REVUE DE STATISTIQUE APPLIQUÉE

J.-M. GROUIN
B. LECOUTRE

Probabilités prédictives : un outil pour la planification des expériences

Revue de statistique appliquée, tome 44, nº 1 (1996), p. 21-35
<http://www.numdam.org/item?id=RSA_1996__44_1_21_0>

Rev. Statistique Appliquée, 1996, **XLIV** (1), 21-35

PROBABILITÉS PRÉDICTIVES : UN OUTIL POUR LA PLANIFICATION DES EXPÉRIENCES*

J.-M Grouin (1), B. Lecoutre (2)

(1) Equipe ERIS, Laboratoire de Psychologie
Université de Rouen
B.P. 108, 76134 Mont-Saint-Aignan Cedex

(2) U.R.A. 1378, Analyse et Modélisation Stochastique
C.N.R.S. et Université de Rouen
Mathématiques - Site Colbert
76821 Mont-Saint-Aignan Cedex

RÉSUMÉ

Cet article présente une des nombreuses applications possibles de l'inférence bayésienne prédictive au contexte des essais planifiés. Nous nous intéressons ici à la planification d'une expérience où l'objectif est la mise en évidence de l'importance d'un effet d'intérêt : compte tenu d'un premier échantillon E_1 de données, on veut planifier une expérience, soit un nouvel échantillon E_2, de manière à avoir de «bonnes chances» de mettre en évidence un effet notable sur ce nouvel échantillon E_2. Après avoir rappelé la procédure générale usuelle, basée sur la notion de puissance, utilisée pour déterminer l'effectif nécessaire de l'échantillon E_2, ainsi que les critiques adressées à cette procédure, nous envisageons le calcul de la probabilité prédictive d'obtenir une conclusion d'effet notable basée sur le futur échantillon à partir de méthodes d'inférence bayésienne. Nous déterminons également l'effectif nécessaire de l'échantillon E_2 correspondant à une probabilité prédictive donnée d'obtenir la conclusion recherchée. Les méthodes présentées sont illustrées dans le cadre de la planification d'une expérience de psychologie.

Mots-clés : *Prédiction, inférence bayésienne, planification, effectif nécessaire, importance d'un effet.*

SUMMARY

This paper presents one use of the bayesian predictive inference in the context of experiments. We are interested here in the design of a trial which aims at showing the importance of an experimental effect : given first data accumulated from a previous trial we want to determine the required sample size of a new trial for achieving with high chances the statistical conclusion of the effect importance obtained from a bayesian inference method and based on the new trial future data. After presenting the usual procedure based on the power for determining the required sample size and assessing this procedure critically, we consider

* Cette recherche a été effectuée alors que les auteurs étaient rattachés à l'U.R.A. 1201, *Groupe Mathématiques et Psychologie*, C.N.R.S et Université de Paris V.

bayesian predictive procedures to determine the required sample size. The methods are applied to design an experiment of Psychology.

***Keywords** : Prediction, bayesian inference, trial design, sample size, effect importance.*

Introduction

De nombreux auteurs ont souligné l'intérêt de l'approche bayésienne prédictive dans le cadre des essais cliniques ou des expériences de psychologie : *cf.* notamment Lecoutre (1984), Choi et Pepple (1989), Berry (1991), Geisser (1993), Lecoutre, Derzko et Grouin (1995).

Dans cet article, nous nous intéressons plus particulièrement à l'utilisation de cette approche dans la planification d'une expérience où l'objectif est la mise en évidence de l'importance d'un effet d'intérêt. Souvent, dans cette situation, une information expérimentale préliminaire est déjà disponible, sous forme d'un «essai pilote», ou d'une première expérience dont on demande une confirmation des résultats. Formellement, nous considérerons la situation suivante : compte tenu d'un premier échantillon E_1 de données, on veut planifier une expérience (soit un nouvel échantillon E_2) de manière à avoir de «bonnes chances» d'obtenir la conclusion recherchée. Mentionnons que l'inférence prédictive peut également s'appliquer à un résultat portant sur l'ensemble des deux échantillons ($E_1 \cup E_2$), celui déjà recueilli et le futur échantillon (Grouin, 1994). Ceci répond à une autre situation expérimentale que nous n'envisagerons pas ici, qui est celle des analyses intermédiaires.

Dans un premier temps, nous rappellerons la procédure générale usuelle, basée sur la notion de *puissance*, utilisée pour déterminer l'effectif nécessaire, ainsi que les critiques adressées à cette procédure.

Ensuite, après avoir exposé quelques résultats généraux préliminaires, nous envisagerons le calcul de la probabilité prédictive d'obtenir une conclusion d'effet notable basée sur le futur échantillon à partir de méthodes d'inférence bayésienne.

Pour simplifier l'exposé, nous nous limiterons ici à considérer l'inférence sur une combinaison linéaire $\delta = \sum_g c_g \mu_g$ des moyennes de groupes indépendants d'observations, sous le modèle équinormal usuel. Mais les procédures développées peuvent être directement généralisées au cas d'une inférence portant sur un effet associé à une comparaison d'intérêt dans un plan expérimental plus complexe (en particulier un plan à mesures répétées, ou un plan *cross-over*) : Lecoutre (1984); Grouin (1994).

