Inférence sur diverses mesures d'association entre variables catégorielles: Méthodes bayésiennes

Jean-Marc Bernard
Laboratoire de Psychologie Environnementale
CNRS UMR 8069 & Université Paris 5

Journée d'études "L'Analyse Statistique des Données Catégorielles en Sciences Humaines et Sociales"

Boulogne-Billancourt, 6 Juin 2006

1. INTRODUCTION

Objectifs de cet exposé

□ Concernant les mesures d'association

- Illustrer la richesse des questions qu'on peut poser, et des conclusions possibles
- Articulation Description / Inférence

□ Concernant l'étape descriptive

- Définir indices adaptés à la structure des données, et à chaque question posée
- En particulier, indices pour répondre au problème de la grandeur des liaisons, des effets
- Conclusions descriptives

□ Concernant l'inférence statistique

- Répond au problème de la genéralisabilité des conclusions descriptives
- Parfaite adéquation des méthodes bayésiennes
- Méthodes bayésiennes permettent de faire de l'inférence sur n'importe quel indice, n'importe propriété globale

Données catégorielles

□ Structure de données

- Unités statistiques (individus, sujets): `I`
- Une ou plusieurs variables dépendantes toutes catégorielles: `J`, `K`, *etc.*.
- Prototype: un *questionnaire*

□ Structures étudiées ici

- `I->J2*K2`
- `I->J*K`
- `I->J2*K2*L2...`

□ Plan

- Inférence bayésienne illustrée sur `I->J3`
- Tableaux de contingence, mesures d'association
- Exemples

2. INFERENCE STATISTIQUE APPROCHE BAYESIENNE

Exemple illustratif

□ **Données**: taille $n = 24$, $K = 3$ catégories.

Catégorie	k_1	k_2	k_3
Effectifs $\boldsymbol{a}$	10	8	6

□ **Fréquences observées et parentes**

Fréquences obs. $\boldsymbol{f}$	0.417	0.333	0.250	1
Fréquences par. $\boldsymbol{\theta}$	θ_1	θ_2	θ_3	1

□ **Question simple**: L'inférence porte sur un paramètre initial ou dérivé d'intérêt.
On a $f_1 = 0.417$. Que dire de θ_1?
On a $f_1 - f_2 = 0.084$. Que dire de $\theta_1 - \theta_2$?

□ **Question complexe**: L'inférence porte sur plusieurs paramètres conjointement. On observe descriptivement une certaine propriété d'intérêt,

$$f_1 > f_2 > f_3$$

Que dire de la version parente de cette propriété,

$$\theta_1 > \theta_2 > \theta_3 \quad ?$$

De la description a l'inférence

□ **Etape descriptive**: Une conclusion descriptive peut toujours s'exprimer par le fait que les fréquences observées $\boldsymbol{f}$ (ou des indices dérivés $g(\boldsymbol{f})$) possèdent une certaine propriété $\mathcal{P}$:

$$\text{Descriptif: } \quad \mathcal{P}(\boldsymbol{f}) \quad \text{est vrai}$$

□ **Etape inductive**: L'inférence statistique consiste à parvenir à une conclusion sur les fréquences $\boldsymbol{\theta}$ (ou les paramètres dérivés $g(\boldsymbol{\theta})$), sur la base des données observées $\boldsymbol{a}$.

De façon générale, la question de l'inférence peut se formuler par: Peut-on généraliser la propriété $\mathcal{P}$ à la population, *i.e.* aux paramètres $\boldsymbol{\theta}$?

$$\text{Inductif: } \quad \mathcal{P}(\boldsymbol{\theta}) \quad \text{est-il vrai?}$$

Méthodes d'inférence statistique

□ Méthodes fréquentistes

- Les plus communes: tests de signification, *e.g.* χ^2
- Limitées concernant les questions qu'elles autorisent: souvent H_0 d'absence de liaison, d'effet
- Pas toutes valides si effectifs petits
- Inférence dépend des données qui auraient pu être observées (stopping-rule)

□ Méthodes bayésiennes

- Moins connues en SHS, mais en plein essor
- Valides pour des petits effectifs
- Permettent de faire des inférences sur n'importe quel paramètre ou propriété globale, notamment sur taille des effets
- Conclusions plus naturelles en termes de probabilités sur les paramètres inconnus

Une probabilité
peut en cacher une autre!

□ **Fréquences observées**, les f

□ **Fréquences parentes** ou vraies, les θ

Ce sont les *paramètres inconnus* sur lesquels porte l'inférence

□ **Probabilités fréquentistes**

Celles mises en jeu dans l'inférence fréquentiste. S'interprètent comme des fréquences que l'on obtiendrait par des répliques (à l'infini) du jeu de données, dans les mêmes conditions de recueil.

□ **Probabilités bayésiennes**

Celles résultant de l'inférence bayésienne. Ce sont des *probabilités épistémiques*, *i.e.* qui décrivent un état d'incertitude et/ou de connaissance.

Quelle est la probabilité qu'il pleuve ce soir?

... qu'il ait plu dans la matinée?

Inférence bayésienne: généralités

Comment répond-on au problème de l'inférence dans le cadre bayésien?

□ **Principe général**: à partir des effectifs $\boldsymbol{a} = (a_1, \ldots, a_K)$, et d'une *distribution initiale* sur les paramètres $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$, on dérive une *distribution finale* sur ces mêmes paramètres.

$$\text{Prob}\,(\boldsymbol{\theta}) \quad \rightsquigarrow \quad \text{Prob}\,(\boldsymbol{\theta} \mid \boldsymbol{a})$$

Les distributions initiale et finale expriment de façon *probabiliste* l'état initial (avant les données) et l'état final (après les données) de connaissance sur $\boldsymbol{\theta}$.

□ **Réponse bayésienne** au problème de l'inférence se déduit de la distribution finale en calculant:

$$\text{Prob}\,(\,\mathcal{P}(\boldsymbol{\theta})\ \text{vraie} \mid \boldsymbol{a})$$

Inférence bayésienne sur des fréquences

□ Distributions de Dirichlet

Famille de distributions initiales privilégiée pour l'inférence sur des fréquences

Distribution initiale:

$$(\theta_1, \theta_2, \theta_3) \sim \mathit{Di}(\alpha_1, \alpha_2, \alpha_3)$$

Distribution finale:

$$(\theta_1, \theta_2, \theta_3)|\boldsymbol{a} \sim \mathit{Di}(\alpha_1 + a_1, \alpha_2 + a_2, \alpha_3 + a_3)$$

□ Forces initiales et finales

L'inférence bayésienne peut être décrite en termes de *forces*: chaque *force initiale*

$$\alpha_k$$

est actualisée en une *force finale*

$$\alpha_k + a_k$$

Force totale: $\nu = \sum \alpha_k \rightsquigarrow \nu + n$

Choix des forces initiales α

□ **Formaliser un état initial d'ignorance**
La distribution finale exprime l'information sur les paramètres apportée par les données:

