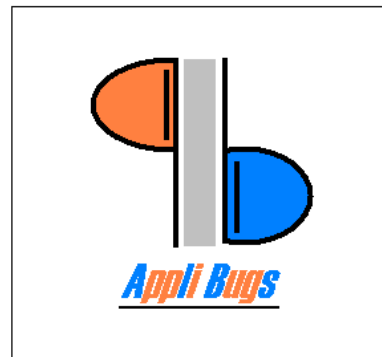


Evaluation de méthodes diagnostiques en l'absence de gold standard sur la base de variables binaires et continues : intérêt de l'approche bayésienne

Thibaut Lurier^{1,2}, Marie Laure Delignette Muller³

Elsa Jourdain¹ et Florence Ayrat²

UMR 0346 EPIA ¹, USC 1233², LBBE³



Sommaire

- Evaluation de tests binaires : **Modèle à classe latente**
- Evaluation de tests quantitatifs : **Modèle de mélange**
- Combinaison des deux approches
- Illustration sur le diagnostic de la dictyocaulose bovine
- Travaux en cours

Evaluation de méthodes diagnostiques

- Considérons une maladie M et un test diagnostique T
- On cherche à estimer la Sensibilité (Se) et la spécificité (Sp) du test :

$$Se = P\left(T^+ | M^+\right) \quad Sp = P\left(T^- | M^-\right)$$

- Par rapport à un Gold Standard de Se et $Sp = 100\%$

$$\hat{Se} = \frac{n_{T^+}}{n_{M^+}} \quad \hat{Sp} = \frac{n_{T^-}}{n_{M^-}}$$

Gold Standard?

- Généralement :
 - le Gold Standard **n'existe pas!**
 - Il est trop cher ou trop invasif
- Autres possibilités :
 - Utiliser un **test de référence de Se et Sp connues**
⇒ Assez rare en médecine vétérinaire
 - Utiliser un **Gold standard imparfait** en tant que Gold standard
⇒ Estimations biaisées
 - Utiliser des **patients de statuts connus** et souvent caricaturaux
⇒ Estimations biaisées (surestimation des Se et Sp)

Modèle à classe latente (*Hui and Walter, 1980*)

- Estimation des caractéristiques de **plusieurs tests binaires** :
 - Considérons 2 tests appliqués à une population dont la prévalence de la maladie est P

	Test 2 positif	Test 2 négatif
Test 1 positif	$p_{11} = Se_1 \times Se_2 \times P + (1 - Sp_1) \times (1 - Sp_2) \times (1 - P)$	$p_{10} = Se_1 \times (1 - Se_2) \times P + (1 - Sp_1) \times Sp_2 \times (1 - P)$
Test 1 négatif	$p_{01} = (1 - Se_1) \times Se_2 \times P + Sp_1 \times (1 - Sp_2) \times (1 - P)$	$p_{00} = (1 - Se_1) \times (1 - Se_2) \times P + Sp_1 \times Sp_2 \times (1 - P)$

- 3 degrés de liberté pour 5 paramètres à estimer ($P \ Se_1 \ Sp_1 \ Se_2 \ Sp_2$)
- En revanche, si on se place dans deux populations de prévalences différentes
 \Rightarrow **6 DL pour 6 paramètres** à estimer ($P_1 \ P_2 \ Se_1 \ Sp_1 \ Se_2 \ Sp_2$)
 \Rightarrow 2 solutions possibles!
- Variantes possibles avec 3 tests et une population (8DL pour 7 paramètres)...

Modèle général

- Si on considère 2 tests appliqués à S populations de tailles respectives N_s et de prévalences P_s alors :

$$n_s \sim \text{multinomiale}(N_s, p_{s11}, p_{s10}, p_{s01}, p_{s00})$$

$$p_{s11} = P_s \times Se_1 \times Se_2 + (1 - P_s) \times (1 - Sp_1) \times (1 - Sp_2)$$

$$p_{s10} = P_s \times Se_1 \times (1 - Se_2) + (1 - P_s) \times (1 - Sp_1) \times Sp_2$$

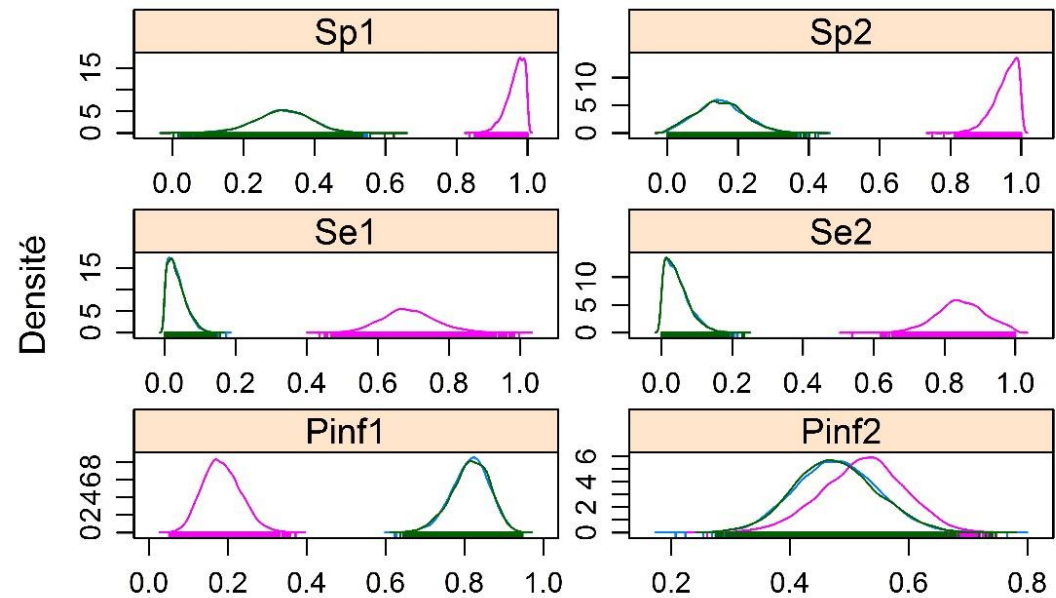
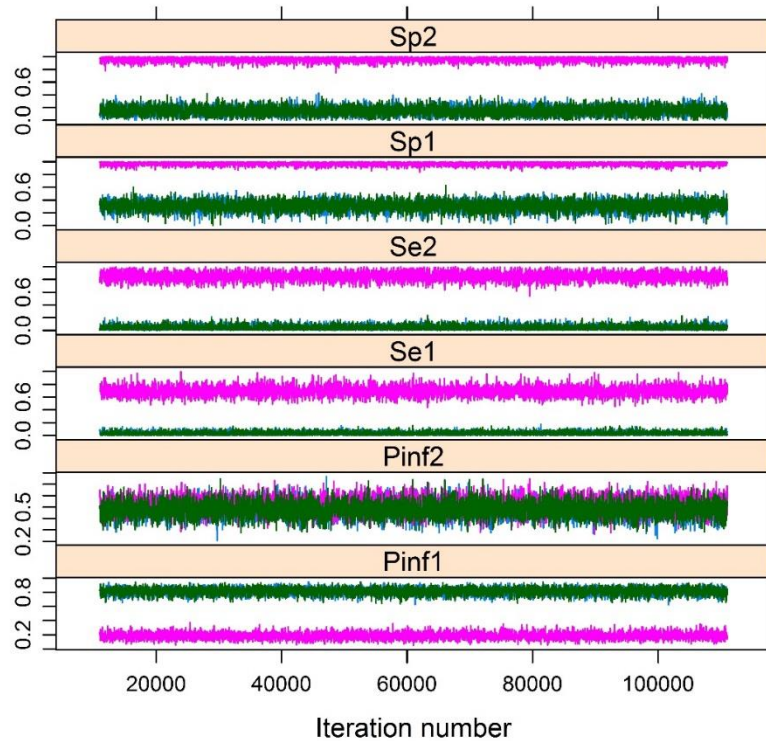
$$p_{s01} = P_s \times (1 - Se_1) \times Se_2 + (1 - P_s) \times Sp_1 \times (1 - Sp_2)$$

$$p_{s00} = P_s \times (1 - Se_1) \times (1 - Se_2) + (1 - P_s) \times Sp_1 \times Sp_2$$

- Inférence possible par maximum de vraisemblance ou par inférence bayésienne mais deux solutions sans priors informatifs

Itération sous JAGS

- Inférence à partir de données simulées :
 - $Sp_1 = 0.99$, $Se_1 = 0.8$, $Se_2 = 0.8$, $Sp_2 = 0.95$, $P_1 = 0.1$, $P_2 = 0.5$, $N_1 = 100$, $N_2 = 100$
 - Priors non informatifs (loi uniforme [0;1])



2 solutions possibles!!!