Nous nous limiterons également au choix de distributions *a priori* non-informatives sur les paramètres. L'inférence mise en œuvre reflète alors l'apport propre aux données; et ce choix se justifie par les motivations d'objectivité ou de neutralité, qui sont sous-jacentes à l'analyse des données expérimentales. En particulier, il n'est pas habituel d'intégrer dans l'analyse les données d'une expérience préliminaire : celles-ci seront donc utilisées uniquement pour planifier l'expérience future. Un intérêt pratique important de l'utilisation de distributions non-informatives est d'englober les procédures fréquentistes usuelles (test de signification et intervalle de confiance), dont l'usage dans la présentation des résultats expérimentaux est souvent incontournable en l'état actuel. Mais ce choix n'est pas limitatif, et l'utilisation, par exemple, de distributions *a priori* conjuguées au modèle d'échantillonnage intégrant des opinions personnelles sur les valeurs possibles des paramètres, conduirait à une formalisation similaire du problème, et à des calculs analogues (Grouin, 1994).

Enfin, les méthodes seront illustrées par l'exemple de la planification d'une expérience de psychologie.

1. De la puissance du test à la probabilité prédictive

L'effectif du futur échantillon est généralement choisi en considérant la puissance du test usuel

$$H_0 : \text{«}\delta = 0\text{»} \text{ versus } H_1 : \text{«}\delta > 0\text{»}$$

Il s'agit de calculer la probabilité $\rho(\Delta)$ de rejeter l'hypothèse nulle $\delta = 0$ au seuil α, si δ était égale à une valeur de référence Δ. Si nous fixons simultanément le seuil α, la valeur Δ et la puissance $\rho(\Delta)$ du test pour cette valeur, l'effectif minimal de l'échantillon nécessaire à l'obtention de la conclusion recherchée («le test est significatif au seuil α») est déterminé.

Une difficulté essentielle est la nécessité de choisir une valeur de référence unique Δ. Un choix souvent effectué, voire recommandé (*cf.* S.J. Pocock, 1983, par exemple) est celui de «la plus petite valeur Δ_0 qui est intéressante, signifiante pour le domaine expérimental considéré», car on ne veut pas «déceler un effet trop faible». Mais, en pratique, on constate que ce choix varie d'un expérimentateur à un autre, en fonction de motivations différentes. En particulier Δ pourra être la valeur préalablement observée dans une expérience préliminaire, ou encore une valeur justifiée par des considérations extra-statistiques, d'ordre économique par exemple.

Pour surmonter cette difficulté, plusieurs auteurs ont proposé de calculer une «puissance moyenne» ρ_{moy} égale à la moyenne des puissances évaluées sous chacune des valeurs possibles de δ, pondérée selon une distribution bayésienne sur δ disponible avant l'expérience considérée (*cf.* Brown *et al.*, 1987, Spiegelhalter et Freedman, 1988). Dans le cadre bayésien, cette puissance moyenne s'interprète comme la probabilité *prédictive* d'obtenir un résultat significatif pour le test effectué sur le futur échantillon.

La procédure usuelle de calcul d'effectif peut en fait faire l'objet d'une critique fondamentale. En effet, puisque l'objectif réel est généralement de pouvoir conclure que δ est supérieur à la borne minimale d'intérêt Δ_0, celle-ci devrait être intégrée dans la procédure statistique elle-même, et le test pertinent devrait être le suivant :

$$H_0 : \text{«}\delta = \Delta_0\text{»} \text{ versus } H_1 : \text{«}\delta > \Delta_0\text{»}$$

Dans une perspective classique de calcul d'effectif, la puissance du test pertinent devrait alors être considérée sous une hypothèse alternative du type $\delta = \Delta_0 + x$ où x désigne un écart positif par rapport à l'hypothèse nulle. Le choix d'une valeur x se révèle être à nouveau une source d'embarras pour l'expérimentateur.

Enfin, une dernière difficulté vient de ce que la puissance dépend de paramètres parasites (variances,...). Il n'est pas satisfaisant, comme on le fait généralement, de se contenter de fixer une valeur plus ou moins arbitraire pour ces paramètres.

Pour résoudre les difficultés précédentes, nous proposons, pour planifier une expérience, d'une part d'utiliser explicitement une procédure statistique de «recherche d'une conclusion d'effet notable» ($\delta > \Delta_0$), et d'autre part d'évaluer la probabilité prédictive d'obtenir cette conclusion sur le futur échantillon, compte tenu de l'information préalable disponible. Le cadre statistique bayésien permet une formulation précise de ce problème. Pour les plans d'expérience usuels, les distributions bayésiennes prédictives mises en jeu pour le calcul des probabilités prédictives peuvent être explicitées.