- $\alpha_k = 0$, Haldane (1948)
- $\alpha_k = 1/K$, Perks (1947)
- $\alpha_k = 1/2$, Jeffreys (1961)
- $\alpha_k = 1$, Bayes-Laplace

Différences faibles (Bernard, 1996)

□ **Solution usuelle (Jeffreys)**: Adoptée ici

$$\alpha_k = 1/2$$

□ **Modèle Dirichlet imprécis (IDM)**
Considérer un *ensemble* de distribution de Dirichlet; conclusions en termes d'*intervalles de probabilité* (Bernard, 1996, 2005; Walley, 1996):

$$0 \leq \alpha_k \leq 1 \quad \text{et} \quad \nu = \sum \alpha_k = 1$$

Distribution finale

☐ **Distribution bayésienne finale**:

$$(\theta_1, \theta_2, \theta_3) \sim \mathit{Di}(10 + 1/2,\ 8 + 1/2,\ 6 + 1/2)$$

☐ **Le simplexe**: Représentation triangulaire

Les 3 sommets correspondent à chacune des 3 compositions en fréquences les plus extrêmes: $(1, 0, 0)$ (gauche), $(0, 1, 0)$ (bas), $(0, 0, 1)$ (droite).

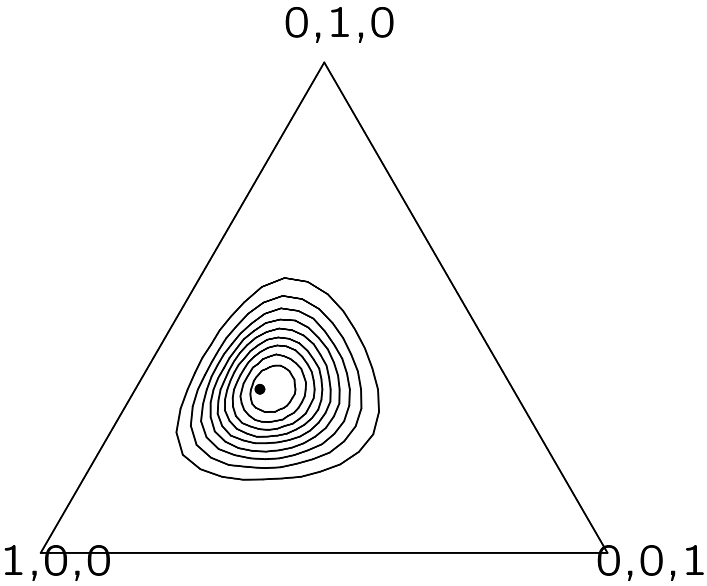
☐ **Caractéristiques de la distribution**

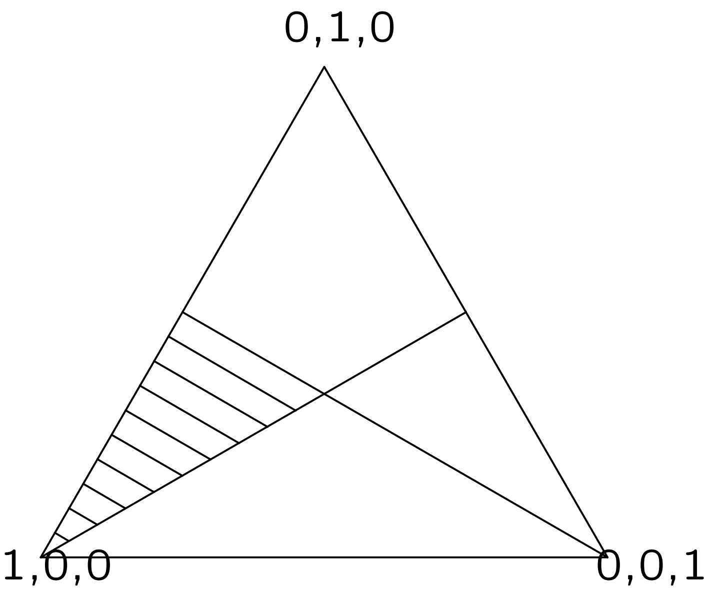
- Le centre de la distribution correspond (à peu près) à la

$$\boldsymbol{f} = (0.42, 0.33, 0.25)$$

- Sa dispersion reflète l'effectif total $n = 24$.

Les deux ingrédients





- **Conclusion inductive**:

$$\text{Prob}(\theta_1 > \theta_2 > \theta_3) \;=\; 0.423$$

3. TABLEAUX DE CONTINGENCE MESURES D'ASSOCIATION

Tableau de contingence: 2 exemples

□ **Données "Civilité":** Observation à l'entrée d'un magasin des individus qui tiennent la porte à la personne qui les suit, selon leur sexe (Moser, Corroyer, 2002)

		Porte?	
		Oui	Non
	Homme	222	144
Sexe	Femme	254	260

□ **Données "Cheveux-Yeux":** Couleur des yeux et Couleur des cheveux, sur $n = 5387$ individus (Fisher).

		Cheveux			
		Noir	Chat	Blond	Roux
	Noir	326	688	343	98
	MarrF	38	116	84	48
Yeux	MarrC	241	584	909	403
	Bleu	110	188	412	681
	Vert	3	4	26	85

Types de questions statistiques

□ Questions symétriques

- Liaison, association entre `J` et `K`
- *Attraction* entre modalités $j \in J$ et $k \in K$, ou *sur-représentation* de certaines cases jk
- *Répulsion* entre modalités j et k, ou *sous-représentation* de certaines cases jk

□ Questions asymétriques

- Problèmes de *comparaison*: La répartition selon `K` diffère t-elle selon les modalités de `J`.
  Y-a t-il un "effet" du `SEXE` sur `PORTE`?
- Y-a t-il des *implications* ou quasi-implications entre certaines modalités, $j \Longrightarrow k$?
  Par exemple: $Yeux : Verts \Longrightarrow Chev : Roux$?

