

REBASTABA : paquet **R** pour la manipulation de

Réseaux Bayésiens

en vue

d'une **Statistique Bayésienne**



Jean-Baptiste Denis & Isabelle Albert

*Mathématiques et Informatique Appliquées
Institut National de Recherche Agronomique*



PLAN

(1) Contexte

- ♦ statistique bayésienne (SB)
- ♦ réseaux bayésiens (RB)
- ♦ SB versus RB

(2) Rebastaba

- ♦ objectifs
- ♦ aperçus
- ♦ idées

(1) Statistique Bayésienne

θ : paramètres d'un modèle

D : des observations disponibles

(1) Statistique Bayésienne

θ : paramètres d'un modèle

$[\theta]$: prior

D : des observations disponibles

$[D | \theta]$: vraisemblance

(1) Statistique Bayésienne

θ : paramètres d'un modèle

$[\theta]$: prior

D : des observations disponibles

$[D | \theta]$: vraisemblance

postérieure : $[\theta | D] = K [\theta] [D | \theta]$

(1) Statistique Bayésienne

θ : paramètres d'un modèle

$[\theta]$: priore

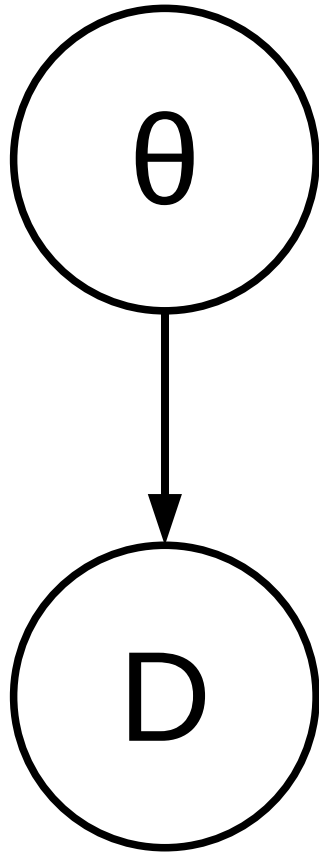
D : des observations disponibles

$[D | \theta]$: vraisemblance

postérieure : $[\theta | D] = K [\theta] [D | \theta]$

mise à jour de connaissances

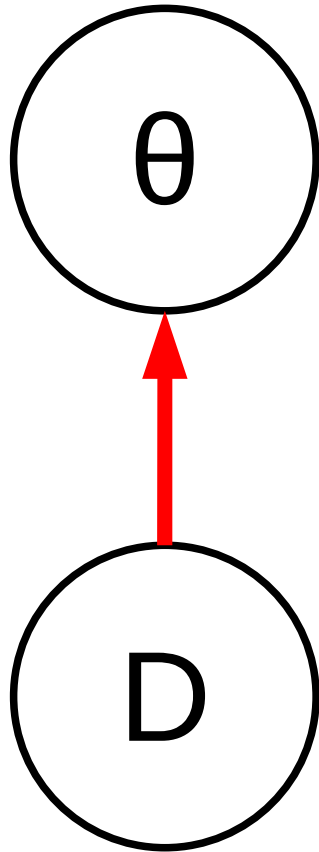
(1) Statistique Bayésienne



$[\theta]$: priore

$[D \mid \theta]$: vraisemblance

(1) Statistique Bayésienne



$[\theta]$: priore

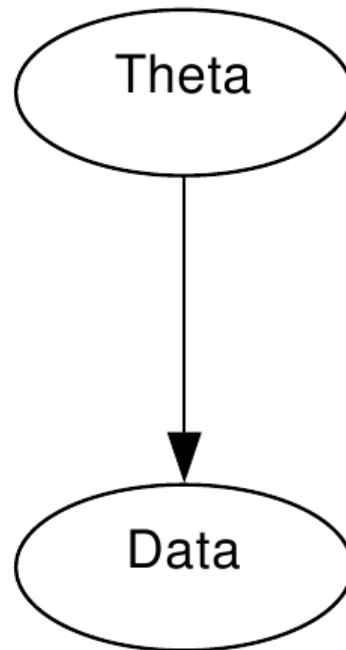
$[D | \theta]$: vraisemblance

postérieure : $[\theta | D] = K [\theta] [D | \theta]$

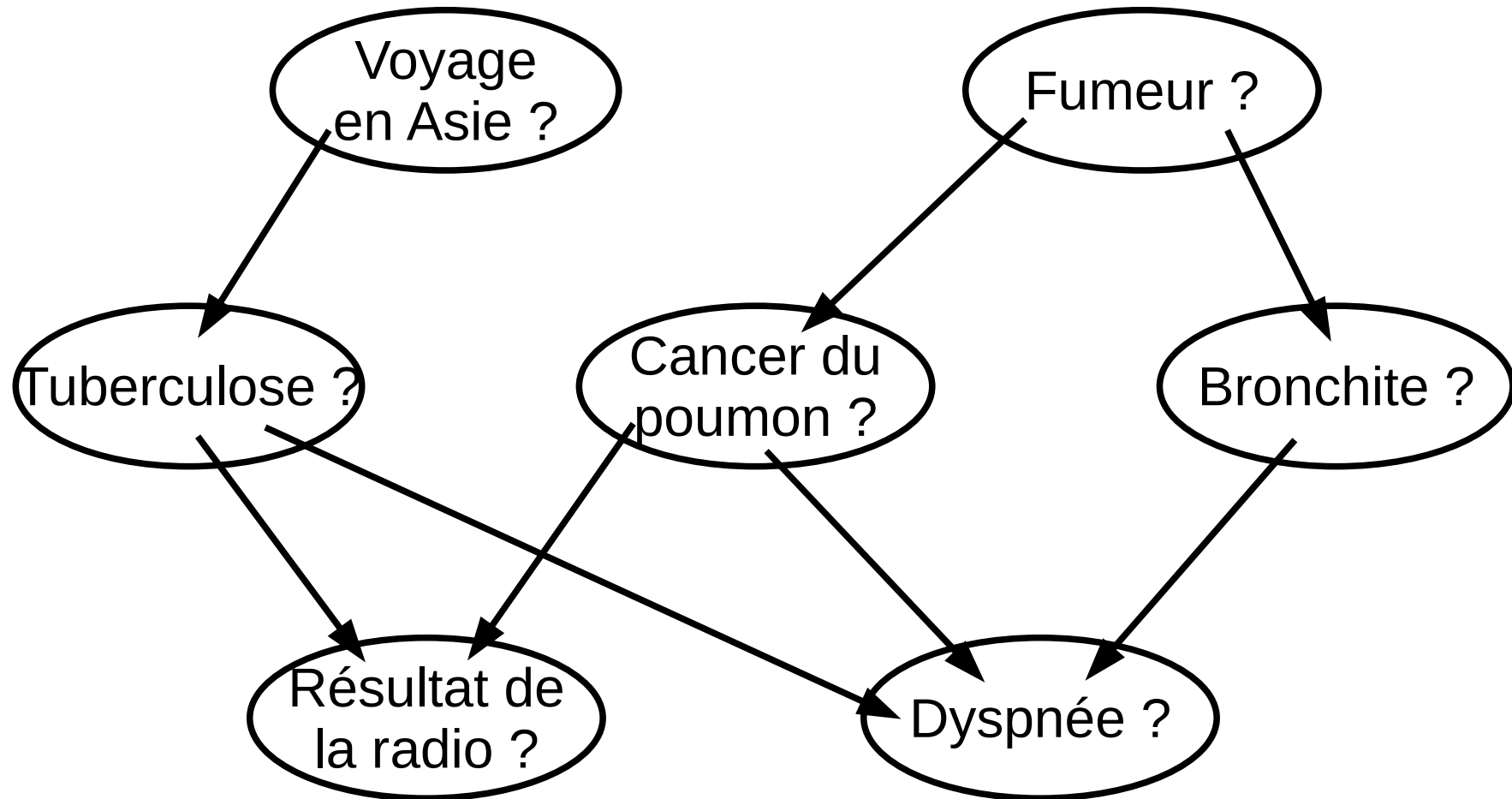
(retourner la flèche)

et $[D]$?