1.1. Notations et résultats préliminaires

d et s^2 désigneront les grandeurs aléatoires observables définies respectivement par la combinaison linéaire $d = \sum c_g \overline{x}_g$ des moyennes et la variance corrigée intragroupe. n_g est l'effectif du groupe g. Nous poserons le modèle équinormal usuel de paramètres μ_g et σ^2 (variance supposée commune aux différents groupes). Sous ce modèle, d et σ^2 sont conjointement exhaustives pour les paramètres $\delta = \sum c_g \mu_g$ et σ^2 et leurs distributions d'échantillonnage sont indépendantes, avec :

$$d|\delta,\ \sigma^2 \sim N(\delta, b^2\sigma^2) \quad \text{où } b^2 = \sum c_g^2/n_g$$

$$s^2|\delta,\ \sigma^2 \sim \sigma^2\chi_q^2/q \quad \text{où } q = \sum (n_g - 1)$$

Les procédures de test fréquentistes utilisent les distributions d'échantillonnage de la statistique $(d-\delta)/bs$, qui est une distribution t usuelle à q degrés de liberté

$$\frac{d-\delta}{bs}|\delta,\ \sigma^2 \sim t_q$$

et de la statistique d/bs, qui est une distribution t *non-centrée*

$$\frac{d}{bs}|\delta,\ \sigma^2 \sim t'_q\left(\frac{\delta}{b\sigma}\right)$$

Pour la distribution *a priori* non-informative usuelle (distribution uniforme pour le couple (δ, (log σ^2)), les distributions bayésiennes *a posteriori* relatives à δ et σ^2 (conditionnelles à un échantillon observé E) sont respectivement des distributions t *généralisée* et *khi-deux inverse* (*cf.* par exemple Box et Tiao, 1992) :

$$\delta|E \sim \delta|d,\ s^2 \sim t_q(d, b^2s^2)$$

$$\sigma^2|E \sim \sigma^2|d,\ s^2 \sim s^2(\chi_q^2/q)^{-1}$$

avec la distribution conditionnelle à σ^2

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

où d et s^2 désignent ici les quantités observées (nous utilisons les mêmes notations que pour les variables pour simplifier les écritures).

Par la suite, nous aurons à considérer les inférences relatives d'une part à l'échantillon déjà recueilli E_1, d'autre part à l'échantillon futur E_2. Nous utiliserons les distributions précédentes, en indiciant par 1 ou par 2 les quantités d, b^2, s^2 et q.

1.2. Recherche d'une conclusion d'effet notable portant sur δ

Il s'agit de déterminer la probabilité prédictive d'une conclusion d'effet notable (soit $\delta > \Delta_0$) pour l'expérience future. Tous les résultats nécessaires sont fournis par l'étude de la distribution que nous appelons *K-prime*, qui est une généralisation, entre autres, de la distribution *t non-centrée* classique, qui intervient dans la puissance du test de Student. Ces résultats sont énoncés en annexe.

Pour la distribution *a priori* non-informative, la distribution *a posteriori* relative à δ conditionnellement à l'échantillon futur E_2 (avec des effectifs fixés) est

$$\delta|E_2 \sim t_{q_2}(d_2, b_2^2 s_2^2)$$

Une conclusion d'effet notable sera obtenue pour cet échantillon futur si la probabilité bayésienne *a posteriori*

$$\Pr(\delta > \Delta_0|E_2) = \Pr\left(t_{q_2} > \frac{\Delta_0 - d_2}{b_2 s_2}\right)$$

où t_{q_2} est la distribution de Student usuelle, est supérieure ou égale à une garantie γ spécifiée à l'avance ($\gamma > 0.5$). De manière équivalente, cette conclusion est obtenue si le test fréquentiste unilatéral de l'hypothèse nulle H_0 : «$\delta = \Delta_0$» (*vs* H_1 : «$\delta > \Delta_0$») est significatif au seuil $\alpha = 1 - \gamma$.

L'échantillon futur conduira donc à une conclusion d'effet notable si l'événement ω suivant se réalise :

$$\frac{d_2 - \Delta_0}{b_2 s_2} \text{ est supérieur à la valeur } \mathrm{t}_{q_2,\gamma} \text{ telle que } \Pr(t_{q_2} < \mathrm{t}_{q_2,\gamma}) = \gamma$$

Le calcul de la probabilité prédictive $\pi(\omega)$ de cet événement revient donc à expliciter la distribution prédictive relative à la statistique $(d_2 - \Delta_0)/b_2 s_2$ étant donné le premier échantillon E_1. Celle-ci se déduit des distributions d'échantillonnage (conditionnelles aux paramètres)

$$d_2|\delta,\ \sigma^2 \sim N(\delta, b_2^2\sigma^2) \text{ et } s_2^2|\delta, \sigma^2 \sim s_2^2|\sigma^2 \sim \sigma^2\chi_{q_2}^2/q_2$$

et des distributions *a posteriori* relatives aux paramètres issues de l'échantillon E_1, caractérisées par

$$\delta|\sigma^2 \sim N(d_1, b_1^2\sigma^2) \quad \text{et} \quad \sigma^2 \sim (\chi^2_{q_1}/q_1)^{-1}$$

Nous en déduisons, en utilisant les propriétés classiques des distributions conditionnelles :

$$d_2|\sigma^2 \sim N(d_1,\ (b_1^2 + b_2^2)\sigma^2)$$

d'où la distribution conditionnelle à σ^2qui est une distribution t *non-centrée*

$$\frac{d_2 - \Delta_0}{b_2 s_2}|\sigma^2 \sim t'_{q_2} \sim t'_{q_2}\left(\frac{d_1 - \Delta_0}{b_2\sigma}, \frac{b_1^2 + b_2^2}{b_2^2}\right)$$

et par suite la distribution marginale, qui est une distribution K*-prime*

$$\frac{d_2 - \Delta_0}{b_2 s_2} \sim K'_{q_1,q_2}\left(\frac{d_1 - \Delta_0}{b_2 s_1}, \frac{b_1^2 + b_2^2}{b_2^2}\right)$$