Notations, Premières statistiques

- **Effectifs observés**: a_{jk} pour chaque case jk

- **Fréquences conjointes et marginales**

  - Fréquences conjointes f_{jk}:

$$f_{jk} \quad = \quad n_{jk}/n$$

  - Fréquences marginales f_j et f_k:

$$f_j = \sum_{k \in K} f_{jk} \qquad\qquad f_k = \sum_{j \in J} f_{jk}$$

- **Exemple: Civilité**

	Porte?		
	Oui	Non	Total
Homme	0.25	0.16	0.42
Femme	0.29	0.30	0.58
Total	0.54	0.46	1.000

Fréquences conditionnelles

□ Fréquences conditionnelles en ligne

$$f_k^j = \frac{f_{jk}}{f_j}$$

	Porte?		
	Oui	Non	Total
Homme	60.7%	39.3%	100%
Femme	49.4%	50.6%	100%
Moy	54.1%	45.9%	100%

□ Fréquences conditionnelles en colonne

$$f_j^k = \frac{f_{jk}}{f_k}$$

	Porte?		
	Oui	Non	Moy
Homme	46.6%	35.6%	41.6%
Femme	53.4%	64.4%	58.4%
Total	100.0%	100.0%	100.0%

Indépendance locale et globale

□ **Indépendance locale** (ou absence de liaison) entre deux modalités j et k, lorsque:

$$\begin{aligned} j \perp\!\!\!\perp k \quad \textit{ssi} \quad & f_j^k = f_j \\ \textit{ssi} \quad & f_k^j = f_k \\ \textit{ssi} \quad & f_{jk} = f_j f_k \end{aligned}$$

□ **Indépendance globale** (ou absence de liaison) entre J et K lorsque: les fréquences conditionnelles en ligne sont les mêmes pour chaque ligne, les fréquences conditionnelles en colonne sont les mêmes pour chaque colonne

$$\begin{aligned} J \perp\!\!\!\perp K \quad \textit{ssi} \quad & \forall j,k, \quad f_j^k = f_j \\ \textit{ssi} \quad & \forall j,k, \quad f_k^j = f_k \\ \textit{ssi} \quad & \forall a,b, \quad f_{jk} = f_j f_k \end{aligned}$$

i.e. lorsque chaque fréquence conjointe f_{jk} est égale à la fréquence-produit $f_j f_k$ correspondante.

□ **Rôle important des fréquences-produits**

$$\widehat{f_{jk}} = f_j f_k$$

Divers indices d'association

□ **Une panoplie d'indices** (Goodman, Kruskal, 1954, 1959)

□ **Indices locaux**

- Comparer une fréquence conjointe à une fréquence de référence, *i.e.* f_{jk} à $\widehat{f_{jk}}$
- Comparer une fréquence conditionnelle à une fréquence marginale, *i.e.* f_k^j à f_k
- Comparer plusieurs fréquences conditionnelles, *i.e.* f_k^j à $f_k^{j'}$
- Comparaison: par différence, rapport, accroissement relatif
- Odds-ratio, un rapport de deux rapports
- PEM (Cibois,2004)

□ **Indices globaux**

- Coefficient de contingence, ϕ ou ϕ^2
- V^2 de Cramer, *etc.*.

Un indice de liaison local

□ **Taux de liaison** entre modalités j et k (*e.g.* Rouanet *et al.*, 1987):

$$t_{jk} = \frac{f_{jk} - \widehat{f_{jk}}}{\widehat{f_{ab}}} \qquad \left[= \frac{O - T}{T}\right]$$

S'interprète comme un taux de sur- ou de sous-représentation par rapport à l'indépendance

	Porte? Oui	Non
Homme	0.12	-0.14
Femme	-0.09	0.10

□ **Sens de la liaison selon** t_{jk}

> 0	attraction, sur-représentation
$= 0$	indépendance locale
< 0	répulsion, sous-représentation
$= -1$	répulsion maximale, incompatibilité stricte

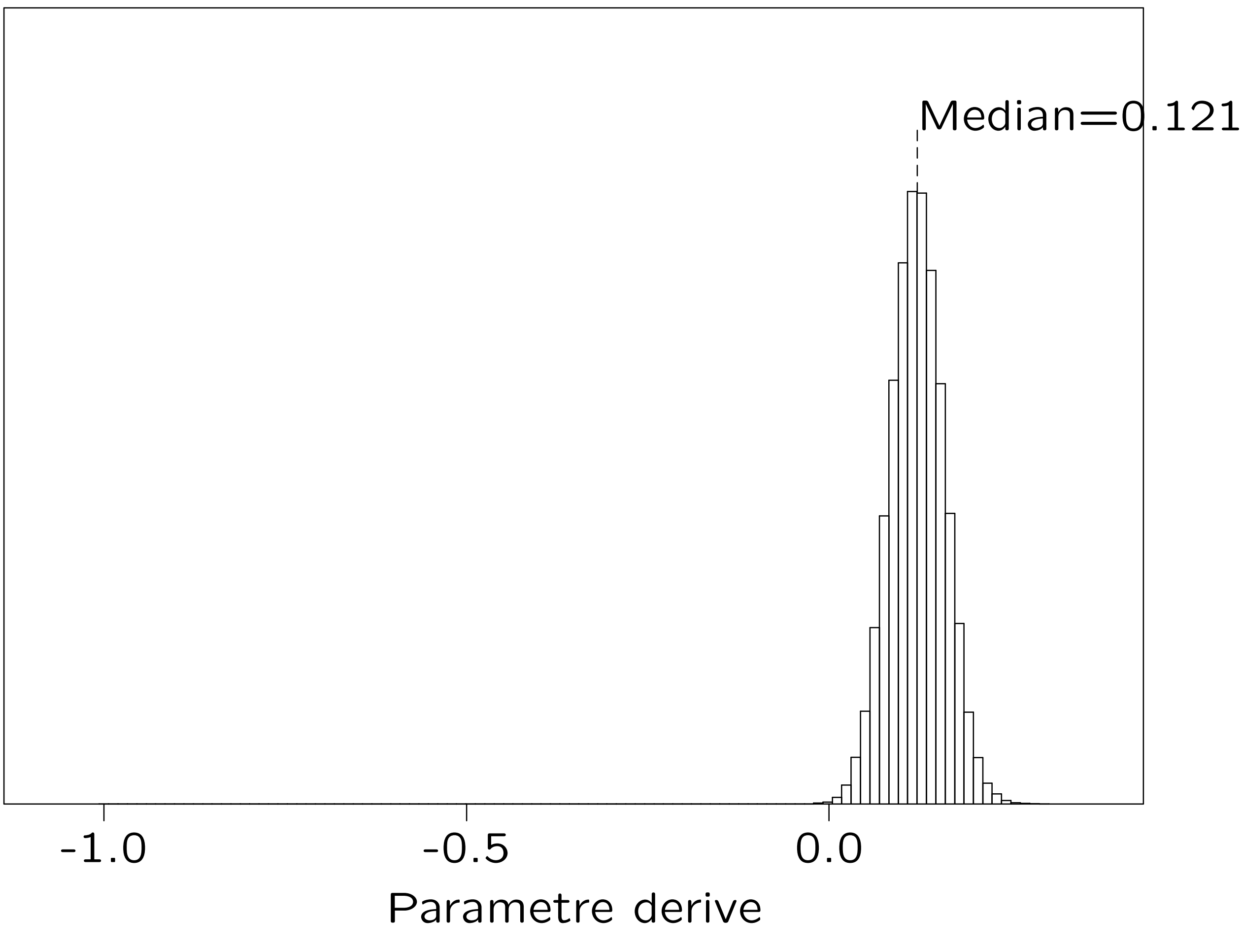
5. DOSSIER “CIVILITÉ”

Civilité: Taux liaison (Hom,Oui)