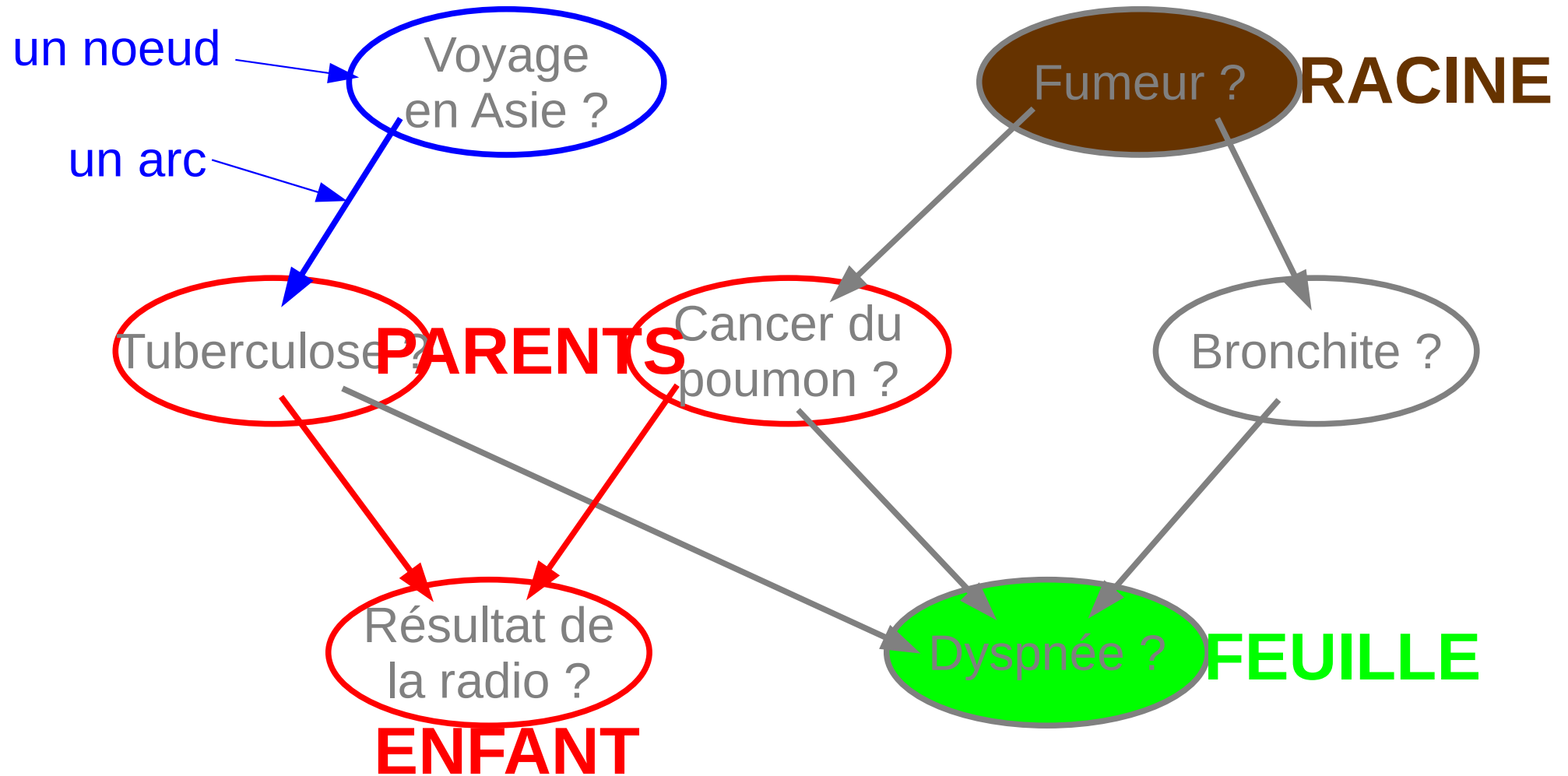
Les “doodles” de WinBugs



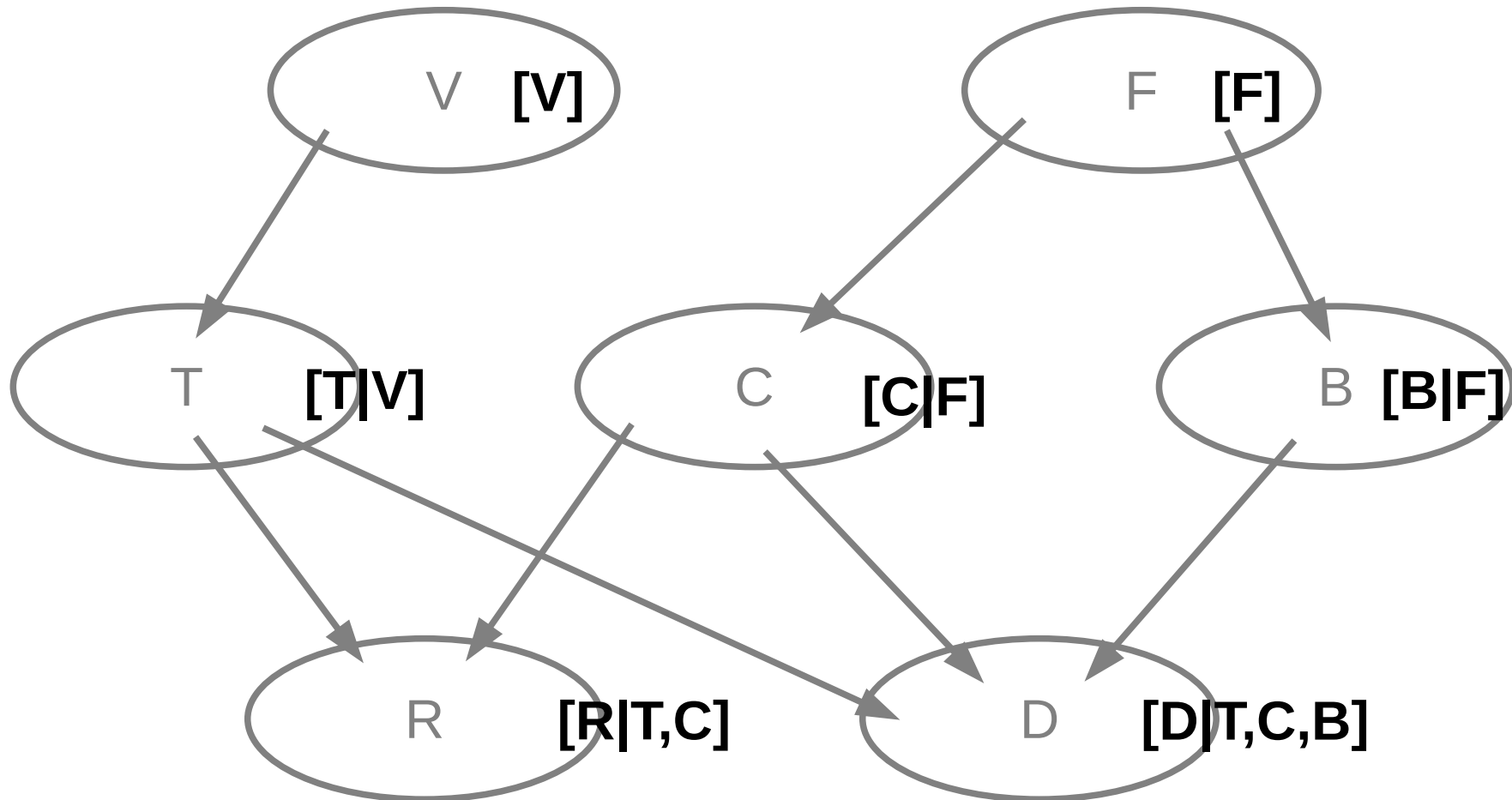
The Asian network (Lauritzen *et al.*, 1988)



(1) Réseaux Bayésiens



Définition d'une probabilité conjointe

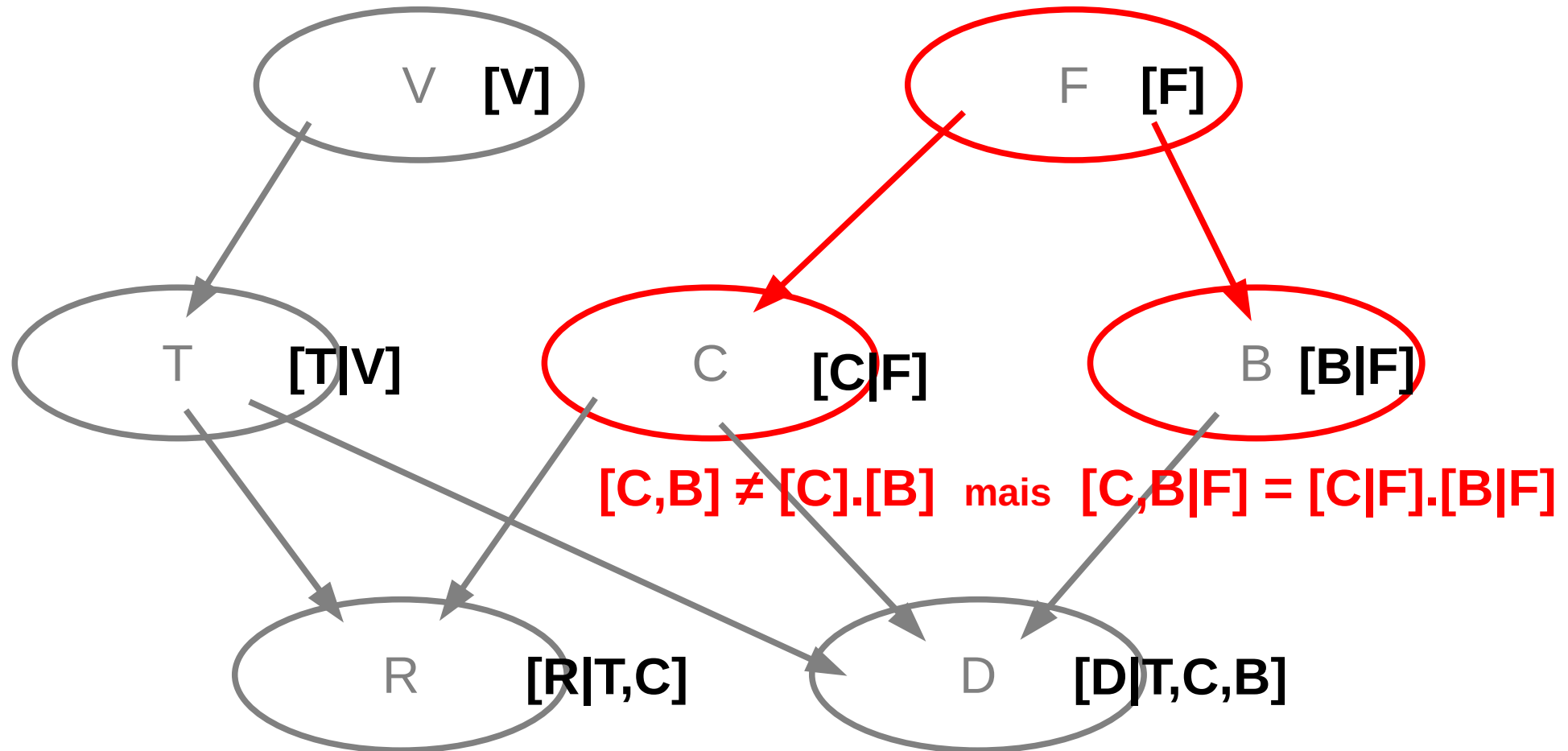


$$[V,F,T,C,B,R,D] = [V].[F].[T|V].[C|F].[B|F].[R|T,C].[D|T,C,B]$$

Pas de cycles !