La probabilité prédictive relative à l'événement recherché (pour des effectifs fixés) est donc :

$$\begin{aligned}\pi(\omega) &= \Pr\left(K'_{q_1,q_2}\left(\frac{d_1 - \Delta_0}{b_2 s_1}, \frac{b_1^2 + b_2^2}{b_2^2}\right) < \mathrm{t}_{q_2,\gamma}\right)\\ &= \Pr\left(K'_{q_2,q_1}\left(\mathrm{t}_{q_2,\gamma}, \frac{b_1^2 + b_2^2}{b_2^2}\right) < \frac{d_1 - \Delta_0}{b_2 s_1}\right)\end{aligned}$$

Celle-ci ne dépend en fait de d_1 et s_1 que par l'intermédiaire de $(d_1 - \Delta_0)/s_1$.

La probabilité prédictive conditionnelle à une valeur δ fixée est obtenue en posant $d_1 = \delta$ et $b_1 = 0$:

$$\pi(\omega|\delta) = \Pr\left(K'_{q_1,q_2}\left(\frac{\delta - \Delta_0}{b_2 s_1}, 1\right) > \mathrm{t}_{q_2,\gamma}\right)$$

Remarques

Cette probabilité prédictive ne peut pas dépasser une borne maximale qui est déterminée par les données du premier échantillon seulement. En effet, quand $q_2 \to \infty$ nous avons $b_2 \to 0$, et par suite

$$\begin{aligned}\pi(\omega) &= \Pr\left(K'_{q_1,q_2}\left(\frac{d_1 - \Delta_0}{s_1},\ b_1^2 + b_2^2\right) > b_2\, \mathrm{t}_{q_2,\gamma}\right)\\ &\to \Pr\left(K'_{q_1,\infty}\left(\frac{d_1 - \Delta_0}{s_1},\ b_1^2\right) > 0\right) = \Pr\left(t_{q_1} > \frac{\Delta_0 - d_1}{b_1 s_1}\right)\end{aligned}$$

Cette borne est la probabilité *a posteriori* pour le premier échantillon que δ soit supérieur à Δ_0.

Par ailleurs, quand $q_1 \to \infty$ (soit $\sigma = s_1$ fixé) et $b_1 = 0$ (soit $\delta = d_1$ fixé), nous retrouvons comme cas limite la distribution d'échantillonnage t *non-centrée*. Rappelons que, dans le cadre fréquentiste, la probabilité $\pi(\omega)$ est dans ce cas la *puissance* du test, c'est-à-dire ici la probabilité de rejeter l'hypothèse nulle d'une différence égale à Δ_0 si la vraie différence vaut δ.

1.3. Recherche d'une conclusion d'effet notable portant sur δ/σ

Dans une expérience de psychologie, il est usuel de spécifier l'importance d'un effet en terme d'effet calibré, c'est-à-dire en fonction du rapport δ/σ (*cf.* Rouanet, Lépine et Pelnard-Considère, 1975; Cohen, 1977; Rouanet, Lecoutre et Bernard, 1987). La recherche d'une conclusion d'effet notable sur le futur échantillon se traduira alors de la façon suivante : «$\delta/\sigma > c$» où c désigne une borne au delà de laquelle l'effet vrai calibré peut être considéré comme important.

Pour une distribution *a priori* non informative, la distribution *a posteriori* relative à δ/σ conditionnellement à l'échantillon E_2 est un cas limite de la distribution K', que nous appelons *lambda-prime* :

$$\frac{\delta}{\sigma}|E_2 \sim K'_{q_2,\infty}\left(\frac{d_2}{s_2}, b_2^2\right) \sim \Lambda'_{q_2}\left(\frac{d_2}{s_2}, b_2^2\right)$$

Une conclusion d'effet calibré notable sera obtenue si la probabilité bayésienne *a posteriori*

$$\Pr(\delta/\sigma > c|E_2) = \Pr\left(K'_{q_2,\infty}\left(\frac{d_2}{s_2}, b_2^2\right) > c\right) = \Pr\left(t'_{q_2}(c, b_2^2) < \frac{d_2}{s_2}\right)$$

est supérieure ou égale à une garantie γ spécifiée à l'avance ($\gamma > 0.5$). De manière équivalente, cette conclusion est obtenue si le test fréquentiste unilatéral de l'hypothèse nulle H_0 : «$\delta/\sigma = c$» (*vs* H_1 : «$\delta/\sigma > c$») est significatif au seuil $\alpha = 1 - \gamma$.

L'échantillon futur conduira donc à une conclusion d'effet calibré notable si l'événement ω suivant se réalise :

d_2/s_2 est supérieure à la valeur x_γ donnée par la fonction de répartition de la distribution t *non-centrée* :

$$\Pr(t'_{q_2}(c, b_2^2) < x_\gamma) = \gamma$$

Le calcul de la probabilité prédictive $\pi(\omega)$ de cet événement revient donc à expliciter la distribution prédictive relative à la statistique d_2/s_2.