- **Descriptivement**: $t_{obs} = 0.12$

- **Inductivement**

$$\begin{aligned} \text{Prob}(t_{par} > 0.0) &= 0.9996 \\ \text{Prob}(t_{par} \in [0.05; 0.19]) &= 0.95 \end{aligned}$$

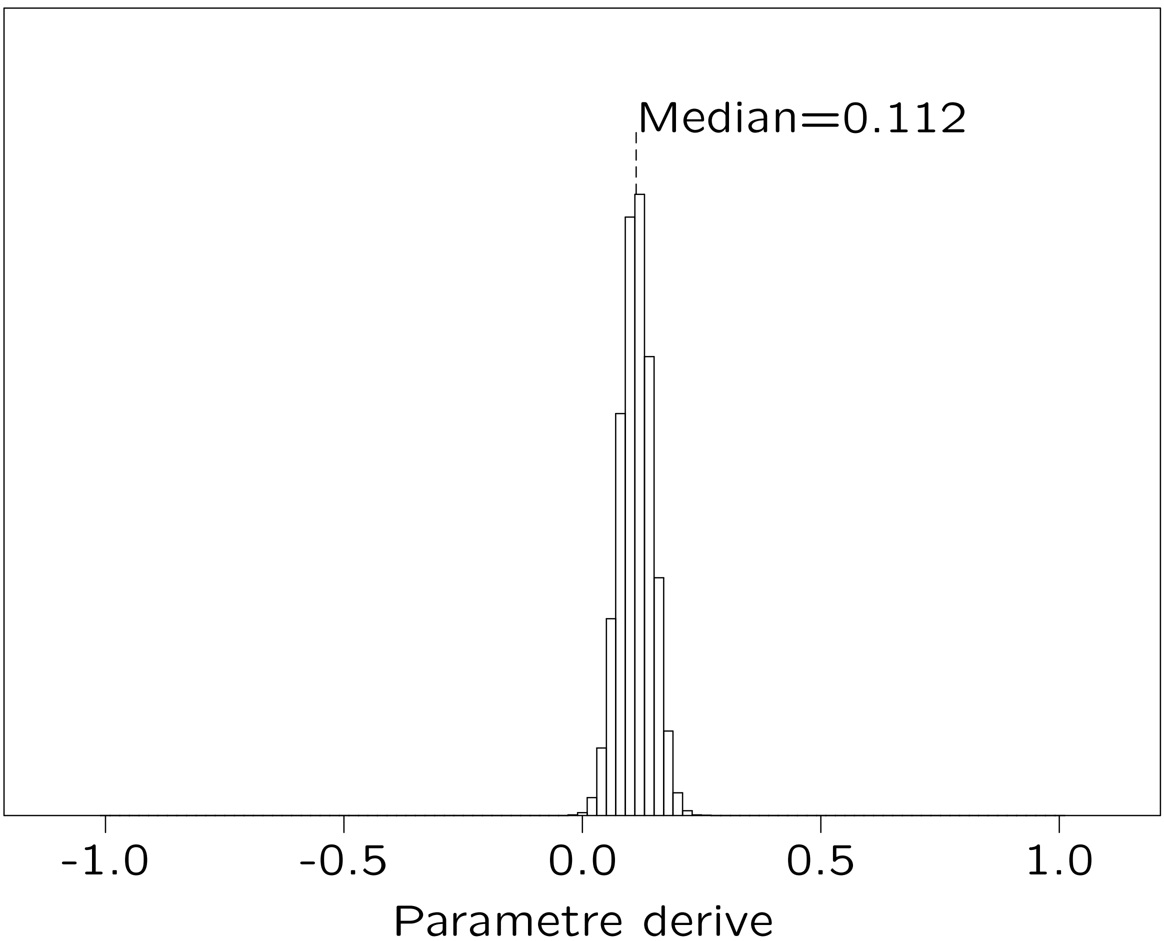


Civilité: Différence (Oui/Hom,Oui/Fem)

- **Descriptivement**: $\mathrm{DIF}_{obs} = 0.11$

- **Inductivement**

$$\begin{aligned} \mathrm{Prob}(\mathrm{DIF}_{par} \in [0.05; 0.18]) &= 0.95 \\ \mathrm{Prob}(-0.15 < \mathrm{DIF}_{par} < 0.15) &= 0.87 \end{aligned}$$

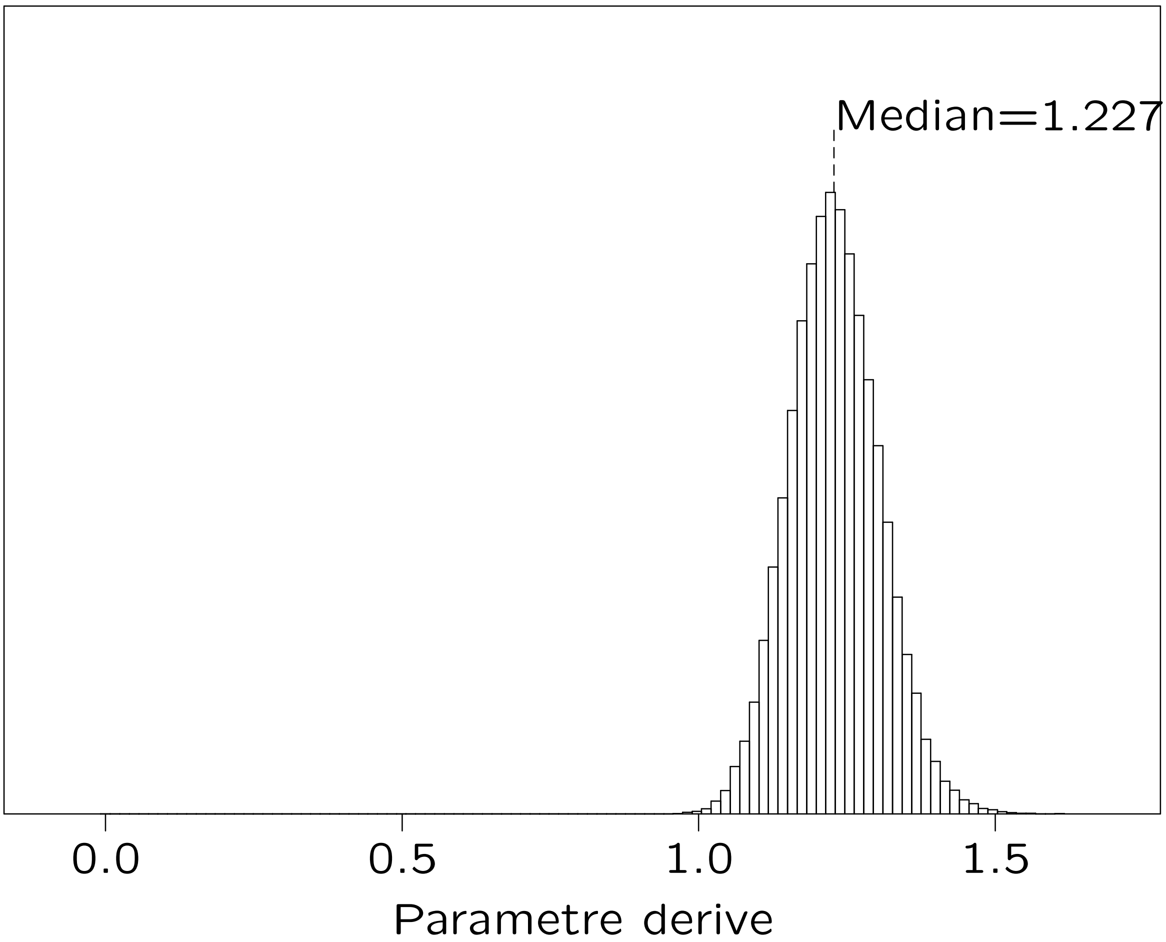


Civilité: Rapport (Oui/Hom,Oui/Fem)