En restreignant les possibles (priors vaguement informatifs)

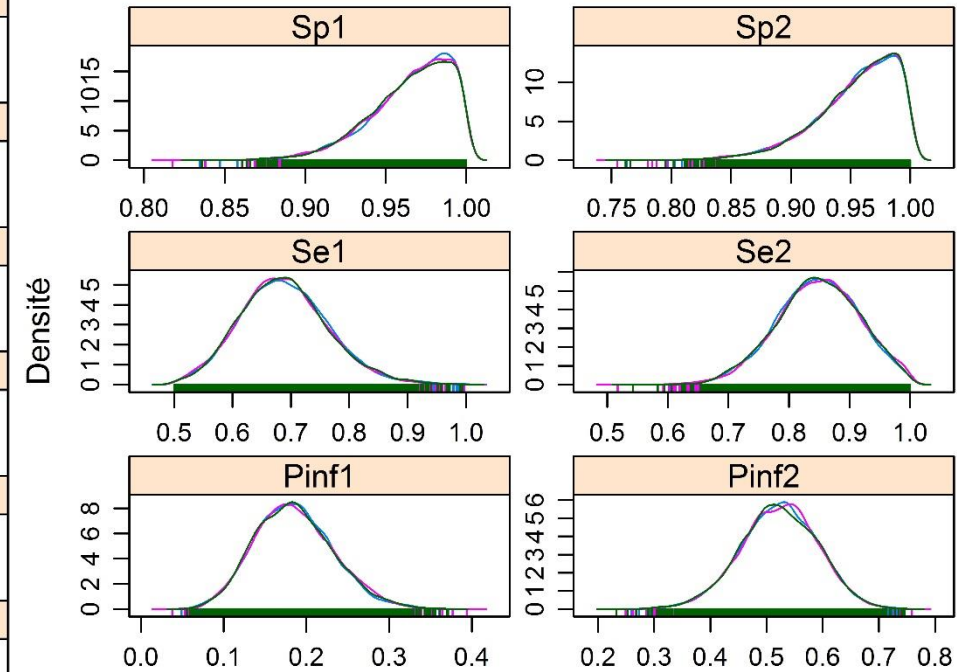
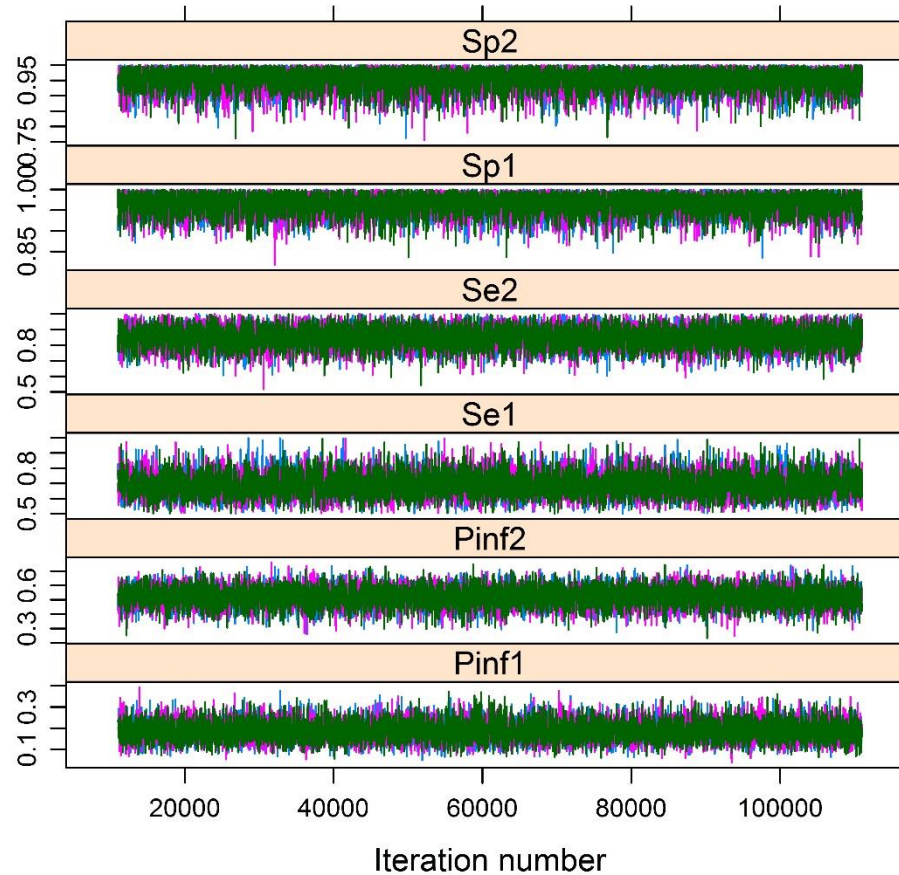
$$Sp_1 > 0.5$$

$$Se_1 > 0.5$$

$$P_1 < 0.5$$

Utilité de loi a priori
« légèrement » informative

- Uniforme [0,5; 1]



Hypothèses du modèle

- **Se et Sp constantes** dans chacune des sous populations
 - Les Se et Sp identiques dans les populations à prévalences différentes?

- **Indépendance conditionnelle** au statut vis-à-vis de la maladie

$$P\left(T_2^+ \mid T_1^+, M^+\right) = P\left(T_2^+ \mid M^+\right)$$

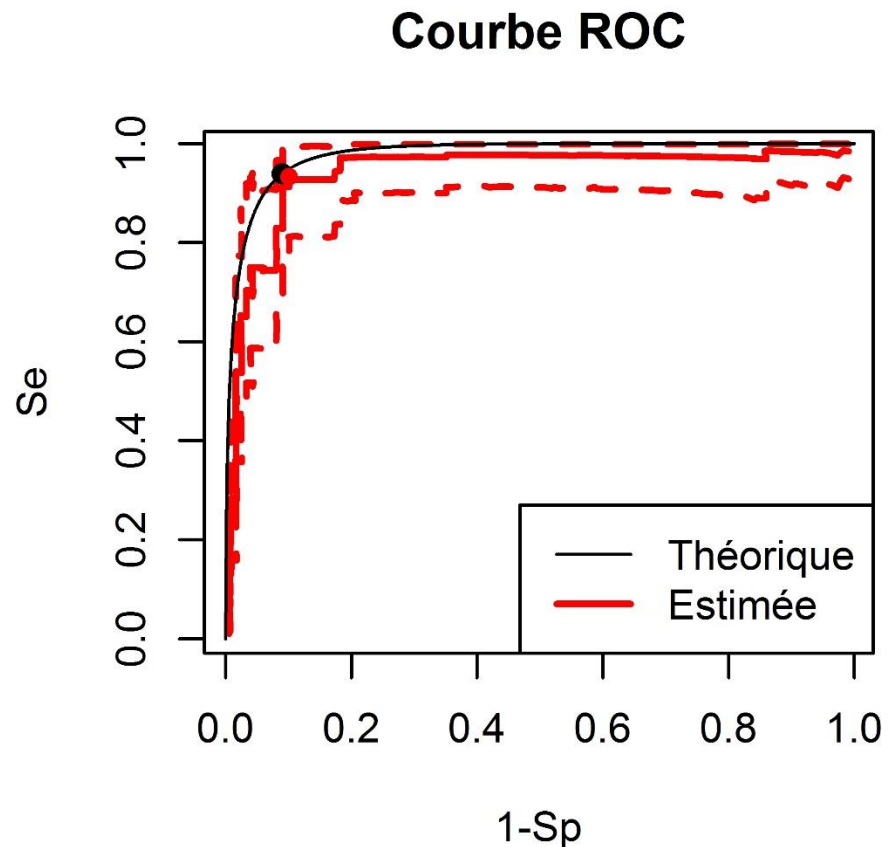
- On considère que l'hypothèse est vérifiée quand les tests reposent sur des principes biologiques différents (PCR qualitative et bactériologie)
 - Possibilité de prendre en compte la dépendance conditionnelle mais alors avec plus de paramètres... (*Hadgu and Qu, 1998; Dendukuri and Joseph, 2001*)
- Le **statut latent est biologiquement similaire** pour les deux tests
 - Immunité pour des sérologies// présence du pathogène pour bactério et PCR
 - Possibilité de modéliser plus de 2 statuts latents (*Dendukuri et al. 2009*)

