constater que conditionnellement à $\varphi_0 = 0.20$ – qui est approximativement la moyenne 0.1974 de la distribution $Bêta(7.5, 23.5)$ – l'intervalle obtenu $[0.647, 0.987]$ est encore proche des précédents.

8 Un modèle multinomial pour un tableau 2×2

8.1 Le problème : Etude d'un modèle logique

Considérons un groupe de n sujets, avec deux ensembles d'attributs (ou variables) binaires, respectivement notés $V = \{v_1, v_0\}$ et $W = \{w_1, w_0\}$. Pour fixer les idées, supposons que W soit le fait qu'un individu soit décédé d'une maladie cardiaque et que V soit le fait qu'il ait eu ou non au préalable un infarctus du myocarde. Considérons l'exemple suivant de "modèle logique" (Lecoutre et Charron, 2000).

Il existe une implication absolue (ou logique) $v1 \Rightarrow w1$ (par exemple) si tous les individus ayant la modalité $v1$ ont aussi la modalité $w1$, alors que l'inverse n'est pas nécessairement vrai.

Mais l'hypothèse d'une implication absolue est de peu d'intérêt en pratique, puisqu'il suffit d'observer une seule fois l'événement $(v1, w0)$ pour la réfuter.

En conséquence, nous devons considérer l'hypothèse plus faible " $v1$ implique dans la plupart des cas $w1$ " ($v1 \leftrightarrow w1$).

Le problème est d'évaluer l'écart au modèle logique d'une implication absolue qui prédit que "la case $[v1, w0]$ est vide". Un indice d'écart peut être défini à partir des proportions des différentes cases

	$w1$	$w0$	
$v1$	φ_{11}	φ_{10}	$\varphi_{1.}$
$v0$	φ_{01}	φ_{00}	$\varphi_{0.}$
	$\varphi_{.1}$	$\varphi_{.0}$	1

comme

$$\frac{\varphi_{10}}{\varphi_{1.}\varphi_{.0}} \quad \text{qui varie de } 0 \text{ à } +\infty$$

En fait il est plus usuel de considérer le complément à 1 de cet indice, qui sera noté $\eta_{v1 \leftrightarrow w1}$ (et qui est donc à proprement parler un indice d'ajustement du modèle)

$$\eta_{v1 \leftrightarrow w1} = 1 - \frac{\varphi_{10}}{\varphi_{1.}\varphi_{.0}} \quad (-\infty < \eta_{v1 \leftrightarrow w1} < +1)$$

Cet indice a été utilisé dans des contextes variés, avec différentes approches. Il peut être vu notamment comme une mesure d'*efficacité prédictive* du modèle pour prédire l'issue de W étant donné v_1 .

La prédiction est parfaite (il y a une implication absolue) quand $\eta_{v1 \leftrightarrow w1} = +1$.

La prédiction est d'autant plus efficace que $\eta_{v1 \leftrightarrow w1}$ est plus proche de +1.

En cas d'indépendance on a $\eta_{v1 \leftrightarrow w1} = 0$.

Une valeur nulle ou négative signifie que le modèle est rejeté.

En conséquence, pour pouvoir conclure à l'efficacité prédictive du modèle, nous devons démontrer que $\eta_{v1 \leftrightarrow w1}$ a une valeur proche de +1. On peut définir de la même manière les indices $\eta_{v1 \leftrightarrow w0}$, $\eta_{w1 \leftrightarrow v1}$ et $\eta_{w0 \leftrightarrow v0}$, ou encore caractériser l'équivalence entre deux modalités. Il y a équivalence (absolue) entre $v1$ et $w1$ (par exemple) si $\eta_{v1 \leftrightarrow w1} = 1$ et $\eta_{w1 \leftrightarrow v1} = 1$ ("les cases $[v1, w0]$ et $[v0, w1]$ sont vides") et on pourra prendre comme indice d'ajustement au modèle d'équivalence le minimum de ces deux indices.

Nous supposons ici un modèle d'échantillonnage multinomial, soit pour un échantillon d'effectif n la probabilité d'observer les effectifs n_{ij}

$$Pr(n_{11}, n_{10}, n_{01}, n_{00} | \varphi_{11}, \varphi_{10}, \varphi_{01}, \varphi_{00}) = \frac{n!}{n_{11}!n_{10}!n_{01}!n_{00}!} \varphi_{11}^{n_{11}} \varphi_{10}^{n_{10}} \varphi_{01}^{n_{01}} \varphi_{00}^{n_{00}}$$

8.2 Solutions fréquentistes

Pour construire un intervalle de confiance, on a proposé des solutions asymptotiques (par exemple, Fleiss, 1981), qui sont manifestement inappropriées pour les petits échantillons. On a également développé des solutions basées sur le test conditionnel de Fisher (Copas et

Loeber, 1990 ; Lecoutre et Charron, 2000). Ce test utilise la distribution d'échantillonnage de n_{11} (par exemple). Un résultat classique est que cette distribution n_{11} , étant donné les marges observées fixées, dépend seulement du produit croisé $\rho = \frac{\varphi_{11}\varphi_{00}}{\varphi_{10}\varphi_{01}}$ (voir Cox, 1970, page 4). On peut ainsi, à partir de cette distribution, tester l'hypothèse nulle $\rho = \rho_0$ contre l'hypothèse alternative $\rho < \rho_0$ (ou contre $\rho > \rho_0$) en utilisant la probabilité que n_{11} dépasse la valeur observée (dans la direction appropriée). On a donc une procédure analogue au test binomial considéré dans le cas de l'inférence sur une proportion, avec de la même manière la possibilité de définir une solution "incluante" et une solution "excluante". On obtient comme cas particulier le "test de permutation" de Fisher pour l'hypothèse nulle $\rho = 1$, c'est-à-dire $\eta_{11} = 0$.

En inversant ce test conditionnel, on peut construire un intervalle de confiance pour le produit croisé ρ . On en déduit un intervalle pour η_{11} en remplaçant ρ par ses limites de confiance dans l'expression suivante qui donne η_{11} en fonction de ρ

$$\eta_{11} = \frac{1 + (\rho - 1)(\varphi_{1.} + \varphi_{.1} - \varphi_{1.}\varphi_{.1} - [(1 + (\varphi_{1.} + \varphi_{.1})(\rho - 1)^2 - 4\varphi_{1.}\varphi_{.1}\rho(\rho - 1)]^{1/2}}{2(\rho - 1)\varphi_{.1}(1 - \varphi_{1.})}$$

Malheureusement ces limites dépendent des "vraies" marges $\varphi_{.1}$ et $\varphi_{1.}$. La procédure la plus courante consiste à simplement remplacer ces *paramètres parasites* par leurs estimateurs $f_{.1}$ et $f_{1.}$. Elle est nettement plus performante que les solutions asymptotiques, mais n'est pas satisfaisante pour des valeurs extrêmes des paramètres. Même si on peut trouver des principes plus efficaces pour traiter les paramètres parasites (par exemple, Toecher, 1950 ; Rice, 1988), on se trouve ici confronté à un éternel problème qui est bien entendu complètement évité par l'approche bayésienne.

8.3 Solution bayésienne

La solution bayésienne pour un modèle d'échantillonnage multinomial est une généralisation immédiate du cas binomial. Choisissons une distribution initiale (conjuguée) de *Dirichlet*, qui est une généralisation multidimensionnelle de la distribution *Bêta*

$$(\varphi_{11}, \varphi_{10}, \varphi_{01}, \varphi_{00}) \sim \text{Dirichlet}(\nu_{11}, \nu_{10}, \nu_{01}, \nu_{00})$$

La distribution finale est également une distribution de *Dirichlet* dans laquelle les poids initiaux sont simplement ajoutés aux effectifs observés

$$(\varphi_{11}, \varphi_{10}, \varphi_{01}, \varphi_{00}) | \text{données} \sim \text{Dirichlet}(n_{11} + \nu_{11}, n_{10} + \nu_{10}, n_{01} + \nu_{01}, n_{00} + \nu_{00})$$

En utilisant les propriétés fondamentales de la distribution de *Dirichlet* (voir par exemple Bernardo & Smith, 1994, page 135), la distribution finale du paramètre dérivé η_{11} peut être caractérisée comme une fonction de trois distributions *Bêta* indépendantes.

$$\begin{aligned} X &= \varphi_{10} | \text{données} \sim \text{Bêta}(n_{10} + \nu_{10}, n_{11} + \nu_{11} + n_{01} + \nu_{01} + n_{00} + \nu_{00}) \\ Y &= \frac{\varphi_{00}}{1 - \varphi_{10}} = \frac{\varphi_{00}}{1 - X} | \text{données} \sim \text{Bêta}(n_{00} + \nu_{00}, n_{11} + \nu_{11} + n_{01} + \nu_{01}) \\ Z &= \frac{\varphi_{11}}{1 - \varphi_{10} - \varphi_{00}} = \frac{\varphi_{11}}{(1 - Y)(1 - X)} | \text{données} \sim \text{Bêta}(n_{11} + \nu_{11}, n_{01} + \nu_{01}) \end{aligned}$$

puisque

$$\eta_{v1 \leftrightarrow w1} = 1 - \frac{X}{(X + Z(1 - Y)(1 - X))(X + Y(1 - X))}$$

nous sommes ramenés au même problème technique que dans la situation précédente du diagnostic médical.

8.3.1 Exemple numérique : Etude de mortalité

Considérons l'implication "Infarctus du myocarde \leftrightarrow Décès cardiaque avant deux ans". Pour des patients qui n'ont subi aucun traitement médical, on dispose des données suivantes relatives à 340 patients "à risque"

		Décès			
		oui	non		
Infarctus du myocarde	oui	20	72	92	[20/92=0.22]
	non	17	231	248	[17/248=0.07]
		37	303	340	

On remarquera que les proportions de décès sont assez faibles (heureusement!) – respectivement 0.22 après un infarctus et 0.07 sans infarctus – de sorte que l'effectif 72 dans la case [oui,non] est en rapport élevé. En conséquence des valeurs relativement faibles de l'indice sont ici "cliniquement significantes". Nous avons ici les valeurs observées de l'indice

- pour l'implication "Infarctus \leftrightarrow Décès" (case [oui,non] vide) : $H_{v1 \leftrightarrow w1} = 0.12$
- pour l'implication "Décès \leftrightarrow Infarctus" (case [non,oui] vide) : $H_{w1 \leftrightarrow v1} = 0.37$

La méthode bayésienne objective, avec la distribution initiale de Jeffreys *Dirichlet* ($1/2, 1/2, 1/2, 1/2$), donne la distribution finale

$$(\varphi_{11}, \varphi_{10}, \varphi_{01}, \varphi_{00}) | \text{données} \sim \text{Dirichlet}(20.5, 72.5, 17.5, 231.5)$$

ce qui conduit aux intervalles de crédibilité suivants

- "Infarctus \leftrightarrow Décès" : $Pr(+0.06 < \eta_{v1 \leftrightarrow w1} < +0.19) = 0.90$
- "Décès \leftrightarrow Infarctus" : $Pr(+0.20 < \eta_{w1 \leftrightarrow v1} < +0.54) = 0.90$

Ces intervalles permettent de conclure à l'existence d'une implication d'importance limitée et indiquent qu'en fait ici le décès est un meilleur pronostic de l'infarctus que l'inverse.

LesProportions

Dans **LePAC** activez le menu **LesBayésiens** et le sous-menu **LesProportions**, ce qui affiche la fenêtre pour l'inférence sur des proportions. Activez le bouton d'option **LesImplications**. Entrez dans les champs appropriés les **effectifs** :

- pour **v1** : **20** (**w1**) et **72** (**w0**)
- pour **v0** : **17** (**w1**) et **231** (**w0**)

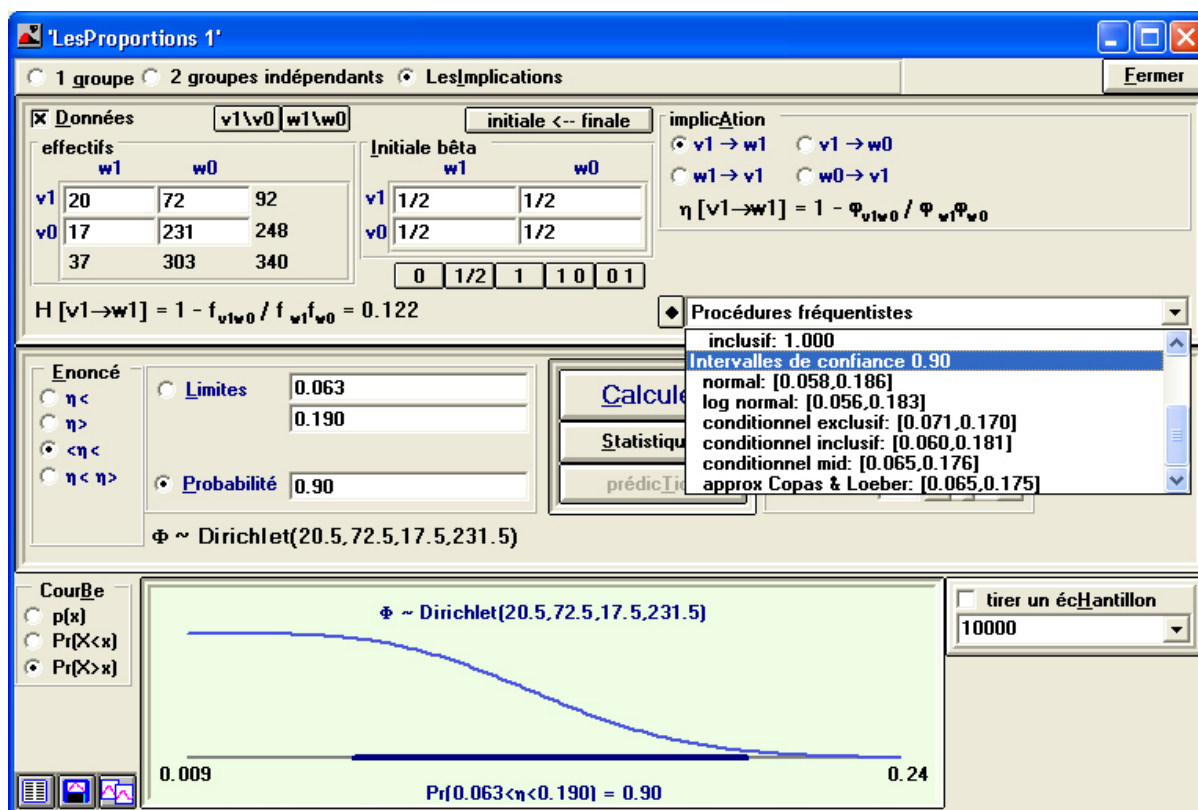
Entrez s'il y a lieu les paramètres de la distribution **initiale bêta** – soit dans chaque cas **1/2** (mais c'est en principe inutile car c'est l'option par défaut).

Sélectionnez les boutons d'option

- pour l'**implication** : **v1** \rightarrow **w1** (c'est l'option par défaut)
- pour l'**énoncé** : $< \eta <$
- pour la **courbe** : $Pr(X > x)$

- probabilité

Entrez dans le champ **probabilité** : **0.90** pour obtenir la figure ci-après (avec **3** décimales pour la limite).



On dispose également des données suivantes relatives à 357 patients ayant reçu un traitement médical à titre préventif

		Décès		
		oui	non	
Infarctus du myocarde	oui	1	78	79 [1/79=0.01]
	non	13	265	278 [13/278=0.05]
		14	343	357

Dans ce cas on espère évidemment que le traitement va réduire le nombre de décès après infarctus. Idéalement, s'il n'y avait aucun décès cardiaque chez les patients traités après un infarctus (case [oui,oui] vide), on aurait l'implication absolue "Infarctus \Rightarrow Non décès". On obtient les résultats suivants pour cette implication

"Infarctus \leftrightarrow Non décès" : $H_{v1 \leftrightarrow w0} = +0.68$ et $Pr(-0.10 < \eta_{v1 \leftrightarrow w0} < +0.94) = 0.90$

qui montrent qu'ici, en dépit d'une valeur observée nettement plus élevée de l'indice, on ne peut pas conclure à l'existence d'une implication. La grande largeur de l'intervalle de crédibilité indique que la précision est mauvaise, ce qui est une conséquence des très faibles proportions de décès observées. Bien entendu, on ne peut pas non plus conclure qu'il n'y a pas d'implication ou que l'implication est faible. On voit ici le danger qu'il y aurait à interpréter le résultat non significatif des "tests d'indépendance" usuels (le khi-deux par exemple) en faveur de l'hypothèse d'indépendance.

8.3.2 Interprétation du seuil observé du test de permutation de Fisher

Profitons en pour illustrer la réinterprétation bayésienne du test de permutation (conditionnel aux marges). Pour le test unilatéral usuel (solution “incluante”), l’hypothèse nulle $H_0 : \eta_{v1 \leftrightarrow w0} = 0$ n’est pas rejetée, le seuil observé étant $p_{inc} = 0.145$. On sait que ce test est conservateur ; mais si on adopte la solution “excluante”, on obtient un seuil nettement plus petit $p_{exc} = 0.028$, ceci étant du à la mauvaise précision expérimentale. Généralisant le cas d’une proportion, il existe deux distributions initiales non informatives extrêmes, qui donnent une interprétation éclairante de ces seuils

$Pr(\eta_{v1 \leftrightarrow w0} < 0) = 0.145 = p_{inc}$ pour la distribution initiale <i>Dirichlet</i> (1, 0, 0, 1) (la plus défavorable à H_0) soit la distribution finale <i>Dirichlet</i> (2, 78, 13, 266)
$Pr(\eta_{v1 \leftrightarrow w0} < 0) = 0.028 = p_{exc}$ pour la distribution initiale <i>Dirichlet</i> (0, 1, 1, 0) (la plus favorable à H_0) soit la distribution finale <i>Dirichlet</i> (1, 79, 14, 265)
$Pr(\eta_{v1 \leftrightarrow w0} < 0) = 0.072 \approx \frac{p_{inc} + p_{exc}}{2} = 0.086$ pour la distribution initiale <i>Dirichlet</i> ($1/2, 1/2, 1/2, 1/2$) soit la distribution finale <i>Dirichlet</i> (1.5, 78.5, 13.5, 265.5)

LesProportions

Dans la configuration de la figure précédente, entrez maintenant les données :

- pour **v1** : **1** (**w1**) et **78** (**w0**)
- pour **v0** : **13** (**w1**) et **265** (**w0**)

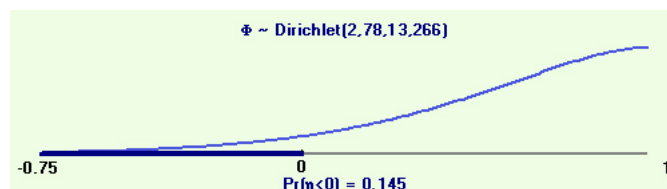
Entrez la distribution **initiale bêta** :

- pour **v1** : **1** (**w1**) et **0** (**w0**)
- pour **v0** : **0** (**w1**) et **1** (**w0**)

Sélectionnez les boutons d’option

- pour l’**implication** : **v1** \rightarrow **w0**
- pour l’**énoncé** : $\eta <$
- pour la **courbe** : $Pr(X > x)$
- **limite**

Entrez dans le champ **limite** : **0** pour obtenir la figure ci-après (avec **3** décimales pour la probabilité).



8.3.3 Avantages de l'inférence bayésienne

Dans toutes les situations l'inférence bayésienne traite de manière explicite et performante le problème des paramètres parasites. Elle prend explicitement en compte les problèmes liés au caractère discret de la distribution et à la possibilité d'observer des effectifs nuls grâce la distribution initiale. Comme pour le cas d'une seule proportion, le choix d'une distribution initiale non informative n'est ni plus ni moins arbitraire ou subjective que les conventions de l'approche fréquentiste.

En outre les probabilités de couverture des intervalles de crédibilité bayésiens se comparent favorablement à celles des intervalles de confiance fréquentistes. Ainsi, une étude de simulation (Lecoutre & Charron, 2000) montre que les deux taux d'erreurs fréquentistes associés aux deux distributions initiales non informatives extrêmes encadrent toujours le taux d'erreur recherché. La méthode bayésienne objective correspondant à la distribution initiale symétrique intermédiaire entre ces deux extrêmes a des propriétés de couverture remarquables. Naturellement l'écart entre ces différentes distributions initiales se réduit jusqu'à disparaître quand la taille de l'échantillon augmente. En ce qui concerne les procédures fréquentistes, il apparaît que les intervalles basés sur l'approche conditionnelle – avec les solutions incluant, excluant, et moyenne – sont moins performants que leurs analogues bayésiens. En particulier les deux taux d'erreurs associés aux deux solutions extrêmes n'encadrent pas toujours le taux recherché.

Il n'y a aucune difficulté à étendre de manière immédiate les procédures bayésiennes à toute autre situation faisant intervenir le modèle multinomial. En particulier ici on obtient très facilement par simulation la distribution du minimum de deux indices pour établir l'équivalence entre deux modalités; on peut aussi aisément effectuer une inférence pour comparer les indices associés à deux groupes indépendants (par exemple ici les patients traités et non traités); etc.