- **Descriptivement**: RAP$_{obs}$ = 1.23

- **Inductivement**

$$\text{Prob}(\text{RAP}_{par} \in [1.10; 1.38]) \;=\; 0.95$$



Indices plus complexes

□ **Odds-ratio**

$$\text{OR}_{obs} = \frac{f_{jk}/f_{jk'}}{f_{j'k}/f_{j'k'}}$$

□ **PEM** (pourcentage de l'écart maximum)

$$\text{PEM}_{obs} = \mathit{max} \left\{ \frac{f_{jk} - \widehat{f_{jk}}}{f_j - \widehat{f_{jk}}}, \frac{f_{jk} - \widehat{f_{jk}}}{f_k - \widehat{f_{jk}}} \right\}$$

□ **Coefficient de contingence ϕ^2**

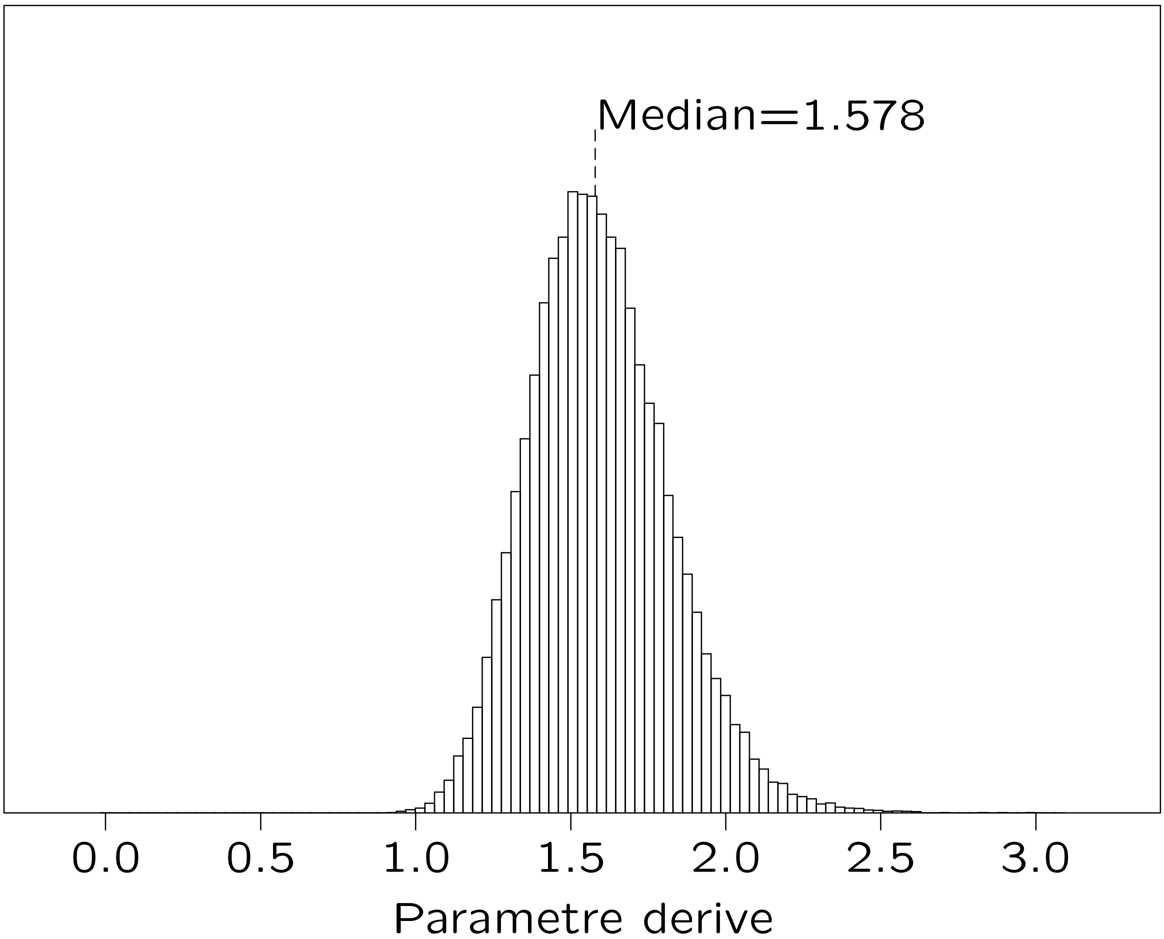
$$\phi^2_{obs} = \sum_{jk} \widehat{f_{jk}} \; t^2_{jk}$$

Civilité: Odds-ratio (Hom,Oui)

☐ **Descriptivement**: $\text{OR}_{obs} = 1.58$

☐ **Inductivement**

$$\begin{aligned} \text{Prob}(\text{OR}_{par} \in [1.21; 2.07]) &= 0.95 \\ \text{Prob}(\text{OR}_{par} \in [1.11; 2.27]) &= 0.99 \end{aligned}$$