Nous obtenons en procédant comme dans la section précédente :

$$\pi(\omega) = \Pr\left(K'_{q_1,q_2}\left(\frac{d_1}{s_1}, b_1^2 + b_2^2\right) > x_\gamma\right) = \Pr\left(K'_{q_2,q_1}(x_\gamma,\ b_1^2 + b_2^2) < \frac{d_1}{s_1}\right)$$

qui ne dépend en fait de d_1 et s_1 que par l'intermédiaire du rapport d_1/s_1. Bien entendu, pour $c = 0$, il est équivalent de considérer l'effet ou l'effet calibré.

Remarques

Cette probabilité prédictive ne peut pas dépasser une borne maximale qui est déterminée par les données du premier échantillon seulement. En effet, quand $q_2 \to \infty$ nous avons $b_2 \to 0$, et par suite

$$x_\gamma \to c \quad \text{et} \quad \pi(\omega) \to \Pr\left(K'_{q_1,\infty}\left(\frac{d_1}{s_1}, b_1^2\right) > c\right)$$

Cette borne est la probabilité *a posteriori* pour le premier échantillon que δ/σ soit supérieur à c.

Par ailleurs, quand $q_1 \to \infty$ et (soit $\sigma = s_1$ fixé) et $b_1 = 0$ (soit $\delta = d_1$ fixé), nous retrouvons encore comme cas limite la distribution d'échantillonnage t *non-centrée*.

2. Application à la planification d'une expérience de psychologie

Cosidérons à titre d'illustration les données extraites de l'expérience de L. Vézin (*in* Ghiglione et Richard, 1993, pages 589-590, Tome 2), dont l'objectif est de comparer plusieurs méthodes dans l'apprentissage d'une règle par différents sujets. La règle utilisée concerne les vitesses de rotation des roues d'un engrenage : le rapport des vitesses est inversement proportionnel au nombre de dents. Les deux méthodes (retenues dans notre exemple) sont les suivantes : une première méthode où l'exploration de l'engrenage par le sujet est libre et une deuxième méthode où l'exploration de l'engrenage par le sujet est guidée avec des questions et des indications quant à la réponse correcte. Le niveau d'acquisition a été mesuré au cours d'une épreuve dite de transfert des connaissances dans laquelle des roues différentes ainsi que des rapports de vitesse différents de ceux de la phase d'apprentissage ont été utilisés. Chaque sujet passe une épreuve comportant 9 problèmes et obtient une note égale au nombre de problèmes correctement résolus (de 0 à 9). Une affectation des sujets au hasard en deux groupes correspondant aux deux méthodes a été effectuée. Les effectifs, moyennes et écart-types corrigés calculés pour les deux méthodes sur le premier échantillon E_1 des données observées sont les suivants :

– Méthode 1 : $n = 16$, $\overline{x} = 3.9$, $s = 2.1$

– Méthode 2 : $n = 18$, $\overline{x} = 5.6$, $s = 1.8$

d'où $d_1 = +1.7$, $s_1^2 = 3.79$, $q_1 = 32$, et $b_1^2 = 17/144$

2.1. Recherche d'une conclusion d'effet notable portant sur δ

La borne minimale Δ_0 au delà de laquelle l'effet d'intérêt est considéré comme important est prise égale ici à un point de moyenne ($\Delta_0 = +1$). Sur le plan descriptif, la différence observée entre les deux moyennes ($d_1 = 1.7$) est donc notable. Prenons la garantie $\gamma = 0.95$.

La probabilité bayésienne *a posteriori* standard, conditionnellement aux données du premier échantillon, que δ soit supérieur à $+1$ est

$$\Pr(\delta > +1|E_1) = \Pr\left(t_{q_1}(d_1, b_1^2 s_1^2) > +1\right) = 0.848 < 0.95$$

Ce résultat est encourageant, mais ne nous permet cependant pas de conclure à un effet notable avec une garantie d'au moins 0.95. Dans une telle situation, on peut envisager de planifier une nouvelle expérience, avec un nombre de sujets par groupe plus élevé.

Considérons d'abord la procédure généralement utilisée pour déterminer le nombre n de sujets par groupe du futur échantillon dans cette situation (voir l'introduction) : on calcule la puissance du test usuel, c'est-à-dire la probabilité de rejeter l'hypothèse nulle $\delta = 0$ (*vs* $\delta > 0$) si la différence δ était égale à $\delta_0 = +1$ (ou éventuellement à la valeur observée $d_1 = +1.7$), en prenant pour la variance σ^2 la valeur observée 3.79. Nous avons par exemple les probabilités

	$\delta = +1$	$\delta = +1.7$
$n = 10$	0.295	0.593
$n = 25$	0.558	0.919
$n = 50$	0.818	0.996
$n = 75$	0.931	0.9999
$n = 100$	0.976	1.0000

La première critique à l'égard de cette procédure est qu'elle ne prend pas directement en considération la conclusion recherchée (différence supérieure à $+1$); la seconde critique est qu'elle ignore notre incertitude sur la vraie valeur de σ^2. Maintenant, compte tenu des résultats de la première expérience, la probabilité prédictive d'obtenir une conclusion d'effet notable ($\delta > 1$), avec une garantie de 0.95 au moins pour un futur échantillon, avec n sujets par groupe, est donnée par :

$$\pi(\omega, n) = \Pr\left(K'_{32,2(n-1)}\left(+0.360\sqrt{n/2}, 1 + \frac{17n}{288}\right) > t_{2(n-1),0,95}\right)$$

avec la probabilité conditionnelle à δ

$$\pi(\omega, n|\delta) = \Pr\left(K'_{32,2(n-1)}\left(\frac{\delta - 1}{1.946}\sqrt{n/2}, 1\right) > t_{2(n-1),0,95}\right)$$

