

Campylobacter / poulet

Réseaux bayésiens

Approche statistique bayésienne

Jean-Baptiste Denis (Mia / MiaJ à Jouy)
Isabelle Albert (Mia / Mét@risk à Paris)

ABARI – 14 octobre 2005

contexte

- Groupe de travail Afssa dans le cadre du *CES* microbiologie
- Groupe *campylo* autour de la thèse d'E. Grenier
 - Isabelle Albert,
 - Jean-Baptiste Denis,
 - Emmanuel Grenier (ISAB),
 - Judith Rousseau (Paris-Dauphine).

PLAN

1. *Campylobacter*, campylobactérioses

2. Réseaux bayésiens

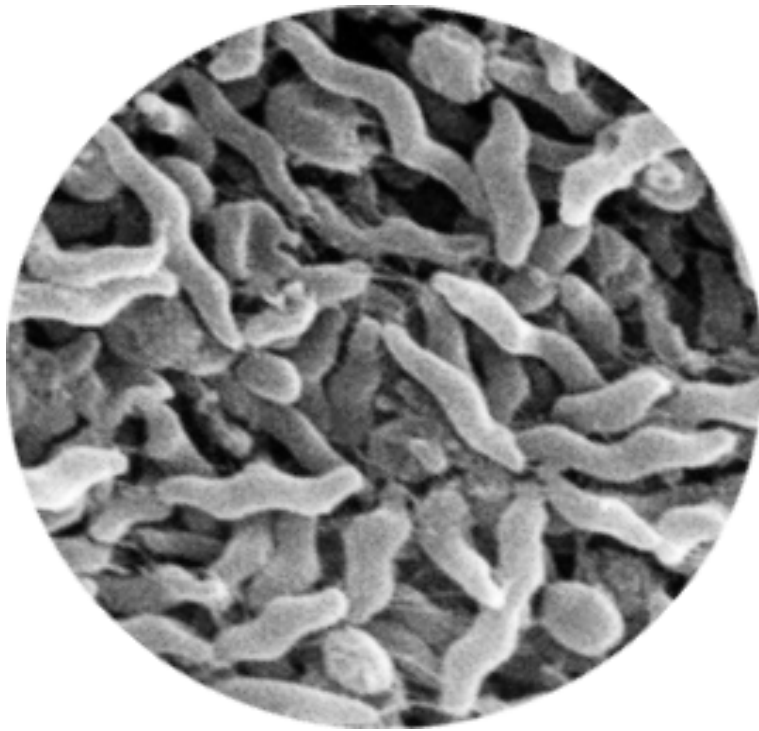
- Généralités
- Modélisation pour la campylobactériose

3. Statistique bayésienne

- Généralités
- Traitement de la chaîne modélisée

4. Logiciels

(1) *Campylobacter* / campylobactérioses



Contamination par campylobacter

- De nombreuses origines mais poulet cru
- Contamination croisée (ou barbecue)



- Quelle incidence sur la population française ?