9 Inférence sur une moyenne

Reprenons l'exemple médical utilisé par Student (1908) dans son article originel sur la procédure qui a été ensuite appelée le "test t de Student".

9.1 Le problème : L'exemple historique de Student

Etant donné, pour chacun des $n=10$ patients, les heures de sommeil supplémentaires procurées par l'utilisation de chacun des deux somnifères "*soporific* [1]" et "*soporific* [2]", Student effectuait une inférence sur la différence des moyennes entre les deux somnifères, en construisant une nouvelle série de données obtenue "en soustrayant [1] de [2]". Les dix différences individuelles ainsi obtenues sont données dans le tableau suivant.

+1.2	+2.4	+1.3	+1.3	0	+1.0	+1.8	+0.8	+4.6	+1.4
------	------	------	------	---	------	------	------	------	------

 $n = 10$

9.2 Solutions fréquentistes : le test t usuel et l'intervalle de confiance

Student calculait alors la moyenne $+1.580$ [d] et l'écart-type (non corrigé) standard 1.167 [soit $s = 1.230$, corrigé pour les degrés de liberté] de cette série. Il concluait alors à partir de sa table de la "distribution t " : "*the probability is .9985 or the odds are about 666 to 1 that 2 is the better soporific*" (ce qui n'est certainement pas une formulation fréquentiste *orthodoxe*!). En termes modernes, nous calculons la statistique de test t pour l'inférence sur une moyenne sous le modèle normal $t = +1.580 / (1.230 / \sqrt{10}) = +4.062$ et nous trouvons le seuil unilatéral $p_{unil} = 0.0014$ (9 *dl*), soit $1 - p_{unil} = 0.9986$ qui correspond à la valeur ".9985" calculée par Student (les calculs étant effectués ici à partir des données avec la précision maximale).

Pour les généralisations ultérieures, il est commode d'introduire la notation $b = 1/\sqrt{n}$. Nous obtenons alors l'intervalle de confiance $1 - \alpha$ à partir du $100(1 - \alpha/2)$ -ème percentile de la distribution de Student à $q = n - 1$ degrés de liberté

$$d \pm bst_q^{1-\alpha/2} \text{ soit ici pour } \alpha = 0.05 \text{ l'intervalle } [+0.70, +2.46]$$

$$\text{avec } b = \frac{1}{\sqrt{n}} = 0.316 \quad q = n - 1 = 9 \quad t_9^{0.975} = +2.262$$

9.3 Solution fiducio-bayésienne

Cet exemple historique de Student est une application typique de l'approche que nous appelons "l'analyse spécifique" (Rouanet & Lecoutre, 1983 ; Lecoutre, 1984, 2005, 2006 ; Lecoutre et al., 2000). Les données de base sont pour chacun des $n=10$ patients la différence entre les heures de sommeil supplémentaires obtenues par l'usage d'un somnifère ("*hyoscyamine hydobromide*"), les heures de sommeil étant mesurées avant et après traitement avec soit [1] "*dextro hyoscyamine hydobromide*" soit [2] "*lævo hyoscyamine hydobromide*" (on notera que ces données de base sont elles-mêmes déjà des données dérivées). L'analyse de Student est un exemple typique d'inférence spécifique : elle est effectuée directement à partir des données *dérivées* pertinentes et ne met en jeu que l'inférence sur la moyenne d'une distribution normale.

Nous pouvons appliquer aux données précédentes l'inférence bayésienne élémentaire sur la moyenne, avec seulement deux paramètres, la différence moyenne de la population δ et la variance σ^2 . Ici d et s^2 sont des statistiques conjointement exhaustives pour le couple (δ, σ^2) ; elles ont les distributions d'échantillonnage indépendantes respectives (avec les notations $b = 1/\sqrt{n}$ et $q = n - 1$ introduites précédemment)

$$d | \delta, \sigma^2 \sim N(\delta, b^2 \sigma^2)$$

$$s^2 | \delta, \sigma^2 \sim \sigma^2 \frac{\chi_q^2}{q} \text{ que l'on notera plus simplement } s^2 | \delta, \sigma^2 \sim \sigma^2 \varphi_q^2$$

La traduction bayésienne de l'exhaustivité est que la distribution finale ne dépend que de ces statistiques. En particulier, la distribution finale *fiducio-bayésienne* (ou "bayésienne objective") marginale de δ est une distribution t *généralisée*. Elle est centrée sur la différence moyenne observée $d = +1.580$ et a pour facteur d'échelle $e = bs = s/\sqrt{n} = 0.389$. Cette distribution a le même nombre de degrés de liberté $q = n - 1 = 9$ que le test t .

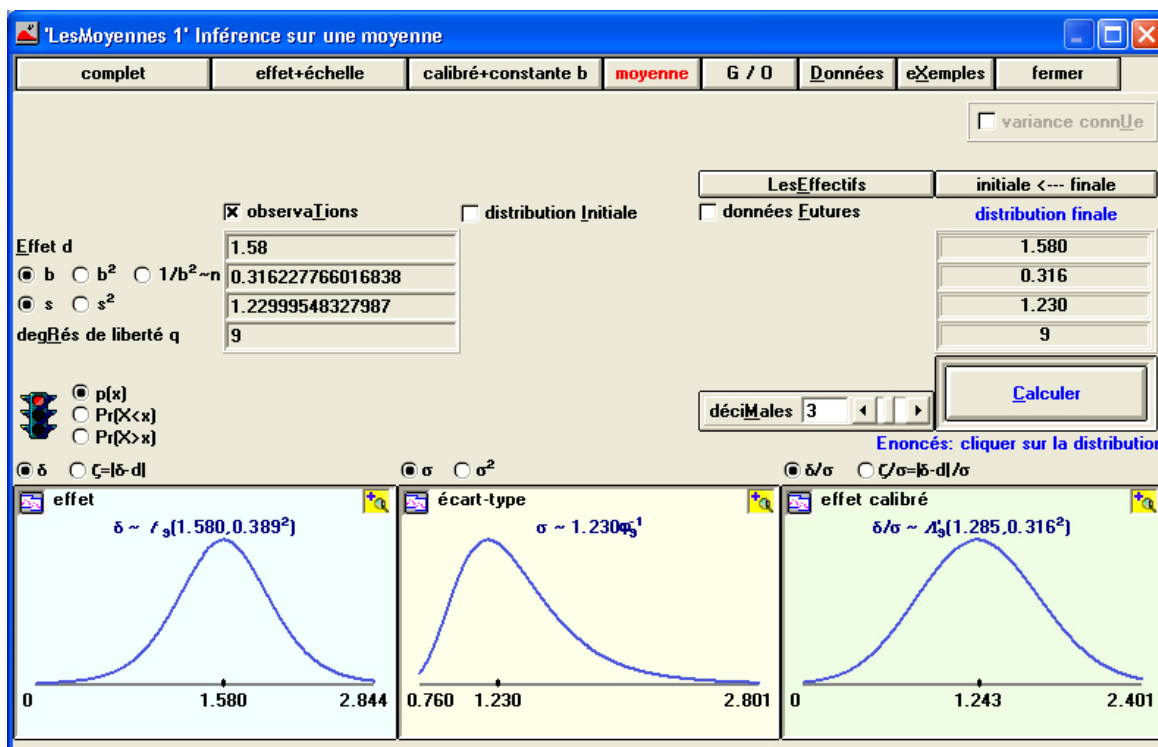
Elle s'écrit $d + et_q$, ou encore par analogie avec la distribution normale

$$\delta \mid \text{données} \sim t_q(d, e^2) \text{ soit ici } \delta \mid \text{données} \sim \delta \mid d, s^2 \sim t_9(+1.580, 0.389^2)$$

Il faut noter que cette distribution ne doit pas être confondue avec la distribution t noncentrée, familière aux utilisateurs de l'analyse de la puissance.

LesMoyennes

Dans **LePAC** activez le menu **LesBayésiens** et le sous-menu **LesMoyennes**, ce qui affiche la fenêtre pour l'inférence sur des moyennes. Cliquez sur le bouton **Exemples** et cliquez sur **Student**, ce qui affiche les valeurs pertinentes pour cet exemple dans une fenêtre **inférence sur une moyenne**. Cliquez sur le bouton **calculer** pour obtenir le test de signification et différentes statistiques descriptives (voir le détail dans la liste déroulante **t - degrés de liberté : q=9**). Puis cliquez sur le bouton **ok** pour obtenir la figure ci-après.



Le facteur d'échelle e est le dénominateur de la statistique de test t usuelle, puisque $t = d/e$ (en supposant $d \neq 0$)

$$\delta \mid \text{données} \sim \delta \mid d, t \sim t_q\left(d, \left(\frac{d}{t}\right)^2\right)$$

Par conséquent, la distribution fiducio-bayésienne de δ peut être directement dérivée de d et de $t=+4.062$. Ce résultat met en évidence la propriété fondamentale de la statistique de test t d'être un estimateur de la précision expérimentale, *conditionnellement à la valeur observée d* . Plus précisément, $(d/t)^2$ estime la variance d'erreur d'échantillonnage $\epsilon^2 = b^2\sigma^2$ de d .

9.4 Interprétation fiducio-bayésienne du seuil unilatéral

Le seuil unilatéral p_{unil} du test t est exactement la probabilité fiducio-bayésienne que la vraie différence δ soit de signe opposé à celui de la différence observée. Pour les données de Student ($p_{unil} = 0.0014$), il y a une probabilité finale 0.14% que la différence soit négative et la probabilité complémentaire 99.86% qu'elle soit positive.

$$Pr(\delta < 0 | \text{données}) = p_{unil} = 0.0014 \text{ et } Pr(\delta > 0 | \text{données}) = 1 - p_{unil} = 0.9986$$

Si l'énoncé de Student "la probabilité est .9985 [...] que [2] est le meilleur somnifère" est **interdit** dans le cadre fréquentiste, il est donc parfaitement justifié dans le cadre fiducio-bayésien.

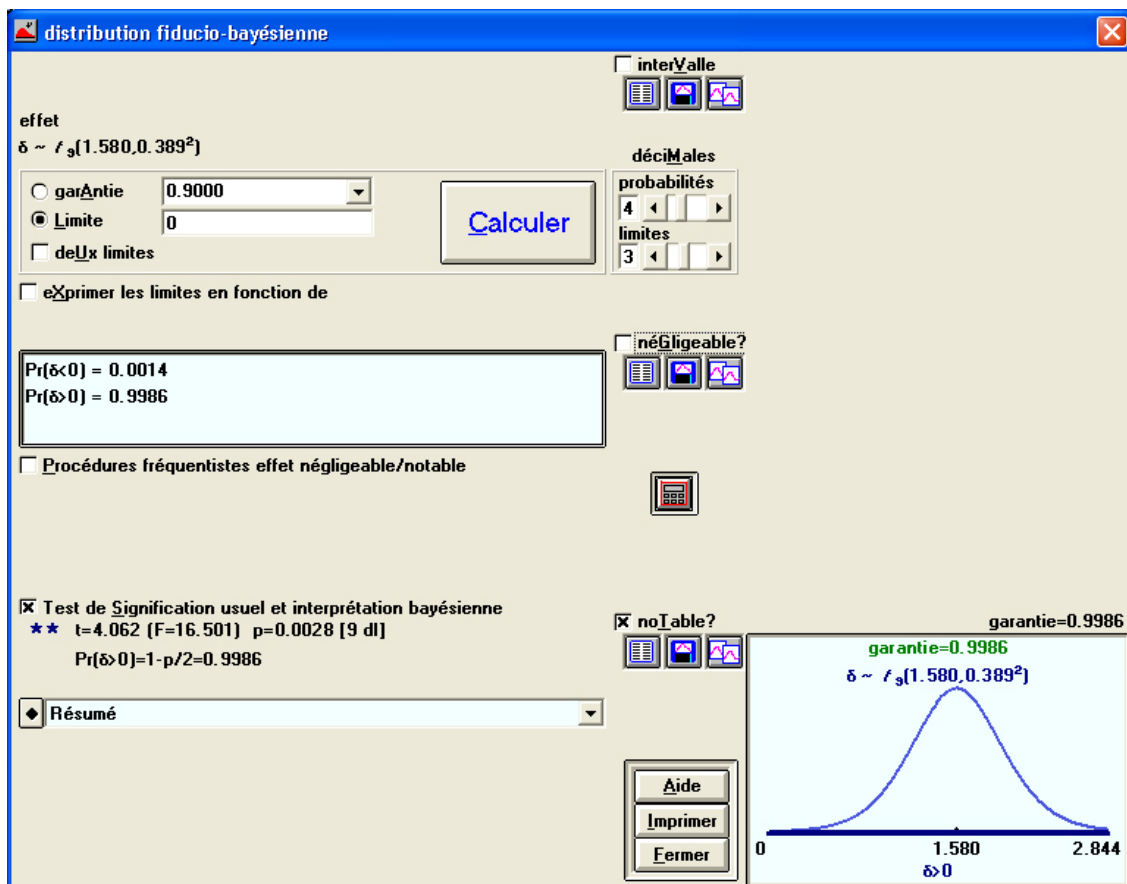
LesMoyennes

Dans la situation de la figure précédente, cliquez sur la courbe de l'effet δ , ce qui affiche la "fenêtre des énoncés".

Sélectionnez le bouton d'option

- **limite**

Entrez dans le champ **limite** : 0 et appuyez sur touche "Entrée" du clavier (ou cliquez sur le bouton **Calculer**) pour obtenir la figure ci-après (avec 4 décimales pour les probabilités). Notez que l'on peut désactiver les cases à cocher **intervalle** et **négligeable** qui ne sont pas pertinentes ici).



De plus cette interprétation met clairement en évidence les insuffisances méthodologiques des tests de signification. Il devient manifeste que le seuil observé *en lui-même* ne dit rien

sur la grandeur de δ . D'une part, un résultat, même "hautement significatif" (p_{unil} "très petit"), permet seulement de conclure que δ a le même signe que la différence observée d . D'autre part, un résultat "non-significatif" n'est en toute rigueur qu'un constat d'ignorance, comme cela est illustré par l'interprétation fiducio-bayésienne $Pr(\delta < 0) = Pr(\delta > 0) = 1/2$ d'un test "parfaitement non-significatif" (soit $d = 0$).

9.5 Interprétation fiducio-bayésienne de l'intervalle de confiance fréquentiste

Il y a une probabilité (ou *garantie*) 95% que δ soit compris entre les limites (fixées) de l'intervalle de confiance fréquentiste – conditionnellement aux données – soit ici entre +0.70 et +2.46 heures

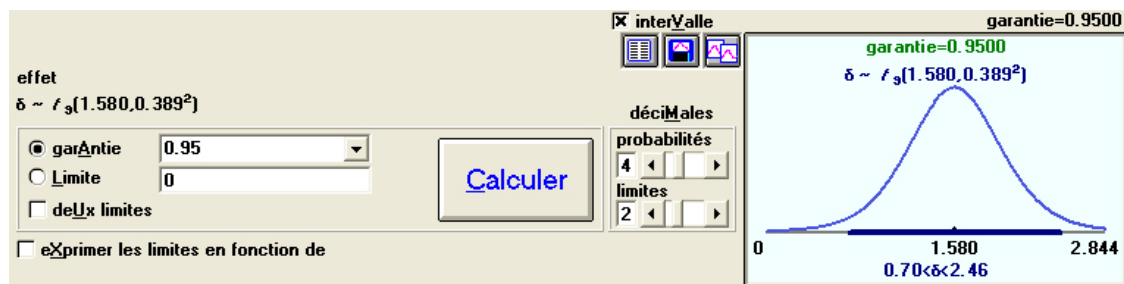
$$Pr(+0.70 < \delta < +2.46 \mid \text{données}) = 0.95$$

LesMoyennes

Dans la situation de la figure précédente, sélectionnez le bouton d'option

- **garantie**

et entrez dans le champ **garantie** : **0.95** (que vous pouvez sélectionner dans la liste déroulante) pour obtenir la figure ci-après (avec **2** décimales pour les limites). Activez si besoin la case à cocher **intervalle**.



9.6 Réponses bayésiennes directes aux questions sur la grandeur des effets

Au delà des réinterprétations des procédures fréquentistes usuelles, d'autres énoncés bayésiens fournissent des réponses directes aux questions sur la grandeur des effets. Nous pouvons calculer la probabilité que δ dépasse un temps de sommeil supplémentaire fixé, plus aisé à interpréter, par exemple une heure

$$Pr(\delta > +1 \mid \text{données}) = 0.915$$

"Il y a une probabilité 91.5% que δ dépasse une heure". Puisque l'unité de mesure est ici signifiante, il est aisé d'apprécier la signification pratique de la grandeur de δ .

Pour résumer les résultats, on peut rapporter : "il y a une probabilité finale 91.5% que la différence soit positive et grande ($\delta > +1$), une probabilité 8.4% qu'elle soit positive mais limitée ($0 < \delta < +1$), et une probabilité 0.14% qu'elle soit négative". Un tel énoncé n'a pas d'équivalent fréquentiste.

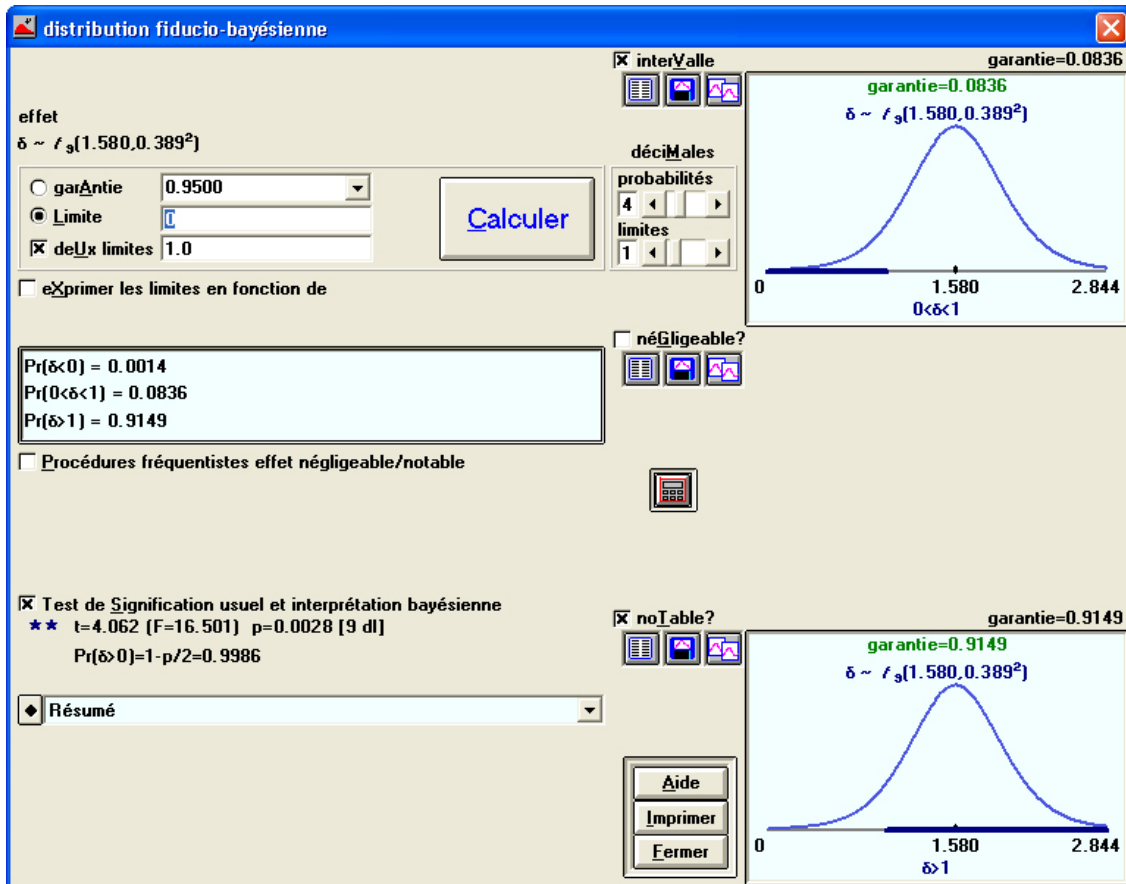
Les Moyennes

Dans la situation de la figure précédente, sélectionnez le bouton d'option

- **limite**

et activez la case à cocher **deux limites**.

Entrez dans les champs **limite** et **deux limites** : **0** et **1** pour obtenir la figure ci-après (avec **1** décimale pour les limites et **4** pour les probabilités).