La figure 1 donne la probabilité prédictive $\pi(\omega, n)$ en fonction de l'effectif n par groupe. Nous considérons également la probabilité prédictive conditionnelle à

$\delta = +1.7$. Nous avons par exemple :

	$\pi(\omega, n)$	$\pi(\omega, n \mid \delta = +1.7)$
$n = 10$	0.241	0.191
$n = 25$	0.399	0.347
$n = 50$	0.525	0.549
$n = 75$	0.590	0.695
$n = 100$	0.629	0.797
$n = 200$	0.705	0.960
$n = 1246$	0.800	1.000
$n \to \infty$	0.848	1

Par conséquent, si nous voulons avoir au moins une chance sur deux de pouvoir conclure à l'existence d'un effet notable avec une garantie supérieure à 0.95 au moins, il faudra un effectif de 44 sujets par groupe. Si nous voulons au moins trois chances sur quatre, cet effectif devra être de 387. Etc. On voit que ces effectifs sont nettement supérieurs à ceux que l'on obtient en utilisant la procédure usuelle.

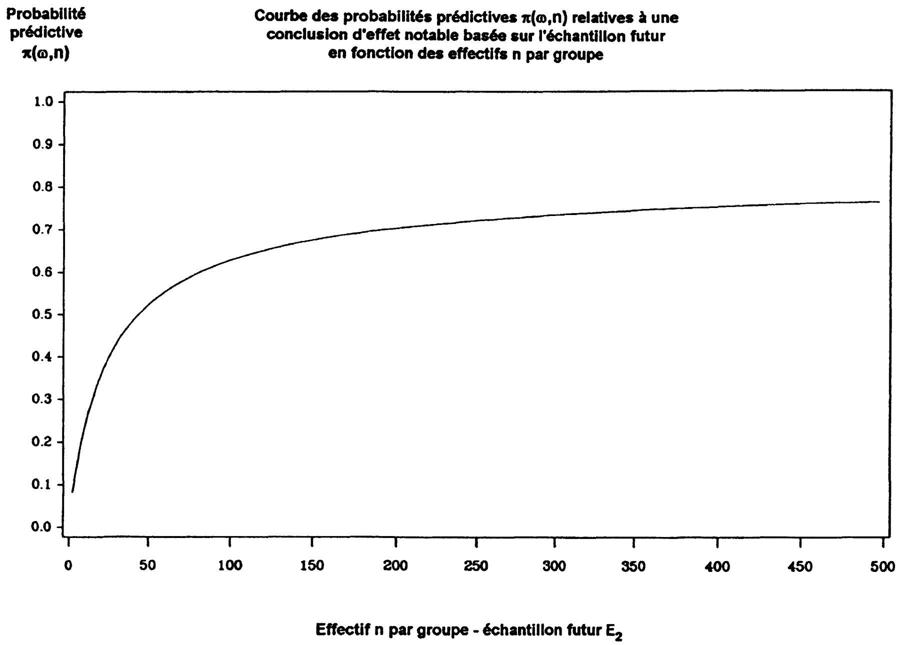

FIGURE 1

2.2. Recherche d'une conclusion d'effet notable portant sur δ/σ

Le problème est comparable à celui traité précédemment. La borne minimale c au delà de laquelle l'effet calibré peut être considéré comme important est prise ici égale à 1/2. La probabilité bayésienne *a posteriori* standard, conditionnellement aux données du premier échantillon, que δ/σ soit supérieur à $+1/2$ est

$$\Pr\left(\frac{\delta}{\sigma} > +\frac{1}{2}|E_1\right) = \Pr\left(\Lambda'_{q_1}\left(\frac{d_1}{s_1}, b_1^2\right) > +\frac{1}{2}\right) = 0.846 < 0.95$$

A nouveau, nous ne pouvons pas conclure à l'existence d'un effet notable avec une garantie de 0.95 au moins à partir des premières données.

La probabilité bayésienne 0.846 est du même ordre que celle calculée précédemment dans le cas d'une recherche de conclusion d'importance sur l'effet δ, ce qui est cohérent avec le choix de la borne minimale d'intérêt $c = 1/2 \approx \Delta_0/s_1$. De même les valeurs des probabilités prédictives $\pi(\omega, n)$ sont sensiblement équivalentes à celles obtenues dans le cas précédent : on trouve par exemple $\pi(\omega, 200) = 0.701$, au lieu de 0.705.

3. Conclusion

En conclusion, nous pensons pouvoir affirmer que les méthodes basées sur l'inférence prédictive devraient avoir une place parmi les méthodes inférentielles usuelles pour l'analyse des données expérimentales. Elles permettent de répondre de façon générale aux interrogations de l'expérimentateur, vis à vis de la stabilité de résultats déjà observés, sur des données futures; et elles peuvent être utilisées avec profit dans un contexte de planification ou d'analyse intermédiaire. Ces méthodes apparaissent particulièrement pertinentes pour déterminer l'effectif nécessaire à l'obtention d'une conclusion souhaitée. L'évaluation du potentiel inductif du futur échantillon à l'aide des probabilités prédictives pourrait ainsi éviter aux expérimentateurs bien des illusions quant aux chances réelles de pouvoir atteindre la conclusion souhaitée.