Le pb campylobacter

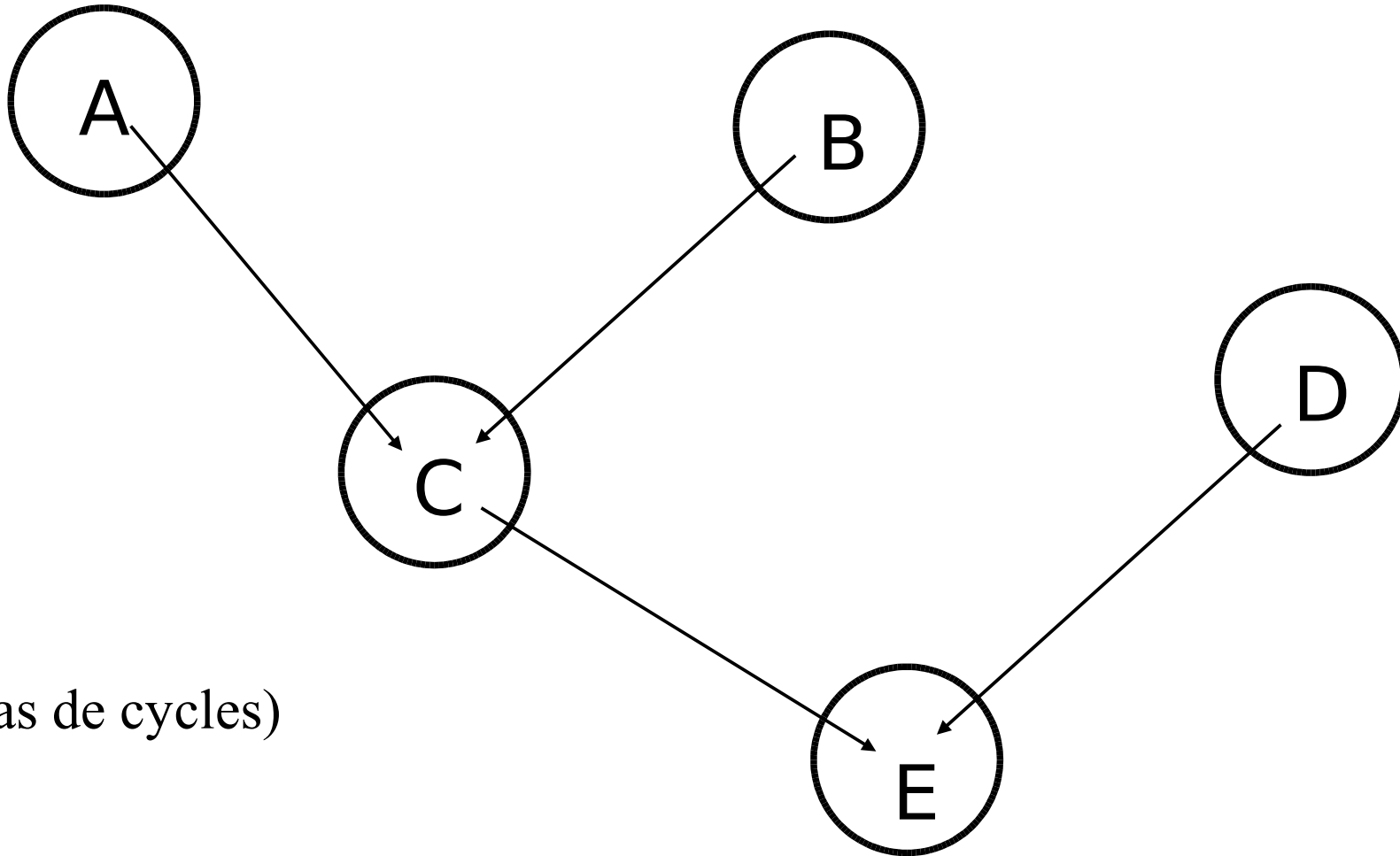
- Cité comme 1ère cause d'entérite bactérienne (2 à 8 millions aux USA, avec ~ 800 morts par an),
- Très mal identifié au niveau épidémiologique,
- Incapacité d'éradiquer le danger au niveau de la production,
- Enjeux économiques majeurs (réglementation, protection sanitaire).

quelle approche de modélisation ?

- peu ou pas de données, hétérogènes.
- phénomènes rares, mal connus variables.
- beaucoup d'interlocuteurs.