9.7 La question de la réplication des observations

Etant donné l'expérience réalisée, la distribution prédictive exprime notre état de connaissance sur des données futures. Par exemple, que pouvons-nous dire de la valeur de la différence d' que nous observerions pour de nouvelles données? La distribution prédictive pour d' dans un échantillon futur d'effectif n' est naturellement plus dispersée que la distribution de δ relative à la population (ceci d'autant plus que l'effectif du nouvel échantillon est plus petit). Ainsi la distribution fiducio-bayésienne finale prédictive pour d' , étant donné la valeur d observée dans les données disponibles, est encore une distribution t généralisée (naturellement centrée sur d) :

$$d' \sim t_q(d, e^2 + e'^2) \text{ où } e' = s/\sqrt{n'}$$

En fait, l'incertitude sur δ conditionnellement aux données disponibles (reflétée par e^2) s'ajoute à l'incertitude sur les résultats de l'échantillon futur quand δ est connue (reflétée par e'^2). Par exemple, à partir des données de Student, la distribution prédictive pour

une unité expérimentale future ($n' = 1$) est $d' \sim t_9(+1.580, 1.290^2)$. Ainsi, pour une unité expérimentale supplémentaire, “il y a une probabilité 87.4% que la différence soit positive et une probabilité 78.8% que la différence dépasse une demi-heure”.

LesMoyennes

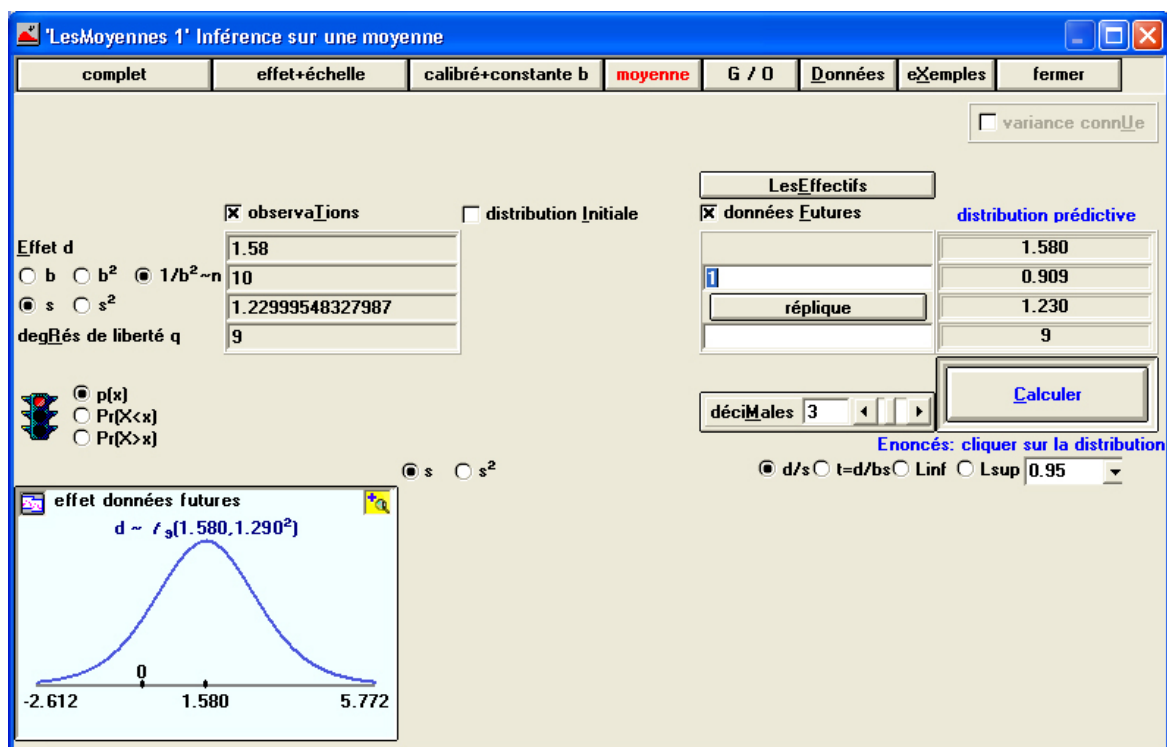
Dans la situation de la figure précédente, fermez la “fenêtre des énoncés” pour revenir à la “fenêtre principale”.

Sélectionnez le bouton d’option

- $1/b^2 \sim n$

ce qui affiche l’effectif 10 au lieu de la constante $b = 1/\sqrt{10}$

Activez la case à cocher **données futures** et entrez dans le champ $1/b^2 \sim n$ pour les données futures l’effectif **1**. Cliquez sur le bouton **calculer** pour obtenir la figure ci-après.



Pour un échantillon futur de taille 10, soit une réplique avec le même effectif ($e' = e$), la distribution prédictive est $d' \sim t_9(+1.580, 0.550^2)$.

On notera que l’on peut de même obtenir la distribution prédictive relative à la statistique de test t ou aux limites de crédibilité fiducio-bayésiennes (soit encore les limites de confiance fréquentistes). Ceci met en jeu une nouvelle distribution appelée K -prime (voir Lecoutre, 1996a, 1999, 2001). Ces distributions permettent notamment de déterminer l’effectif nécessaire pour obtenir une conclusion voulue (par exemple $\delta > 0.5$) avec une garantie fixée.

LesMoyennes

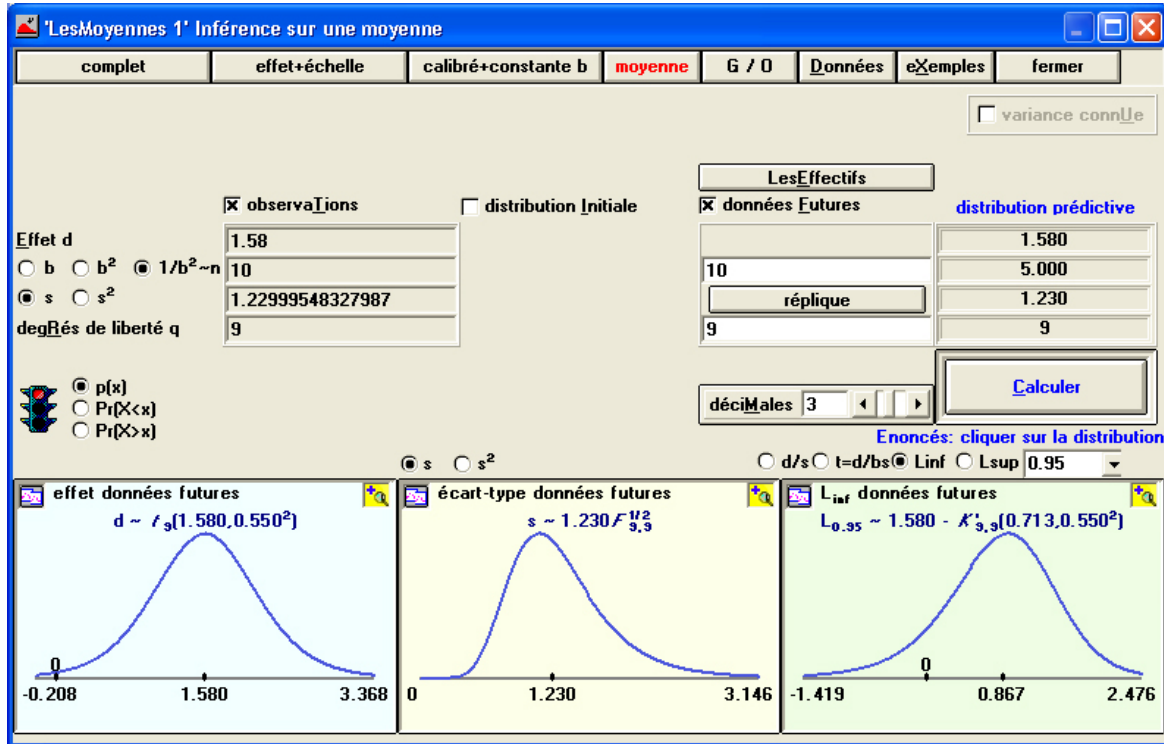
Dans la situation de la figure précédente, entrez maintenant pour les données futures : l’effectif **n : 10**

les **degrés de liberté q** : 9

ce que vous pouvez afficher directement en cliquant sur le bouton **réplique**.
Sélectionnez le bouton d'option

- **Linf**

et cliquez sur le bouton **calculer** pour obtenir la figure ci-après.



On trouve ici pour la différence de l'échantillon futur : “une probabilité prédictive 0.991 qu'elle soit positive et une probabilité 0.959 qu'elle dépasse une demi-heure”.

LesMoyennes

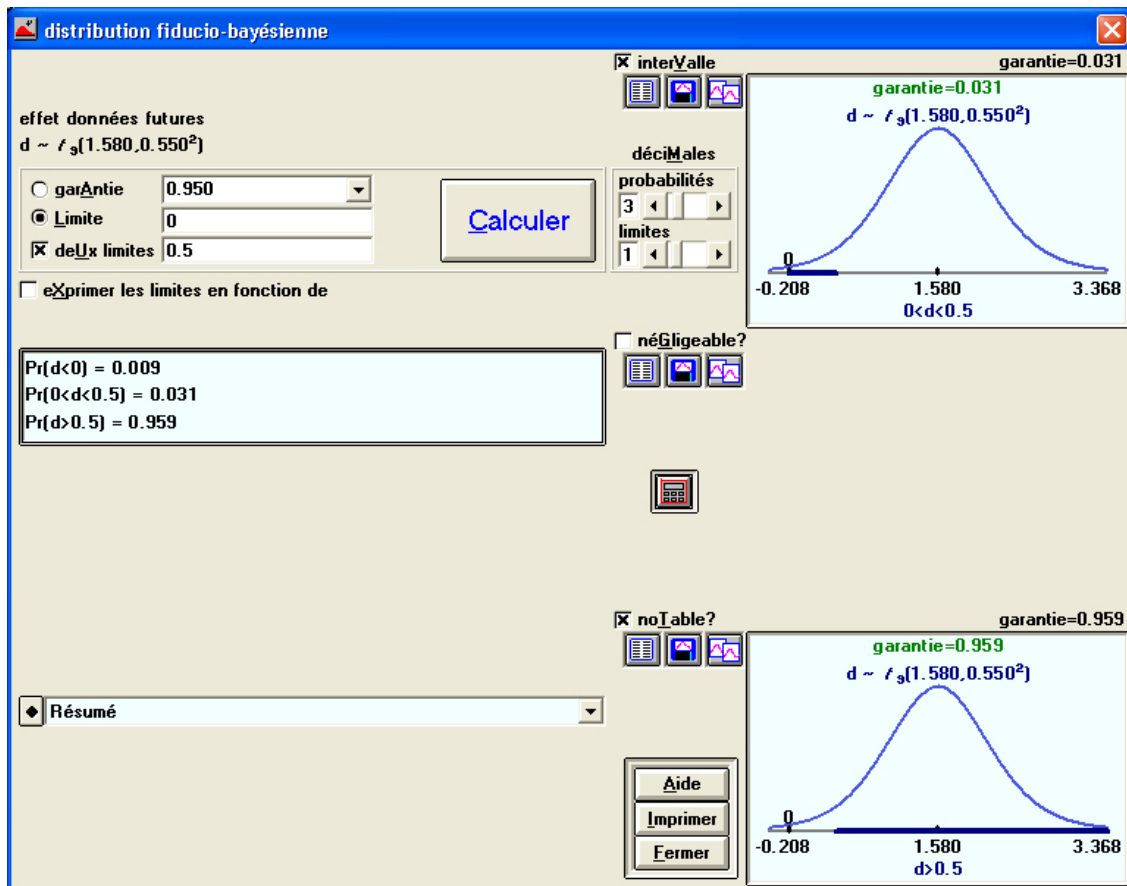
Dans la situation de la figure précédente, cliquez sur la courbe de l'**effet données futures d**, ce qui affiche la “fenêtre des énoncés”.

Sélectionnez le bouton d'option

- **limite**

et activez la case à cocher **deux limites**.

Entrez dans les champs **limite** et **deux limites** : 0 et 0.5 pour obtenir la figure ci-après (avec 1 décimale pour les limites et 3 pour les probabilités). La case à cocher **négligeable** qui n'est pas pertinentes ici peut être désactivée.



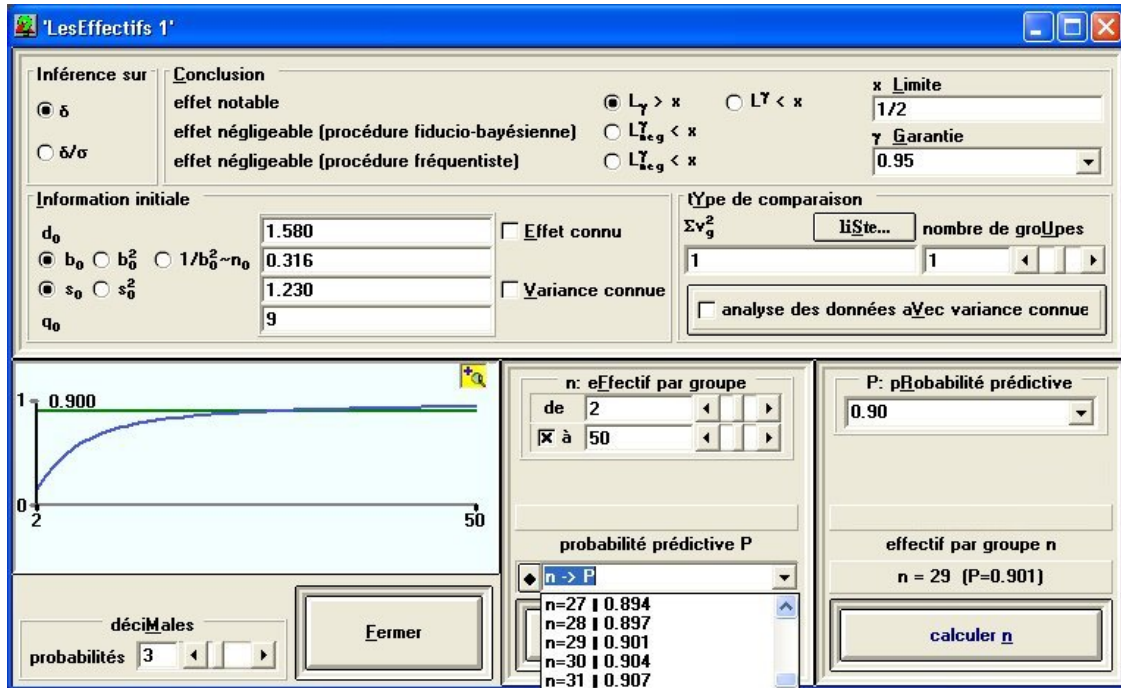
On trouverait de même pour la limite de crédibilité inférieure 0.95 : “une probabilité prédictive 0.893 qu’elle soit positive et une probabilité 0.715 qu’elle dépasse une demi-heure”. Pour obtenir une probabilité prédictive 0.90 que la limite dépasse une demi-heure, il faudrait un effectif $n' = 29$. On notera qu’un “calcul de puissance” traditionnel utilisant pour les paramètres les données de l’échantillon initial, soit $\delta = 1.580$ et $\sigma = 1.230$ qui seraient supposées des valeurs *connues* ne tiendrait pas compte de l’incertitude sur ces valeurs ; on trouverait donc sans surprise que sous ces hypothèses un effectif $n' = 13$ serait suffisant pour obtenir une probabilité prédictive 0.90 que la limite dépasse une demi-heure.

LesEffectifs

Dans la situation de la figure précédente, fermez la “fenêtre des énoncés” pour revenir à la “fenêtre principale”. Cliquez sur le bouton **LesEffectifs** pour afficher la fenêtre de ce programme. L’**information initiale** est automatiquement affichée.

Sélectionnez le bouton d’option **effet notable** $L_\gamma > x$ (c’est l’option par défaut) et entrez la limite x : $1/2$.

Entrez pour l’**effectif** n : **2** ; activez la case à cocher **à** et entrez l’effectif : **50** ; puis cliquez sur le bouton **calculer P** pour obtenir la figure ci-après. Les probabilités pour chaque n sont affichées dans la liste déroulante $n \rightarrow P$. Par exemple, pour un effectif futur de $n = 27$, on a une probabilité 0.894 d’obtenir, avec une garantie 0.95, la conclusion d’un effet (“notable”) supérieur à une demi-heure, soit $L_{0.95} > 1/2$.



9.8 Autres analyses

On peut aussi facilement effectuer des inférences sur l'écart-type σ (ou sur la variance) et sur la différence *calibrée* (ou *standardisée*) $\frac{\delta}{\sigma}$ (voir Lecoutre, 1996a), ou encore une prédiction sur l'analyse finale à une étape intermédiaire (Lecoutre, 2001), etc.

Les mêmes techniques peuvent être utilisées avec une distribution initiale conjuguée, puisque la distribution finale est du même type que la distribution fiducio-bayésienne.

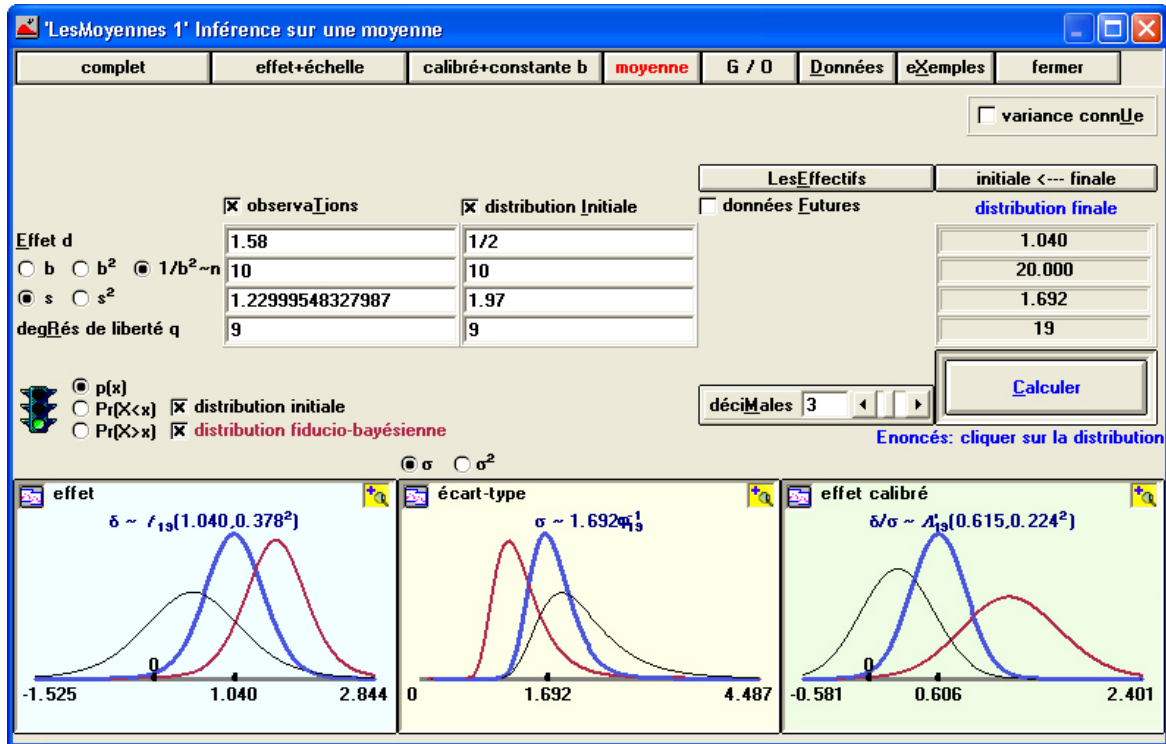
LesMoyennes

Dans la situation de la figure précédente, fermez le programme LesEffectifs pour revenir à la "fenêtre principale" du programme LesMoyennes (si besoin, activez ce programme, cliquez sur le bouton **Exemples** puis sélectionnez **Student**, ce qui affiche les valeurs pertinentes pour cet exemple dans une fenêtre **inférence sur une moyenne**. cliquez ensuite sur le bouton **ok**).

Activez la case à cocher **distribution initiale** (sur la même ligne que **observations** et entrez les valeurs appropriées :

- pour l'effet **d** : **1/2**
- pour **1/b² ~ n** (activez s'il y a lieu cette option) : **1/2**
- pour l's : **1.97**
- pour les **degrés de liberté q** : **9**

Activez les cases à cocher **distribution initiale** (au dessus des courbes) et **distribution fiducio-bayésienne** et cliquez sur le bouton **calculer** pour obtenir la figure ci-après.



10 Inférence sur une combinaison linéaire de moyennes

Cet exemple étend de manière immédiate les solutions précédentes à l'inférence sur une combinaison linéaire de moyennes. On examine ici la compatibilité d'un modèle d'ajustement polynomial avec les données en analyse de variance.