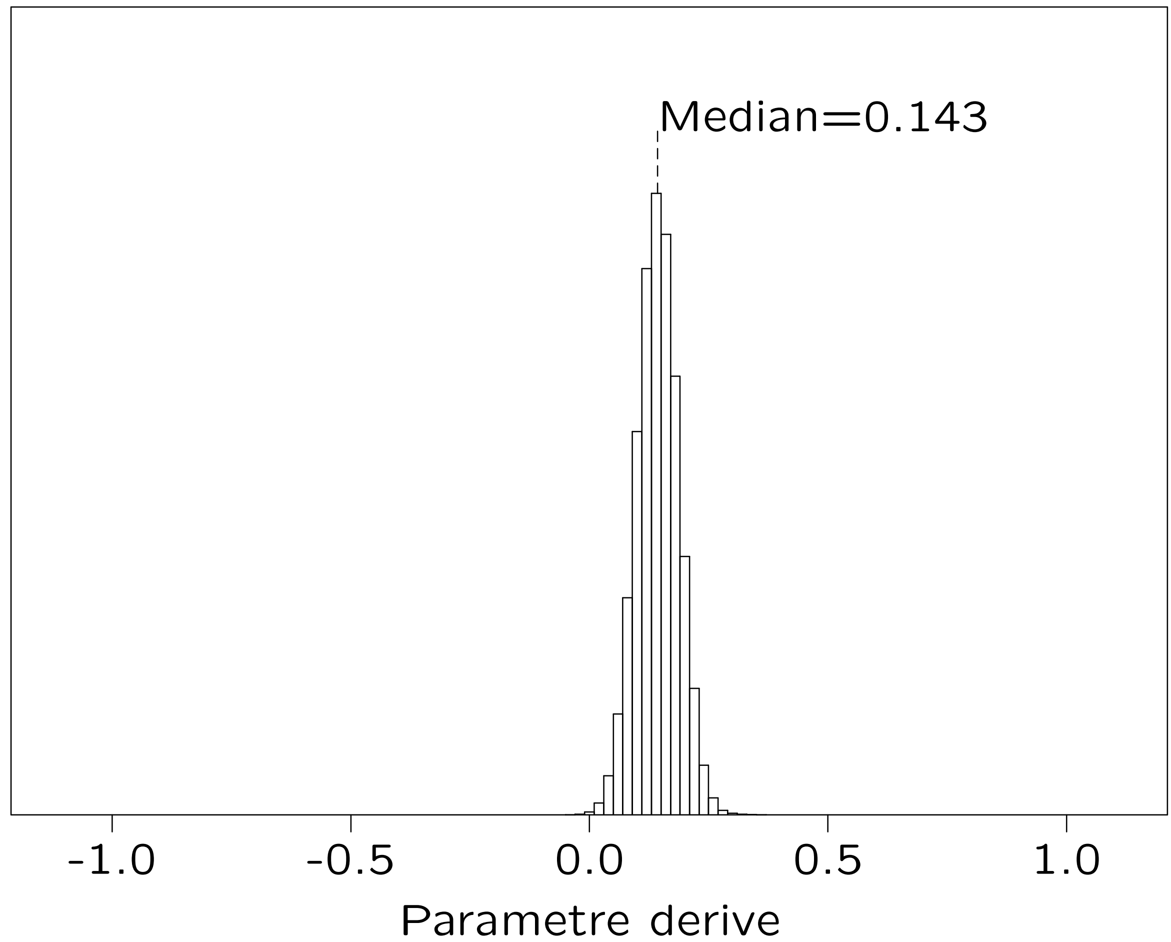


Civilité: PEM (Hom,Oui)

- **Descriptivement**: $\text{PEM}_{obs} = 0.14$

- **Inductivement**

$$\text{Prob}(\text{PEM}_{par} \in [0.06; 0.23]) \quad = \quad 0.95$$

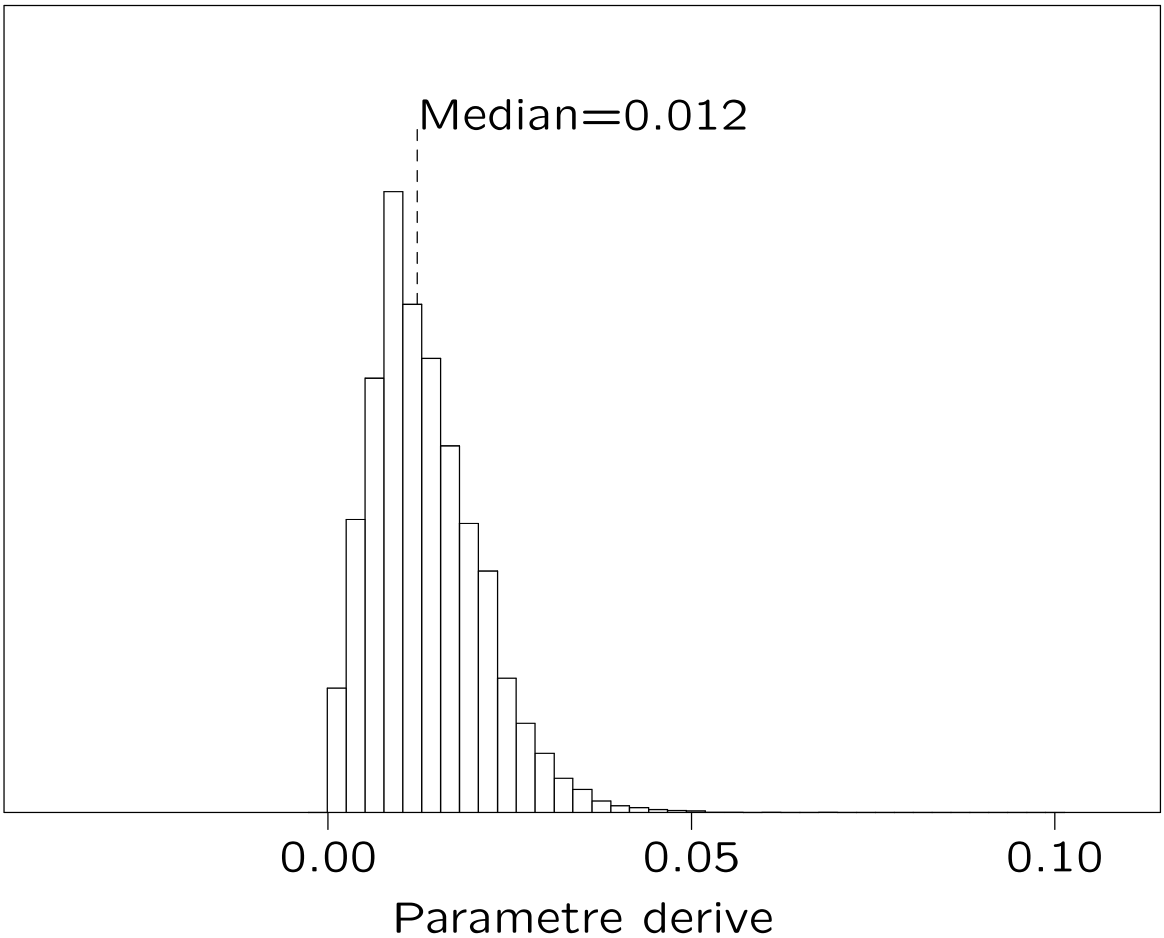


Civilité: Coefficient de contingence

- **Descriptivement**: $\phi^2_{obs} = 0.012$

- **Inductivement**

$$\text{Prob}(\phi^2_{par} \in [0.002; 0.031]) \;=\; 0.95$$



6. DOSSIER “CHEVEUX-YEUX”

Dossier "Cheveux-Yeux"

□ Données

Couleur des yeux et Couleur des cheveux, relevée sur $n = 5387$ individus (Fisher).