Quid des tests quantitatifs?

- Possibilité de **réduire le test quantitatif à un test binaire** :
 - > ou < seuil de positivité
 - **Choix du seuil de positivité souvent arbitraire**
 - **Perte d'information** : même valeur diagnostique si la valeur est juste au dessus ou très au dessus du seuil
- Possibilité de réitérer le modèle à classe latente pour chaque valeur possible du test quantitatif
 - Réalisation de nombreuses mesures avec des variations des Se et Sp non uniformes d'un seuil à l'autre
 - Dépendance conditionnelle variable en fonction des seuils

Exemple sur données simulées

- Un modèle pour chacune des 200 valeurs simulées
- Priors vaguement informatifs



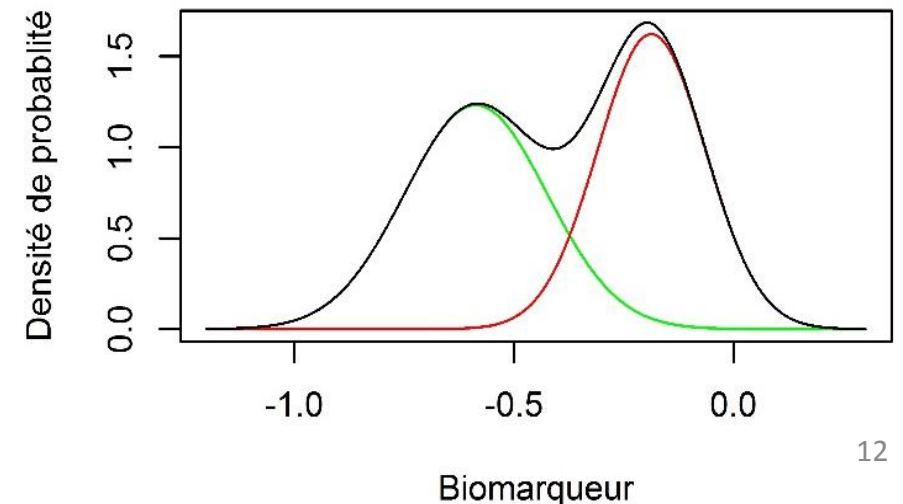
Modèle de mélange (*Hanley, 1996; Marin et al., 2005*)

- On suppose que la distribution du biomarqueur suit une loi bi-normale avec **deux distributions latentes normales** chez les malades et non-malades
- Un seul test quantitatif dont les deux distributions latentes sont estimables
- Avec P =Prévalence de la maladie, M_i = Statut vis-à-vis de la maladie et T_i = résultat du test

$$M_i \sim \text{Bernoulli}(P)$$

$$(T_i | M_i = 1) \sim N(\mu_{M^+}, \sigma_{M^+})$$

$$(T_i | M_i = 0) \sim N(\mu_{M^-}, \sigma_{M^-})$$



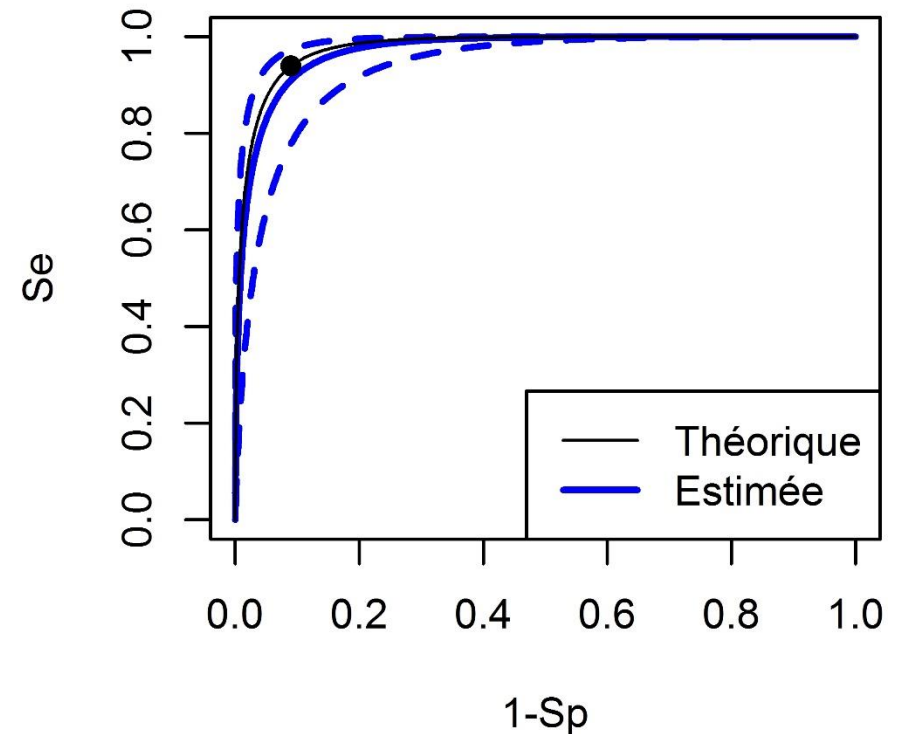
Avantages

- Modèle simple
- Courbe ROC facilement modélisable
 - avec pour chaque seuil s

$$Se(s) = 1 - \Phi\left(\frac{s - \mu_{M^+}}{\sigma_{M^+}}\right); \quad Sp(s) = \Phi\left(\frac{s - \mu_{M^-}}{\sigma_{M^-}}\right)$$