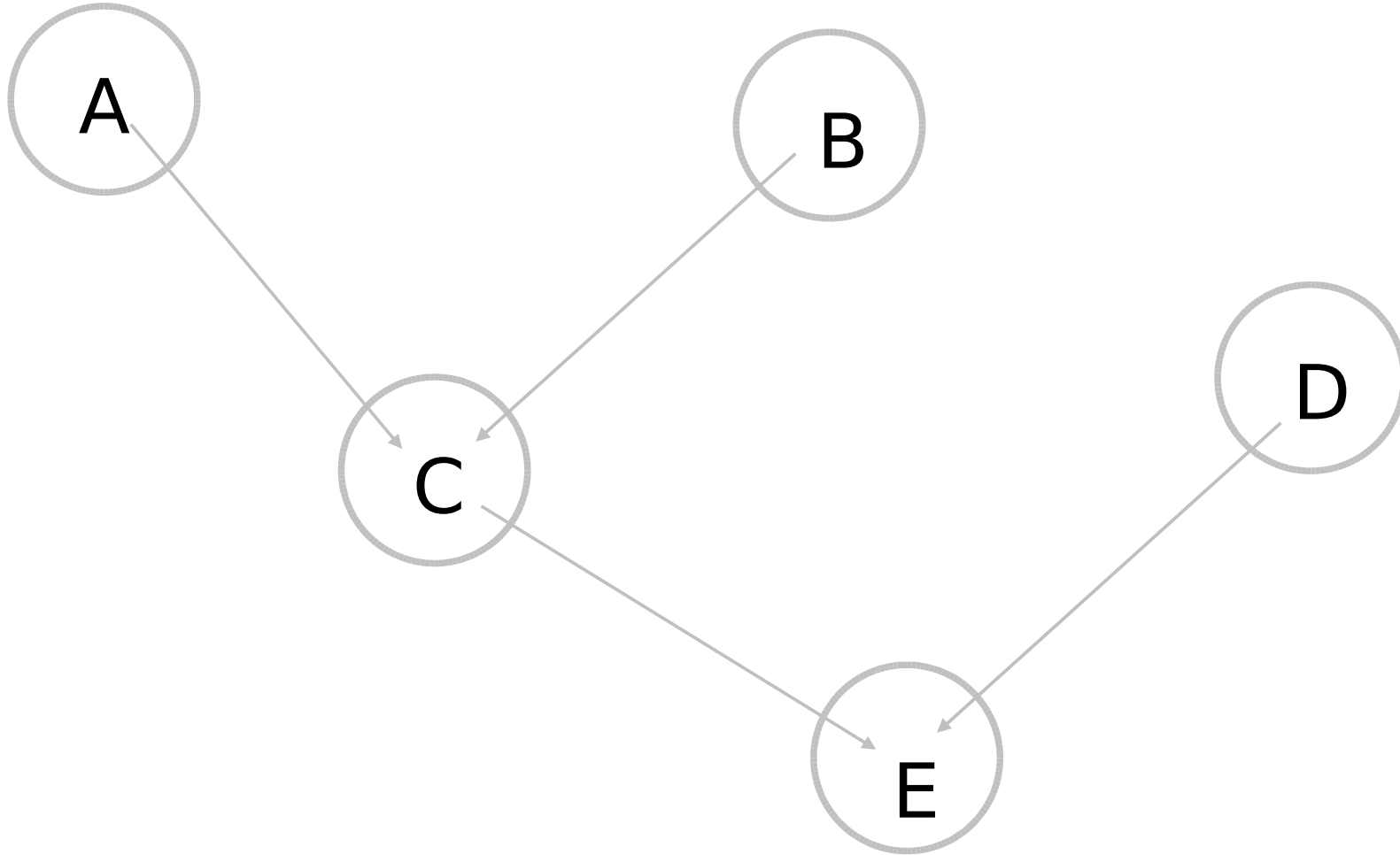
**modélisation probabiliste, globale
et doublement bayésienne !**