		Cheveux			
		Noir	Chat	Blond	Roux
Yeux	Noir	326	688	343	98
	MarrF	38	116	84	48
	MarrC	241	584	909	403
	Bleu	110	188	412	681
	Vert	3	4	26	85

□ Problème

Résumer les attractions et répulsions entre modalités lignes et modalités colonnes, de façon descriptive et inductive

Analyse descriptive

□ Tableau des taux de liaison

	Noir	Chat	Blond	Roux
Noir	**0.68**	**0.61**	-0.28	-0.72
MarrF	-0.00	**0.38**	-0.11	-0.31
MarrC	-0.15	-0.07	**0.29**	-0.23
Bleu	-0.41	-0.54	-0.10	**1.01**
Vert	-0.81	-0.88	-0.33	**1.95**

□ Attractions / répulsions

	Noir	Chat	Blond	Roux
Noir	$\oplus\oplus$	$\oplus\oplus$	$\odot$	$\odot\odot$
MarrF	$=$	$\oplus$	$\odot$	$\odot$
MarrC	$\odot$	$=$	$\oplus$	$\odot$
Bleu	$\odot$	$\odot\odot$	$\odot$	$\oplus\oplus$
Vert	$\odot\odot$	$\odot\odot$	$\odot$	$\oplus\oplus$

Analyse inductive: Attractions / Répulsions

□ **Probabilités bayésiennes**, Prob($\tau_{jk} > 0$)

	Noir	Chat	Blond	Roux
Noir	**1.00**	**1.00**	0.00	0.00
MarrF	0.50	**1.00**	0.09	0.00
MarrC	0.00	0.01	**1.00**	0.00
Bleu	0.00	0.00	0.00	**1.00**
Vert	0.00	0.00	0.00	**1.00**

□ **Attractions / répulsions** (garantie 0.95)

	Noir	Chat	Blond	Roux
Noir	$\oplus$	$\oplus$	$\odot$	$\odot$
MarrF	?	$\oplus$	?	$\odot$
MarrC	$\odot$	$\odot$	$\oplus$	$\odot$
Bleu	$\odot$	$\odot$	$\odot$	$\oplus$
Vert	$\odot$	$\odot$	$\odot$	$\oplus$

Analyse inductive: Sur-représentations élevées

- **Probabilités bayésiennes**, Prob($\tau_{jk} > 0.50$)

	Noir	Chat	Blond	Roux
Noir	**0.99**	**0.99**	0.00	0.00
MarrF	0.00	0.10	0.00	0.00
MarrC	0.00	0.00	0.00	0.00
Bleu	0.00	0.00	0.00	**1.00**
Vert	0.00	0.00	0.00	**1.00**

- **Sur-représentations** > 0.50 (garantie 0.95)

	Noir	Chat	Blond	Roux
Noir	$\oplus$	$\oplus$	$\circ$	$\circ$
MarrF	$\circ$	?	$\circ$	$\circ$
MarrC	$\circ$	$\circ$	$\circ$	$\circ$
Bleu	$\circ$	$\circ$	$\circ$	$\oplus$
Vert	$\circ$	$\circ$	$\circ$	$\oplus$

Et avec 10 fois moins de données?

- **Sur-représentations** > 0.50, *Bayésien*

	Noir	Chat	Blond	Roux
Noir	?	?	∘	∘
MarrF	?	?	∘	∘
MarrC	∘	∘	∘	∘
Bleu	∘	∘	∘	⊕
Vert	∘	∘	∘	⊕

- **Sur-représentations** > 0.50, *IDM*

	Noir	Chat	Blond	Roux
Noir	?	?	∘	∘
MarrF	∘	?	∘	∘
MarrC	∘	∘	∘	∘
Bleu	∘	∘	∘	⊕
Vert	?	∘	?	⊕

De l'association à l'implication

□ **Jusqu'ici**:

- Intérêt: Ecarts à l'indépendance
- Point de référence implicite: indépendance, *i.e.* $t_{jk} = 0$

□ **Maintenant**:

- Intérêt: Implications ou Quasi-implications
- Point de référence: sous-représentation maximale, *i.e.* $t_{jk} = -1$
- Si $t_{jk} = -1$, alors $j \Longrightarrow$ non-k *(implication stricte)*
- Si $t_{jk} \approx -1$, alors $j \longrightarrow$ non-k *(quasi-implication)*

□ **Idée: confronter les données à un modèle logique**

7. DOSSIER “FRACTIONS”

Données "Fractions"

□ Données

Extrait d'une expérience de Psychologie Développementale sur la maitrise des fractions chez l'enfant ($n = 165$ sujets).

Deux épreuves J et K consistent à déterminer une quantité à l'aide d'une fraction qui exprime: un rapport partie-partie (épreuve J), un rapport partie-tout (épreuve K):

		K: k	K: k'
J	j	36	3
	j'	36	90

□ Analyse descriptive à vue

- En général, la réussite à J implique la réussite à K (36/39). Il y a quasi-implication: $j \longrightarrow k$.
- La réciproque $k \longrightarrow j$ est loin d'être vérifiée.

Analyse descriptive

□ **Problème.** Peut-on conclure à une *quasi-implication*, $j \longrightarrow k$?

□ **Indice descriptif.** L'indice "*Del*" fournit une mesure de l'écart des fréquences observées à un modèle logique quelconque (une ou plusieurs *cases d'erreur*) (Loevinger, 1947; Hildebrand *et al.* 1977; Bernard, Charron, 1996).