10.1 Le problème : Evaluation du "0.05 cliff effect"

Rosenthal et Gaito ont réalisé une des premières expériences sur l'interprétation du seuil observé d'un test de signification (Rosenthal & Gaito, 1963, 1964). Poitevineau & Lecoutre (2001) ont effectué une réplique de cette expérience auprès de 18 chercheurs en psychologie. On demandait au chercheur de se placer dans la situation suivante : il avait réalisé une expérience pour tester l'efficacité d'un traitement et avait effectué un test "t de Student" pour groupes appariés avec un échantillon de n sujets. On lui demandait, en fonction du seuil observé p , d'indiquer son degré de confiance en l'hypothèse (alternative) selon laquelle le traitement a réellement un effet. Chaque sujet répondait sur une échelle de 0 à 1 pour 24 cas présentés dans un ordre aléatoire et correspondant à douze valeurs de p

0.001 0.01 0.03 0.05 0.07 0.10 0.15 0.20 0.30 0.50 0.70 0.90

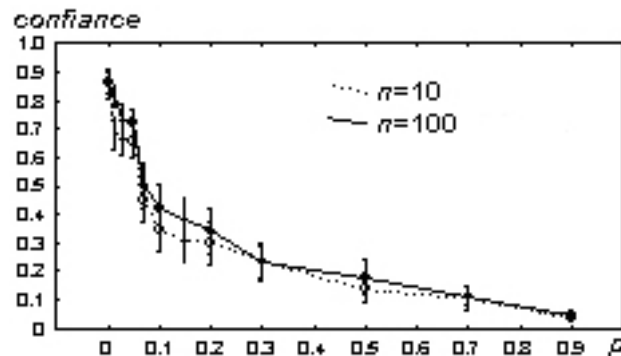
et à deux valeurs de n

10 100

Dans l'expérience initiale, Rosenthal & Gaito rapportaient une "chute de confiance" (*cliff effect*), c'est-à-dire une forte diminution de la confiance dès que p était supérieur au seuil fatidique 0.05.

10.2 Résultats numériques

Les courbes moyennes de la confiance en fonction du seuil p obtenues dans la réplique sont semblables à celles de l'expérience de Rosenthal et Gaito.



Cependant les données individuelles sont hétérogènes. On peut facilement identifier trois catégories distinctes de courbes :

- (1) des courbes exponentielles décroissantes ;
- (2) des courbes linéaires négatives ;
- (3) des courbes en tout-ou-rien ;

et les 18 chercheurs peuvent en conséquence être classés en trois groupes clairement distincts.

Le problème considéré ici est celui de l'évaluation de la "chute de confiance". Nous devons d'abord nous donner une définition explicite de l'absence de chute. Nous adopterons la définition utilisée par Nelson, Rosenthal et Rosnow (1986)

"The patterns of the parent means associated with the four consecutive p-values .03, .05, .07, and .10 and the predicted means based on a second-degree polynomial equation are the same"

Suivant cette définition, on considère le polynôme du second degré qui ajuste le mieux les données pour les quatre seuils considérés (0.03, 0.05, 0.07, et 0.10). La moyenne quadratique des résidus donne un indice d'ajustement ("*goodness of fit*") pour évaluer la chute de confiance.

On obtient par exemple pour le groupe "exponentiel" les résultats suivants.

LesMoyennes

Dans **LePAC** activez le menu **LesBayésiens** et le sous-menu **LesMoyennes**, ce qui affiche la fenêtre pour l'inférence sur des moyennes. Cliquez sur le bouton **Données** pour afficher l'éditeur de données. Sélectionnez le **nombre de groupes** : **1** et le **nombre d'occasions** (nombre de répétitions pour chaque sujet) : **4**. Entrez les données pour le groupe exponentiel que vous trouverez dans le tableau suivant (0.8625 0.7900 0.7200 0.5850, etc.) en les séparant par des espaces ou un passage à la ligne.

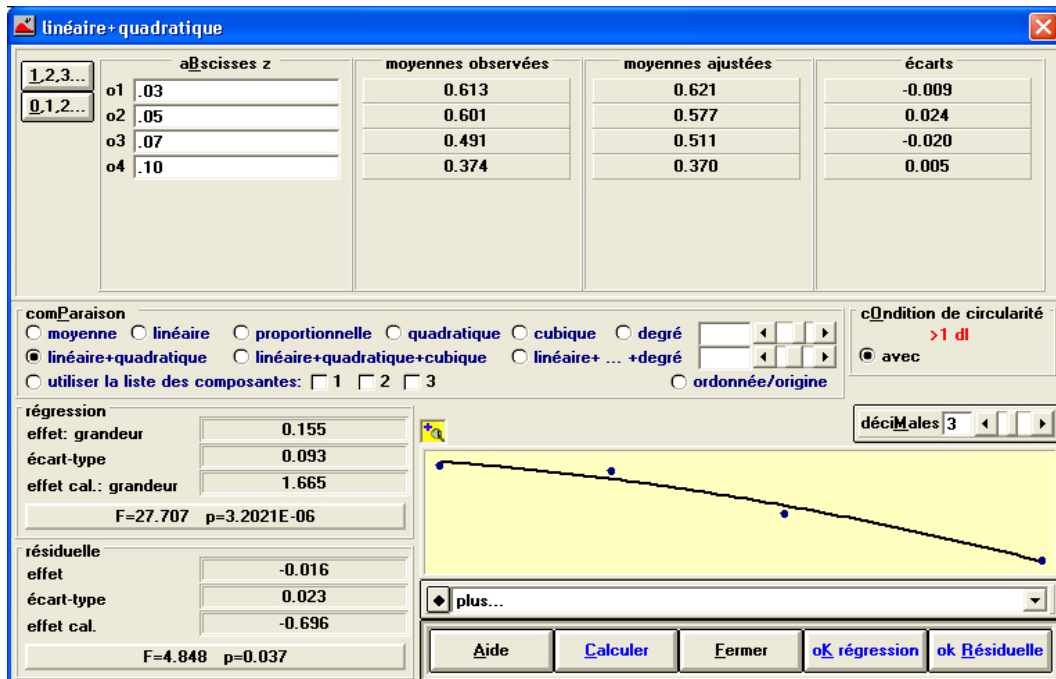
Cliquez ensuite sur le bouton **calculer** (vous pouvez enregistrer les données si vous le souhaitez) pour afficher la fenêtre correspondant à la **structure S×0** avec les statistiques

pertinentes (moyennes, écarts-types, covariances). Cliquez sur le bouton **Régression** où sont affichées les **moyennes observées**.

Entrez les **abscisses z** pour la régression : respectivement **0.03, 0.05, 0.07, 0.10** pour **o1** à **o4**. Sélectionnez le bouton d'option :

- **linéaire+quadratique**

et cliquez sur le bouton **calculer** pour obtenir la figure ci-après.



En cliquant sur le bouton **ok résiduelle** vous obtenez les inférences sur l'effet δ présentées plus loin.

Les tableaux ci-après fournissent les données individuelles (moyennées pour chaque p sur les deux valeurs de n), ainsi que les moyennes observées et ajustées et les résidus pour les trois groupes de sujets.

groupe “exponentiel” (10 sujets)

p	0.03	0.05	0.07	0.10	<i>contraste pertinent</i>
	0.8625	0.7900	0.7200	0.5850	+0.0024
	0.9450	0.9250	0.9325	0.8725	+0.0086
	0.4850	0.4500	0.1775	0.0400	-0.0501
	0.6475	0.6400	0.4925	0.4400	-0.0301 $d = -0.016$
	0.1900	0.1850	0.1400	0.1500	-0.0107 $s = 0.023$
	0.4875	0.5000	0.2775	0.2225	-0.0503 $e = s/\sqrt{10} = 0.0074$
	0.4100	0.4850	0.4900	0.1400	+0.0158 $t = d/e = -2.202$
	0.6500	0.5925	0.4200	0.3300	-0.0269
	0.7300	0.7150	0.6675	0.5300	+0.0037
	0.7200	0.7275	0.5900	0.4325	-0.0224
moyenne observée	0.613	0.601	0.491	0.374	
moyenne ajustée	0.621	0.577	0.511	0.370	
écart (résidu)	-0.009	+0.024	-0.020	+0.005	

groupe “linéaire” (4 sujets)

p	0.03	0.05	0.07	0.10	<i>contraste pertinent</i>
	0.9025	0.8475	0.9125	0.7925	+0.0235 $d = +0.007$
	0.8500	0.7950	0.6250	0.7150	-0.0073 $s = 0.032$
	0.8500	0.7800	0.7800	0.6825	+0.0427 $e = s/\sqrt{4} = 0.016$
	0.9350	0.8725	0.9050	0.8300	+0.0256 $t = d/e = +0.457$
moyenne observée	0.884	0.824	0.806	0.755	
moyenne ajustée	0.881	0.834	0.797	0.757	
écart (résidu)	+0.004	-0.011	+0.009	-0.002	

groupe “tout-ou-rien” (4 sujets)

p	0.03	0.05	0.07	0.10	<i>contraste pertinent</i>
	0.8275	0.8800	0.2475	0.0800	-0.1443 $d = -0.159$
	0.2075	0.4050	0.1800	0.0200	-0.0677 $s = 0.071$
	0.8625	0.8200	0	0	-0.1878 $e = s/\sqrt{4} = 0.036$
	1	1	0	0	-0.2358 $t = d/e = -4.455$
moyenne observée	0.724	0.776	0.107	0.025	
moyenne ajustée	0.808	0.543	0.301	-0.019	
écart (résidu)	-0.083	+0.233	-0.194	+0.044	

L’analyse de la comparaison résiduelle conduit à considérer le contraste *cubique* avec des coefficients appropriés

$$[+0.1310, -0.3668, +0.3057, -0.0699]$$

Nous avons par exemple pour le groupe “tout-ou-rien” l’effet observé associé au contraste précédent

$$d = 0.1310 \times 0.724 - .3668 \times 0.776 + .3057 \times 0.107 - .0699 \times 0.025 = -0.159$$

Par construction 0.159 est la moyenne quadratique des résidus

$$0.159 = \sqrt{\frac{(-0.083)^2 + (+0.233)^2 + (-0.194)^2 + (+0.044)^2}{4}} = |d|$$

On notera que les signes des coefficients ont été arbitrairement choisis de sorte qu'une valeur de d négative représente une chute de confiance, une valeur positive traduisant au contraire une augmentation.

La situation générale traitée ici est l'inférence sur une combinaison linéaire de moyennes δ . Nous avons deux objectifs distincts selon les groupes :

- pour les groupes "exponentiel" et "linéaire", démontrer que l'effet de chute est petit (négligeable) et donc montrer que δ est proche de zéro ($|\delta| < \epsilon$);
- pour le groupe "tout-ou-rien", démontrer que l'effet de chute est grand (notable) et donc montrer que δ est supérieur à une valeur positive jugée suffisante ($\delta > \epsilon'$).

10.3 Solution fiducio-bayésienne

A partir des données individuelles constituées par le contraste *pertinent* (voir tableau 1), nous sommes ramenés à l'inférence sur une moyenne traitée précédemment. Nous en déduisons les distributions fiducio-bayésiennes pour δ

$$\begin{aligned} \text{Groupe exponentiel} & \quad t \sim t_9(-0.016, 0.007^2) \\ \text{Groupe linéaire} & \quad t \sim t_4(+0.007, 0.016^2) \\ \text{Groupe tout-ou-rien} & \quad t \sim t_4(-0.159, 0.036^2) \end{aligned}$$

et les énoncés pertinents sur l'effet de chute de confiance

- pour démontrer que l'effet de chute est négligeable

$$\begin{aligned} \text{Groupe exponentiel} (n = 10) & \quad Pr(|\delta| < 0.030) = 0.95 \\ \text{Groupe linéaire} (n = 4) & \quad Pr(|\delta| < 0.052) = 0.95 \end{aligned}$$

- pour démontrer que l'effet de chute est notable

$$\text{Groupe tout-ou-rien} (n = 4) \quad Pr(\delta < -0.075) = 0.95$$

Il est manifeste que le groupe minoritaire tout-ou-rien est largement responsable de l'effet de chute moyen. La méthodologie bayésienne offre une grande souplesse en permettant de choisir le type d'énoncé le plus approprié au vu des données. Un énoncé particulier sur un paramètre n'est qu'une des façons possibles de résumer la distribution finale. Cela la rend idéalement adaptée pour étendre une analyse de données exploratoire.

Bien entendu, on pourrait calculer une probabilité conjointe relative aux trois groupes. Cela pourrait notamment être approprié si l'on objectait que ces groupes ont été constitués au vu des données et que cela impose donc d'utiliser une méthode de comparaison multiple. Cela ne soulève pas de problème conceptuel, mais nécessite le recours à une méthode d'intégration numérique (facile à mettre en œuvre par simulation).

Ici encore, d'autres analyses sont faciles à envisager.

PARTIE III - LES ASPECTS TECHNIQUES : QUELQUES OUTILS DE BASE

11 Dérivation de la distribution finale

11.1 Echantillonnage binomial

Pour le modèle binomial de paramètre φ , pour un échantillon de taille n avec a succès et $b = n - a$ échecs, la distribution d'échantillonnage est

$$Pr(a | \varphi) = \frac{n!}{a!b!} \varphi^a (1 - \varphi)^b, \quad 0 \leq a \leq n$$

et la vraisemblance est

$$v(\varphi | a) \propto \varphi^a (1 - \varphi)^b$$

11.1.1 Distribution initiale *Bêta*

La distribution initiale a pour densité

$$p(\varphi) = \frac{1}{B(a_0, b_0)} \varphi^{a_0-1} (1 - \varphi)^{b_0-1}$$

où la fonction *Bêta complète* $B(a_0, b_0)$ est définie à partir de la fonction *Gamma*

$$B(a_0, b_0) = \frac{\Gamma(a_0)\Gamma(b_0)}{\Gamma(a_0 + b_0)}$$

De $\int_0^1 p(\varphi) d\varphi = 1$, on déduit l'intégrale *Bêta* qui servira pour dériver la distribution prédictive

$$\int_0^1 \varphi^{u-1} (1 - \varphi)^{v-1} d\varphi = B(u, v)$$

Cette intégrale permet également de dériver les moments de la distribution *Bêta*

$$\text{Moy}(\varphi^r) = \int_0^1 \frac{1}{B(a_0, b_0)} \varphi^{a_0+r-1} (1 - \varphi)^{b_0-1} d\varphi = \frac{B(a_0 + r, b_0)}{B(a_0, b_0)} = \prod_{j=0}^{r-1} \frac{a_0 + j}{a_0 + b_0 + j}$$

cette formule étant valide pour tout r réel tel que $r > -a_0$

La densité de la distribution finale est proportionnelle au produit $v(\varphi | a)p(\varphi)$, soit

$$p(\varphi | a) \propto \varphi^{a_0+a-1} (1 - \varphi)^{b_0+b-1}$$

Par comparaison avec les termes en φ de la densité initiale, on voit qu'il s'agit encore d'une distribution *Bêta* obtenue en remplaçant simplement a_0 par $a_0 + a$ et b_0 par $b_0 + b$ soit

$$\varphi | a \sim \text{Bêta}(a_1, b_1) \text{ avec } a_1 = a_0 + a \text{ et } b_1 = b_0 + b$$

On peut voir que cette dérivation est valide même si la densité initiale est impropre ($a_0 = 0$ ou $b_0 = 0$). La densité finale n'est impropre que si $a_1 = 0$ ou $b_1 = 0$.

11.1.2 Distribution initiale mélange de densités *Bêta*

Soit la distribution initiale $\lambda_1^0 \text{Bêta}(a_1^0, b_1^0) \oplus \dots \oplus \lambda_J^0 \text{Bêta}(a_J^0, b_J^0)$ de densité

$$p(\varphi) = \sum_{j=1}^{j=J} \frac{\lambda_j^0}{B(a_j^0, b_j^0)} \varphi^{a_j^0-1} (1 - \varphi)^{b_j^0-1} \quad \sum_{j=1}^{j=J} \lambda_j^0 = 1$$

La densité de la distribution finale est proportionnelle au produit $v(\varphi | a)p(\varphi)$, soit

$$p(\varphi | a) \propto \sum_{j=1}^{j=J} \frac{\lambda_j^0}{B(a_j^0, b_j^0)} \varphi^{a_j^0+a-1} (1-\varphi)^{b_j^0+b-1}$$

Pour chaque j les termes en φ sont ceux de la distribution $B\hat{e}ta(a_j^0 + a, b_j^0 + b)$, qui a pour densité

$$\frac{1}{B(a_j^0 + a, b_j^0 + b)} \varphi^{a_j^0+a-1} (1-\varphi)^{b_j^0+b-1}$$

On en déduit donc que la distribution finale est le mélange

$$\lambda_1 B\hat{e}ta(a_1^0 + a, b_1^0 + b) \oplus \dots \oplus \lambda_J B\hat{e}ta(a_J^0 + a, b_J^0 + b)$$

avec

$$\lambda_j = \frac{\lambda_j^0 B(a_j^0 + a, b_j^0 + b) / B(a_j^0, b_j^0)}{\sum_{k=1}^{k=J} \lambda_k^0 B(a_k^0 + a, b_k^0 + b) / B(a_k^0, b_k^0)}$$

11.2 Echantillonnage de Pascal

Pour le modèle binomial négatif (échantillonnage de Pascal) dans lequel on arrête l'expérience quand on a observé un nombre fixé à l'avance a de succès, c'est le nombre d'observations n qui est une variable aléatoire. Sa distribution d'échantillonnage s'écrit

$$Pr(n | \varphi) = \frac{(n-1)!}{(a-1)!(n-a)!} \varphi^a (1-\varphi)^{n-a}, \quad n \geq a$$

et la vraisemblance est encore

$$v(\varphi | a) \propto \varphi^a (1-\varphi)^{n-a}$$

d'où la même procédure.

12 Dérivation de la distribution prédictive

12.1 Echantillonnage binomial

Pour la distribution initiale $\varphi \sim B\hat{e}ta(a_0, b_0)$, la distribution prédictive du nombre de succès a pour un échantillon de taille n est

$$Pr(a) = \int_0^1 Pr(a | \varphi) p(\varphi) d\varphi = \frac{(n-1)!}{(a-1)!(n-a)!} \frac{1}{B(a_0, b_0)} \int_0^1 \varphi^{a-1} (1-\varphi)^{b_1} d\varphi$$

d'où, en utilisant l'intégrale $B\hat{e}ta$ définie précédemment, la distribution

$$Pr(a) = \frac{n!}{a!(n-a)!} \frac{B(a_0 + a, b_0 + b)}{B(a_0, b_0)}, \quad 0 \leq a \leq n$$

qui est la distribution $B\hat{e}ta$ -binomiale, ou distribution de *Pólya*.

12.2 Echantillonnage binomial négatif

La distribution prédictive du nombre d'observations n pour un échantillon avec a succès s'obtient de la même manière. On a dans ce cas

$$Pr(n) = \frac{(n-1)!}{(a-1)!(n-a)!} \frac{B(a_0 + a, b_0 + n - a)}{B(a_0, b_0)}, \quad n \geq a$$

qui est la distribution $B\hat{e}ta$ -Pascal.

13 Calcul des moments par mélange – Espérance conditionnelle

Les calculs par *mélange* sont caractéristiques de l'inférence bayésienne et fournissent des principes généraux de démonstration. En particulier, il est souvent possible ainsi d'explicitier les moments des distributions finales, ce qui fournira en particulier des approximations de ces distributions. Quand on doit recourir à des méthodes d'intégration numérique, il sera utile de contrôler les résultats à l'aide de ces approximations.

13.1 Exemple 1 – Moments de la distribution prédictive

Pour le modèle binomial, on a la moyenne de la distribution d'échantillonnage (conditionnelle à φ)

$$\text{Moy}(a \mid \varphi) = n\varphi$$

d'où la moyenne de la distribution prédictive

$$\text{Moy}(a) = n\text{Moy}(\varphi) = \frac{na_0}{a_0 + b_0}$$

On peut calculer de la même manière les autres moments, par exemple

$$\text{Moy}(a^2 \mid \varphi) = n\varphi(n\varphi + 1 - \varphi)$$

d'où

$$\text{Moy}(a^2) = \frac{n(n-1)a_0(a_0+1)}{(a_0+b_0)(a_0+b_0+1)} + \frac{na_0}{a_0+b_0}$$

Etc.

13.2 Exemple 2 – Moments de la distribution du rapport $\tau = \varphi_1/\varphi_2$

Dans le cas de deux échantillons binomiaux indépendants (ou dans la règle "rejouez le gagnant"), on est amené à considérer deux distributions *Bêta* indépendantes $\varphi_1 \sim \text{Bêta}(a_1, b_1)$ et $\varphi_2 \sim \text{Bêta}(a_2, b_2)$. Cherchons par exemple les moments du rapport $\tau = \varphi_1/\varphi_2$.