(2) Réseau bayésien



(pas de cycles)

$$[A,B,C,D,E] = [A] \cdot [B] \cdot [C|A,B] \cdot [D] \cdot [E|C,D]$$

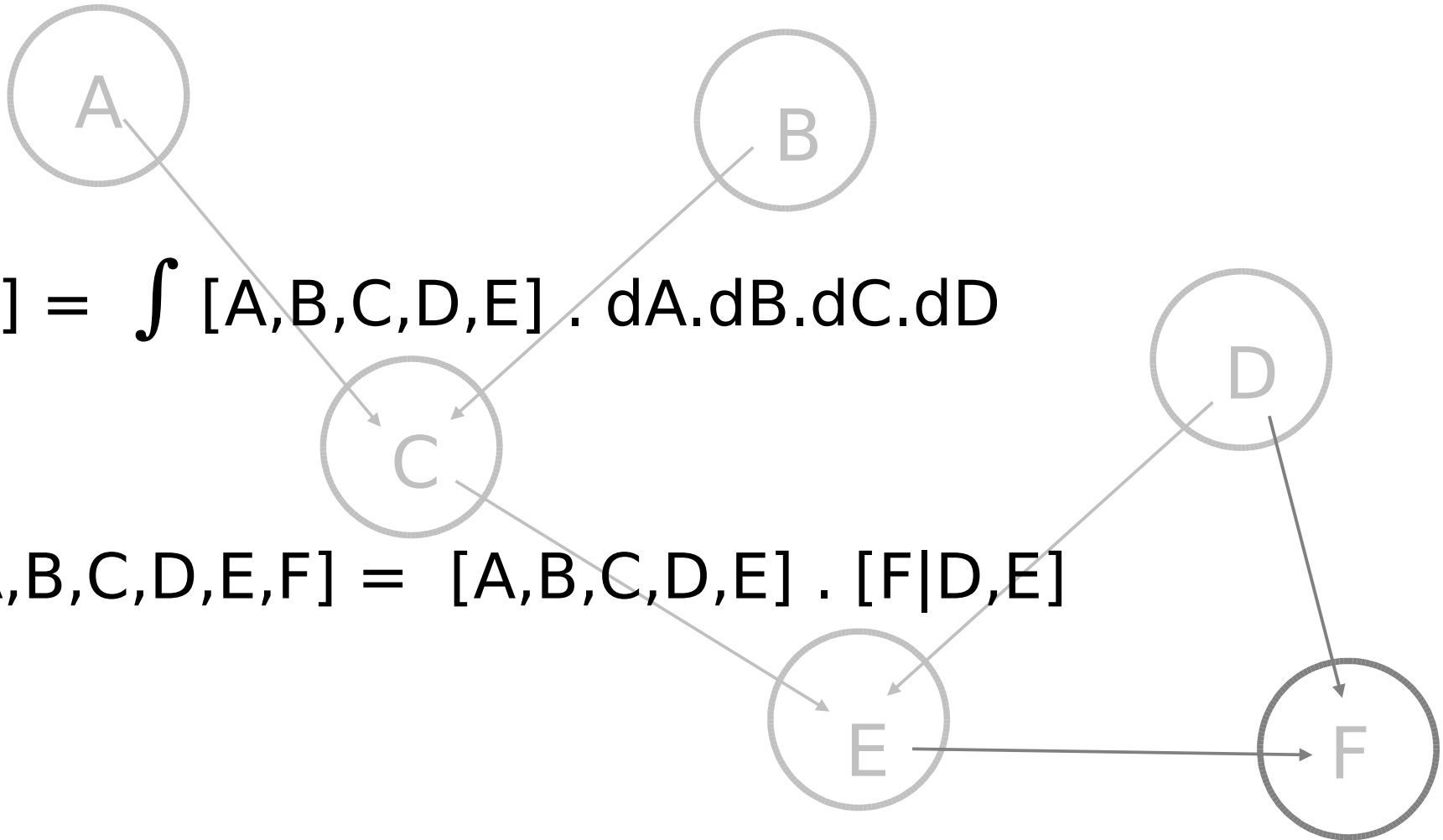


$$[E|A,B,C,D] = [E|C,D]$$

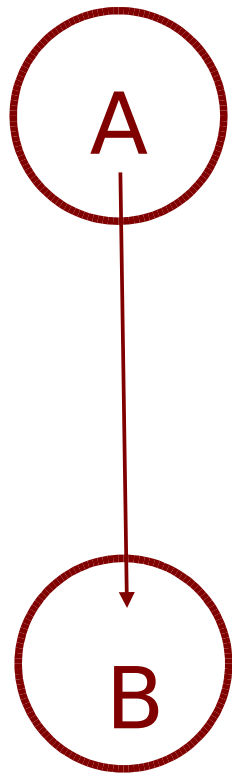
$$[E|A,B,D] = \int [E|C,D] \cdot [C|A,B] \cdot dC$$