Les solutions développées ici sont relativement simples à mettre en œuvre. Des programmes informatiques généraux ont été réalisés et seront prochainement diffusés.

Nous noterons encore que, lorsque des informations expérimentales préliminaires ne sont pas disponibles, les mêmes procédures peuvent encore être utilisées, en remplaçant la distribution *a posteriori* issue de l'échantillon E_1 par une distribution «subjective», de la même famille, qui pourra par exemple résulter de la consultation d'experts du domaine concerné.

Références Bibliographiques

Berry D.A. (1991) – Experimental design for drug development : A Bayesian approach. *Journal of Biopharmaceutical Statistics* **1**, 81-102.

Box G.E.P., Tiao G.C. (1992) – *Bayesian Inference in Statistical Analysis* (deuxième édition). Wiley : New York.

Brown B.W., Herson J., Atkinson E., Rozell, M.E. (1987) – Projection from previous studies : a bayesian and frequentist compromise. *Controlled Clinical Trials* **8**, 29-49.

Choi S.C., Pepple P.A. (1989) – Monitoring clinical trials based on predictive probability of significance. *Biometrics* **45** , 317-323.

Cohen J. (1977) – *Statistical Power Analysis for the Behavioral Sciences*. New York : Academic Press.

Fisher R.A. (1990) – *Statistical Methods, Experimental Design and Scientific Inference* (réédition). Oxford : Oxford University Press.

Geisser S. (1993) – *Predictive Inference : An Introduction*. New York : Chapman et Hall.

Ghiglione R., Richard J.-F. (1993) – *Cours de Psychologie*. Paris : Dunod, Tome 2.

Grouin J.-M. (1994) – *Procédures bayésiennes prédictives pour les essais expérimentaux*. Thèse de doctorat de Mathématiques Appliquées. Paris : Université René Descartes.

Lecoutre, B. (1984) – *L'Analyse Bayésienne des Comparaisons*. Lille : Presses Universitaires de Lille.

Lecoutre B., Derzko G., Grouin J.-M. (1995) – Bayesian predictive approach for inference about proportions. *Statistics in Medicine, 14,* 1057-1063.

Lecoutre, B., Guigues, J.-L., Poitevineau, J. (1992) – Distribution of quadratic forms of multivariate Student variables. *Applied Statistics* **41**, 617-627.

Pocock S.J. (1983) – Clinical Trials, a practical approach. Wiley.

Rouanet H., Lépine D., Pelnard-Considère J. (1975) – Bayes-fiducial procedures as practical substitutes for misplaced significance tests : an application to educational data. In D.N.M. De Gruitjer and L.J.Th. Vander Kamp (Eds), Advances in Psychological and Educational Measurement, New York : Wiley, 33- 50.

Rouanet H., Lecoutre B., Bernard J.-M. (1987) – L'inférence fiducio-bayésienne comme méthode d'analyse de données : Un exemple d'application à des données psychométriques. *Statistique et Analyse des Données* **11**, 58-74.

Spiegelhalter D.J., Freedman, L.S. (1988) – Bayesian Approaches to Clinical Trials. Bayesian Statistics 3, *Oxford University Press* p. 453-477.

Annexe : Distribution K-prime

Nous énonçons dans cette annexe les résultats concernant la distribution K-prime, qui permettent de justifier et de mettre en œuvre les procédures développées dans cet article.

Caractérisation

La distribution *K-prime* a été introduite dans Lecoutre (1984) avec la caractérisation suivante :

$$\text{si } x|y^2 \sim t_{q+r}\left(ay,\, b^2\frac{qy^2+r}{q+r}\right) \; (y>0)$$

(i.e. $x|y^2$ suit une distribution du t *généralisée* avec par définition :
$t_{q+r}\left(ay,\, b^2\frac{qy^2+r}{q+r}\right) = ay + b\left(\frac{qy^2+r}{q+r}\right)^{1/2} t_{q+r}$ où t_{q+r} désigne la distribution usuelle du t de Student à $q+r$ degrés de liberté)

$$\text{et } y^2 \sim F_{q,r}$$

(distribution F usuelle)

alors x a la distribution *K-prime* à q et r degrés de liberté, d'indice d'excentricité a et d'échelle b :

$$x \sim K'_{q,r}(a, b^2)$$

Cette distribution inclut comme cas particuliers :

• pour $q = \infty$, la distribution t *non centrée*, qui intervient dans la puissance du test de Student

$$K'_{\infty,r}(a, b^2) = b\, t'_r\left(\frac{a}{b}\right)$$

distribution souvent limitée au cas $b = 1$ et notée alors : $t'_r(a)$.