On a la distribution conditionnelle à φ_2

$$\tau = \frac{\varphi_1}{\varphi_2} \mid \varphi_2 \sim \frac{1}{\varphi_2} \text{Bêta}(a_1, b_1) \quad 0 < \tau < \frac{1}{\varphi_2}$$

d'où les moments conditionnels qui se déduisent des moments de la distribution *Bêta*(a_1, b_1)

$$\text{Moy}(\tau^r \mid \varphi_2) = \frac{\text{B}(a_1 + r, b_1)}{\text{B}(a_1, b_1)} \varphi_2^{-r}$$

et les moments voulus qui se déduisent des moments de la distribution $B\hat{e}ta(a_2, b_2)$

$$\text{Moy}(\tau^r) = \frac{B(a_1 + r, b_1)}{B(a_1, b_1)} \text{Moy}(\varphi_2^{-r}) = \frac{B(a_1 + r, b_1)}{B(a_1, b_1)} \frac{B(a_2 - r, b_2)}{B(a_2, b_2)} \quad a_2 > r$$

Ainsi, dans l'exemple considéré pour la règle d'expérimentation "rejouez le gagnant", on a les deux distributions finales indépendantes $\varphi_1 \sim B\hat{e}ta(74.5, 20.5)$ et $\varphi_2 \sim B\hat{e}ta(35.5, 21.5)$, d'où notamment

$$\text{Moy}(\tau) = \frac{a_1}{a_1 + b_1} \frac{a_2 + b_2 - 1}{a_2 - 1} = 1.2729$$

$$\text{Moy}(\tau^2) = \frac{a_1 + 1}{a_1 + b_1 + 1} \frac{a_2 + b_2 - 2}{a_2 - 2} \text{Moy}(\tau) = 1.6436 \quad \text{d'où } \text{ety}(\tau) = 0.1525$$

14 Calcul des densités et fonctions de répartitions par mélange

Le même principe de *mélange* s'applique aux densités et aux fonctions de répartition.

14.1 Exemple 1 – Dérivation de la distribution du t de Student

Ainsi la distribution du t de Student est "classiquement" définie comme le rapport $t = x/y$ où x et y ont des distributions indépendantes, respectivement $N(0, 1)$ et $(\chi_q^2/q)^{1/2}$. Plutôt que d'expliciter la distribution conjointe du couple (X, Y) et d'effectuer un changement de variable, écrivons

$$t | y \sim N(0, \frac{1}{y})$$

Nous en déduisons directement

$$p(t) = \int_0^{+\infty} p(t | y)p(y)dy$$

On peut ainsi (même si le calcul est formellement équivalent) éviter le calcul du *Jacobien* lié à un changement de variable et obtenir plus facilement des formules appropriées. En outre la définition de la distribution de Student comme mélange est celle qui intervient naturellement dans l'inférence bayésienne.

Bien entendu les moments s'obtiennent encore par mélange, par exemple

$$\text{Moy}(t^2 | y) = y^{-2} \quad \text{d'où } \text{Moy}(t^2) = \text{Moy}(y^{-2}) = \frac{q}{q-2} \quad q > 2$$

On trouvera des démonstrations détaillées et des applications à d'autres distributions dans Lecoutre (1996a, 2000).

14.2 Exemple 2 – Inférence sur le rapport de deux distributions *Bêta* de paramètres entiers

Soit φ_1 et φ_2 indépendamment distribuées *Bêta*($a_1, n_1 - a_1$) et *Beta*($a_2, n_2 - a_2$). On suppose a_i et $n_i > a_i$ entiers positifs. On veut calculer $P = Pr(\frac{\varphi_1}{\varphi_2} > u)$. Pour $0 < u \leq 1$, on peut utiliser le résultat suivant

$$P = \sum_{j=0}^{a_1-1} \frac{\Gamma(a_2 + j)}{\Gamma(a_2)} \frac{\Gamma(n_2)}{\Gamma(n_2 + j)} \binom{n_1-1}{j} u^j {}_2F_1(j - n_1 + 1, j + a_2, j + n_2, u) \quad 0 < u \leq 1$$

où ${}_2F_1(a, b, c, u)$ est la fonction hypergéométrique

Pour $u > 1$, on peut calculer de la même manière $P = 1 - Pr(\frac{\varphi_2}{\varphi_1} > \frac{1}{u})$.

Preuve

Conditionnellement à φ_2

$$Pr\left(\frac{\varphi_1}{\varphi_2} > u \mid \varphi_2\right) = Pr\left(\text{Bêta}(a_1, n_1 - a_1) > u\varphi_2\right) = Pr\left(\text{Bin}(u\varphi_2, n_1 - 1) < a_1\right)$$

où on utilise la relation fondamentale entre la fonction de répartition de la distribution *Bêta* et celle de la distribution *Binomiale* : $Pr(\text{Bêta}(a, b) < \varphi) = Pr(\text{Bin}(\varphi, a + b - 1) > a - 1)$ (Johnson, Kotz & Kemp, 1992). On en déduit

$$Pr\left(\frac{\varphi_1}{\varphi_2} > u \mid \varphi_2\right) = \sum_{j=0}^{a_1-1} Pr\left(\text{Bin}(n_1 - 1, u\varphi_2) = j\right) = \sum_{j=0}^{a_1-1} \binom{n_1-1}{j} u^j \varphi_2^j (1 - u\varphi_2)^{n_1-j-1}$$

La probabilité marginale $P = Pr(\frac{\varphi_1}{\varphi_2} > u)$ s'obtient par le "mélange" par la densité $p(\varphi_2)$ de la distribution *Bêta*($a_2, n_2 - a_2$)

$$P = \sum_{j=0}^{a_1-1} \binom{n_1-1}{j} u^j \int_0^1 \varphi_2^j (1 - u\varphi_2)^{n_1-j-1} p(\varphi_2) d\varphi_2$$

dont on déduit le résultat

$$\begin{aligned} P &= \sum_{j=0}^{a_1-1} \binom{n_1-1}{j} u^j \frac{\Gamma(n_2)}{\Gamma(a_2)\Gamma(n_2 - a_2)} \int_0^1 \varphi_2^{j+a_2-1} (1 - \varphi_2)^{n_2-a_2-1} (1 - u\varphi_2)^{n_1-j-1} d\varphi_2 \\ &= \sum_{j=0}^{a_1-1} \frac{\Gamma(a_2 + j)}{\Gamma(a_2)} \frac{\Gamma(n_2)}{\Gamma(n_2 + j)} \binom{n_1-1}{j} u^j {}_2F_1(j - n_1 + 1, j + a_2, j + n_2, u) \end{aligned}$$

Voir par exemple <http://functions.wolfram.com/HypergeometricFunctions/Hypergeometric2F1/> pour la définition et les propriétés de la fonction hypergéométrique.

15 Méthodes numériques

Si les méthodes bayésiennes reposent sur le principe général selon lequel la densité finale est proportionnelle au produit de la densité initiale et de la vraisemblance

$$p(\theta | \text{données}) = c v(\theta | \text{données}) p(\theta)$$

il n'y a évidemment pas de moyen à la fois simple et universel de “normaliser” ce produit, c'est-à-dire de trouver la constante c telle que $\int p(\theta | \text{données}) d\theta = 1$.

En dehors des cas où l'on peut “identifier” une distribution connue – et conduisant à des calculs “faciles”⁸ – il faut donc recourir à des techniques d'intégration numérique, méthodes *déterministes classiques* ou méthodes de *Monte-Carlo*. Ce sont surtout ces dernières qui ont reçu l'attention des statisticiens bayésiens, en raison de leur possibilité de traiter les problèmes les plus complexes, même (et surtout) quand le problème nécessite de calculer une intégrale de dimension élevée. Il s'agit de *simuler* la distribution de probabilité voulue, donc de programmer un algorithme qui génère (*pseudo*) aléatoirement des valeurs distribuées selon cette distribution. L'appellation de *Monte-Carlo* provient naturellement de la roulette de son casino, considérée comme un mécanisme type pour générer des nombres au hasard. Ces méthodes sont de plus en plus répandues et de nombreuses variantes – plus ou moins sophistiquées selon la nature du problème – ont été mises au point (voir notamment Tierney, 1994; Chib & Greenberg, 1995; Gilks, Richardson & Spiegelhalter, 1996; Gamerman, 1997; Robert & Casella, 2004). Elles peuvent être virtuellement appliquées à n'importe quelle analyse bayésienne. L'intérêt croissant pour ces techniques – et leur rôle de plus en plus important – est par exemple attesté par la place (le chapitre 9) qui leur est maintenant consacrée dans la troisième édition du livre de Lee (2004). Pour le lecteur intéressé, ce chapitre pourra constituer un “tutoriel” de base.

Ici je donnerai simplement des exemples de techniques de simulations élémentaires. Je décrirai aussi une méthode (déterministe) simple et efficace, mais souvent méconnue.

15.1 Des simulations élémentaires

Considérons à titre d'illustration le calcul de $Pr(\frac{\varphi_1}{\varphi_2} > u)$ dans l'exemple de la règle d'expérimentation “rejouez le gagnant”. On a les deux distributions finales indépendantes $\varphi_1 \sim \text{Bêta}(74.5, 20.5)$ et $\varphi_2 \sim \text{Bêta}(35.5, 21.5)$. Supposons que l'on veuille calculer $Pr(\frac{\varphi_1}{\varphi_2} > 1.25)$. Il suffit de simuler un échantillon de taille N du couple (φ_1, φ_2) , donc d'effectuer N tirages selon la distribution conjointe. On en déduira un échantillon de n'importe quel paramètre dérivé auquel on s'intéresse – par exemple la différence, le rapport, la moyenne, le maximum, etc. – permettant de calculer une probabilité relative à ce paramètre (et bien d'autres quantités).

On trouvera ci-après un exemple de programme, écrit dans un *pseudo langage*, effectuant le calcul de $Pr(\frac{\varphi_1}{\varphi_2} > 1.25)$. Il suffit de disposer d'une fonction $r\text{Beta}(A, B)$ qui retourne une valeur aléatoire pour la distribution $\text{Bêta}(A, B)$. P est une approximation de la probabilité cherchée.

⁸Obtenir des formules explicites, comme dans l'exemple précédent du rapport deux distributions *Bêta* ne résoud pas nécessairement tous les problèmes de calcul. Ainsi, le calcul de la fonction hypergéométrique ${}_2F_1$ nécessite lui-même des méthodes numériques... En outre, ces formules ne sont valables que pour des entiers; dans le cas de réels, on peut encore obtenir des formules “explicites” pour les densités, mais le calcul de la fonction de répartition nécessite encore le calcul d'une intégrale.

```

N = 100000
A = 74.5 : B = 20.5 : C = 35.5 : D = 21.5
U = 1.25
I = 0
M1 = 0 : M2 = 0
Faire
  X = rBeta(A,B)
  Y = rBeta(C,D)
  R = X/Y
  Si R > U alors P = P+1
  M1 = M1 + R
  M2 = M2 + R*R
  I = I+1
Jusqu'à ce que I > N
P = P/N
M1= M1/N
M2= M2/N

```

Pour $N=100\,000$ tirages, j'ai trouvé $P \approx 0.523$ (la valeur exacte avec 3 décimales étant 0.521). Pour donner une idée de la variabilité, les valeurs obtenues par tranches successives de 10 000 tirages étaient les suivantes

0.524 0.523 0.518 0.519 0.524 0.529 0.529 0.514 0.521 0.531

Par comparaison, en effectuant 9 fois 100 000 autres tirages, j'ai obtenu successivement

0.520 0.521 0.522 0.519 0.521 0.522 0.520 0.520 0.518

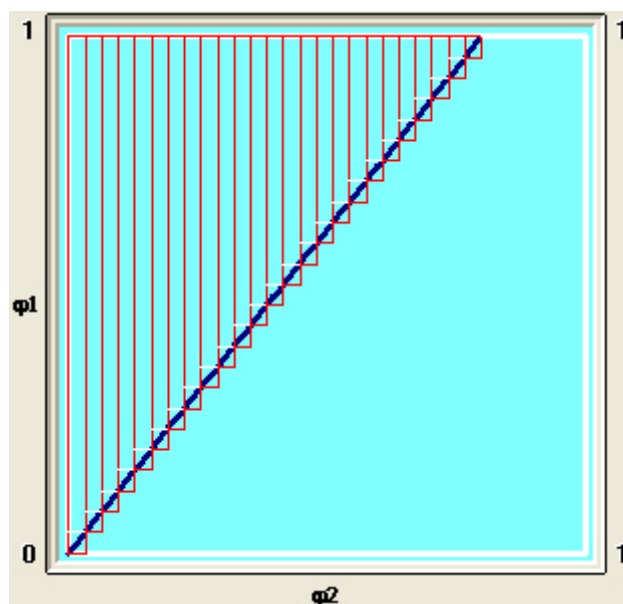
On peut de cette manière obtenir les percentiles de la distribution. Par exemple, pour calculer u tel que $Pr\left(\frac{\chi^2_1}{\chi^2_2} > u\right) = 0.95$, la méthode la plus simple consiste à garder les valeurs R calculées à chaque tirage dans un tableau, puis d'ordonner ce tableau et de prendre comme valeur approchée la $0.05 \times N$ -ième valeur (la 50ème pour $N=10\,000$). Si N est grand, ordonner le tableau peut être très long, mais on peut alors trouver des procédures plus efficaces. On peut évidemment obtenir une figuration approchée de la densité de la distribution, de sa fonction de répartition, etc. Cette figuration sera utile pour apprécier la plausibilité des résultats et détecter une anomalie éventuelle.

Il est également utile, quand cela est possible de calculer des quantités qui peuvent être vérifiées indépendamment. Ainsi, dans le programme ci-dessus, j'ai inclus le calcul approché des moments $M1 \approx \text{Moy}(\tau)$ et $M2 \approx \text{Moy}(\tau^2)$. Pour $N=100\,000$ tirages, les valeurs obtenues, $\text{Moy}(\tau) \approx 1.2726$ et $\text{ety}(\tau) \approx 0.1523$, sont très proches des valeurs exactes avec 4 décimales 1.2729 et 0.1525 calculées précédemment. C'est notamment dans les cas particuliers où la densité des distributions en jeu n'est pas bornée – par exemple une distribution du type *Bêta* $(0.5, b)$ – que la méthode doit être utilisée avec précaution et que de telles vérifications seront utiles.

La généralisation aux situations considérées dans les sections “une généralisation du modèle binomial avec trois proportions” et “un modèle multinomial pour le tableau 2×2 ” est immédiate.

15.2 Une méthode simple pour calculer certaines probabilités

Novick et Jackson (1974, pages 338-342), dans un livre malheureusement maintenant difficile à trouver, décrivent une méthode simple pour calculer la probabilité $Pr(\varphi_1 - \varphi_2 > u)$ (avec les notations utilisées ici). Cette méthode est facilement généralisable ; reprenons ainsi l'exemple précédent du calcul de $Pr(\frac{\varphi_1}{\varphi_2} > 1.25)$. La situation est représentée dans la figure ci-après : il s'agit de calculer la probabilité du *triangle supérieur* délimité par la droite *bleue* d'équation $\varphi_1 = 1.25\varphi_2$.



On voit que cette probabilité peut être encadrée :

- pour une majoration, par la somme des probabilités associées aux *rectangles rouges* ;
- pour une minoration, par la somme des probabilités associées aux rectangles limités à la base par les traits *blancs*.

La probabilité d'un rectangle est simplement le produit des probabilités marginales données par les distributions finales *Bêta* indépendantes. On peut donc théoriquement approximer la probabilité cherchée avec une précision fixée à l'avance.

Comme illustration (grossière), supposons que l'on "découpe" l'ensemble des valeurs possibles de φ_2 (l'axe des abscisses) – soit $[0, 0.80]$ ($0.80 = 1/1.25$) – en 10 intervalles égaux $[x, y]$ (sur la figure il y en a 25). On calcule les probabilités des rectangles correspondants, d'où les résultats suivants.

		majoration			minoration	
x	y	$Pr(\varphi_2 \in [x, y])$	$\times Pr(\varphi_1 > 1.25x)$	=	$\times Pr(\varphi_1 > 1.25y)$	=
0	0.08	0.0000	$\times 1.0000$	= 0.0000	$\times 1.0000$	= 0.0000
0.08	0.16	0.0000	$\times 1.0000$	= 0.0000	$\times 1.0000$	= 0.0000
0.16	0.24	0.0000	$\times 1.0000$	= 0.0000	$\times 1.0000$	= 0.0000
0.24	0.32	0.0000	$\times 1.0000$	= 0.0000	$\times 1.0000$	= 0.0000
0.32	0.40	0.0003	$\times 1.0000$	= 0.0003	$\times 1.0000$	= 0.0003
0.40	0.48	0.0144	$\times 1.0000$	= 0.0144	$\times 0.9999$	= 0.0144
0.48	0.56	0.1489	$\times 0.9999$	= 0.1489	$\times 0.9706$	= 0.1445
0.56	0.64	0.4335	$\times 0.9706$	= 0.4207	$\times 0.3697$	= 0.1603
0.64	0.72	0.3431	$\times 0.3697$	= 0.1269	$\times 0.0006$	= 0.0002
0.72	0.80	0.0586	$\times 0.0006$	= 0.00003	$\times 0.0000$	= 0.0000
		<i>somme</i>		0.7113		0.3197

Compte tenu du très petit nombre d'intervalles, on obtient sans surprise ici pour la probabilité $Pr(\frac{\varphi_1}{\varphi_2} > 1.25)$ un encadrement très large : $[0.3197, 0.7113]$, mais si on retient la moyenne de ces deux probabilités (soit une *interpolation linéaire*), on trouve une estimation de la probabilité $(0.3197+0.7113)/2 = 0.516$, qui est étonnamment proche de la valeur exacte (avec 3 décimales) 0.521.

Cette méthode est donc efficace et reste rapide même si on augmente sensiblement le nombre d'intervalles. Bien entendu, comme toute méthode d'intégration numérique, elle peut être raffinée.

A partir de l'exemple précédent, on voit immédiatement une amélioration très simple à apporter. On tronquera l'ensemble des valeurs possibles de φ_2 de manière à éliminer celles qui ne modifient pas la probabilité conjointe, pour une précision donnée que l'on peut contrôler. Ainsi ici, si on se limite à l'ensemble $[0.32, 0.80]$, on sait que l'erreur sur la probabilité conjointe est inférieure à $Pr(\varphi_2 \in [0, 0.32]) = 0.000001$. En découpant l'ensemble $[0.32, 0.80]$ en 1 000 intervalles, on obtient l'encadrement $[0.5194, 0.5221]$, d'où la moyenne exacte avec 3 décimales 0.521 (si on n'avait pas tronqué l'ensemble, on aurait obtenu pour le même nombre d'intervalles un encadrement moins précis $[0.5185, 0.5229]$).

On trouvera ci-après un exemple de programme effectuant le calcul précédent, écrit dans un *pseudo langage*. Il suffit de disposer d'une fonction $Beta(X, A, B)$ qui retourne la probabilité qu'une variable de distribution $Bêta(A, B)$ soit inférieure à X . L'encadrement cherché est $[P1, P2]$.

```

N = 1000
A = 74.5 : B = 20.5 : C = 35.5 : D = 21.5
U = 1.25
X = 0.32 : Y = 0.80
L = (Y-X) / N
P1 = 0 : P2 = 0
PY = 1 - Beta(U*X,A,B)
Faire
  XX = X + L
  PX = Beta(XX,C,D) - Beta(X,C,D)
  P2 = P2 + PX * PY
  PY = 1 - Beta(U*XX,A,B)
  P1 = P1 + PX * PY
  X = XX
Jusqu'à ce que X > Y

```

Un raffinement utile consiste à diviser l'ensemble des valeurs φ_2 en un certain nombre de zones en fonction de la forme de la densité, et à utiliser des intervalles de largeur variable selon la zone. On peut également envisager un autre type d'interpolation que l'interpolation linéaire. Il peut être également préférable dans certains cas d'inverser le rôle de φ_1 et de φ_2 . Etc. Comme pour les simulations, ce sont notamment les cas particuliers où la densité de la distribution marginale n'est pas bornée qui peuvent nécessiter de tels raffinements.

Cette méthode peut être aisément généralisée. Elle peut encore être utilisée pour une fonction de trois paramètres, comme dans les problèmes "diagnostic médical" et "étude d'un modèle logique" – qui conduit à calculer les probabilités de *parallépipèdes rectangles* – avec des temps de calcul qui restent très raisonnables.

PARTIE IV - RETOUR SUR LES ASPECTS CONCEPTUELS L'INTERFACE DE L'INFÉRENCE FRÉQUENTISTE ET DE L'INFÉRENCE BAYÉSIENNE

16 Propriétés fréquentistes de l'inférence bayésienne sur une proportion : c'est simple !

16.1 Modèle binomial

16.1.1 Un exemple typique

Le tableau ci-après donne un exemple typique, avec $\varphi = 0.50$ et $n = 10$. Pour chaque résultat possible $f = a/n$ ($0 \leq a \leq 10$) le tableau donne :

- la probabilité d'échantillonnage $p(f | \varphi = 0.50)$, donnée par la distribution $Bin(0.50, 10)$;
- la probabilité cumulée correspondante $B(f)$;
- la probabilité bayésienne finale $Pr(\varphi < 0.50 | f)$ pour les trois distributions initiales $Bêta(0, 1)$, $Bêta(1/2, 1/2)$, $Bêta(1, 0)$, d'où les distributions finale respectives

$$Bêta(a, n - a + 1) \quad Bêta(a + 1/2, n - a + 1/2) \quad Bêta(a + 1, n - a)$$

Notons respectivement ces trois probabilités finales

$$Pr(\varphi < 0.50 | f) = \begin{array}{|c|c|c|c|} \hline \text{Initiale} & Bêta(0, 1) & Bêta(1/2, 1/2) & Bêta(1, 0) \\ \hline & Pr^+(\varphi | f) & Pr(\varphi | f) & Pr^-(\varphi | f) \\ \hline \end{array}$$

Bien entendu

$$Pr^+(\varphi | f) > Pr(\varphi | f) > Pr^-(\varphi | f)$$

Remarque : on considère ici que les distribution $Bêta(a, 0)$ et $Bêta(0, a)$ sont les distributions ponctuelles aux points 0 et 1.