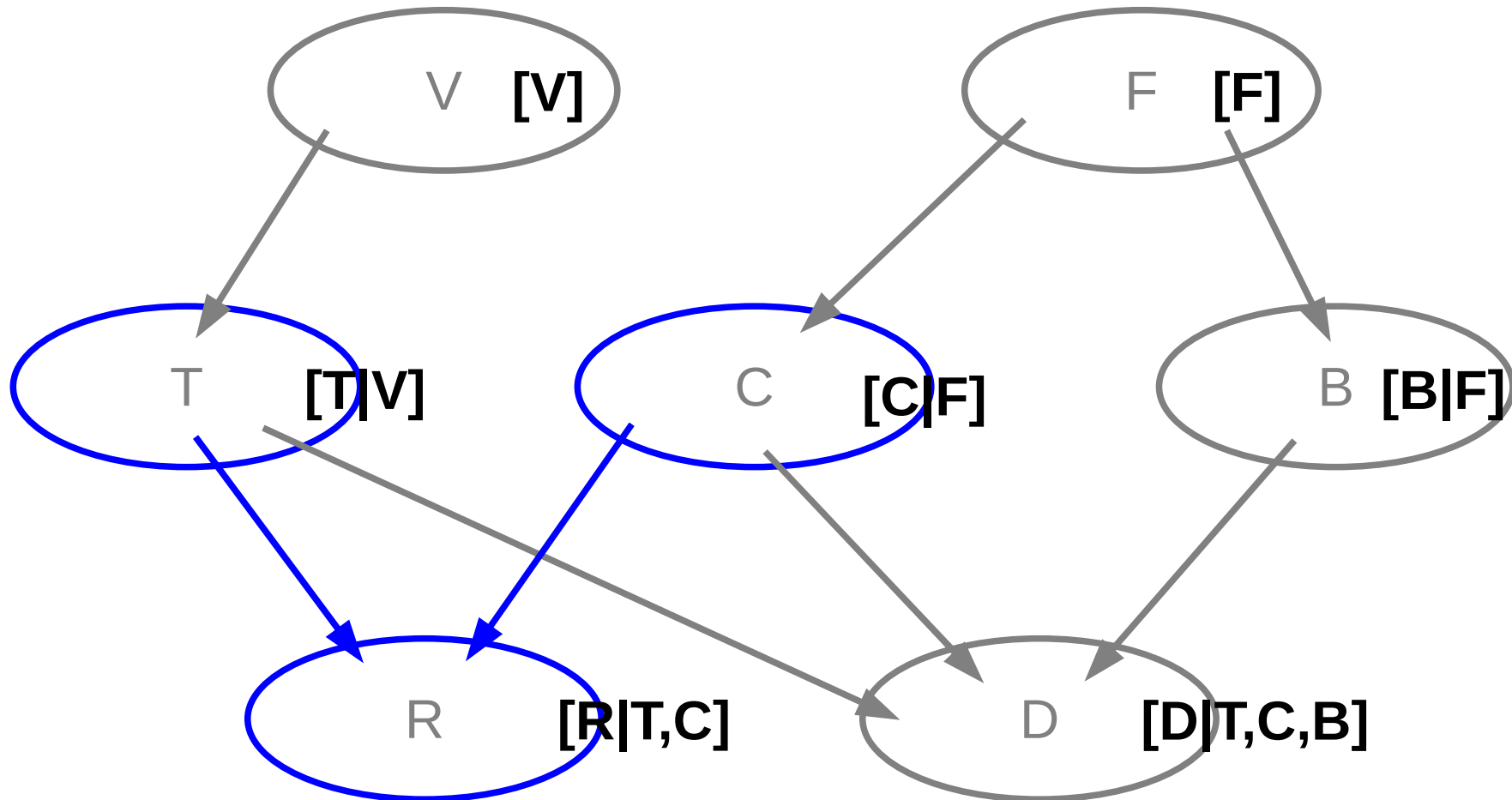
(1) Réseaux Bayésiens

Indépendances conditionnelles lisibles sur le graphe



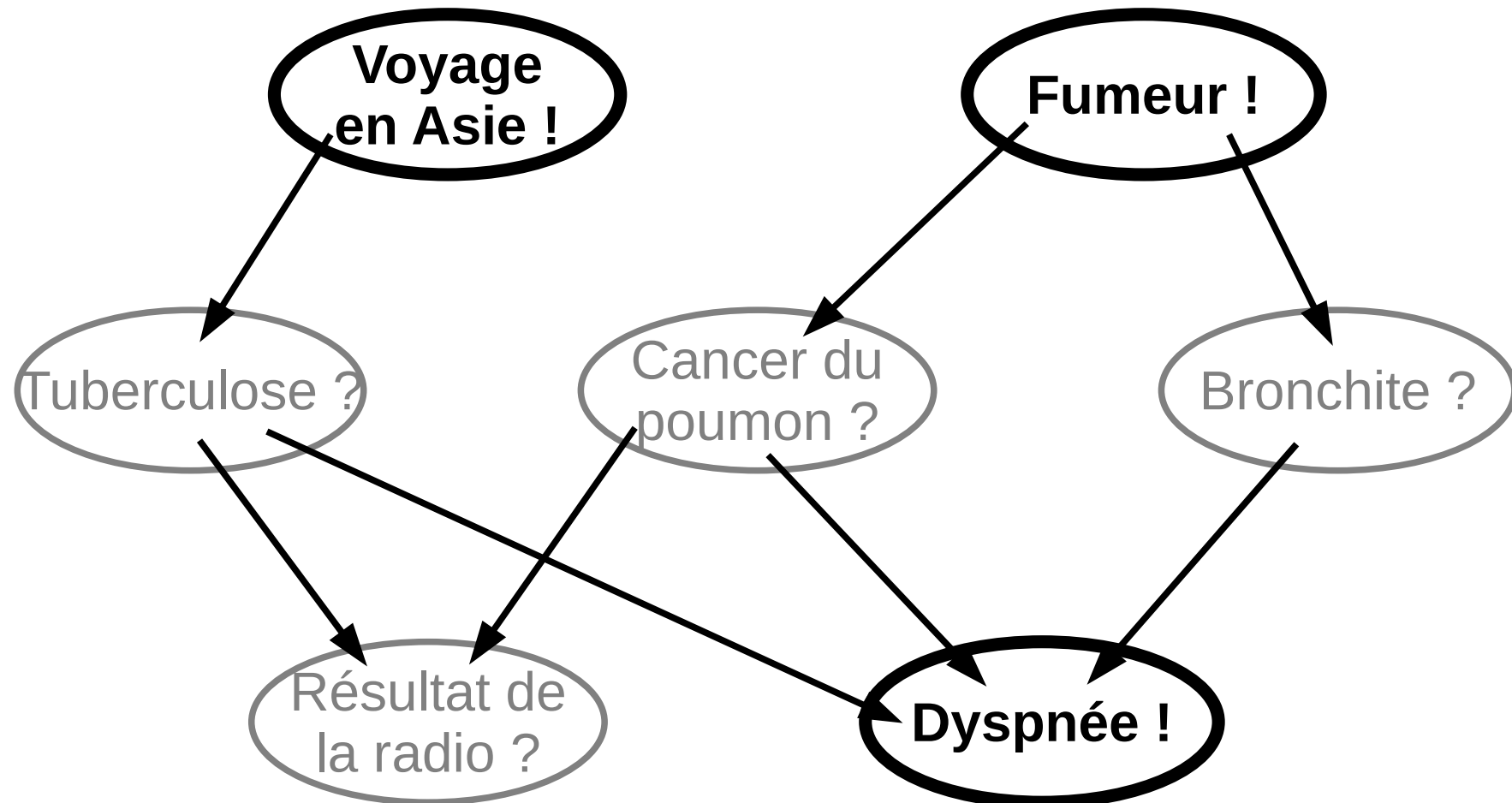
(1) Réseaux Bayésiens

Indépendances conditionnelles lisibles sur le graphe



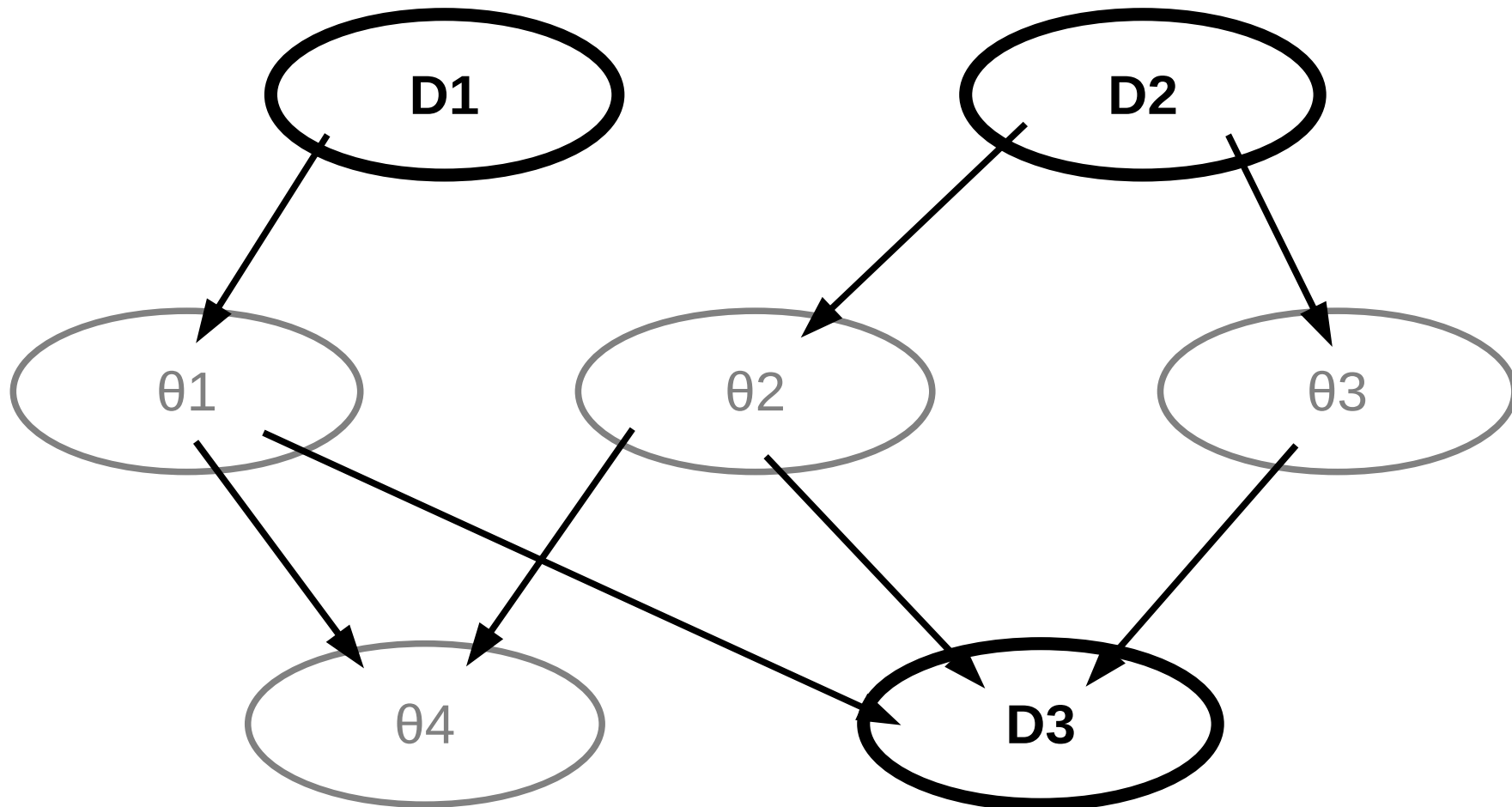
$[T,C] = [T].[C]$ mais $[T,C|R] \neq [T|R].[C|R]$