- Un seul test nécessaire
- Exemple sur données simulées

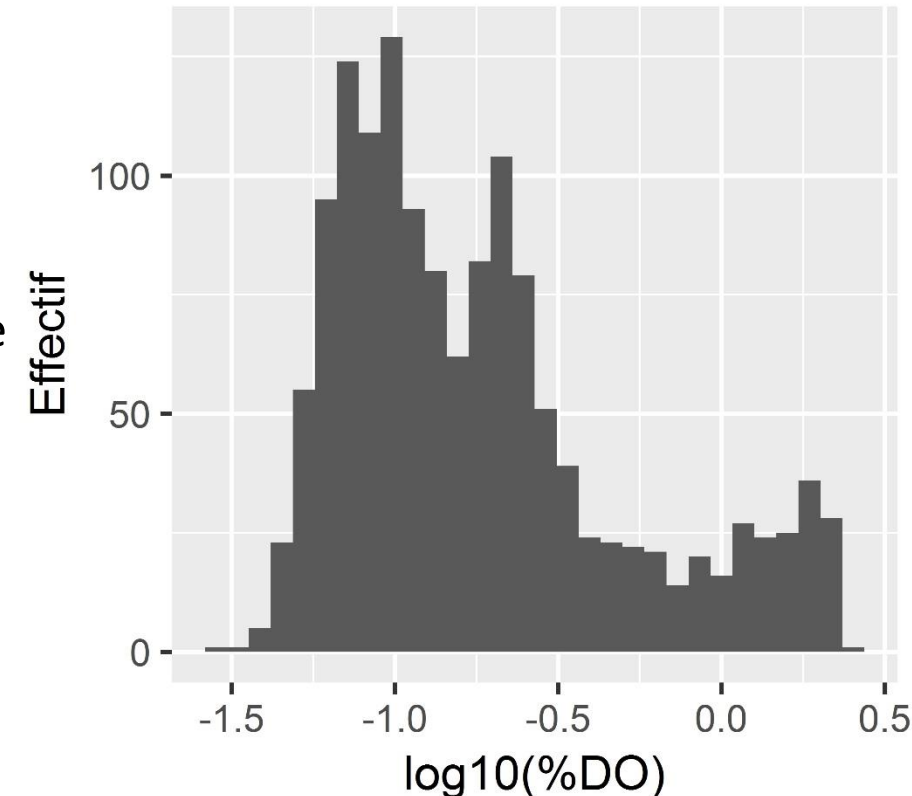
Courbe ROC bi-normale



Limites et extension

- Si un seul test
 - Estimation de la **distribution chez les négatifs et les positifs au test** et pas nécessairement de celle des malades et non malades (*ex : sérologie mesure la présence d'une réaction immunitaire et pas directement la maladie*)
- Nécessite une « **normalisation** » des données
 - Pas toujours possible (*cf. figure fièvre Q bovin*)
 - Possibilité de prendre en compte des covariables comme l'âge (*Gay, 1996*) ou la durée de lactation (*Nielsen et al., 2007*)
- Autres approches paramétriques proposées
 - Plusieurs distributions latentes (*Baughman et al., 2006*)
 - Distributions Gamma (*Jafarzadeh et al., 2010*)
- Approches non paramétriques
 - Mélange semi-paramétrique (*Branscum et al., 2008*)
 - Processus de Dirichlet (*Ladouceur et al., 2011*)

Sérologie FQ bovins



Données, Elodie Rousset, LNR Sophia Antipolis, ANSES

Si plusieurs tests quantitatifs (*Choi et al., 2006*)

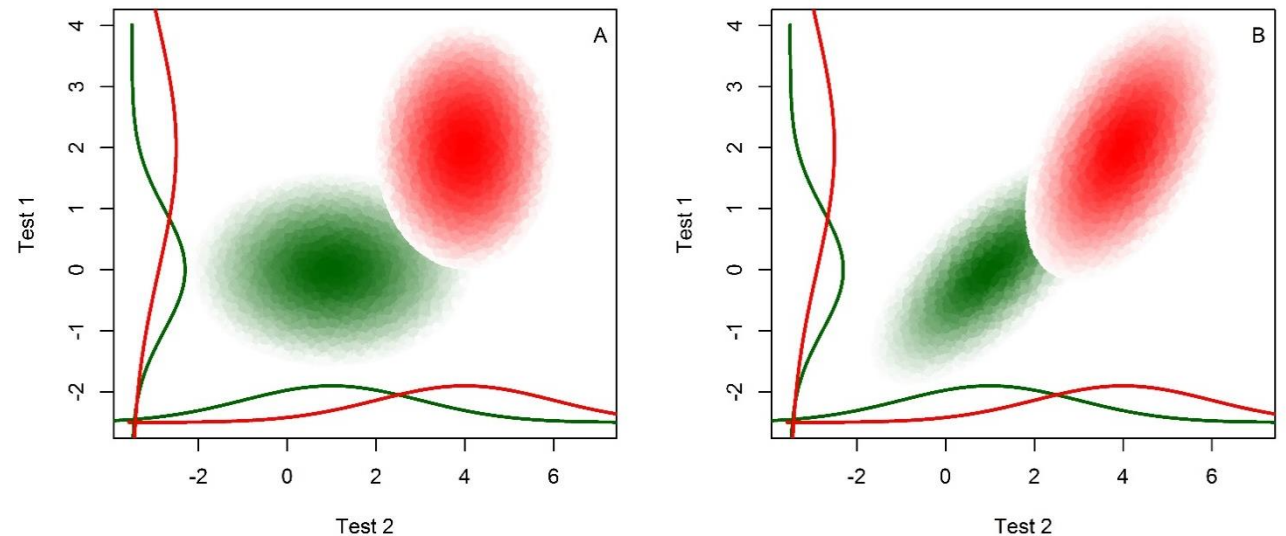
- Approche facilement transposable avec des lois multi-normales
- Possibilité de prendre en compte la dépendance conditionnelle entre les deux tests

$M_i \sim \text{Bernoulli}(P)$

$$T_{iM^+} = \begin{pmatrix} T_{1iM^+} \\ T_{2iM^+} \end{pmatrix} \sim N_2(\mu_{M^+}, \Sigma_{M^+}), \quad T_{iM^-} = \begin{pmatrix} T_{1iM^-} \\ T_{2iM^-} \end{pmatrix} \sim N_2(\mu_{M^-}, \Sigma_{M^-});$$

$$\mu_{M^+} = \begin{pmatrix} \mu_{1M^+} \\ \mu_{2M^+} \end{pmatrix} \quad \Sigma_{M^+} = \begin{pmatrix} \sigma_{1M^+}^2 & \rho_{12M^+} \sigma_{1M^+} \sigma_{2M^+} \\ \rho_{12M^+} \sigma_{1M^+} \sigma_{2M^+} & \sigma_{2M^+}^2 \end{pmatrix}$$

$$\mu_{M^-} = \begin{pmatrix} \mu_{1M^-} \\ \mu_{2M^-} \end{pmatrix} \quad \Sigma_{M^-} = \begin{pmatrix} \sigma_{1M^-}^2 & \rho_{12M^-} \sigma_{1M^-} \sigma_{2M^-} \\ \rho_{12M^-} \sigma_{1M^-} \sigma_{2M^-} & \sigma_{2M^-}^2 \end{pmatrix}$$



- Même contrainte de normalité des distributions latentes
- A ce jour, aucune approche non paramétrique ne permet de prendre en compte la dépendance conditionnelle (*Ladouceur et al., 2011*)

Et pour un cas mêlant test binaire et test quantitatif?

- Possibilité de combiner les deux approches (*Branscum et al., 2008; Scott et al., 2008; Wang et al., 2006*)
- Avec les mêmes notations :

$$M_i \sim \text{Bernoulli}(P)$$

$$(T_{1i} | M_i = 1) \sim N(\mu_{M^+}, \sigma_{M^+}) ; (T_{1i} | M_i = 0) \sim N(\mu_{M^-}, \sigma_{M^-})$$

$$(T_{2i} | M_i = 1) \sim \text{Bernoulli}(Se_{T_2}) ; (T_{2i} | M_i = 0) \sim \text{Bernoulli}(1 - Sp_{T_2})$$