$\varphi = 0.50$					
$f = a/n$	$p(f \varphi)$	$B(f)$	$Pr^+(\varphi f)$	$Pr(\varphi f)$	$Pr^-(\varphi f)$
0/10	0.00097656	0.00097656	1.00000000*	0.99983854*	0.99902344*
$F^- = 1/10$	0.00976562	0.01074219	0.99902344*	0.99631011*	0.98925781*
$F^+ = 2/10$	0.04394531	0.05468750	0.98925781*	0.97396339*	0.94531250
3/10	0.11718750	0.17187500	0.94531250	0.89798452	0.82812500
4/10	0.20507812	0.37695312	0.82812500	0.73517267	0.62304688
5/10	0.24609375	0.62304687	0.62304688	0.50000000	0.37695312
6/10	0.20507812	0.82812500	0.37695312	0.26482733	0.17187500
7/10	0.11718750	0.94531250	0.17187500	0.10201548	0.05468750
8/10	0.04394531	0.98925781	0.05468750	0.02603661	0.01074219
9/10	0.00976562	0.99902344	0.01074219	0.00368989	0.00097656
10/10	0.00097656	1.00000000	0.00097656	0.00016146	0
			$Bêta(0, 1)$	$Bêta(1/2, 1/2)$	$Bêta(1, 0)$

On peut observer dans ce tableau les résultats remarquables (et généraux) suivants.

$$p(f | \varphi) = Pr^+(\varphi | f) - Pr^-(\varphi | f) \quad [1]$$

est la conséquence de la relation fondamentale entre les fonctions de répartition des distributions $Bêta$ et $Binomiale$ (voir Section précédente).

$$Pr^+(\varphi | f) = Pr^-(\varphi | f - \frac{1}{n}) \quad [2]$$

résulte du fait qu'il s'agit dans les deux cas de la même distribution finale.

Considérons maintenant le taux de couverture fréquentiste (pour $\varphi = 0.50$) associé à la limite supérieure de l'intervalle unilatéral de crédibilité $1 - \alpha$. Il y a couverture si la limite est supérieure ou égale à 0.50, donc si $Pr(\varphi < 0.50 | f) \leq 1 - \alpha$. Dans le tableau les "erreurs" (non couverture) pour $\alpha = 0.05$ sont indiquées par *.

Notons F^- la plus grande valeur de f telle que $Pr^-(\varphi | f) \geq 1 - \alpha$; cette valeur est donc telle que $Pr^+(\varphi | f + 1) = Pr^-(\varphi | f) \geq 1 - \alpha$. On a ici $F^- = 1/10$. En raison de [2] $F^+ = F^- + \frac{1}{n}$ est la plus grande valeur de f telle que $Pr^+(\varphi | f) \geq 1 - \alpha$. On a ici $F^+ = 2/10$.

En conséquence le taux de couverture associé à la distribution initiale $B\hat{e}ta(1, 0)$ est $1 - B(F^-)$, et en raison de [1] et [2] on a simplement

$$B(F^-) = Pr^+(\varphi | 0) - Pr^-(\varphi | F^-) = 1 - Pr^-(\varphi | F^-)$$

d'où le taux de couverture

$$1 - B(F^-) = Pr^-(\varphi | F^-) \geq 1 - \alpha \quad \text{avec ici } 1 - B(1/10) = 0.98925781$$

De même le taux de couverture associé à la distribution initiale $B\hat{e}ta(0, 1)$ est $1 - B(F^+)$

$$1 - B(F^+) = Pr^-(\varphi | F^+) < 1 - \alpha \quad \text{avec ici } 1 - B(2/10) = 0.94531250$$

En conséquence le taux de couverture associé à la distribution initiale $B\hat{e}ta(1/2, 1/2)$ ne peut être que $B(F^-)$ ou $B(F^+)$. De plus ce résultat est vrai aussi pour toute distribution initiale $Beta(a_0, b_0)$ avec $0 \leq a_0 \leq 1$ et $0 \leq b_0 \leq 1$.

On a bien entendu des résultats analogues pour la limite inférieure, les rôles des distributions $B\hat{e}ta(0, 1)$ et $B\hat{e}ta(1, 0)$ étant inversés.

16.1.2 En conclusion

La famille de distributions initiales $B\hat{e}ta(a_0, b_0)$ avec $0 \leq a_0 \leq 1$ et $0 \leq b_0 \leq 1$ correspond à une "zone d'ignorance" (Bernard, 1996). Il y a deux distributions "extrêmes", $B\hat{e}ta(1, 0)$ et $B\hat{e}ta(0, 1)$, l'une assurant que le taux d'erreur fréquentiste est toujours plus petit que α et l'autre qu'il est toujours plus grand que α (le sens de l'inégalité dépendant du fait que l'on considère la limite inférieure ou supérieure). Le taux de couverture associé aux autres distributions de la famille étant toujours égal à l'un des deux taux associés à ces distributions extrêmes (dépendant de φ et n).

Comparons cette situation à ce qui se passe dans le cas continu, par exemple pour l'inférence sur la moyenne μ d'une distribution *Normale* $N(\mu, 1)$ à partir d'un échantillon de taille n . On a pour la moyenne observée la distribution d'échantillonnage *Normale* $N(\mu, 1/n)$. On remarque que l'on a dans ce cas, pour une distribution initiale uniforme, la distribution finale $N(m, 1/n)$, d'où l'égalité des densités

$$p(\mu | m) = p(m | \mu)$$

En fait, dans le cas de l'inférence sur une moyenne, la densité de la distribution d'échantillonnage de la statistique m est une fonction symétrique de m et du paramètre μ , ce que l'on peut voir intuitivement comme un argument de type *fiduciaire* justifiant le choix de la distribution initiale uniforme. Dans le cas de l'inférence sur une proportion, on ne peut pas avoir directement un tel argument de symétrie puisque la statistique est discrète et le paramètre continu, mais l'égalité précédente est remplacée par

$$Pr^+(\varphi | f) - Pr^-(\varphi | f) = p(f | \varphi)$$

qui en constitue une extension naturelle.

Ceci conduit à considérer ici comme “probabilité fiducio-bayésienne”, non pas une seule valeur, mais un intervalle constitué par les deux probabilités associées aux deux distributions extrêmes. On rejoint ici la notion de *probabilité imprécise* (voir pour les applications en inférence statistique Walley, 1996), comme conséquence du caractère discret de la distribution d'échantillonnage.

Ceci suggère encore que si l'on veut retenir une seule distribution “non informative”, celle-ci doit être choisie comme “intermédiaire” entre $B\hat{e}ta(1, 0)$ et $B\hat{e}ta(0, 1)$. Un choix privilégié est alors $B\hat{e}ta(1/2, 1/2)$, qui est la solution de Jeffreys. C'est effectivement la distribution initiale qui paraît donner sur l'ensemble des valeurs de φ et de n le meilleur taux de couverture fréquentiste, meilleur que celui obtenu pour la plupart des méthodes fréquentistes d'intervalles de confiance (Cai, 2005).

16.2 Mod\ele binomial n\egatif

On considère toujours l'inf\erence sur une proportion, mais \a partir d'un \echantillon dont le nombre de succ\es a est fix\e avant l'exp\erience (on s'arr\ete d\es que ce nombre est atteint). On a dans ce cas pour le nombre total d'observations $n \geq a$ la distribution d'\echantillonnage $BinNeg(\varphi, a)$.

Remarque : la distribution d'\echantillonnage *Binomiale n\egative* est souvent d\efinie comme la distribution de $b = n - a$, le nombre d'\echecs observ\es quand a est atteint.

16.2.1 Un exemple typique

Le tableau ci-apr\es donne un exemple typique, avec $\varphi = 0.20$ et $a = 3$. Pour chaque r\esultat possible $f = a/n$ ($n \geq 3$) le tableau donne :

- la probabilit\e d'\echantillonnage $p(f | \varphi = 0.20)$, donn\ee par la distribution $BinNeg(0.20, 3)$;
- la probabilit\e cumul\ee correspondante $B(f)$;
- la probabilit\e bay\esienne finale $Pr(\varphi > 0.20 | f)$ pour les quatre distributions initiales $B\hat{e}ta(1, 0)$, $B\hat{e}ta(0, 0)$, $B\hat{e}ta(0, 1/2)$, $B\hat{e}ta(0, 1)$, soit les distributions finales respectives

$$B\hat{e}ta(a + 1, n - a) \quad B\hat{e}ta(a, n - a) \quad B\hat{e}ta(a, n - a + 1/2) \quad B\hat{e}ta(a, n - a + 1)$$

Notons maintenant les probabilit\es finales

$$Pr(\varphi < 0.50 | f) = \begin{array}{|c|c|c|} \hline \text{Initiale} & B\hat{e}ta(0, 0) & B\hat{e}ta(0, 1/2) & B\hat{e}ta(0, 1) \\ \hline & Pr^+(\varphi | f) & Pr(\varphi | f) & Pr^-(\varphi | f) \\ \hline \end{array}$$

Bien entendu

$$Pr^+(\varphi | f) > Pr(\varphi | f) > Pr^-(\varphi | f)$$

$$\varphi = 0.20$$

$f = a/n$	$p(f \varphi = 0.20)$	$B(f)$		$Pr^+(\varphi f)$	$Pr(\varphi f)$	$Pr^-(\varphi f)$
3/3	0.00800000	0.00800000	1.00000000*	1.00000000*	0.99728632*	0.99200000*
3/4	0.01920000	0.02720000	0.99840000*	0.99200000*	0.98386991*	0.97280000*
3/5	0.03072000	0.05792000	0.99328000*	0.97280000*	0.95882595*	0.94208000
3/6	0.04096000	0.09888000	0.98304000*	0.94208000	0.92276264	0.90112000
3/7	0.04915200	0.14803200	0.96665600*	0.90112000	0.87742592	0.85196800
3/8	0.05505024	0.20308224	0.94371840	0.85196800	0.82503681	0.79691776
3/9	0.05872026	0.26180250	0.91435827	0.79691776	0.76788506	0.73819750
3/10	0.06039798	0.32220047	0.87912612	0.73819750	0.70809554	0.67779953
3/11	0.06039798	0.38259845	0.83886080	0.67779953	0.64750882	0.61740155
3/12	0.05905580	0.44165425	0.79456895	0.61740155	0.58763489	0.55834575
3/13	0.05669357	0.49834782	0.74732431	0.55834575	0.52965171	0.50165218
3/14	0.05360119	0.55194901	0.69818988	0.50165218	0.47442964	0.44805099
3/15	0.05002778	0.60197679	0.64816210	0.44805099	0.42256891	0.39802321
3/16	0.04617949	0.64815628	0.59813433	0.39802321	0.37444215	0.35184372
3/17	0.04222125	0.69037753	0.54887620	0.35184372	0.33023683	0.30962247
3/18	0.03828060	0.72865812	0.50102546	0.30962247	0.28999475	0.27134188
3/19	0.03445254	0.76311066	0.45508874	0.27134188	0.25364706	0.23688934
3/20	0.03080462	0.79391528	0.41144886	0.23688934	0.22104429	0.20608472
3/21	0.02738189	0.82129717	0.37037603	0.20608472	0.19198125	0.17870283
3/22	0.02421135	0.84550852	0.33204139	0.17870283	0.16621725	0.15449148
etc.						
			$B\hat{e}ta(1, 0)$	$B\hat{e}ta(0, 0)$	$B\hat{e}ta(0, 1/2)$	$B\hat{e}ta(0, 1)$

Considérons maintenant le taux de couverture fréquentiste (pour $\varphi = 0.20$) associé à la limite inférieure de l'intervalle unilatéral de crédibilité $1 - \alpha$. Il y a couverture si la limite est inférieure ou égale à 0.20, donc si $Pr(\varphi > 0.20 | f) \leq 1 - \alpha$. Dans le tableau les "erreurs" (non couverture) pour $\alpha = 0.05$ sont indiquées par *.

16.2.2 En conclusion

On a clairement des résultats analogues à ceux énoncés pour le modèle binomial. La famille de distributions initiales $B\hat{e}ta(a_0, b_0)$ avec $0 \leq a_0 \leq 1$ et $0 \leq b_0 \leq 1$ correspond à une "zone d'ignorance". Les deux distributions "extrêmes", $B\hat{e}ta(1, 0)$ et $B\hat{e}ta(0, 1)$ assurent encore, l'une que le taux d'erreur fréquentiste est toujours plus petit que α et l'autre qu'il est toujours plus grand que α (le sens de l'inégalité dépendant du fait que l'on considère la limite inférieure ou supérieure).

On a encore pour le modèle binomial négatif une égalité du type

$$Pr^+(\varphi | f) - Pr^-(\varphi | f) = p(f | \varphi)$$

mais dans ce cas Pr^+ correspond à la distribution initiale $B\hat{e}ta(0, 0)$ et Pr^- à la distribution initiale $B\hat{e}ta(0, 1)$

Ceci suggère que si l'on veut retenir une seule distribution "non informative", celle-ci doit être choisie comme "intermédiaire" entre $B\hat{e}ta(0, 0)$ et $B\hat{e}ta(0, 1)$. Un choix "naturel" est alors $B\hat{e}ta(0, 1/2)$, qui est la solution de Jeffreys. C'est effectivement la distribution initiale qui paraît donner sur l'ensemble des valeurs de a et de φ le meilleur taux de couverture fréquentiste, meilleur que celui obtenu pour la plupart des méthodes fréquentistes d'intervalles de confiance (Cai, 2005).

Notons encore que le taux de couverture associé à la distribution $B\acute{e}ta(0, 1/2)$ est toujours égal à l'un des deux taux associés aux distributions $B\acute{e}ta(0, 0)$ et $B\acute{e}ta(0, 1)$ (dépendant de φ et a), et non comme pour le modèle binomial à ceux associés aux deux distributions extrêmes.

16.3 Généralisation aux tableaux 2×2 et tests conditionnels “à la Fisher”

L'exemple “un modèle multinomial pour le tableau 2×2 ” a permis d'illustrer la réinterprétation bayésienne du test de permutation de Fisher. Ceci constitue une extension des résultats précédents pour une proportion. On a la même réinterprétation dans le cas de deux groupes indépendants avec chacun une distribution d'échantillonnage *Binomiale* ou chacun une distribution *Binomiale négative*, ou encore pour un *modèle multinomial négatif*. Illustrons la dans les deux premiers cas.

Considérons l'exemple du tableau de données suivant, avec respectivement 3 et 2 observations et un succès observé dans chaque groupe.

	1	0	
g1	$n_{11} = 1$	$n_{10} = 2$	$n_{1.} = 3$
g2	$n_{21} = 1$	$n_{20} = 1$	$n_{2.} = 2$
	$n_{.1} = 2$	$n_{.0} = 3$	$n = 5$

16.4 Le test conditionnel aux marges (test de permutation)

16.4.1 Cas binomial

Dans le cas binomial on retient l'un des effectifs – par exemple n_{11} – comme statistique de test, et on calcule la proportion de tableaux possibles avec les mêmes marges qui sont “plus extrêmes” que le tableau observé. Il y a trois types de tableaux de données possibles

0	3	3		1	2	3		2	1	3
2	0	2		1	1	2		0	2	2
2	3	5		2	3	5		2	3	5
$1 \times 1 = 1$				$3 \times 2 = 6$				$3 \times 1 = 3$		

la dernière ligne indiquant le nombre de tableaux du type. Par exemple, pour le tableau du milieu, avec 3 sujets dans le groupe 1 il y a 3 façons possibles d'avoir $n_{11} = 1$ et $n_{12} = 2$, et avec 2 sujets dans le groupe 2 il y a 2 façons possibles d'avoir $n_{21} = 1$ et $n_{22} = 1$, donc $3 \times 2 = 6$ tableaux de ce type.

Si on veut montrer que $\varphi_1 < \varphi_2$, le seuil observé du test unilatéral (*inchant*) est la proportion des tableaux pour lesquels $n_{11} \leq 1$, soit

$$p_{inc} = (1 + 6)/10 = 0.70$$

Si on exclut la valeur observée (soit $n_{11} < 1$) on obtient le seuil *exchant*

$$p_{exc} = 1/10 = 0.10$$

Formellement, ces seuils sont donnés par la distribution d'échantillonnage de n_{11} sous l'hypothèse nulle $H_0 : \varphi_1 = \varphi_2$, conditionnellement aux marges, qui est une distribution *Hypergéométrique*

$$n_{11} | n_{.1}, n_{1.}, n_{.2}, n_{2.}, [H_0] \sim Hyp(n, n_{.1}, n_{1.}) \quad \text{soit ici } Hyp(5, 2, 3)$$

d'où

$$p_{inc} = Pr(Hyp(n, n_{.1}, n_{1.}) \leq n_{11}) = 0.70 \quad \text{et} \quad p_{exc} = Pr(Hyp(n, n_{.1}, n_{1.}) < n_{11}) = 0.10$$

16.4.2 Cas binomial négatif

Dans le cas binomial négatif on retient par exemple n_1 comme statistique de test et on calcule la proportion de tableaux possibles avec les mêmes valeurs n , n_{11} et n_{21} (et donc $n_{.1}$ et $n_{.0}$) qui sont "plus extrêmes" que le tableau observé. Il y a quatre types de tableaux de données possibles

$\mathbf{1}$	$\mathbf{0}$	1	$\mathbf{1}$	$\mathbf{1}$	2	$\mathbf{1}$	$\mathbf{2}$	3	$\mathbf{1}$	$\mathbf{3}$	4
$\mathbf{1}$	$\mathbf{3}$	4	$\mathbf{1}$	$\mathbf{2}$	3	$\mathbf{1}$	$\mathbf{1}$	2	$\mathbf{1}$	$\mathbf{0}$	1
$\mathbf{2}$	$\mathbf{3}$	$\mathbf{5}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{5}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{5}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{5}$
$1 \times 4 = 4$			$2 \times 3 = 6$			$3 \times 2 = 6$			$4 \times 1 = 4$		

la dernière ligne indiquant le nombre de tableaux du type. Dans ce cas, si on veut montrer que $\varphi_1 < \varphi_2$, le seuil observé du test unilatéral (*inchant*) est la proportion des tableaux pour lesquels $n_{1.} \geq 3$, soit

$$p_{inc} = (6 + 4)/20 = 0.50$$

Si on exclut la valeur observée (soit $n_{1.} > 3$) on obtient le seuil *excluant*

$$p_{exc} = 4/20 = 0.25$$

Formellement, ces seuils sont donnés par la distribution d'échantillonnage de n_1 sous l'hypothèse nulle $H_0 : \varphi_1 = \varphi_2$, conditionnellement à n , n_{11} et n_{21} , qui est une distribution *Hypergéométrique négative*

$$n_{1.} | n, n_{11}, n_{21}, [H_0] \sim HypNEG(n, n_{11}, n_{21}) \quad \text{soit ici } HypNEG(5, 1, 1)$$

d'où

$$p_{inc} = Pr(HypNEG(n, n_{11}, n_{21}) \geq n_{1.}) = 0.50 \quad \text{et} \quad p_{exc} = Pr(HypNEG(n, n_{11}, n_{21}) > n_{1.}) = 0.25$$

16.5 Réinterprétation bayésienne

16.5.1 Cas binomial

Soit la distribution initiale $\varphi_1 \sim Bêta(\nu_{11}, \nu_{10})$ et $\varphi_2 \sim Bêta(\nu_{21}, \nu_{20})$ avec φ_1 et φ_2 indépendantes. On suppose que les (ν_{ij}) sont des entiers positifs ou nuls. On a le résultat suivant

$$\begin{aligned} & Pr(\varphi_1 > \varphi_2 | (n_{ij})) \\ &= Pr(Hyp(n + \nu_{11} + \nu_{10} + \nu_{21} + \nu_{20} - 2, n_{.1} + \nu_{11} + \nu_{21} - 1, n_{.1} + \nu_{11} + \nu_{10} - 1) < n_{11} + \nu_{11}) \end{aligned}$$

soit, si $\nu_{11} + \nu_{10} = 1$, $\nu_{11} + \nu_{21} = 1$ et $\nu_{21} + \nu_{20} = 1$

$$Pr(\varphi_1 > \varphi_2 | (n_{ij})) = Pr(Hyp(n, n_{.1}, n_{1.}) < n_{11} + \nu_{11})$$

Par conséquent, pour le cas binomial on a la réinterprétation du seuil

$$Pr(\varphi_1 > \varphi_2 | (n_{ij})) = p_{inc} = 0.70 \quad \text{si } \nu_{11} = \nu_{20} = 1 \text{ et } \nu_{10} = \nu_{21} = 0$$

$$Pr(\varphi_1 > \varphi_2 | (n_{ij})) = p_{exc} = 0.10 \quad \text{si } \nu_{11} = \nu_{20} = 0 \text{ et } \nu_{10} = \nu_{21} = 1$$

et, si l'on veut retenir une seule distribution "non informative", ceci suggère comme choix privilégié la solution intermédiaire $\nu_{11} = \nu_{10} = \nu_{21} = \nu_{20} = 1/2$, qui est la solution de Jeffreys.