$$[E] = \int [A,B,C,D,E] \cdot dA \cdot dB \cdot dC \cdot dD$$

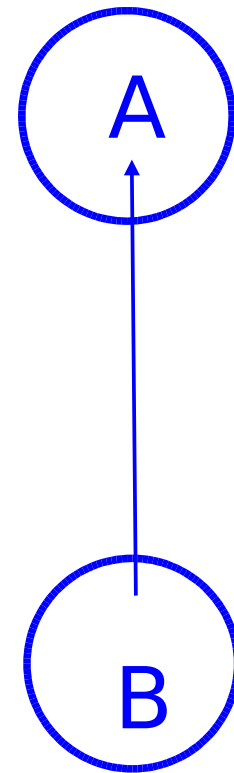
$$[A,B,C,D,E,F] = [A,B,C,D,E] \cdot [F|D,E]$$



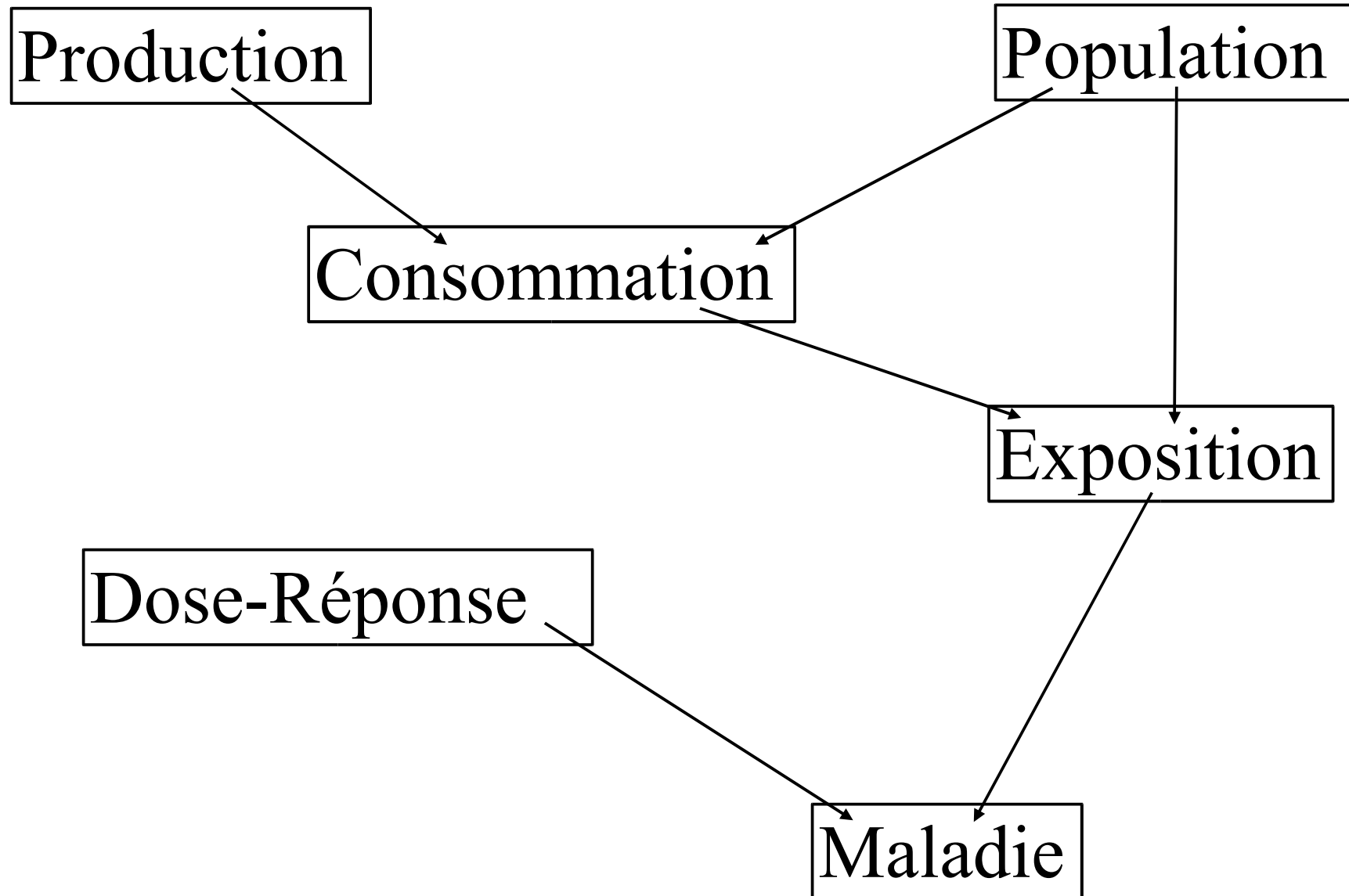
FLÈCHES \neq CAUSE

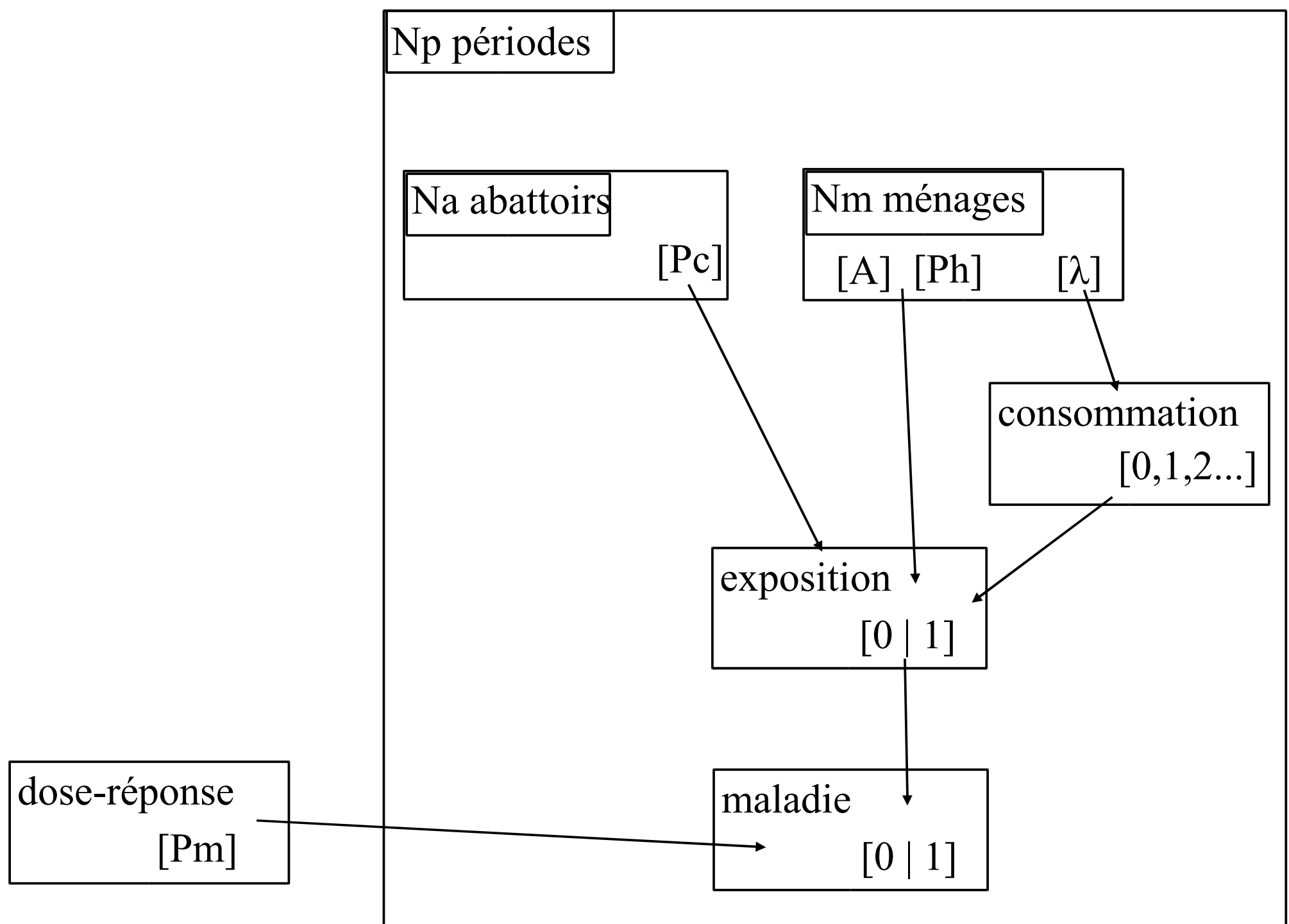


$$\begin{aligned} [A].[B|A] \\ = \\ [A,B] \\ = \\ [B].[A|B] \end{aligned}$$



Modélisation pour la campylobactériose





Exemples de distribution

$\text{logit}(Pc(a,p)) \sim N(\mu_a, \sigma)$ indépendantes

$\text{expo}(m,p) \sim \text{Binomiale}(\text{conso}(m,p), Pc(A(m),p)Ph(m))$ indépendantes

Des parties non immédiates

— *raccrocher les enquêtes épidémiologiques* —

	Malade	Non	Total
exposé Poulet	P_{PM}	P_{PN}	P_{P+}
exposé Autres causes	P_{AM}	P_{AN}	P_{A+}
Jamais exposé	P_{JM}	P_{JN}	P_{J+}
Total	P_{+M}	P_{+N}	$P_{++=1}$

modélisé avec trois paramètres :

- u : probabilité d'être exposé au poulet
- v : probabilité d'être malade si exposé
- w : probabilité d'être contaminé si pas par le poulet

	Malade	Non	Total
exposé Poulet	$v.u$	-	u
exposé Autres causes	$v.w(1-u)$	-	$w(1-u)$
Jamais exposé	0	-	-
Total	-	-	1

(3) Approche statistique bayésienne

$[\Theta]$: la modélisation (ou *prior*)