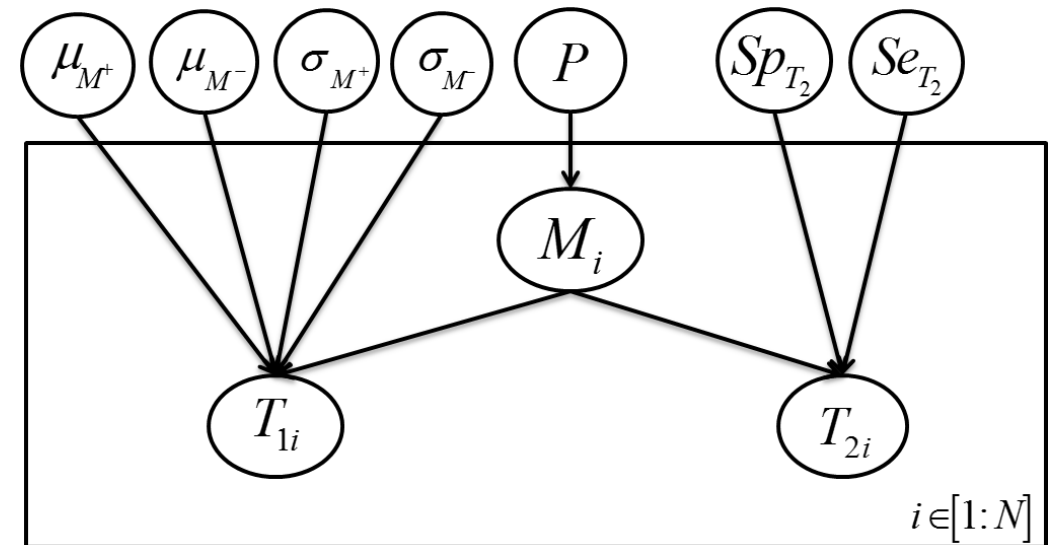
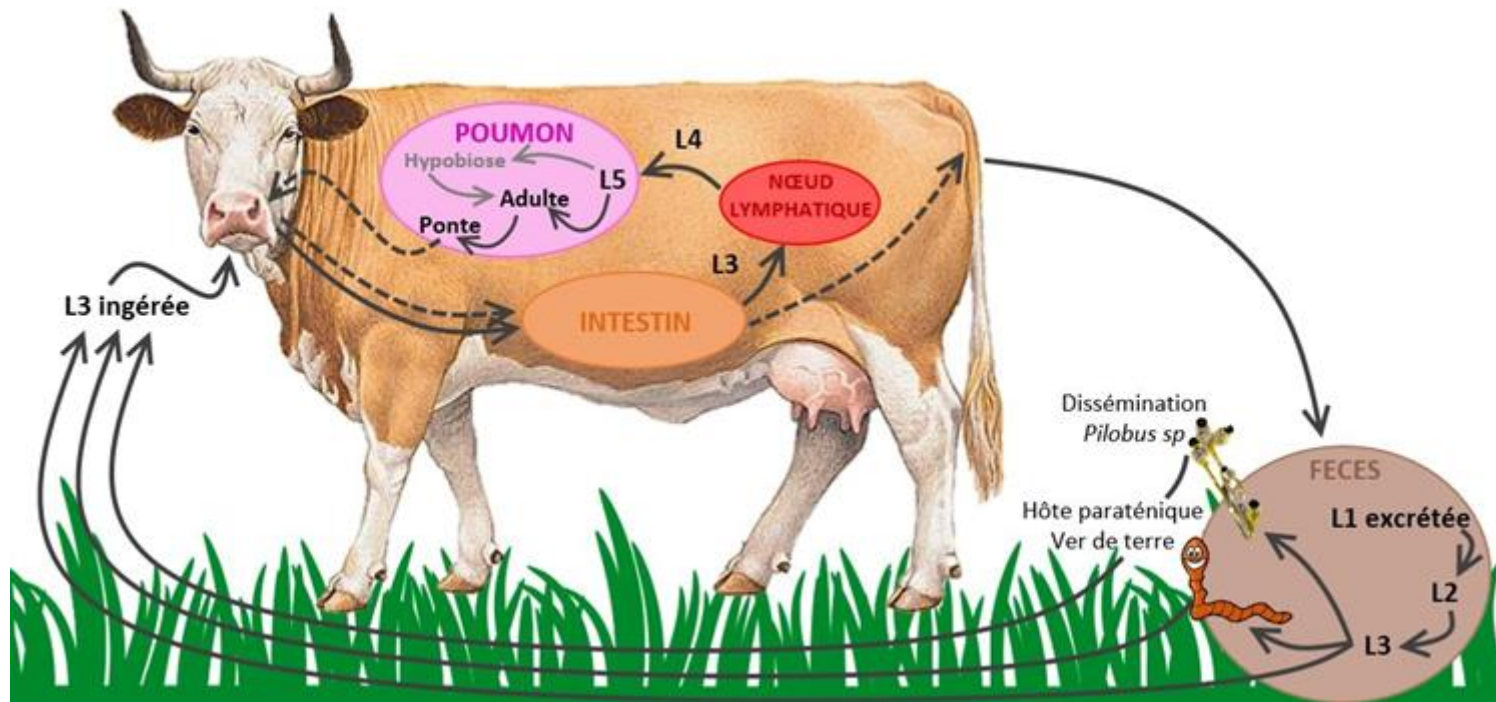


Illustration sur la dictyocaulose (*Lurier et al., 2018*)

- Dictyocaulose = Maladie parasitaire pulmonaire des bovins



(Source : M.A Arcangioli, pathologie du bétail, Vetagro Sup)

Distinction entre infestation pré-patente et patente

Méthodes diagnostiques

- Coproscopie de Baermann (BS), méthode de routine

Sp = 100%, Se inconnue (Eysker et al. 1997)

- Sérologie (ODR), pas encore disponible en France

Se et Sp non évaluées en conditions réelles, ou avec BS en gold standard, antigène spécifique des adultes

Seuil fixé « arbitrairement » (Von Holtum et al. 2008)

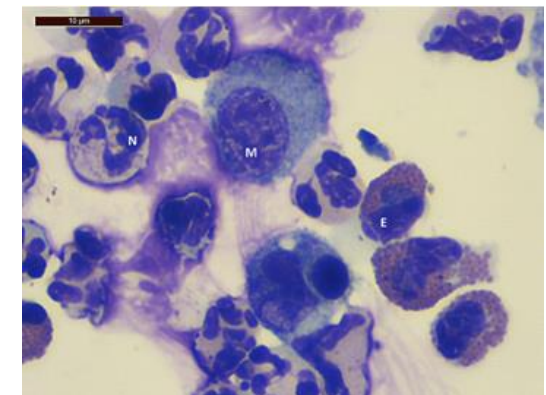
Nouvelles méthodes explorées

- Mise en évidence du parasite dans le liquide de lavage broncho-alvéolaire (BAL)

Sp = 100%

- Cytologie du liquide de LBA : Eosinophilie pulmonaire (EP)
 - Comptage des éosinophiles sur 400 cellules EC

Pas de gold standard et seuils de positivité à définir pour EP et ODR, mesures de statuts différents