Utilisation à des fins de diagnostic



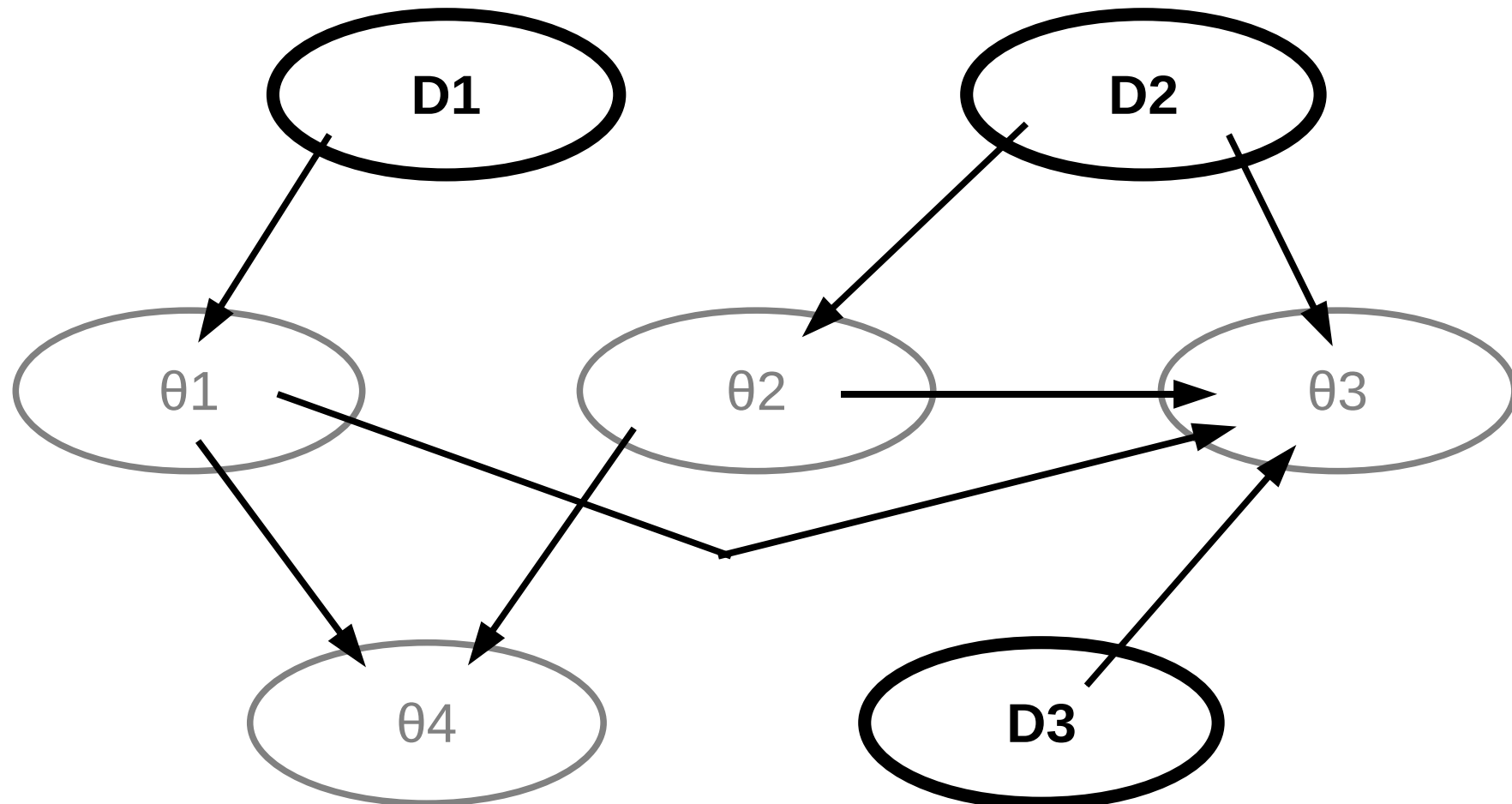
[T,C,B|V,F,D] ou [T|V,F,D], [C|V,F,D] & [B|V,F,D]

même démarche que la stat. bayésienne



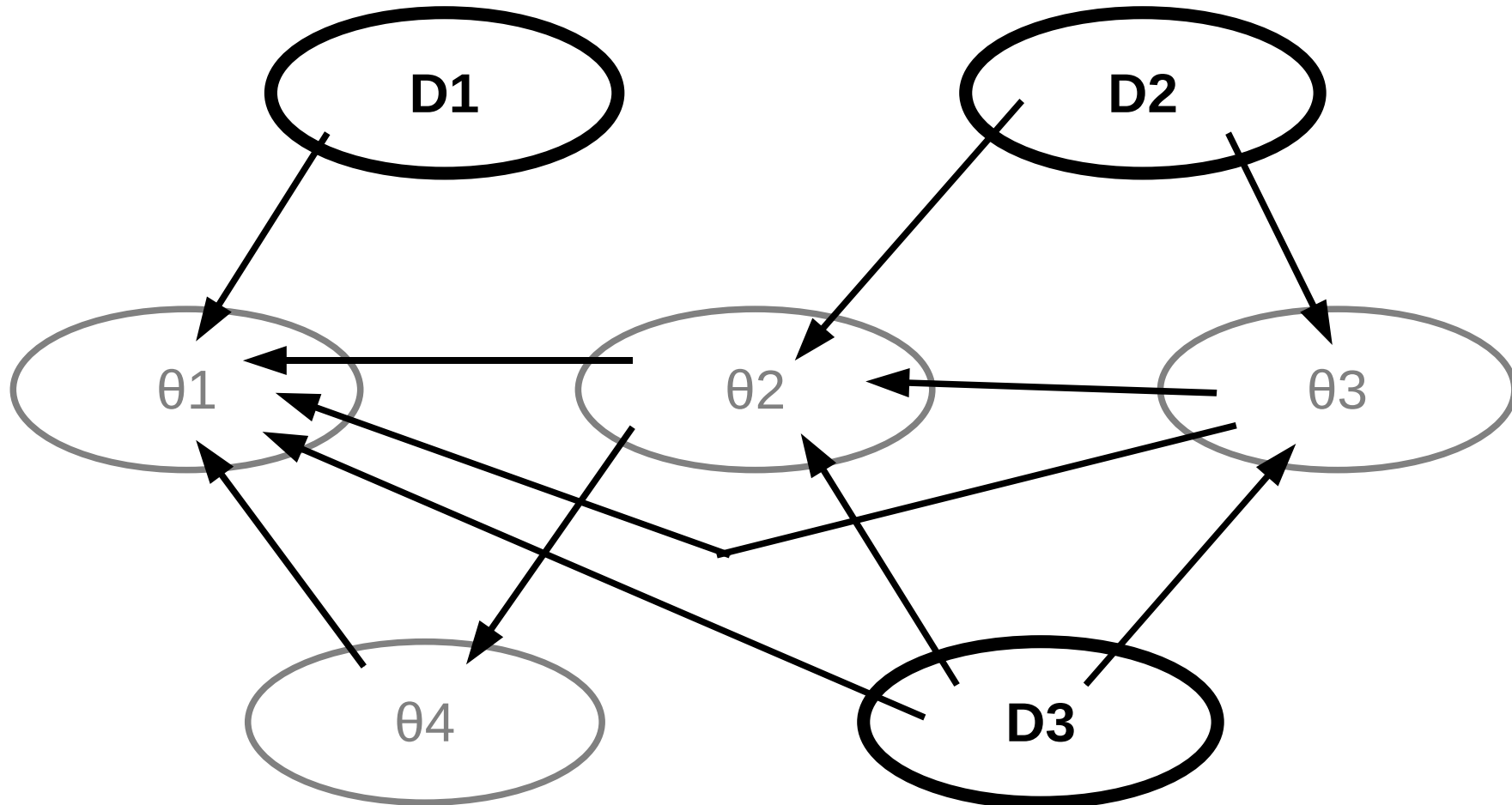
sauf que les arcs sont à adapter au conditionnement.

Une première possibilité



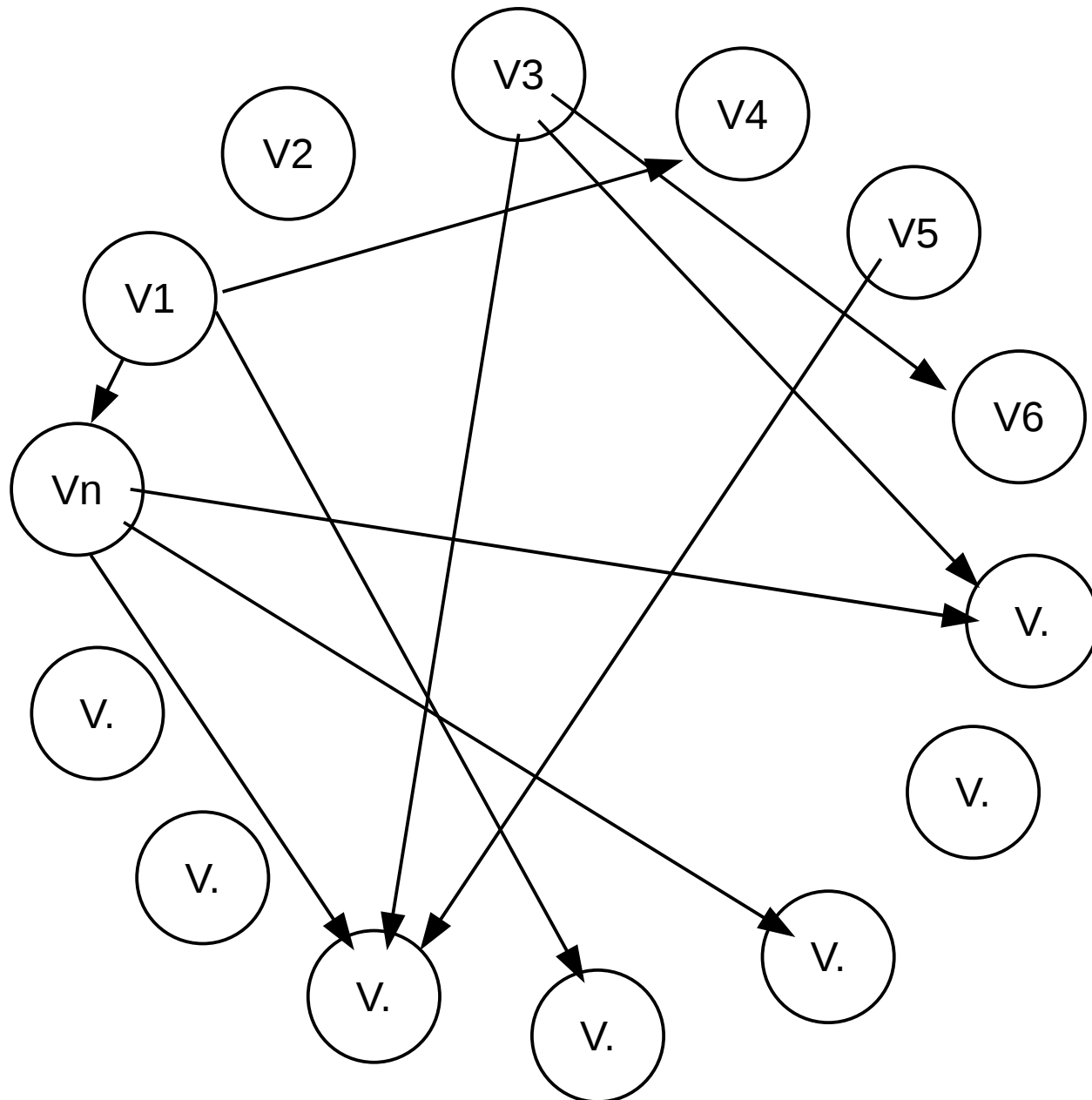
On prend un ordre $(\theta_1, \theta_2, \theta_4, \theta_3)$ des v.a. et récrit les jointes emboîtées par conditionnements successifs...

Une seconde possibilité



Avec l'ordre : $(\theta_3, \theta_2, \theta_4, \theta_1)$...