Le lien entre le seuil du test conditionnel et la probabilité bayésienne finale $Pr(\varphi_1 > \varphi_2 | (n_{ij}))$ résulte encore de la relation fondamentale entre les fonctions de répartition des distributions *Bêta* et *Binomiale*. Il se démontre "par mélange". Conditionnellement à φ_2 , on a

$$\begin{aligned} Pr(\varphi_1 > \varphi_2 | \varphi_2, (n_{ij})) &= Pr(\text{Bêta}(n_{11} + \nu_{11}, n_{10} + \nu_{10}) > \varphi_2) \\ &= Pr(\text{Bin}(\varphi_2, n_{11} + \nu_{11} + n_{10} + \nu_{10} - 1) < n_{11} + \nu_{11}) = \\ &Pr(\text{Bin}(\varphi_2, n_{1.} + \nu_{11} + \nu_{10} - 1) < n_{11} + \nu_{11}) \end{aligned}$$

La probabilité marginale – mélange de la probabilité binomiale par la distribution *Bêta* (ν_{21}, ν_{20}) – est une distribution *Bêta-Binomiale* (voir la section "dérivation de la distribution prédictive")

$$Pr(\varphi_1 > \varphi_2 | (n_{ij})) = Pr(\text{Bêta-Bin}(n_{21} + \nu_{21}, n_{20} + \nu_{20}; n_{1.} + \nu_{11} + \nu_{10} - 1) < n_{11} + \nu_{11})$$

et en utilisant une autre relation remarquable, entre la distribution *Bêta-Binomiale* et la distribution *hypergéométrique* (Johnson, Kotz & Kemp, 1992).

$$Pr(\text{Bêta-Bin}(a, b; n) < u) = Pr(\text{Hyp}(a + b + n - 1, a + u - 1, n) < u)$$

on déduit le résultat énoncé.

16.5.2 Cas binomial négatif

Avec la même distribution initiale que dans le cas précédent, en utilisant une autre relation, entre la distribution *Bêta-Binomiale* et la distribution *hypergéométrique négative* (Johnson, Kotz & Kemp, 1992).

$$Pr(\text{Bêta-Bin}(a, b; n) < u) = Pr(\text{HypNEG}(a + b + n, u, a) > n)$$

on obtient

$$\begin{aligned} Pr(\varphi_1 > \varphi_2 | (n_{ij})) \\ = Pr(\text{HypNEG}(n + \nu_{11} + \nu_{10} + \nu_{21} + \nu_{20} - 1, n_{11} + \nu_{11}, n_{21} + \nu_{21}) > n_{1.} + \nu_{11} + \nu_{10} - 1) \end{aligned}$$

soit, si $\nu_{11} = \nu_{21} = 0$ et $\nu_{10} + \nu_{20} = 1$,

$$Pr(\varphi_1 > \varphi_2 | (n_{ij})) = Pr(\text{HypNEG}(n, n_{11}, n_{21}) > n_{1.} + \nu_{10} - 1)$$

Par conséquent, pour le cas binomial négatif on a la réinterprétation du seuil

$$Pr(\varphi_1 > \varphi_2 | (n_{ij})) = p_{inc} = 0.50 \quad \text{si } \nu_{11} = \nu_{21} = \nu_{10} = 0 \text{ et } \nu_{20} = 1$$

$$Pr(\varphi_1 > \varphi_2 | (n_{ij})) = p_{exc} = 0.25 \quad \text{si } \nu_{11} = \nu_{21} = \nu_{20} = 0 \text{ et } \nu_{10} = 1$$

et, si l'on veut retenir une seule distribution "non informative", ceci suggère comme choix privilégié la solution intermédiaire $\nu_{11} = \nu_{21} = 0$ et $\nu_{10} = \nu_{20} = 1/2$, qui est la solution de Jeffreys.

17 Règle de Jeffreys

17.1 Information de Fisher et distribution non informative de Jeffreys

Les considérations précédentes fournissent une justification intuitive de la distribution *non informative* de Jeffreys. La justification formelle est que celle-ci est basée sur l'information de Fisher (1922, 1925). Dans le cas d'un seul paramètre – typiquement la proportion φ , l'information de Fisher $I(\varphi)$ apportée par un échantillon \mathbf{y} est définie, à partir de la vraisemblance $v(\varphi | \mathbf{y})$ comme

$$I(\varphi | \mathbf{y}) = E_{\varphi} \left[\left(\frac{\partial}{\partial \varphi} \log v(\varphi | \mathbf{y}) \right)^2 \right] = -E_{\varphi} \left[\frac{\partial^2}{\partial^2 \varphi} \log v(\varphi | \mathbf{y}) \right]$$

et la densité de la distribution de Jeffreys est proportionnelle à sa racine carrée

$$p_J(\varphi) \propto \sqrt{I(\varphi)}$$

17.2 Deux distributions initiales pour les échantillonnages binomial et de Pascal

17.2.1 Echantillonnage binomial

$$v(\varphi | \mathbf{y}) = v(\varphi | a) = \frac{n!}{a!(n-a)!} \varphi^a (1-\varphi)^{n-a}$$

et

$$E(a | \varphi) = n \varphi$$

On en déduit

$$\log v(\varphi | a) = a \log \varphi + (n-a) \log(1-\varphi) + k$$

où k est une constante, et

$$I(\varphi | a) = -E \left(-\frac{a}{\varphi^2} - \frac{n-a}{1-\varphi^2} \right) = \frac{n}{\varphi} + \frac{n}{1-\varphi} = n\varphi^{-1}(1-\varphi)^{-1}$$

donc

$$p_J(\varphi) \propto \varphi^{-1/2}(1-\varphi)^{-1/2} \quad \text{soit } \varphi \sim \text{Bêta}(1/2, 1/2)$$

17.2.2 Echantillonnage de Pascal (distribution Binomiale négative)

$$v(\varphi | \mathbf{y}) = v(\varphi | n) = \frac{(n-1)!}{(a-1)!(n-a)!} \varphi^a (1-\varphi)^{n-a}, \quad n \geq a$$

et

$$E(n | \varphi) = a \frac{1-\varphi}{\varphi}$$

On déduit en procédant de la même manière

$$I(\varphi | n) = n\varphi^{-2}(1-\varphi)^{-1}$$

donc

$$p_J(\varphi) \propto \varphi^{-1}(1-\varphi)^{-1/2} \quad \text{soit } \varphi \sim \text{Bêta}(0, 1/2)$$

18 Une nécessité pour les analyses intermédiaires : concilier l'approche bayésienne avec les desiderata fréquentistes

La plupart des expérimentateurs considèrent que la possibilité dans un plan d'expérience de s'arrêter avant le terme prévu, en effectuant une analyse *intermédiaire*, ne doit pas être ignorée. La raison est que l'arrêt peut induire un biais sur l'inférence qu'ils souhaitent explicitement corriger. Les procédures fréquentistes introduisent précisément une telle correction. C'est pourquoi de nombreux expérimentateurs sont désappointés par le fait que les méthodes bayésiennes, contrairement à la pratique fréquentiste, ignorent généralement ce desiderata, au nom du *principe de vraisemblance*. En conséquence, ceci constitue un frein sérieux à l'utilisation des méthodes bayésiennes dans l'analyse des données expérimentales.

18.1 Les principes de vraisemblance et des règles d'arrêt

"If the experiment is changed, then the expression of relative ignorance can be expected to change correspondingly." (Box & Tiao, 1992, page 46)

Box et Tiao (1992) ont discuté le fait – mis en évidence dans la section précédente – que la règle de Jeffreys aboutit à deux distributions initiales différentes pour les échantillonnages binomial et de Pascal, bien que ces deux types d'expériences conduisent à la même vraisemblance. Ils ont mis en avant l'idée que "l'ignorance" avant l'expérience est relative au plan d'expérience et qu'une distribution non informative appropriée doit donc dépendre du modèle d'échantillonnage. Mais cela est contraire au "principe de vraisemblance" (considéré comme une conséquence du théorème de Bayes), auquel la plupart des bayésiens vouent une stricte obéissance.

Principe de vraisemblance - "Si x_1 et x_2 sont tels qu'il existe une constante c telle que, pour tout φ , $v(\varphi|x_1) = cv(\varphi|x_2)$, ils apportent la même information sur φ et doivent conduire à la même inférence." (adapté de Robert, 1992, page 23)

Une conséquence du principe de vraisemblance est le principe des règles d'arrêt.

Principe des règles d'arrêt - “*once the data have been obtained, the reasons for stopping experimentation should have no bearing on the evidence reported about unknown model parameters.*” (Bayarri & Berger, 2004, page 77)

Ce principe affirme que l'on devrait faire la même inférence si l'on s'arrête en ayant observé a succès après n (fixé) observations (échantillonnage binomial) ou si l'on s'arrête après avoir effectué n observations pour obtenir a (fixé) succès (échantillonnage de Pascal). Ce sont là deux exemples simples de règles d'arrêt ; mais des règles plus complexes – qui ont une grande importance en pratique – sont fournies par les analyses intermédiaires (voir l'exemple “Inférence sur une proportion”). Au nom de ces principes, les tentatives précédentes pour incorporer ces règles dans une distribution initiale objective (Ye, 1993) ont été plus ou moins abandonnées, ignorant les arguments avancés par Box & Tiao.

18.2 Une nouvelle approche prometteuse

La plupart des expérimentateurs ne sont pas convaincus par ces principes. La situation ne devrait cependant pas rester bloquée, puisque de Cristofaro (1996, 2004, 2006) a ouvert la possibilité de concilier l'approche bayésienne avec les desiderata fréquentistes. Il argumente que le plan expérimental (incluant la règle d'arrêt) est antérieur à l'information de l'échantillon et que l'information sur le plan constitue une partie de l'évidence. Il en déduit que la formule de Bayes doit intégrer cette information, et que par suite les principes de vraisemblance et des règles d'arrêt n'en sont plus une conséquence automatique. Une idée de base est de considérer que l'information du plan ignorée dans la vraisemblance peut être recouverte dans l'information de Fisher, et par suite dans la distribution initiale de Jeffreys ; celle-ci trouve ainsi une légitimité nouvelle et apporte une réponse au problème de l'évidence implicitement contenue dans le plan d'expérience. Dans ce cadre, nous pouvons obtenir une réponse bayésienne cohérente et pleinement justifiée au problème des analyses intermédiaires.

Cette approche nouvelle a été développée par Bunouf (2006) dans le cas de l'inférence sur une proportion avec analyses intermédiaires. Elle apparaît prometteuse et offre des avantages décisifs, à la fois d'un point de vue bayésien et d'un point de vue fréquentiste (Bunouf & Lecoutre, 2006).

REMARQUES FINALES ET QUELQUES THÈMES POUR ALLER PLUS LOIN

Le moment semble venu de parvenir à des procédures d'analyse des données expérimentales qui permettent de remédier de manière constructive aux mauvais usages des tests de signification usuels. Il y a indéniablement une reconnaissance grandissante que l'inférence bayésienne peut être idéalement adaptée à cet objectif. Elle satisfait les demandes des scientifiques : procédures objectives (incluant les seuils p traditionnels) ; procédures sur les grandeurs des effets (allant au delà des seuils p) ; procédures pour planifier et conduire les expériences.

Alors, pourquoi les scientifiques, et en particulier les expérimentateurs, alors qu'ils apparaissent réellement souhaiter un autre type d'inférence statistique, semblent souvent réticents à mettre en pratique les procédures bayésiennes ? Dans un papier particulièrement lucide Winkler écrivait il y a plus de 30 ans :

“this state of affairs appears to be due to a combination of factors including philosophical conviction, tradition, statistical training, lack of ‘availability’, computational difficulties, reporting difficulties, and perceived resistance by journal editors.” (Winkler, 1974, page 129)

ajoutant que si on laissait de côté les aspects philosophiques, aucun de ces arguments n'était réellement déterminant. Cette analyse n'en reste pas moins d'actualité ; on peut simplement ajouter qu'un nouvel obstacle important en pratique est que les logiciels statistiques standard qui sont de nos jours largement utilisés continuent à ignorer les méthodes bayésiennes.

“We [statisticians] will all be Bayesians in 2020, and then we can be a united profession.” (Lindley, in Smith, 1995, page 317)

En fait nous sommes vraisemblablement dans une période cruciale, car la masse des travaux théoriques sur l'inférence bayésiennes qui ont été effectués donne lieu à un nombre sans cesse croissant d'applications. L'un des facteurs décisifs pourrait être le récent *“draft guidance document”* de l'*US Food and Drug Administration* (FDA, 2006). Ce document, qui discute *“the least burdensome way of addressing the relevant issues related to the use of Bayesian statistics in medical device clinical trials”*, ouvre la possibilité pour les expérimentateurs (au moins dans le domaine des essais cliniques) d'être réellement bayésiens en pratique.

18.3 Quelques avantages de l'inférence bayésienne

18.3.1 Une meilleure compréhension des procédures fréquentistes

“Students [exposed to a Bayesian approach] come to understand the frequentist concepts of confidence intervals and P values better than do students exposed only to a frequentist approach.” (Berry, 1997)

Les méthodes bayésiennes objectives permettent aux utilisateurs de surmonter des difficultés usuelles rencontrées avec approche fréquentiste. En particulier, il est très naturel pour eux d'utiliser les interprétations bayésiennes des tests de signification et des intervalles de confiance dans le langage des probabilités sur les paramètres inconnus. En retour les mauvais usages et les abus des tests de signification de l'hypothèse nulle sont beaucoup mieux compris. En particulier les utilisateurs des méthodes bayésiennes deviennent très vite avertis du fait que les résultats non significatifs ne peuvent pas être interprétés comme “preuve d'absence d'effet.”

18.3.2 Combiner différentes sources d'information

Une analyse de données expérimentales devraient toujours inclure une analyse bayésienne objective pour exprimer l'apport propre des données – ce que les données ont à dire – indépendamment de toute information extérieure. Cependant, des *“a priori informatifs”* ont aussi un rôle important à jouer. Ils peuvent aider à raffiner l'inférence et à apprécier la sensibilité des conclusions vis-à-vis d'informations supplémentaires.

“an objective scientific report is a report of the whole prior-to-posterior mapping of a relevant range of prior probability distributions, keyed to meaningful uncertainty interpretations.” (Dickey, page 135)

En regard du besoin d’objectivité des scientifiques, on pourrait même argumenter sur la nécessité d’explorer l’impact de différentes distributions initiales pertinentes.

Les techniques bayésiennes informatives sont idéalement adaptées pour *combiner les informations* des données d’une expérience avec celles d’autres études, et par suite pour planifier une série d’expériences. Des utilisations plus ou moins convaincantes ont été proposées. Par exemple, Irony et Pennello (2001) discutent la manière d’introduire ces techniques dans les essais cliniques en médecine.

De façon idéale, quand une “bonne information initiale” est disponible, elle peut (doit ?) être utilisée pour obtenir la même conclusion qu’une “analyse bayésienne objective”, mais avec un plus petit échantillon de données. Naturellement, cela suppose une connaissance réelle basée sur des données plutôt que des opinions d’experts, qui sont généralement sujettes à controverses. Cependant mon opinion est que l’utilisation de ces techniques doit être plus systématiquement explorée avant de pouvoir apprécier ce que devrait être leur contribution précise à l’analyse des données expérimentales.

18.3.3 Les probabilités prédictives : Un outil très séduisant

“An essential aspect of the process of evaluating design strategies is the ability to calculate predictive probabilities of potential results.” (Berry, 1991, page 81)

Un apport majeur du paradigme bayésienne est la facilité de faire des prédictions sur des observations futures. L’idée prédictive est centrale dans les études expérimentales, car

“the essence of science is replication : a scientist should always be concerned about what would happen if he or another scientist were to repeat his experiment.” (Guttman, 1983)

Les procédures bayésiennes prédictives permettent aux utilisateurs de répondre à des questions essentielles telles que : “quel doit être l’effectif de l’expérience pour avoir des chances raisonnables de démontrer la conclusion recherchée ?” ; “à partir des données actuellement disponibles, quelles sont les chances que le résultat final permette de conclure, ou au contraire ne le permette pas ?” Ces questions sont non conditionnelles en ce qu’elles nécessitent de considérer toutes les valeurs possibles des paramètres. Tandis que la pratique fréquentiste traditionnelles ne traite pas ces questions, les probabilités prédictives leur donnent des réponses directes et naturelles.

En particulier, à partir d’une étude pilote, les probabilités prédictives sur les limites de crédibilité sont un résumé particulièrement utiles pour aider au choix de la taille de l’échantillon d’une expérience (pour un parallèle entre les méthodes bayésiennes et fréquentistes, voir Inoue, Berry & Parmigiani, 2005).

L’approche prédictive est une méthode très séduisante (Baum, Houghton & Abrams, 1989) pour aider à la décision d’interrompre une expérience à une étape intermédiaire. D’une part, une probabilité prédictive faible que l’expérience soit une réussite peut être utilisée comme une règle pour abandonner l’expérience (“pour futilité”). D’autre part, une probabilité prédictive suffisamment élevée suggère d’interrompre l’expérience avant son terme (“pour succès”).

Les probabilités prédictives sont aussi un outil important pour traiter les données manquantes. Les analyses intermédiaires reviennent d'ailleurs à évaluer des données manquantes. Le cas de données de survie censurées est particulièrement illustratif. Au moment d'une analyse intermédiaire les données disponibles sont divisées en trois catégories : (1) les patients inclus pour lesquels l'événement auquel on s'intéresse a déjà été observé ; (2) les patients inclus définitivement censurés ; (3) les patients inclus pour lesquels la période d'observation maximale n'est pas encore atteinte. En conséquence les données manquantes à prédire concernent respectivement les patients de cette dernière catégorie pour lesquels nous avons une information *partielle* et les futurs patients dont l'inclusion est planifiée pour lesquels nous n'avons pas d'information directe. L'approche bayésienne traite cette situation de manière directe et efficace (Lecoutre, Mabika & Derzko, 2002).

On peut encore souligner le fait que les distributions prédictives sont également un outil utile pour construire une distribution initiale subjective, du fait qu'il est généralement plus facile d'exprimer une opinion relative à des données.

18.4 Calculs bayésiens et logiciels statistiques

Il y a actuellement de plus en plus d'applications de l'inférence bayésienne à généraux l'analyse des données expérimentales. Mais un obstacle à une utilisation routinière des méthodes bayésiennes objectives est l'absence de logiciels généraux faciles d'utilisation et conviviaux qui seraient une contrepartie des logiciels fréquentistes standard. On peut espérer que cet obstacle sera levé dans le futur.

Quelques programmes ont été développés dans le but d'enseigner l'inférence bayésienne élémentaire : voir notamment *First Bayes* (O'Hagan, 1996) et un ensemble de *macros* pour Minitab (Albert, 1996). Les logiciels que nous avons développés – et illustrés en partie ici – ont une perspective plus ambitieuse. En particulier PAC est une programme général d'analyse de variance qui inclut à la fois les pratiques fréquentistes traditionnelles (tests de signification, intervalles de confiance) et des procédures bayésiennes de routine (initiales non informatives et conjuguées). Ces procédures sont applicables à des plans d'expérience généraux (en particulier, plans à mesures répétées), équilibrés ou déséquilibrés, avec des données univariées ou multivariées, et des covariables.

A un niveau plus avancé et plus général, WinBUGS (une partie du "BUGS project") est un logiciel général souple et efficace. Son objectif déclaré est de rendre la pratique des méthodes MCMC (*Monte Carlo par Chaînes de Markov*) disponibles aux statisticiens. Il contribue indéniablement à l'utilisation croissante de l'inférence bayésienne pour des applications réelles. Il peut être gratuitement téléchargé à l'adresse Internet : <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>. Cependant, il peut être difficilement recommandé à des débutants s'ils ne sont pas fortement motivés.

18.5 Quelques thèmes pour aller plus loin

Il n'est pas dans mon intention de donner ici une liste de thèmes exhaustive, mais seulement de présenter quelques domaines de recherche qui me semblent particulièrement importants pour le développement méthodologique de l'analyse bayésienne objective pour l'analyse des données expérimentales.