Caractérisation des méthodes et estimation des seuils de positivité

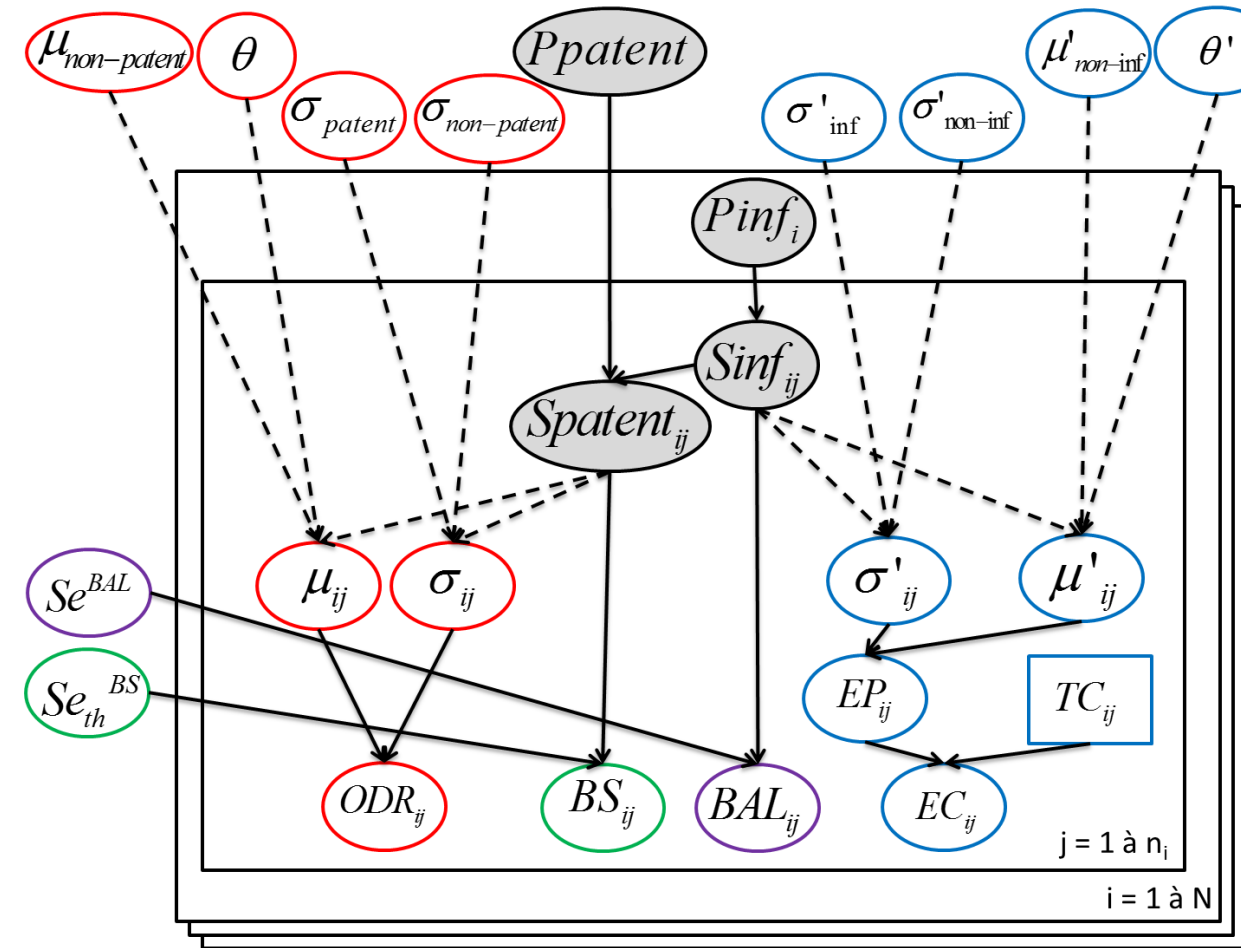
Données

- 11 élevages suspects de dictyocaulose (5 à 6 vaches par troupeau)
+ 10 vaches de TP = 70 vaches
- Pour chaque individu, observation des résultats des quatre méthodes :
BS - **ODR** - **BAL** - **EC**

Modélisation simultanée de l'ensemble des données

- Modèle à 3 classes latentes (**non infesté, pré-patent, patent**)
- Prise en compte simultanée des données qualitatives (**BS** et **BAL**) et quantitatives (**ODR** et **EC**)
- Estimation des seuils de positivité via courbes ROC

Modèle 1 = modèle de base

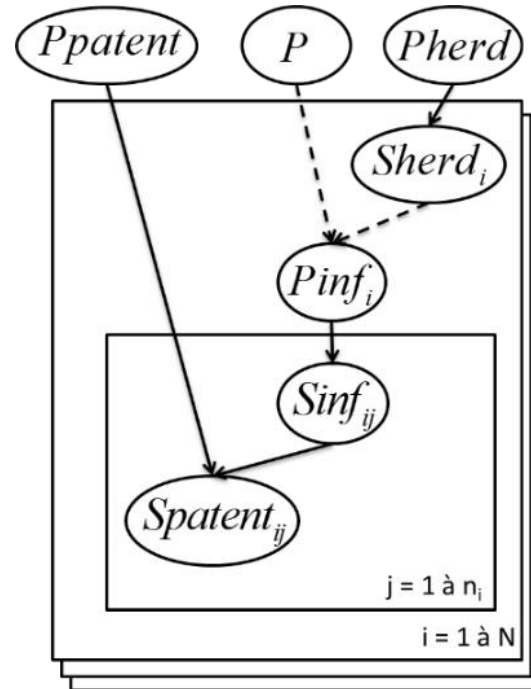


Nœuds	Type	Définition
$S_{inf_{ij}}$	Stochastique	$S_{inf_{ij}} \sim \text{Bern}(P_{inf_i})$
$S_{patent_{ij}}$	Stochastique	$S_{patent_{ij}} \sim \text{Bern}(P_{patent} \times S_{inf_{ij}})$
μ_{ij}	Logique	$\mu_{ij} = \mu_{non_patent} + \theta \times S_{patent_{ij}}$
σ_{ij}	Logique	$\sigma_{ij} = \sigma_{non_patent} \times (1 - S_{patent_{ij}}) + \sigma_{patent} \times S_{patent_{ij}}$
μ'_{ij}	Logique	$\mu'_{ij} = \mu_{non_inf} + \theta' \times S_{inf_{ij}}$
σ'_{ij}	Logique	$\sigma'_{ij} = \sigma_{non_inf} \times (1 - S_{inf_{ij}}) + \sigma_{inf} \times S_{inf_{ij}}$
ODR_{ij}	Stochastique	$\log_{10}(ODR_{ij}) \sim N(\mu_{ij}, \sigma_{ij})$
EP_{ij}	Stochastique	$\text{logit}(EP_{ij}) \sim N(\mu'_{ij}, \sigma'_{ij})$
EC_{ij}	Stochastique	$EC_{ij} \sim \text{Binom}(TC_{ij}, EP_{ij})$
BS_{ij}	Stochastique	$BS_{ij} \sim \text{Bern}(Se_{th}^{BS} \times S_{patent_{ij}})$
BAL_{ij}	Stochastique	$BAL_{ij} \sim \text{Bern}(Se^{BAL} \times S_{inf_{ij}})$

Modèles alternatifs : 2 exemples parmi 6

- Modèle 3

Prévalence « Zero inflated »



$$Sherd_i \sim \text{Bern}(Pherd)$$

$$Pinf_i = P \times Sherd_i$$

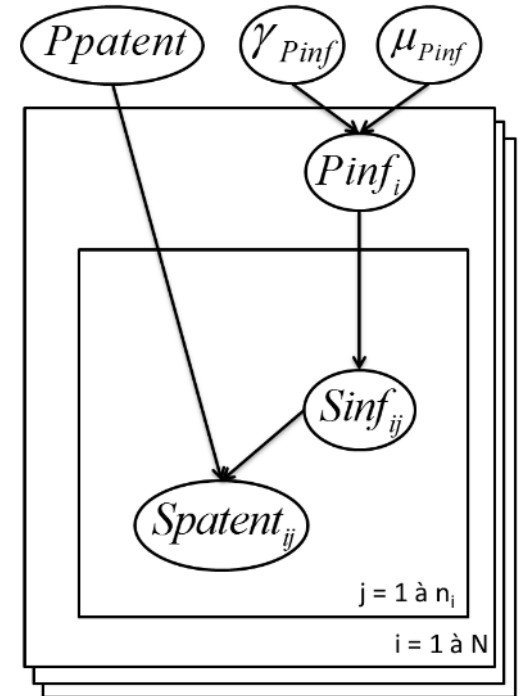
$$Sinf_{ij} \sim \text{Bern}(Pinf_i)$$

$$Spatent_{ij} \sim \text{Bern}(Ppatent \times Sinf_{ij})$$

- Modèle 4

Hierarchique beta-binomial (Dolzhenko and Smith, 2014)

$$\mu = \frac{\alpha}{\alpha + \beta} \quad \gamma = \frac{1}{\alpha + \beta + 1}$$



$$Pinf_i \sim \text{beta} \left(\mu_{Pinf} \times \frac{(1 - \gamma_{Pinf})}{\gamma_{Pinf}}, (1 - \mu_{Pinf}) \times \frac{(1 - \gamma_{Pinf})}{\gamma_{Pinf}} \right)$$

$$Sinf_{ij} \sim \text{Bern}(Pinf_i)$$

$$Spatent_{ij} \sim \text{Bern}(Ppatent \times Sinf_{ij})$$

Prior et inférence

- Priors peu informatifs
- Inférence avec **JAGS**
 - 3 chaînes indépendantes avec initialisations différentes
 - Phase de chauffe de 10 000 itérations
 - Puis 100 000 tirages avec thin de 20 pour éviter l'autocorrélation
- Convergence vérifiée avec l'examen des traces
- Autocorrélation avec la statistique de Gelman et Rubin modifiée (Brooks and Gelman, 1998) et le diagnostic de Raftery (Raftery and Lewis, 1992)