$[Y | \Theta]$: la vraisemblance

$[\Theta | Y]$: la *postérieure*

Schéma binomial : r succès pour n essais

$$[\Theta] : p \sim \text{Bêta}(\alpha, \beta)$$

$$[Y | \Theta] : r | p \sim \text{Bino}(n, p)$$

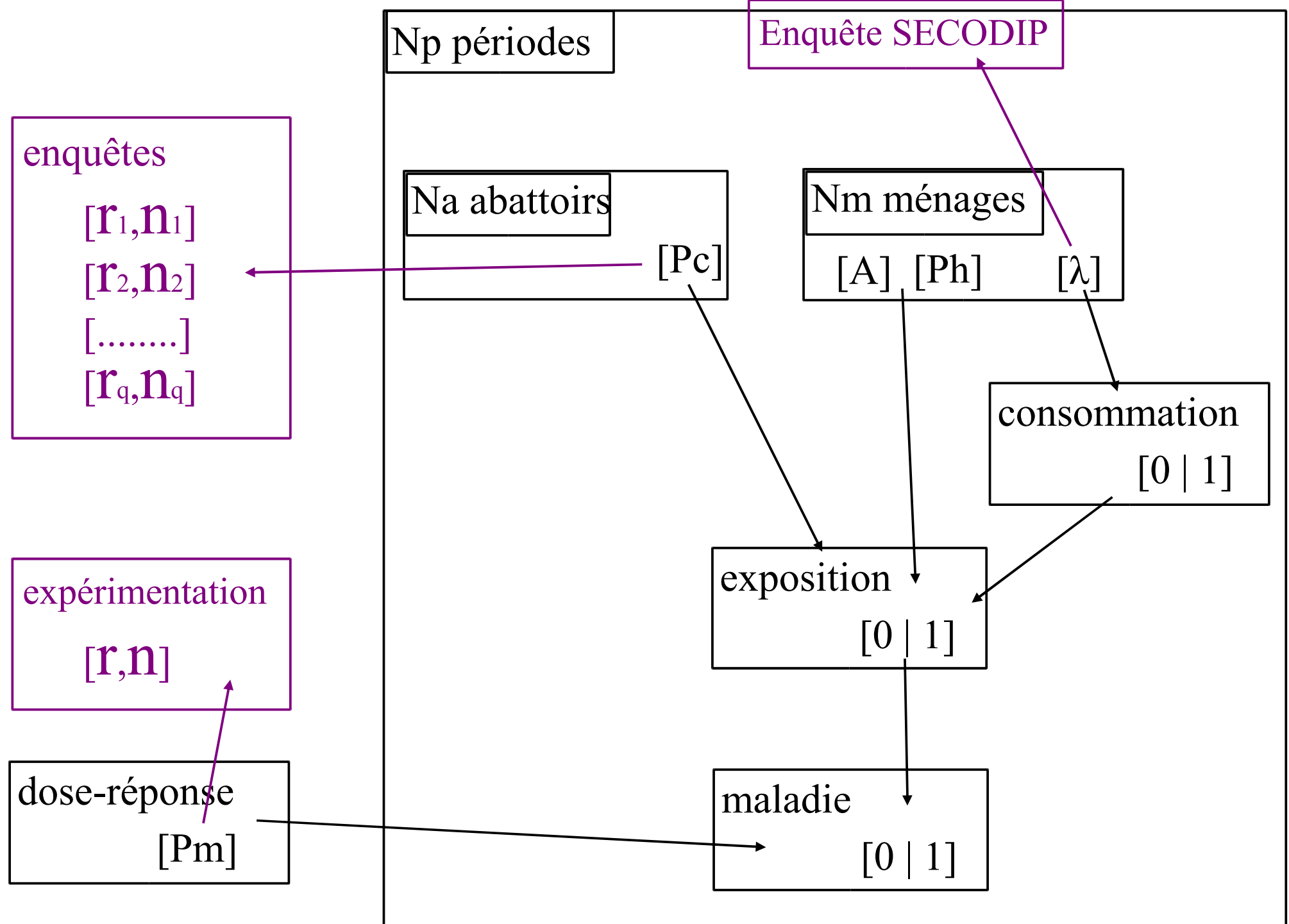
$$[\Theta | Y] : p | r \sim \text{Bêta}(\alpha + r, \beta + (n - r))$$

Traitement de la chaîne modélisée

La démarche est simple :

- Recenser toutes les données pertinentes,
- Les raccrocher à la modélisation,
- Calculer la postérieure des paramètres.

(dire d'experts : en priore ou en données ?)



(4) Logiciels



Open-Bugs

Paquets R

Programmation
Spécifique

En guise de conclusion

- approche globale lourde mais payante,
- dans le détail tout est compliqué,
- réseaux bayésiens
 - outil de dialogue,
 - efficacité dans la modélisation
- statistique bayésienne
 - se concentrer sur le phénomène, pas sur les données
 - logique d'apprentissage
- des problèmes nouveaux sur le bord de la route,