Recherche de structure du graphe



Pas de cycles !

n	nb
5	29281
6	3781503
7	1138779265
8	783702329343
9	1.213442e+15

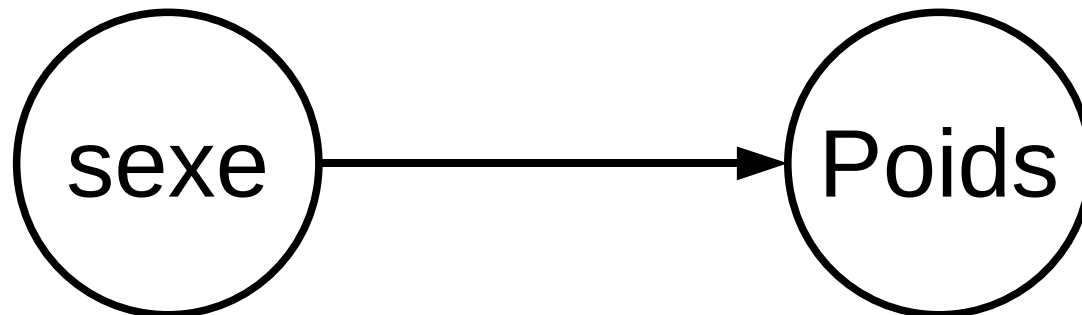
Nature des noeuds

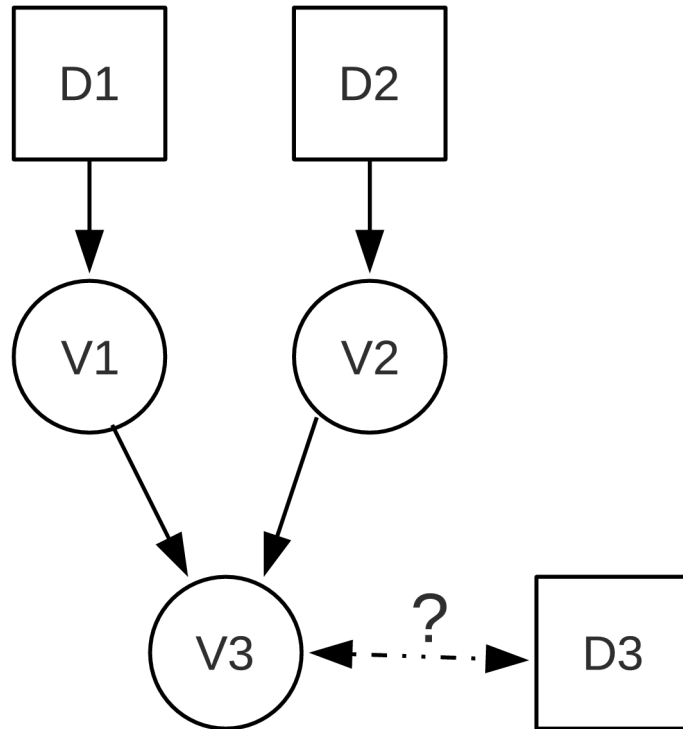
discrète

Poids \sim multinom. $(p_{\text{mai}|\text{sexe}}, p_{\text{nor}|\text{sexe}}, p_{\text{gros}|\text{sexe}})$

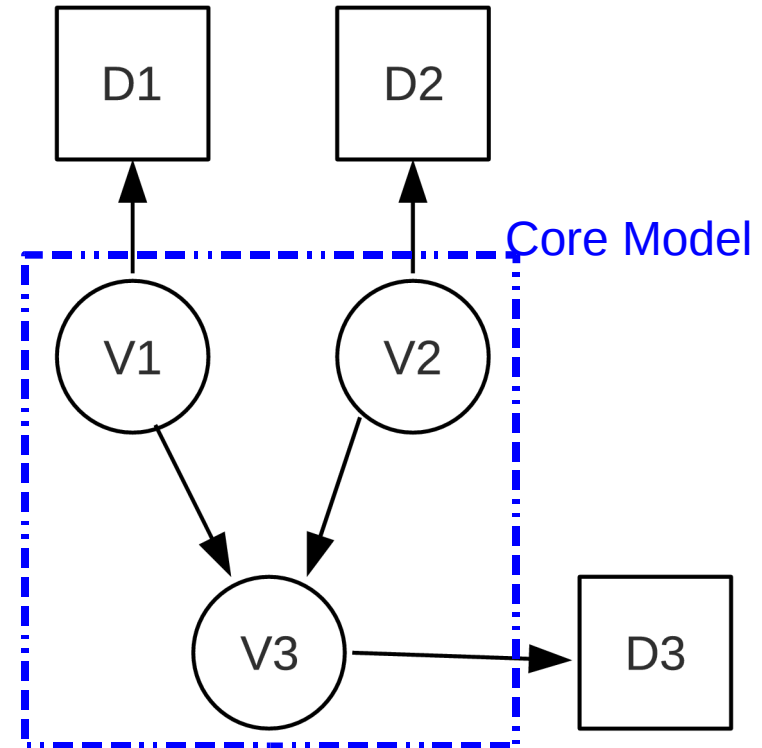
ou

continue

Poids \sim Normale $(\mu_{\text{sexe}}, \sigma^2)$ 



simulation Monte-Carlo
par RB



méta-analyse statistique
par SB

En bref

RB pour définir une distribution
conjointe d'un ensemble de variable
(= modèle)

et

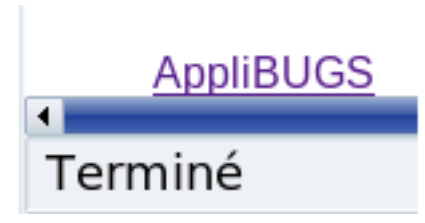
SB pour extraire l'information de
données relative aux paramètres
d'un modèle

(2) ReBaStaBa



REBASTABA : paquet R
projet de

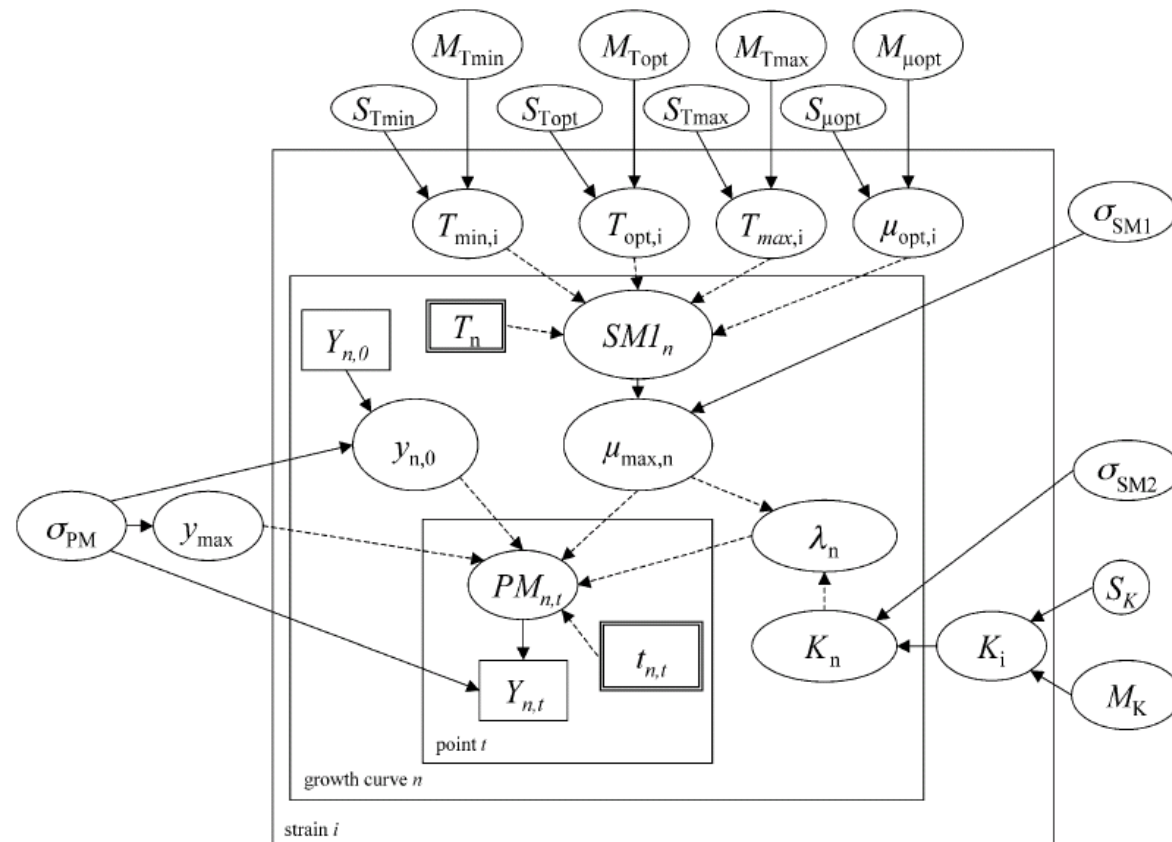
[\[PDF\] rebastaba means seaux yésiens traités par tistique yésienne, that ...](#)
Format de fichier: PDF/Adobe Acrobat - [Version HTML](#)



RB bien commode pour définir un modèle,
SB (et MCMC) aussi pour en faire l'inférence.

(2) objectifs

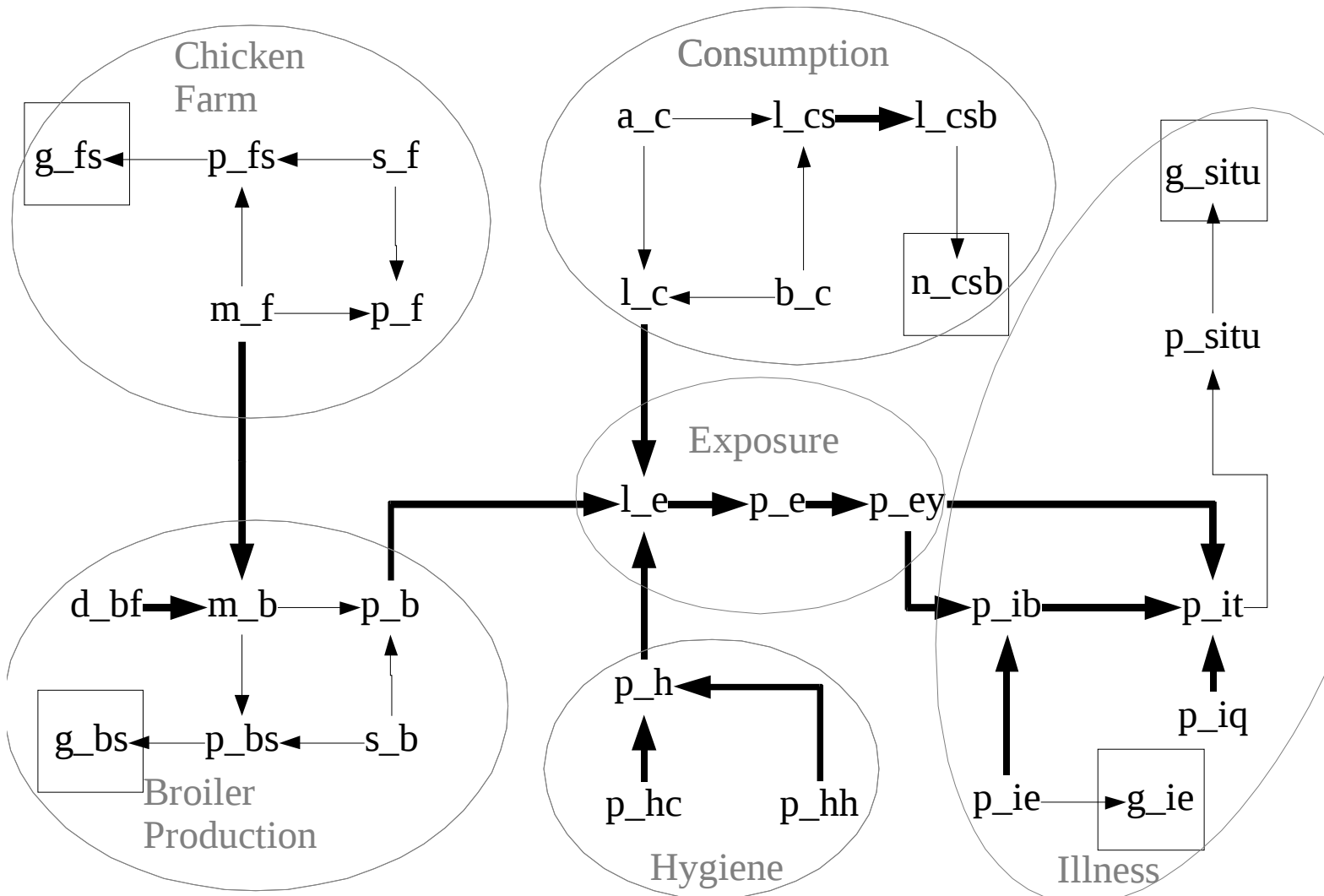
Croissance bactérienne : modélisation hiérarchique