Nœuds	Prior
$\mu_{Pinf}, \mu_{Ppatent}$	unif (0,1)
$\gamma_{Pinf}, \gamma_{Ppatent}$	unif (0,1)
$P_{inf_i}, P_{patent}, Se_{th}^{BS}, Se^{BAL}$	beta $\left(\frac{1}{2}, \frac{1}{2}\right)$
$\mu_{non_patent}, \mu_{non_inf}$	unif (-10,10)
θ, θ'	unif (0,10)
$\sigma_{non_patent}, \sigma_{non_inf}, \sigma_{patent}, \sigma_{inf}$	unif (0,5)

Sélection des modèles

- Calcul du DIC non réalisable avec JAGS (dans ce cas)
- Posterior Predictive Check
- Utilisation de **score de prédiction**
 - Variable binaire : **Score de Brier** (*Brier, 1950*)
 - Variable quantitative : **CRPS (Continuous Ranked Probability Score)**
(*Gneiting and Raftery, 2007; Gschlößl and Czado, 2007; Krüger et al., 2016; Matheson and Winkler, 1976*)

Score de Brier

- Score pour variable binaire :

$$BRIER(Y_k) = -\left(\left(P_{Y_k pred+} - I(Y_k +) \right)^2 + \left(P_{Y_k pred-} - I(Y_k -) \right)^2 \right)$$

- Si l'évènement est observé alors :

$$BRIER(Y_k^+) = -2P_{Y_k pred-}^2$$

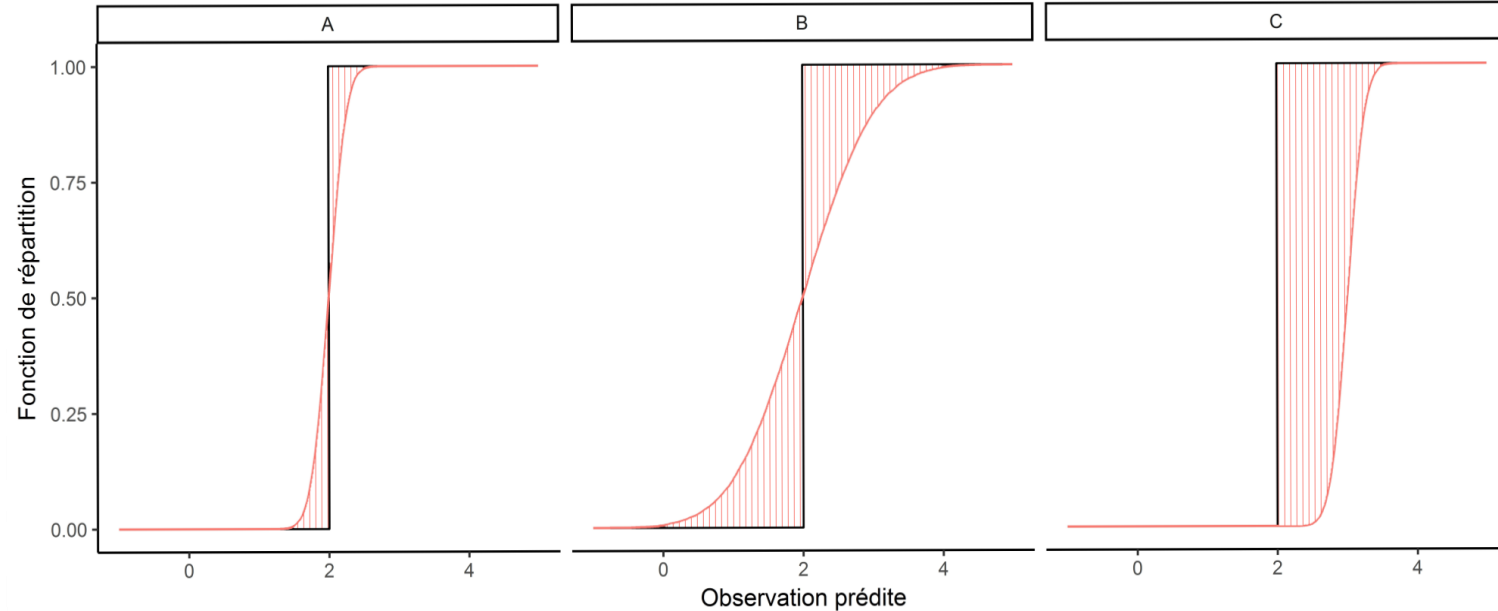
- Sinon

$$BRIER(Y_k^-) = -2P_{Y_k pred+}^2$$

Il est d'autant plus grand que la **probabilité prédite de survenue** de l'évènement est grande pour les évènements qui sont observés et qu'elle est faible pour les évènements non observés.

CRPS

- Pour chaque observation
 - Surface de la différence entre la fonction de répartition prédite et l'indicatrice de l'observation
 - Plusieurs façons de la calculer (notamment Gschlößl and Czado, 2007)
 - Facilement calculable à partir des MCMC



$$CRPS(P_\theta, Y_k) = -\int_{-\infty}^{\infty} (F_k(y|P_\theta) - I(y \geq Y_k))^2 dy$$

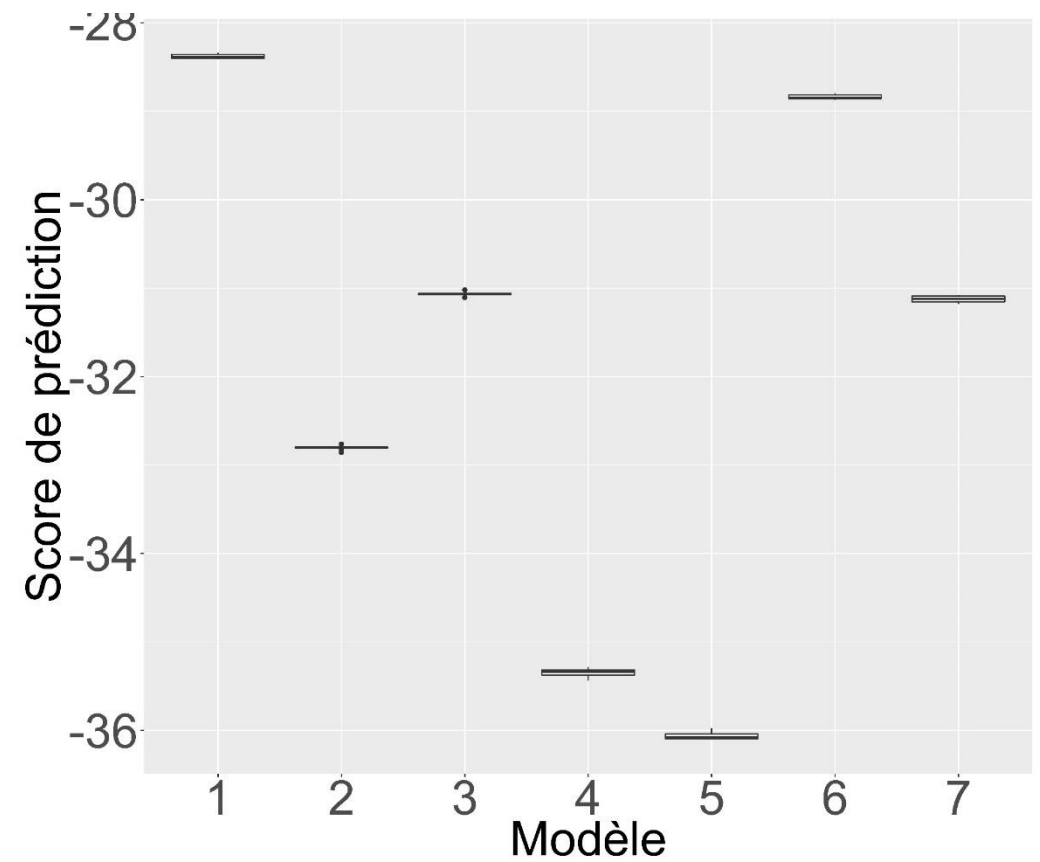
$$CRPS(P_\theta, Y_k) = \frac{1}{2} \mathbb{E} |Y_{pred1,k} - Y_{pred2,k}| - \mathbb{E} |Y_{pred1,k} - Y_k|$$

$$CRPS(P_\theta, Y_k) = \frac{1}{2} \times \frac{1}{Nmcmc} \times \sum_{m=1}^{Nmcmc} |y_{pred1,k}^m - y_{pred2,k}^m| - \frac{1}{Nmcmc} \times \sum_{m=1}^{Nmcmc} |y_{pred1,i}^m - Y_k|$$