Pour le modèle "$j \Longrightarrow k$" dans un tableau 2×2, l'indice *Del* vaut

$$d_{j \Longrightarrow k} = 1 - \frac{f_{jk'}}{f_j f_{k'}}$$

soit $1 - t_{jk'}$

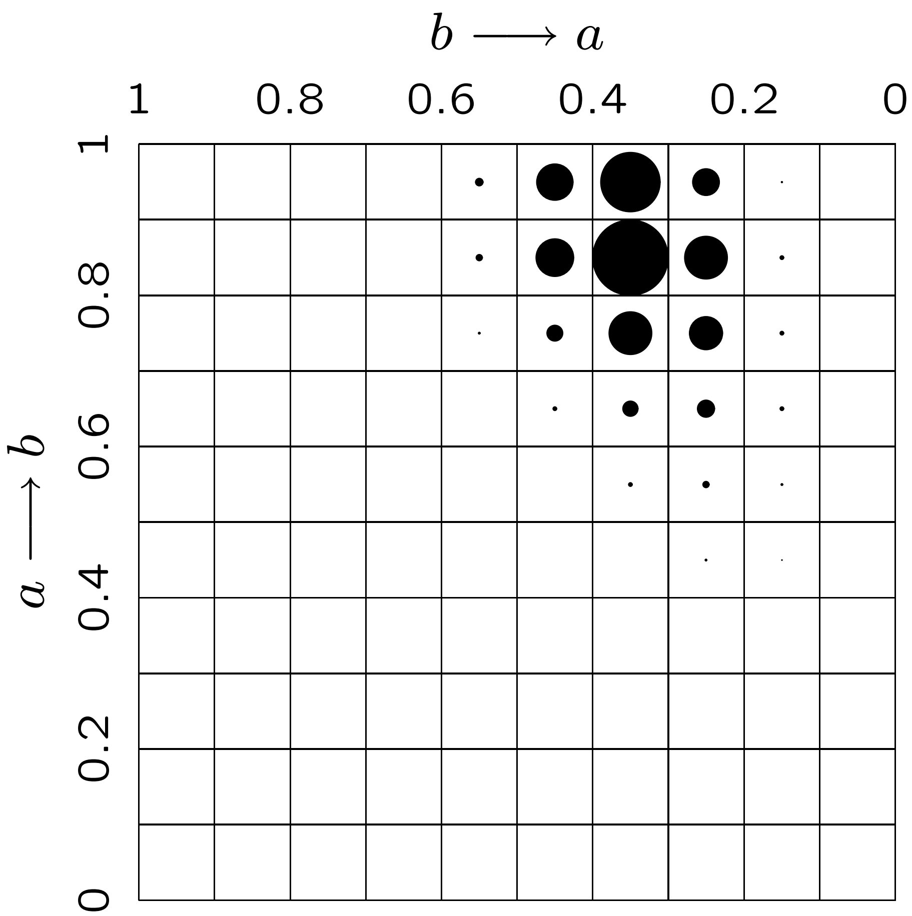
Indice *Del* varie de 0 (indépendance) à 1 (modèle logique strictement vérifié).

□ **Pour les données Fractions**

$$d_{a \Longrightarrow b} = 0.86$$
$$d_{b \Longrightarrow a} = 0.35$$

Analyse inductive

□ **Distribution bayésienne** conjointe des deux indices *Del* parents: $\delta_{j \Longrightarrow k}$ et $\delta_{j \Longrightarrow k}$.



□ **Résumé inductif**

$$Prob^{\star}(\delta_{a \Longrightarrow b} > 0.70 \ \& \ \delta_{b \Longrightarrow a} < 0.50) = 0.96$$

Autre exemple, “Stades” modèle logique

☐ **Les données:** Recherche sur les stades de Piaget (Jamison, 1977): $n = 101$ enfants soumis à 2 épreuves, avec 3 degrés de réussite chacune:

		Inclusion		
		$k1$	$k2$	$k3$
Sériation	$j1$	14	0	0
	$j2$	15	5	2
	$j3$	19	20	26

☐ **Hypothèse d'intérêt** peut s'exprimer par soit comme un modèle logique

$$\mathcal{M} \quad = \quad (k3 \Longrightarrow j3 \ \text{ et } \ k2 \Longrightarrow (j2 \text{ ou } j3))$$

soit comme un ensemble de cases “d'erreur”

☐ **Analyse statistique** (Danis *et al.*, 2000)

- Indice *Del* moyen pour l'ensemble des cases d'erreur
- Propriété globale à partir des indices *Del* pour chaque case d'erreur

8. DOSSIER “RELIGION”

Dossier “Religion”
(Degenne, Lebeaux, 1996)

A	Prier (Oui=a / Non=a')
B	Aller à l’église (id.)
C	Croire au paradis (id.)
D	Education religieuse aux enfants (id.)

Profil				Effectif
A	B	C	D	
a	b	c	d	100
a	b	c	d'	2
a	b	c'	d	14
a	b	c'	d'	2
a	b'	c	d	9
a	b'	c	d'	0
a	b'	c'	d	17
a	b'	c'	d'	**0**
a'	b	c	d	302
a'	b	c	d'	7
a'	b	c'	d	89
a'	b	c'	d'	15
a'	b'	c	d	172
a'	b'	c	d'	16
a'	b'	c'	d	**455**
a'	b'	c'	d'	324
				$n = 1524$

Indice implicatif multivarié

□ **Indice implicatif multivarié** (Bernard, 2002)

Pour chaque profil $p = ijkl$, mesure le degré d'accord avec le modèle $(p \Longrightarrow \emptyset)$ (p est absent):

$$d_{p \Longrightarrow \emptyset} = 1 - \frac{f_{ijkl}}{f_i f_j f_k f_l}$$

□ **Généralise** l'indice "*Del*" et le taux de liaison

□ **Interprétation**

- $d_{p \Longrightarrow \emptyset} = 0$: indépendance locale
- $d_{p \Longrightarrow \emptyset} = 1$; profil absent, donc implication stricte
- Indice négatif si profil sur-représenté

Des effectifs à l'indice $d_{p \Longrightarrow \emptyset}$

□ Tableau des effectifs

	c		c'	
	d	d'	d	d'
ab	100	2	14	2
ab'	9	0	17	0
$a'b$	302	7	89	15
$a'b'$	172	16	455	324

□ Tableau des $d_{p \Longrightarrow \emptyset}$

	c		c'	
	d	d'	d	d'
ab	*-5.58*	**0.58**	**0.39**	**0.72**
ab'	**0.68**	**1.00**	**0.60**	**1.00**
$a'b$	*-1.07*	**0.85**	**0.59**	**0.78**
$a'b'$	**0.37**	**0.81**	*-0.11*	*-1.50*

Quasi-implications (descriptif)
seuil $d_{quasi} = 0.50$

- **Tableau des $d_{p \Longrightarrow \emptyset}$** (en gras si > 0.50)

	c		c'	
	d	d'	d	d'
ab	*-5.58*	**0.58**	0.39	**0.72**
ab'	**0.68**	**1.00**	**0.60**	**1.00**
$a'b$	*-1.07*	**0.85**	**0.59**	**0.78**
$a'b'$	0.37	**0.81**	*-0.11*	*-1.50*

- **Résumé descriptif**

Prier(a) $\longrightarrow$ Eglise(b) $\longrightarrow$ Educ.(d)

Paradis(c) $\longrightarrow$ Educ.(d)

Eglise(b) $\longrightarrow$ (Prier(a) OU Paradis(c))

Quasi-implications (inductif)
seuil $d_{quasi} = 0.50$

☐ **Table des** $Prob(\delta_{p \Longrightarrow \emptyset} \geq 0.50)$ *(IDM)*

	c		c'	
	d	d'	d	d'
ab	0.00	0.43	0.18	0.71
ab'	**0.91**	**0.99**	0.82	**1.00**
$a'b$	0.00	**1.00**	**0.99**	**1.00**
$a'b'$	0.00	**1.00**	0.00	0.00

☐ **Résumé inductif**, à la garantie $\gamma = 0.90$,

Prier(a) ET Paradis(c) $\longrightarrow$ Eglise(b)

Prier(a) $\longrightarrow$ (Eglise(b) OU Educ.(d))

Paradis(c) $\longrightarrow$ (Prier(a) OU Educ.(d))

Eglise(b) $\longrightarrow$ (Prier(a) OU Paradis(c))