RB bien commode pour définir un modèle,
SB (et MCMC) aussi pour en faire l'inférence.

(2) objectifs

Campylobactériose en France : modélisation d'une succession d'événements



Nous utilisons Jags et autres OpenBugs, mais ils ne nous suffisent pas complètement :

- Besoin de nombreuses variations (parties du modèle, sous-ensembles de données)
- Besoin de représentation graphique du graphe
- Besoin de construction automatique
- Besoin d'interfacer avec d'autres applications
- ...

Mise en chantier d'une programmation expérimentale, en fonction des besoins rencontrés.

Une mise en oeuvre laborieuse, des résultats aussi conceptuels.

```
RB <- zero.bn("RB1");  
#  
RB <- add.nd(RB, "SEXE",  
            ltype="Bernoulli",  
            lpara=list(p=0.5));  
#  
RB <- add.nd(RB, "TAILLE",  
            ltype="normal",  
            lpara=list(mu="160+12*({{SEXE}}==1)", sigma=18),  
            ltransfo="[[0]]");  
#  
RB <- add.nd(RB, "POIDS",  
            ltype="normal",  
            lpara=list(mu="{{TAILLE}}-100", sigma=10),  
            ltransfo="[[1]]",  
            lpod=list(c(60, 80)));
```

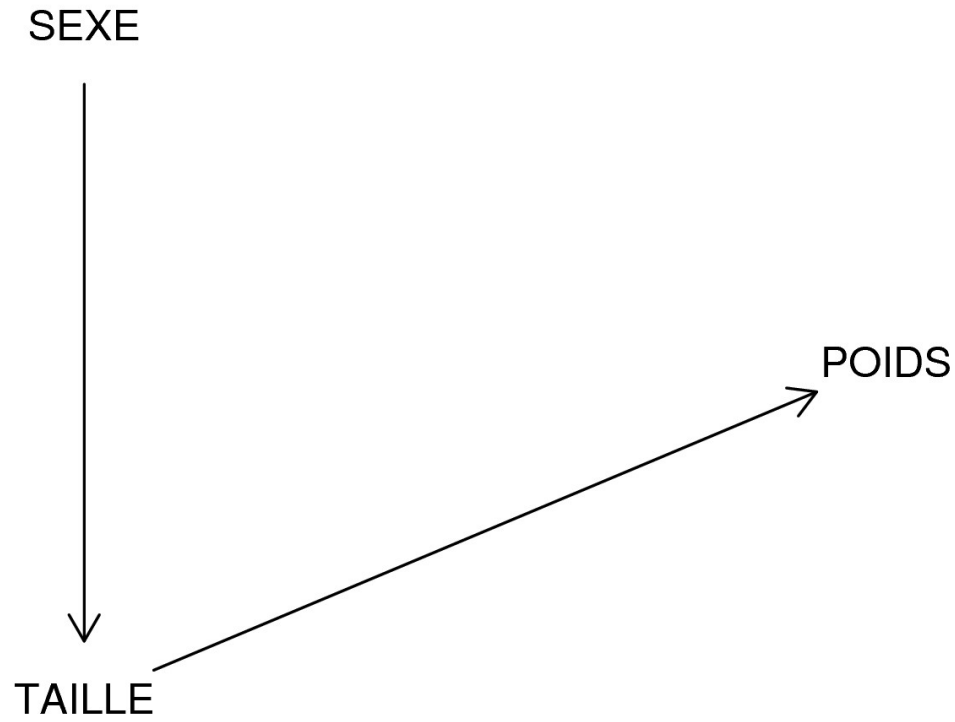
```
RB <- zero.bn("RB1");  
#  
RB <- add.nd(RB, "SEXE",  
             ltype="Bernoulli",  
             lpara=list(p=0.5));  
#  
RB <- add.nd(RB, "TAILLE",  
             ltype="normal",  
             lpara=list(mu="160+12*({{SEXE}}==1)", sigma=18),  
             ltransfo="[[0]]");  
#  
RB <- add.nd(RB, "POIDS",  
             ltype="normal",  
             lpara=list(mu="{{TAILLE}}-100", sigma=10),  
             ltransfo="[[1]]",  
             lpod=list(c(60, 80)));
```

```
RB <- zero.bn("RB1");  
#  
RB <- add.nd(RB, "SEXE",  
            ltype="Bernoulli",  
            lpara=list(p=0.5));  
#  
RB <- add.nd(RB, "TAILLE",  
            ltype="normal",  
            lpara=list(mu="160+12*({{SEXE}}==1)", sigma=18),  
            ltransfo="[[0]]");  
#  
RB <- add.nd(RB, "POIDS",  
            ltype="normal",  
            lpara=list(mu="{{TAILLE}}-100", sigma=10),  
            ltransfo="[[1]]",  
            lpod=list(c(60, 80)));
```

```
RB <- zero.bn("RB1");  
#  
RB <- add.nd(RB, "SEXE",  
            ltype="Bernoulli",  
            lpara=list(p=0.5));  
#  
RB <- add.nd(RB, "TAILLE",  
            ltype="normal",  
            lpara=list(mu="160+12*({{SEXE}}==1)", sigma=18),  
            ltransfo="[[0]]");  
#  
RB <- add.nd(RB, "POIDS",  
            ltype="normal",  
            lpara=list(mu="{{TAILLE}}-100", sigma=10),  
            ltransfo="[[1]]",  
            lpod=list(c(50, 100)));
```

`plot(RB);`

RB1



```
print(RB);
```

```
Object bn 'RB1' with 3 nodes
```

```
Node List
```

```
=====
```

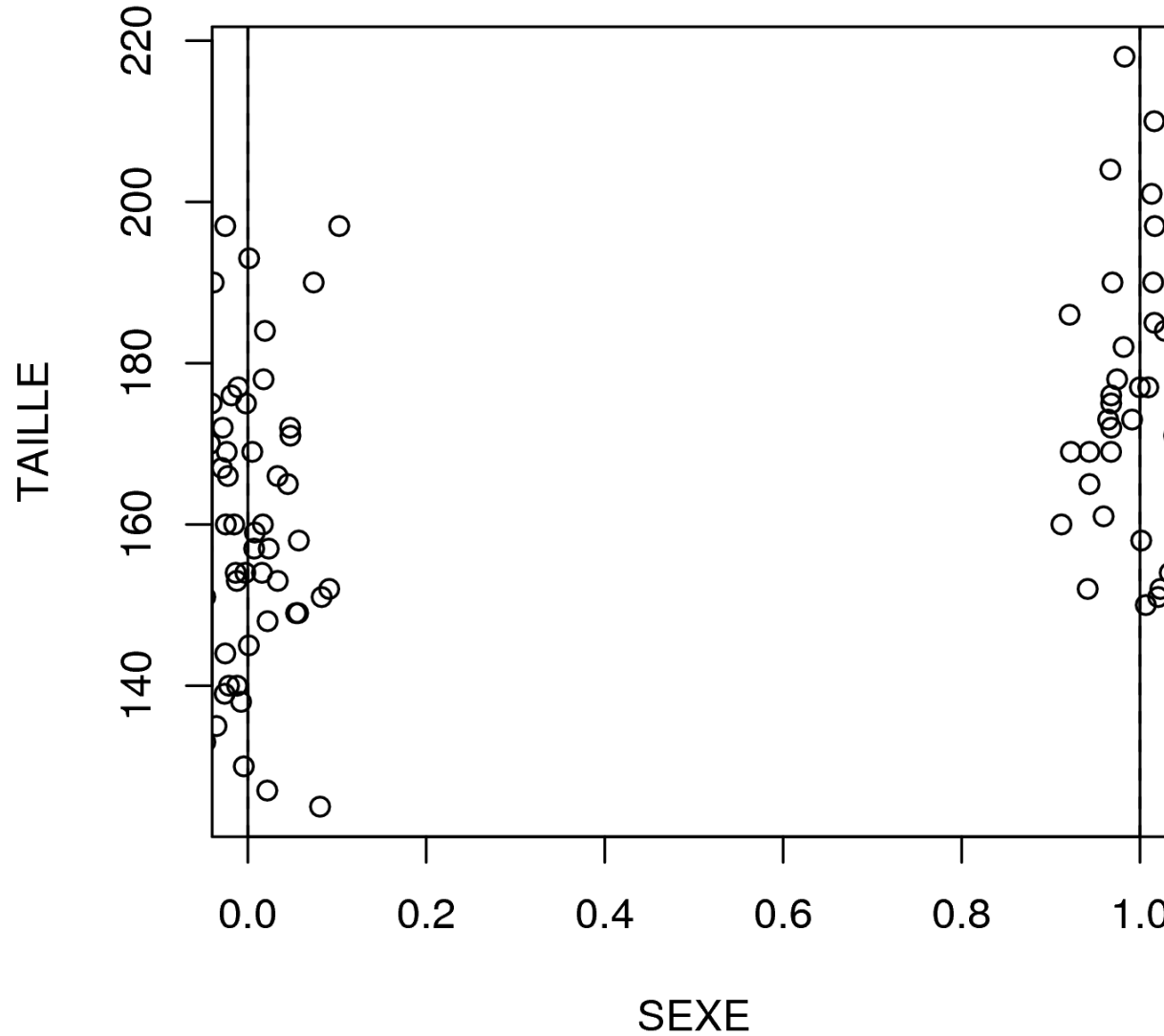
	Name	nb.Var	ltype	Parent(s)
[1,]	"SEXE"	"1"	"Bernoulli"	"- NO -"
[2,]	"TAILLE"	"1"	"normal"	"{{SEXE}}"
[3,]	"POIDS"	"1"	"normal"	"{{TAILLE}}"


```
SIMU <- bn2dn(RB); print(SIMU);
```