• pour $r = \infty$, la distribution *lambda-prime*, qui intervient dans la distribution bayésienne *a posteriori* relative au rapport μ/σ de la moyenne à l'écart-type sous le modèle normal (cette distribution a notamment été utilisée par Fisher, 1990, pages 126-127)

$$K'_{q,\infty}(a, b^2) = \Lambda'_q(a, b^2)$$

Caractérisations équivalentes

Si l'un des trois ensembles de conditions suivants est rempli ($a, b > 0$ et $q > 0$ et $r > 0$ sont des réels) :

• $x|y^2 \sim \Lambda'_q(ay, b^2y^2)$ avec $y > 0$ et $y^2 \sim (\chi^2_r/r)^{-1}$

• $x = w/u$, w et u étant indépendantes

$$\text{avec } w \sim \Lambda'_q(a, b^2) \text{ et } 1/u^2 \sim (\chi^2_r/r)^{-1}$$

• $x|y^2 \sim t'_r(a/y, b^2)$ $(b > 0)$ et $y^2 \sim (\chi_q^2/q)^{-1}$
alors $x \sim K'_{q,r}(a, b^2)$.

Propriétés

$$K'_{q,r}(a,b^2) = bK'_{q,r}(a/b,1)$$
$$K'_{q,r}(0,b^2) = t_r(0,b^2)$$
$$\Pr(K'_{q,r}(-x,b^2) < -a) = \Pr(K'_{q,r}(x,b^2) > a)$$
$$\Pr(K'_{q,r}(a,b^2) < 0) = \Pr(\Lambda'_q(a,b^2) < 0) = \Pr\left(t_q > \frac{a}{b}\right)$$

Relation remarquable

Les fonctions de répartition des distributions K-*prime* à q et r degrés de liberté et K-*prime* à r et q degrés de liberté sont liées par la relation remarquable :

$$\Pr(K'_{q,r}(a,b^2) < x) = \Pr(K'_{r,q}(x,b^2) > a)$$

avec le cas particulier $\Pr(t'_q(a,b^2) < x) = \Pr(\Lambda'_q(x,b^2) > a)$

Densité

$$p(x) = \frac{1}{\sqrt{\pi r}} \frac{1}{\Gamma(q/2)\Gamma(r/2)} b^{q+r} \left(\frac{q}{qb^2+a^2}\right)^{q/2} \left(\frac{r}{rb^2+x^2}\right)^{(r+1)/2}$$
$$\times \sum_{j=0}^{+\infty} \frac{1}{j!}\Gamma((q+j)/2)\Gamma((r+j+1)/2) \left(\frac{4}{(qb^2+a^2)(rb^2+x^2)}\right)^{j/2} \left(a\frac{x}{b}\right)^j$$

Moments

$$\text{moy}\,(x) = ak \quad [q > 1]$$
$$\text{moy}\,(x^2) = \frac{r}{r-2}(a^2+b^2) \quad [r > 2]$$
$$\text{var}\,(x) = \left(\frac{r}{r-2}k^2\right) a^2 + \frac{r}{r-2}b^2 \quad [r > 2]$$
$$\text{moy}\,(x^3) = \frac{r}{r-3}\left(\frac{q+1}{q}a^3 + 3ab^2\right) k \quad [r > 3]$$
$$\text{moy}\,(x^4) = \frac{r^2}{(r-2)(r-4)}\left(\frac{q+2}{q}a^4 + 6a^2b^2 + 3b^4\right) \quad [r > 4]$$

$$\text{avec } k = \frac{\Gamma((q+1)/2)\Gamma((r-1)/2)}{\Gamma(q/2)\Gamma(r/2)} \left(\frac{r}{q}\right)^{1/2}$$

Fonction de répartition

Nous donnons la fonction de répartition pour $b = 1$. Pour b quelconque on utilisera :

$$\Pr(K'_{q,r}(a, b^2) < x) = \Pr(K'_{q,r}(a/b, 1) < x/b)$$

- Si $a = 0$ on utilisera

$$\Pr(K'_{q,r}(0, 1) < x) = \Pr(t_r(0, 1) < x)$$

- Si $a > 0$ et $x < 0$

$$F(x) = \sum_{j=0}^{+\infty} (-1)^j e_j \, \mathrm{IB}_{x^2/(r+x^2)}((j+1)/2,\ r/2)$$

où IB est la fonction *Bêta incomplète*

$$\text{et } e_j = \frac{1}{2} \frac{\Gamma((q+j)/2)}{\Gamma(q/2)\Gamma(1+j/2)} \left(\frac{q}{q+a^2}\right)^{q/2} \left(\frac{a^2}{q+a^2}\right)^{j/2} \quad \text{avec } \sum_{j=0}^{+\infty} e_j = 1$$

- Si $a > 0$ et $x > 0$

$$F(x) = \Pr(t_q > a) + \sum_{j=0}^{+\infty} e_j \, \mathrm{IB}_{x^2/(r+x^2)}((j+1)/2,\ r/2)$$

- Si $a < 0$ on utilisera

$$\Pr(K'_{q,r}(a, 1) < x) = \Pr(K'_{q,r}(-a, 1) > -x) = 1 - \Pr(K'_{q,r}(-a, 1) < -x)$$

Les séries précédentes peuvent être calculées par une double récurrence sur les coefficients e_j et sur les termes de la fonction Bêta incomplète, en utilisant un algorithme analogue à celui développé pour la distribution *psi-deux* dans Lecoutre, Guigues et Poitevineau (1992).