18.5.1 L'interface des inférences fréquentiste et bayésienne

Même si ce tutoriel défend l'idée que l'approche bayésienne doit être privilégiée, cela devrait nous inviter à ne pas radicaliser l'opposition entre les inférences bayésienne et fréquentiste, mais plutôt à considérer leur *interface*. Bayarri et Berger (2004) font une présentation particulièrement intéressante de cette interface. Ils mettent en avant le fait que l'argument fréquentiste traditionnel, qui met en jeu "des répétitions du même problème avec différentes données" ne correspond pas à ce qui est fait en pratique. En conséquence c'est "un principe fréquentiste-bayésien conjoint" qui est pertinent : une procédure donnée (par exemple un intervalle de confiance 95% pour une moyenne sous le modèle Normal) est en pratique utilisé sur "une série de différents problèmes mettant en jeu une série de différentes moyennes de distributions Normales avec une série correspondante de données" (page 60).

Plus généralement, ils passent en revue les questions actuelles pour une synthèse bayésienne-fréquentiste dans une perspective méthodologique. Leur conclusion, qui semble raisonnable, est d'espérer une unification méthodologique, mais pas une unification philosophique.

"Philosophical unification of the Bayesian and frequentist positions is not likely, nor desirable, since each illuminates a different aspect of statistical inference. We can hope, however, that we will eventually have a general methodological unification, with both Bayesians and frequentists agreeing on a body of standard statistical procedures for general use". (Bayarri and Berger, 2004, page 78)

Dans cette perspective, un domaine de recherche actif a pour objectif de trouver des "*probability matching priors*" pour lesquels les probabilités finales de certains ensembles spécifiés sont égales (au moins approximativement) à leur probabilités de couverture : voir Fraser, et al., 2003 ; Sweeting, 2005.

On notera encore, qu'en relevant le défi d'une unification méthodologique, Berger (2003) discute "*the conditional frequentist approach to testing*", dont il argue qu'elle fournit précisément (du moins en ce qui concerne les tests d'hypothèses) la base d'une unification méthodologique des approches de Fisher, Jeffreys et Neyman.

18.5.2 Echangeabilité et modèles hiérarchiques

Sans entrer dans le détail, des événements aléatoires sont *échangeables* "si nous attribuons la même probabilité à une assertion sur n'importe quel nombre donné d'entre eux" (de Finetti, 1972, page 213). C'est une notion clé de l'inférence statistique. Par exemple des sujets futurs doivent être supposés échangeables avec les sujets qui ont déjà été observés pour rendre les probabilités prédictives raisonnables. De même, des expériences semblables doivent être supposés échangeables pour une intégration cohérente des informations.

La notion d'échangeabilité est très importante et utile dans le cadre bayésien. En utilisant la spécification de distribution initiales à différents niveaux, elle permet une modélisation souple de dispositifs expérimentaux liés au moyen de *modèles hiérarchiques* (Bernardo, 1996).

"If a sequence of observations is judged to be exchangeable, then any subset of them must be regarded as a random sample from some model, and there

exist a prior distribution on the parameter of such model, hence requiring a Bayesian approach.” (Bernardo, 1996, page 5)

Les modèles hiérarchiques sont importants pour utiliser de manière appropriée les données d’expériences multicentriques. Ils ont aussi particulièrement adaptés pour les méta-analyses où nous avons des données d’un certain nombre d’études pertinentes qui peuvent être échangeables à certains niveaux mais pas à d’autres (Dumouchel, 1990). Dans tous les cas, le problème peut être décomposé en une série de modèles conditionnels plus simples, en utilisant la méthodologie bayésienne hiérarchique (Good, 1980).

RÉFÉRENCES

References

- Agresti, A. & Min, Y. (2005). Frequentist performance of Bayesian confidence intervals for comparing proportions in 2×2 contingency tables. *Biometrics* **61**, 515–523.
- Albert, J. (1996). *Bayesian Computation Using Minitab*. Wadsworth Publishing Company, Belmont.
- American Psychological Association (2001). *Publication Manual of the American Psychological Association* (5ème édition). Author, Washington, DC.
- Baum, M., Houghton, J., & Abrams, K. R. (1989). Early stopping rules : clinical perspectives and ethical considerations. *Statistics in Medicine* **13**, 1459–1469.
- Bayarri, M. J. & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science* **19**, 58–80.
- Bayes (1763). Essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* **53**, 370–418. [Reproduit dans *Biometrika* **45**, 293–315 (1958)]
- Berger, J. O. (2003). Could Fisher, Jeffreys and Neyman have agreed on testing? [with discussion]. *Statistical Science* **18**, 1–32.
- Berger, J. (2004). The case for objective Bayesian analysis. *Bayesian Analysis* **1**, 1–17.
- Berger, J. O. & Bernardo, J. M. (1992). On the development of reference priors [with discussion]. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics 4. Proceedings of the Fourth Valencia International Meeting*, 35–60. Oxford Univ. Press, Oxford, England.
- Bernard, J.-M. (1996). Bayesian interpretation of frequentist procedures for a Bernoulli process. *The American Statistician* **50**, 7–13.
- Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference [with discussion]. *Journal of the Royal Statistical Society, Series B, Methodological* **41**, 113–147.
- Bernardo, J. M. (1996). The concept of exchangeability and its applications. *Far East Journal of Mathematical Sciences* **4**, 111–121.
- Bernardo, J. & Smith, A. F. M. (1994). *Bayesian Theory*. John Wiley & Sons, New York.

- Bernoulli, J. (1713). *Ars Conjectandi*, (English translation by Bing Sung as Technical report No. 2 of the Department of Statistics of Harvard University, February 12, 1966), Basel, Switzerland.
- Berry, D. A. (1991). Experimental design for drug development : a Bayesian approach. *Journal of Biopharmaceutical Statistics* **1**, 81-101.
- Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician* **51**, 241-246.
- Berry, G. & Armitage, P. (1995). Mid- P confidence intervals : a brief review. *The Statistician* **44**, 417-423.
- Box, G. E. P. & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, Addison Wesley, MA.
- Brown, L. D., Cai, T., & DasGupta, A. (2001). Interval estimation for a binomial proportion [with discussion]. *Statistical Science* **16**, 101-133.
- Bunouf, P. (2006). *Lois Bayésiennes a Priori dans un Plan Binomial Séquentiel*, Thèse de doctorat en mathématiques, Université de Rouen.
- Bunouf P. & Lecoutre B. (2006). Bayesian priors in sequential binomial design. *Comptes Rendus de L'Académie des Sciences Paris, Série I* **343**, 339-344.
- Cai, T. (2005). One-sided confidence intervals in discrete distributions. *Journal of Statistical Planning and Inference* **131**, 63-88.
- Chib, S & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician* **49**, 327-335.
- Copas, J. B. & Loeber, R. (1990). Relative improvement over chance (RIOCI) for 2×2 tables. *British Journal of Mathematical and Statistical Psychology* **43**, 293-307.
- Cox, D. R. (1970). *The Analysis of Binary Data*. Methuen, Londres.
- de Cristofaro, R. (1996). L'influence du plan d'échantillonnage dans l'inférence statistique. *Journal de la Société Statistique de Paris* **137** 23-34.
- de Cristofaro, R. (2004). On the foundations of likelihood principle. *Journal of Statistical Planning and Inference* **126** 401-411.
- de Cristofaro, R. (2006). Foundations of the 'Objective Bayesian Inference'. *First Symposium on Philosophy, History and Methodology of ERROR*. Virginia Tech., Blacksburg VA.
- Dickey J. M. (1986). Discussion of Racine, A., Grieve, A. P., Flühler, H. & Smith, A. F. M., Bayesian methods in practice : Experiences in the pharmaceutical industry. *Applied Statistics* **35**, 93-150.
- Diaconis, P. & Ylvisaker, D. (1985). Quantifying prior opinion. In J. M. Bernardo, D. V. Lindley & A. F. M. Smith (Eds.), *Bayesian Statistics 2*, 133-156. North-Holland, Amsterdam.
- Dumouchel, W. (1990). Bayesian meta-analysis. In D. Berry (Ed.), *Statistical Methodology in Pharmaceutical Science*, 509-529. Marcel-Dekker, New York.
- Efron, B. (1998). R.A. Fisher in the 21st century [with discussion]. *Statistical Science* **13**, 95-122.
- ElQasyr, K. (2006). THèse de doctorat de mathématiques en cours. Université de Rouen.
- FDA (2006). *Guidance for the use of Bayesian statistics in medical device, draft guidance for industry and FDA staff*, U.S. Department of Health and Human Services, Food

- and Drug Administration, Center for Devices and Radiological Health, Rockville MD.
<http://www.fda.gov/cdrh/osb/guidance/1601.html>.
- de Finetti, B. (1972). *Probability, Induction and Statistics : The art of guessing*. John Wiley & Sons, Londres.
- de Finetti, B. (1974). *Theory of probability*, vol.1. John Wiley & Sons, New York.
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society, Series A* **222**, 309–368.
- Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophic Society*, *22*, 700–725.
- Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (2ème édition). John Wiley & Sons, New York.
- Fraser, D. A. S., Reid, N., Wong, A. & Yi, G. Y. (2003). Direct Bayes for interest parameters. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, & M. West (Eds.), *Bayesian Statistics 7*, 529–534. Oxford University Press, Oxford, England.
- Gamerman, D. (1997). *Markov chain Monte Carlo : Stochastic simulation for Bayesian inference*. Chapman & Hall, Londres.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, Londres.
- Good, I. J. (1980). Some history of the hierarchical Bayesian methodology. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian Statistics*, 489–519. Valencia University Press, Valencia. Oxford University Press, Oxford, England.
- Guttman, L. (1983). What is not what in statistics? *The Statistician* **26**, 81–107.
- Haldane, J. B. S. (1948). The precision of observed values of small frequencies. *Biometrika* **35**, 297–300.
- Inoue, L. Y. T., Berry, D. A., & Parmigiani, G. (2005). Relationship between Bayesian and frequentist sample size determination. *The American Statistician* **59**, 79–87.
- Irony, T. Z. & Pennello, G. A. (2001). Choosing an appropriate prior for Bayesian medical device trials in the regulatory setting. In *American Statistical Association 2001 Proceedings of the Biopharmaceutical Section*. American Statistical Association, Alexandria, VA.
- Jaynes, E. T. (2003). *Probability Theory : The Logic of Science* (Edited by G. L. Bretthorst), Cambridge University Press, Cambridge.
- Jeffreys, H. (1961). *Theory of Probability* (3ème édition). Clarendon, Oxford (1st edition : 1939).
- Johnson, Kotz, & Kemp, (1992). *Univariate Discrete Distributions*, 2ème édition. John Wiley and Sons.
- Kirk, R. E. (1982). *Experimental Design. Procedures for the Behavioral Sciences*. Brooks /Cole, Pacific Grove, CA.
- Laplace, P.-S (1774). Mémoire sur la probabilité des causes par les événements. *Savants Étranges* **6**, 621–656 [publié en anglais : Memoir on the probability of the causes of events, *Statistical Science* **1**, 364–378 (1986)].
- Laplace, P.-S (1812). *Théorie Analytique des Probabilités*. Courcier, Paris.

- Laplace, P.-S. (1986/1825). *Essai Philosophique sur les Probabilités* (Reproduction de la 5ème édition, 1825). Christian Bourgois, Paris.
- Lecoutre, B. (1984). *L'Analyse Bayésienne des Comparaisons*. Presses Universitaires de Lille, Lille.
- Lecoutre, B. (1996). *Traitement statistique des données expérimentales : des pratiques traditionnelles aux pratiques bayésiennes*. Avec programmes Windows par B. Lecoutre et J. Poitevineau. DECISIA Editions, Levallois-Perret.
- Lecoutre, B. (1996). Au delà du test de signification ou l'inférence statistique sans tables (à la suite d'Alain Morineau. *La Revue de Modulad* **17**, 98–100.
- Lecoutre, B. (1997/2005). Et si vous étiez un bayésien “qui s'ignore” ? *La Revue de Modulad* **18**, 81–87. Réédition (complétée) : *La Revue de Modulad* **32**, 92–105.
- Lecoutre, B. (1999). Two useful distributions for Bayesian predictive procedures under normal models. *Journal of Statistical Planning and Inference* **77**, 93–105.
- Lecoutre, B. (2000). From significance tests to fiducial Bayesian inference. In H. Rouanet, J.-M. Bernard, M.-C. Bert, B. Lecoutre, M.-P. Lecoutre, & B. Le Roux, *New ways in statistical methodology : From significance tests to Bayesian inference* (2ème édition.), 123–157. Peter Lang, Bern, SW.
- Lecoutre, B. (2001). Bayesian predictive procedure for designing and monitoring experiments. In *Bayesian Methods with Applications to Science, Policy and Official Statistics*. Office for Official Publications of the European Communities, Luxembourg, 301–310.
- Lecoutre, B. (2005). Former les étudiants et les chercheurs aux méthodes bayésiennes pour l'analyse des données expérimentales. *La Revue de Modulad* **33**, 85–107.
- Lecoutre, B. (2006). Training students and researchers in Bayesian methods for experimental data analysis. *Journal of Data Science* **4**, 207–232.
- Lecoutre, B. & Charron, C. (2000). Bayesian procedures for prediction analysis of implication hypotheses in 2×2 contingency tables. *Journal of Educational and Behavioral Statistics* **25**, 185–201.
- Lecoutre, B., Derzko, G., & Grouin, J.-M. (1995). Bayesian predictive approach for inference about proportions. *Statistics in Medicine* **14**, 1057–1063.
- Lecoutre, B. & ElQasyr, K. (2005). Play-the-winner rule in clinical trials : models for adaptative designs and Bayesian methods. In J. Janssen & P. Lenca (Eds.), *Applied Stochastic Models and Data Analysis Conference 2005 Proceedings, Part X. Health*, 1039–1050. ENST Bretagne, Brest.
- Lecoutre, B. & Faure, S. (2007). A note on new confidence intervals for the difference between two proportions based on an Edgeworth expansion. *Journal of Statistical Planning and Inference* **137**, 355–356.
- Lecoutre, B., Lecoutre, M.-P., & Poitevineau, J. (2001). Uses, abuses and misuses of significance tests in the scientific community : won't the Bayesian choice be unavoidable ? *International Statistical Review* **69**, 399–418.
- Lecoutre, B., Mabika, B., Derzko, G. (2002). Assessment and monitoring in clinical trials when survival curves have distinct shapes in two groups : a Bayesian approach with Weibull modeling. *Statistics in Medicine* **21**, 663–674.
- Lecoutre, B. & Poitevineau, J. (1992). PAC (*Programme d'Analyse des Comparaisons*) : *Guide d'utilisation et manuel de référence*. CISIA-CERESTA, Montreuil.
- Lecoutre, B. & Poitevineau, J. (2005). Le logiciel “LePAC”. *La Revue de Modulad* **33**.

- Lecoutre, B., Poitevineau, J., Derzko, G., & Grouin, J.-M. (2000). Désirabilité et faisabilité des méthodes bayésiennes en analyse de variance : application à des plans d'expérience complexes utilisés dans les essais cliniques. In I. Albert & B. Asselain (Eds.), *Biométrie et Méthodes bayésiennes* **14**, Société Française de Biométrie, Paris, 1–23.
- Lecoutre, B., Poitevineau, J., & Lecoutre, M.-P. (2005). Une raison pour ne pas abandonner les tests de signification de l'hypothèse nulle / A reason why not to ban Null Hypothesis Significance Tests. *La Revue de Modulad* **33**, 243–253.
- Lecoutre, M.-P. (2000). And... What about the researcher's point of view. In H. Rouanet, J.-M. Bernard, M.-C. Bert, B. Lecoutre, M.-P. Lecoutre, & B. Le Roux, *New ways in statistical methodology : From significance tests to Bayesian inference* (2ème édition.), 65–95. Peter Lang, Bern, SW.
- Lecoutre, M.-P., Poitevineau, J., & Lecoutre, B. (2003). Even statisticians are not immune to misinterpretations of Null Hypothesis Significance Tests. *International Journal of Psychology* **38**, 37–45.
- Lee, P. (2004). *Bayesian Statistics : An Introduction* (3rd edition). Oxford University Press, New York.
- Lhoste, E. (1923). Le Calcul des probabilités appliqué l'artillerie, lois de probabilité a priori. *Revue d'artillerie* **91**.
- Lindley, D. V. (1993). The analysis of experimental data : The appreciation of tea and wine. *Teaching Statistics* **15**, 22–25.
- Morrison, D.E. & Henkel, R.E. (1969). Significance tests reconsidered. *The American Sociologist* **4**, 131–140.
- Mossman, D. & Berger, J. (2001). Intervals for post-test probabilities : a comparison of five methods. *Medical Decision Making* **21**, 498–507.
- Nelson, N., Rosenthal, R., & Rosnow, R. L. (1986). Interpretation of significance levels and effect sizes by psychological researchers. *American Psychologist* **41**, 1299–1301.
- Novick, M. R., & Jackson, P. H. (1974). *Statistical Methods for Educational and Psychological Research*. McGraw-Hill, New York.
- O'Hagan, A. (1996). *First Bayes* [Teaching package for elementary Bayesian Statistics]. <http://www.tonyohagan.co.uk/1b/>.
- Pagano, R. R. (1990). *Understanding statistics in the behavioral sciences* (3ème édition). West, St. Paul, MN.
- Perks, W. (1947). Some observations on inverse probability including a new indifference rule. *Journal of the Institute of Actuaries* **73**, 285–312.
- Poitevineau, J., & Lecoutre, B. (2001). The interpretation of significance levels by psychological researchers : The .05-cliff effect may be overstated. *Psychonomic Bulletin and Review* **8**, 847–850.
- Rice, W. R. (1988). A new probability model for determining exact P value for 2×2 contingency tables. *Biometrics* **44**, 1–22.
- Robert, C. P. (1992). *L'Analyse Statistique Bayésienne*. Economica, Paris.
- Robert, C. P. & Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York.
- Rosenthal, R. & Gaito, J. (1963). The interpretation of levels of significance by psychological researchers. *Journal of Psychology* **55**, 33–38.
- Rosenthal, R. & Gaito, J. (1964). Further evidence for the cliff effect in the interpretation of levels of significance. *Psychological Reports* **15**, 570.

- Rouanet, H. & Lecoutre, B. (1983). Specific inference in ANOVA : From significance tests to Bayesian procedures. *British Journal of Mathematical and Statistical Psychology* **36**, 252–268.
- Routledge, R. D. (1994). Practicing safe statistics with the mid- p^* . *The Canadian Journal of Statistics* **22**, 103–110.
- Rozeboom, W. W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin* **57**, 416–428.
- Savage, L. (1954). *The Foundations of Statistical Inference*. John Wiley & Sons, New York.
- Smith, A. (1995). A conversation with Dennis Lindley. *Statistical Science* **10**, 305–319.
- Spiegelhalter, D. J., Freedman, L. S., & Parmar, M. K. B. (1994). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society, Series A* **157**, 357–416.
- Student (1908). The probable error of a mean. *Biometrika* **6**, 1–25.
- Sweeting, T. J. (2005). On the implementation of local probability matching priors for interest parameters. *Biometrika* **92**, 47–57.
- Tan, S. B., Chung, Y. F. A., Tai, B. C., Cheung, Y. B., & Machin, D. (2003). Elicitation of prior distributions for a phase III randomized controlled trial of adjuvant therapy with surgery for hepatocellular carcinoma. *Controlled Clinical Trials* **24**, 110–21.
- Tierney, L. (1994). Markov chains for exploring posterior distributions [with discussion]. *The Annals of Statistics* **21**, 1701–1762.
- Toecher, K. D. (1950). Extension of the Neyman-Pearson theory of tests to discontinuous variables. *Biometrika* **37**, 130–144.
- Walley, P. (1996). Inferences from multinomial data : learning about a bag of marbles [with discussion]. *Journal of the Royal Statistical Society B* **58**, 3–57.
- Wilkinson, L. and Task Force on Statistical Inference, APA Board of Scientific Affairs (1999). Statistical Methods in Psychology Journals : Guidelines and Explanations. *American Psychologist* **54**, 594–604.
- Winkler, R. L. (1974). Statistical analysis : theory versus practice. In *The Concept of Probability in Psychological Experiments*, Ed. C.-A.S. Staël Von Holstein, 127–140. D. Reidel, Dordrecht, Pays-Bas.
- Ye, K. (1993). Reference priors when the stopping rule depends on the parameter of interest. *Journal of the American Statistical Association* **88**, 360–363.
- Zaykin, D. V., Meng, Z., & Ghosh, S. K. (2004). Interval estimation of genetic susceptibility for retrospective case-control studies. *BMC Genetics* **5 :9**, 1–11.
- Zelen, M. (1969). Play the winner rule and the controlled clinical trial. *Journal of the American Statistical Association* **64**, 131–146, 1969.
- Zhou, X.-H. & Qin, G. (2004). New intervals for the difference between two independent binomial proportions. *Journal of Statistical Planning and Inference* **123**, 97–115.
- Zhou, X.-H. & Qin, G. (2005). A new confidence interval for the difference between two binomial proportions of paired data. *Journal of Statistical Planning and Inference* **128**, 527–542.
- Zhou, X.-H. & Qin, G. (2007). A supplement to : “A new confidence interval for the difference between two binomial proportions of paired data”. *Journal of Statistical Planning and Inference* **137**, 357–358.