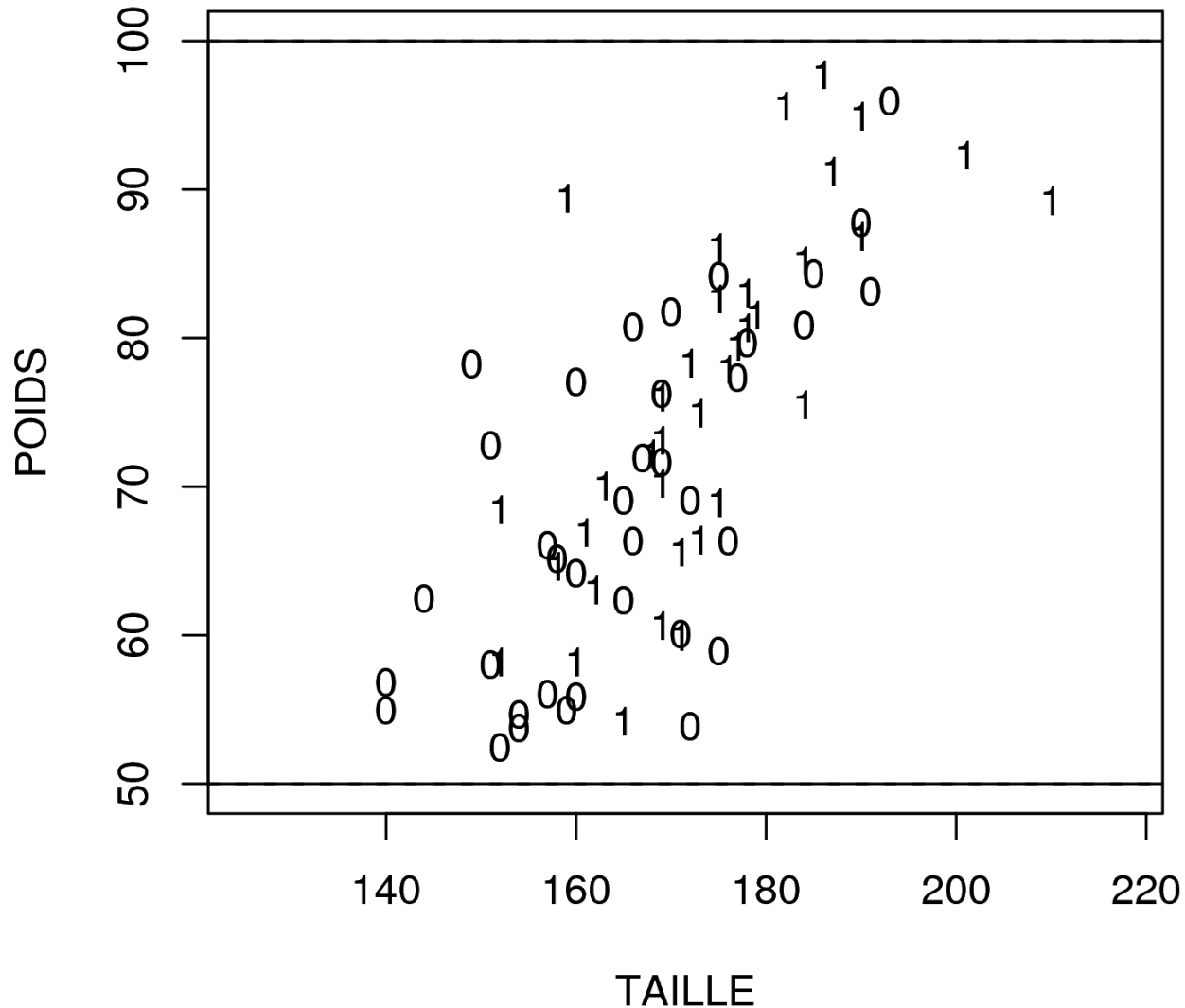
```
-----  
dn object of name "RB1"  
-----  
Main univariate characteristics are:  
      SEXE  TAILLE  POIDS  >?<  
number  100.00  100.000  73.000  100.000  
missing   0.00    0.000  27.000   0.000  
mean      0.45  166.840  72.352   2.700  
std-dev.  0.50   18.421  12.234   4.462  
min       0.00  125.000  52.400   0.000  
Q01%     0.00  126.980  53.336   0.000  
Q05%     0.00  137.850  54.440   0.000  
Q10%     0.00  144.900  55.840   0.000  
Q25%     0.00  153.750  62.400   0.000  
median   0.00  168.500  71.900   0.000  
Q75%     1.00  177.250  81.500  10.000  
Q90%     1.00  190.000  88.900  10.000  
Q95%     1.00  197.000  93.340  10.000  
Q99%     1.00  210.080  96.404  10.000  
MAX      1.00  218.000  97.700  10.000  
-----
```

plot(SIMU);

RB1 (1/1) {100 - 0 / 100 pts}



```
plot(SIMU, lab="SEXE");  
RB1 (1/1) {100 - 27 / 100 pts}
```



```

b9 <- zero.bn("Faire du Jags");
b9 <- add.nd(b9, "A", "normal", lvar=1:4,
            lpara=list(mu=0, sigma=1),
            ltransfo="");
b9 <- add.nd(b9, "B", "normal", lvar=1:4,
            lpara=list(mu="sqrt({{A}}+5)",
                      sigma="1+abs({{A}})"),
            ltransfo="[[2]]");

#
j9 <- bn2bugs(b9);
form.file(j9, "rbsb.b9.jam", app=FALSE);
#
j9s <- bn2jagscript(b9, "rbsb.b9.jam",
                  "rbsb.b9A",
                  iter=c(100, 100, 1));
form.file(j9s, "rbsb.b9.jas", app=FALSE);

```

```

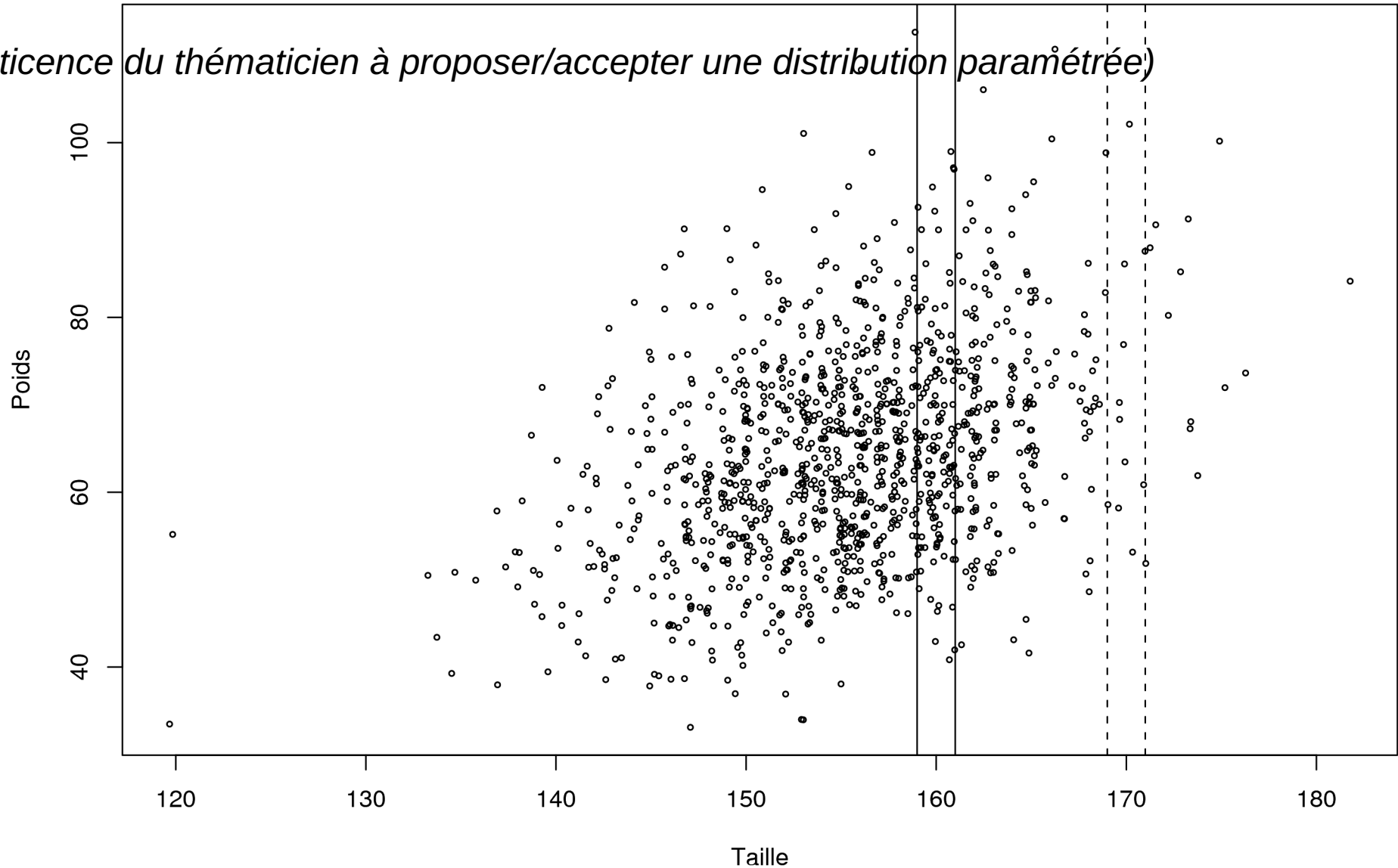
### Beginning of Model from bn "Faire du Jags"
model {
  # ----- A
  for (i4A in 1:4) {
    A[i4A] ~ dnorm(0, 1/(1)^2);
  }
  # ----- B
  for (i4z4B in 1:4) {
    z4B[i4z4B] ~ dnorm(sqrt(A[i4z4B]+5),
                        1/(1+abs(A[i4z4B]))^2);
  }
  B <- round(100*z4B)/100;
}
### End of Model from bn "Faire du Jags"

```

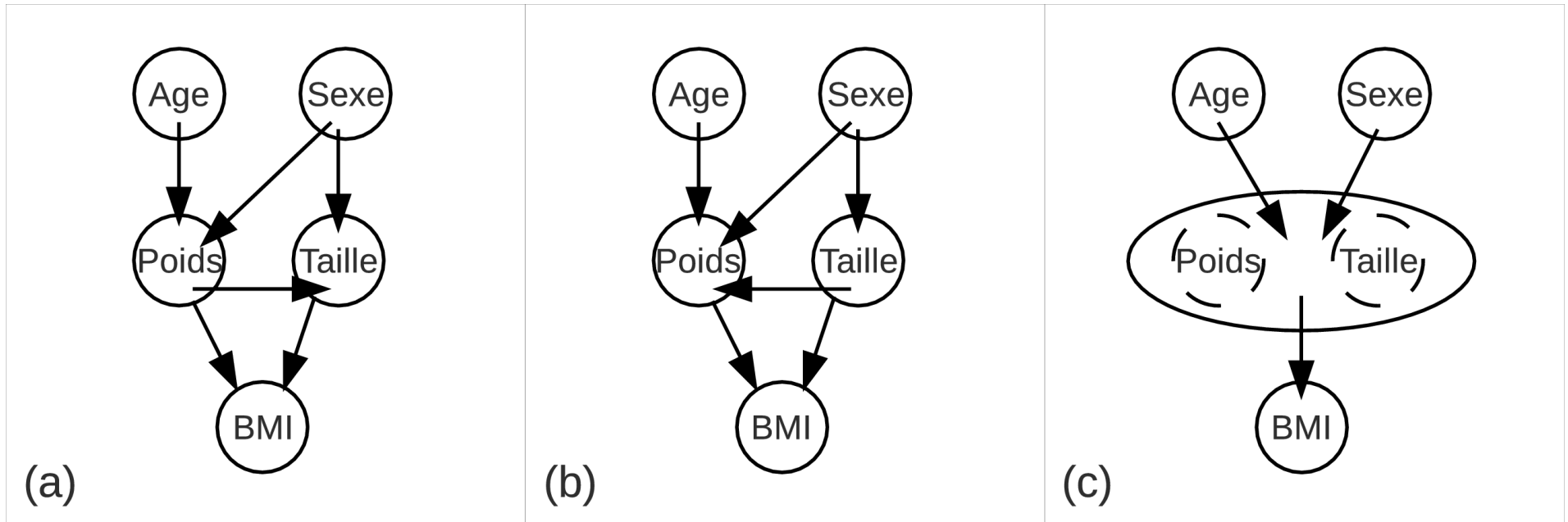
```
model in rbsb.b9.jam
compile
initialize
update 100
monitor set A
monitor set B
update 100
coda *, stem("rbsb.b9A")
```

Poids | Taille (N = 1151)

(réticence du thématicien à proposer/accepter une distribution paramétrée)

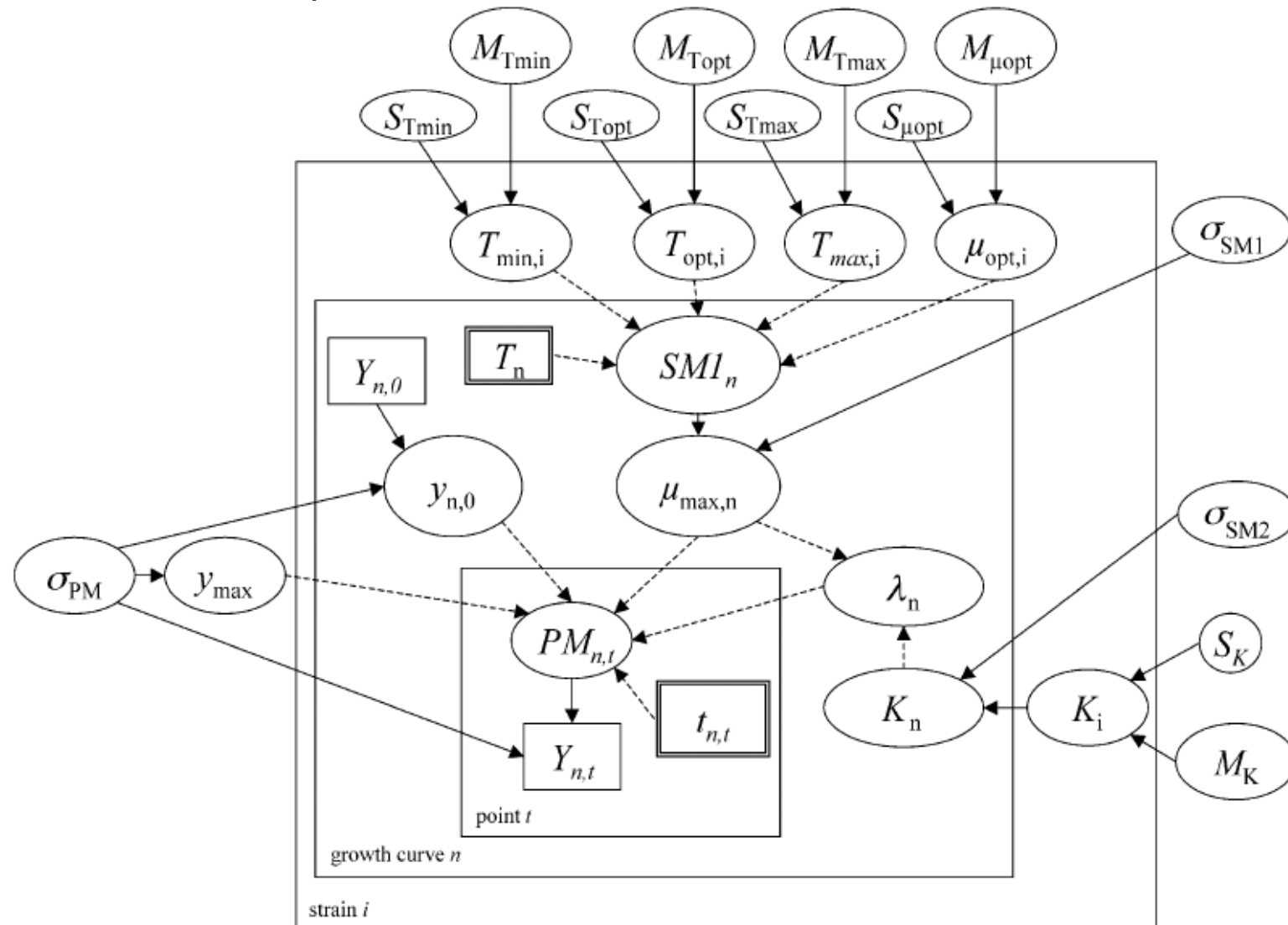


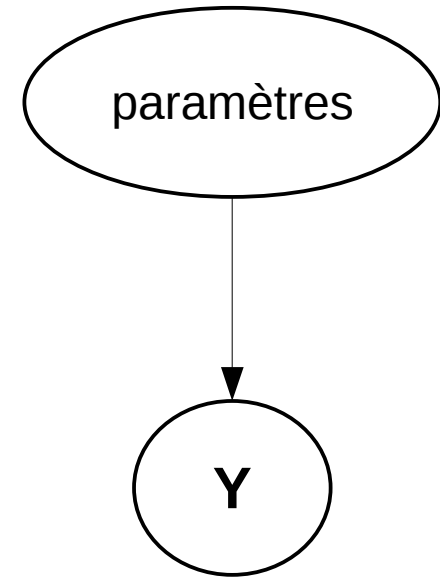
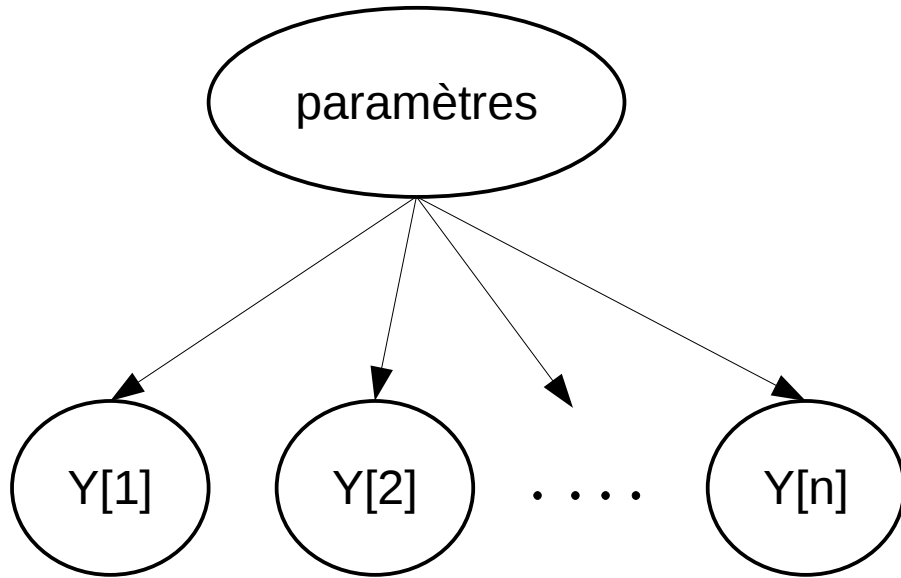
Comment reproduire la corrélation entre Poids et Taille chez les seniors.



Equivalence entre modèles

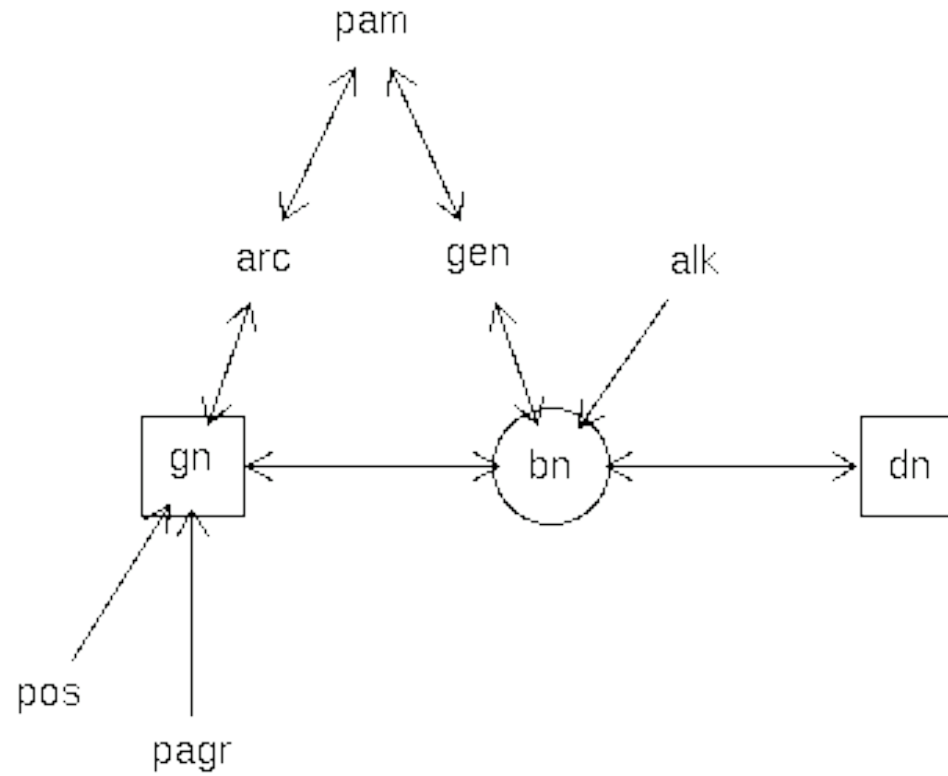
Modélisations hiérarchiques





Conditionnellement indépendants

Distributions similaires mais pas forcément identiques.





Jean-Baptiste.Denis@Jouy.Inra.Fr
Isabelle.Albert@Paris.Inra.Fr

Merci

de
votre
attention