Résultats

- Calcul pour chaque variable observée puis somme
 - 10 calculs pour chaque modèle
 - Variabilité relativement faible
 - Pas de modèle alternatif décrivant mieux les données que le modèle 1
- ⇒ Conservation du modèle 1

Modèle	<i>BS</i>	<i>BAL</i>	<i>ODR</i>	<i>EC</i>
1	-3.74	-12.62	-7.84	-4.14
2	-3.88	-14.19	-9.79	-4.95
3	-3.87	-13.61	-9.41	-4.21
4	-3.89	-14.26	-11.42	-5.88
5	-3.87	-14.82	-11.36	-6.04
6	-3.68	-13.08	-7.88	-4.14
7	-3.80	-13.07	-10.09	-4.13



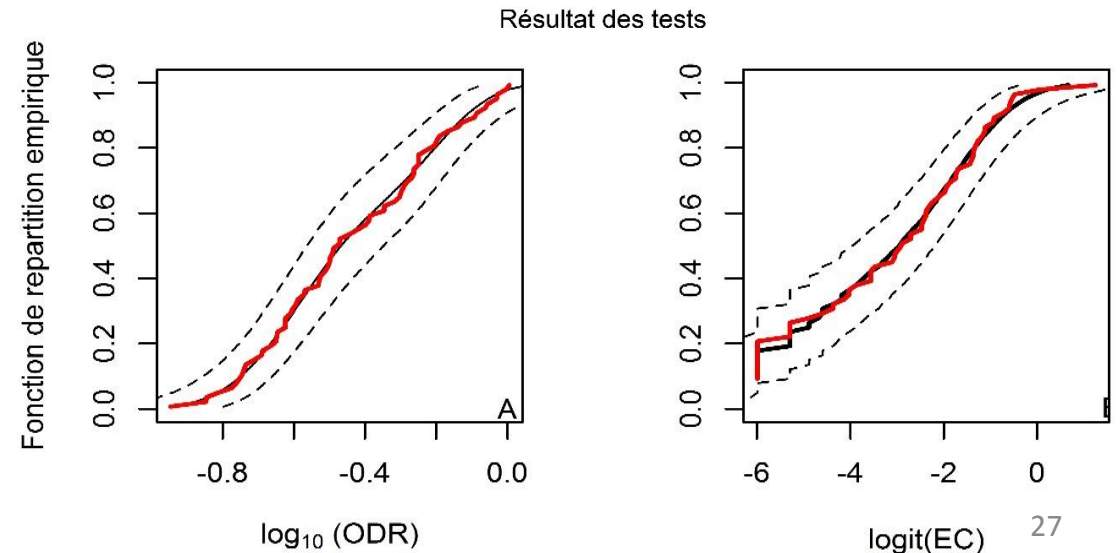
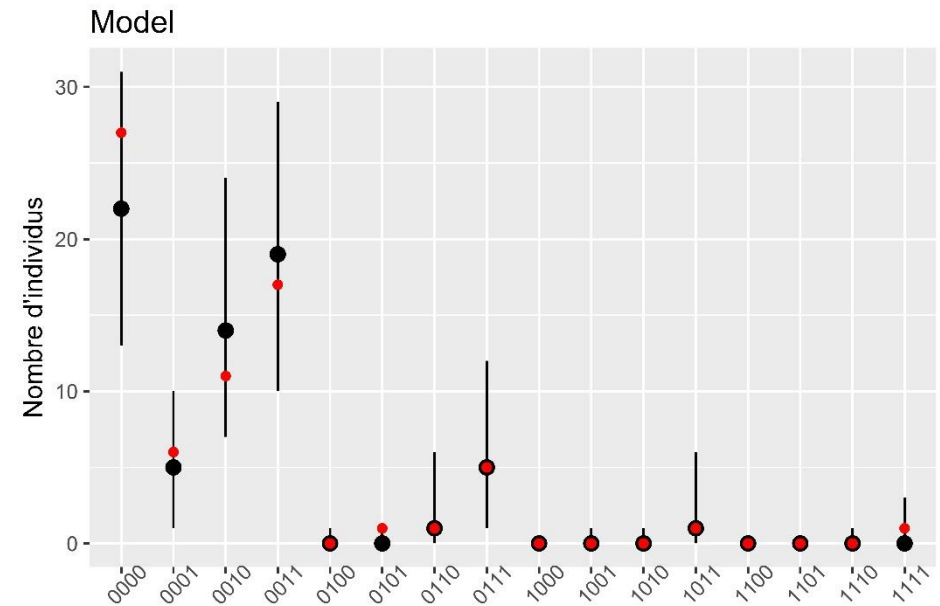
Validation : Posterior Predictive Check

- Nombre d'individus dans chacune des 16 combinaisons des résultats des 4 tests

⇒ Pas de dépendance conditionnelle

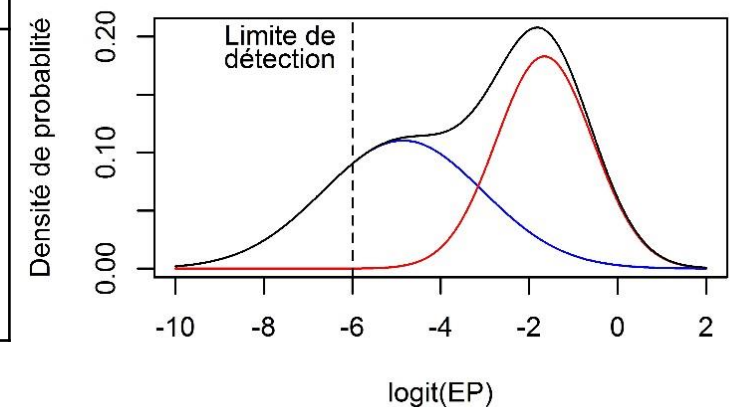
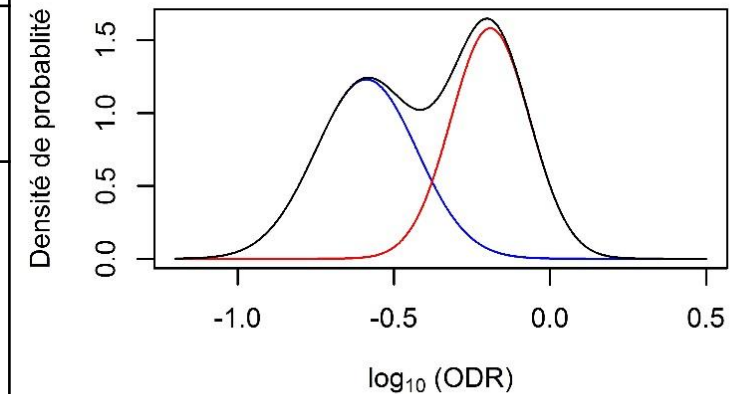
- Fonctions de répartition empiriques pour les deux variables quantitatives

⇒ Bonne adéquation des mélanges de bi normaux



Estimation des paramètres du modèle

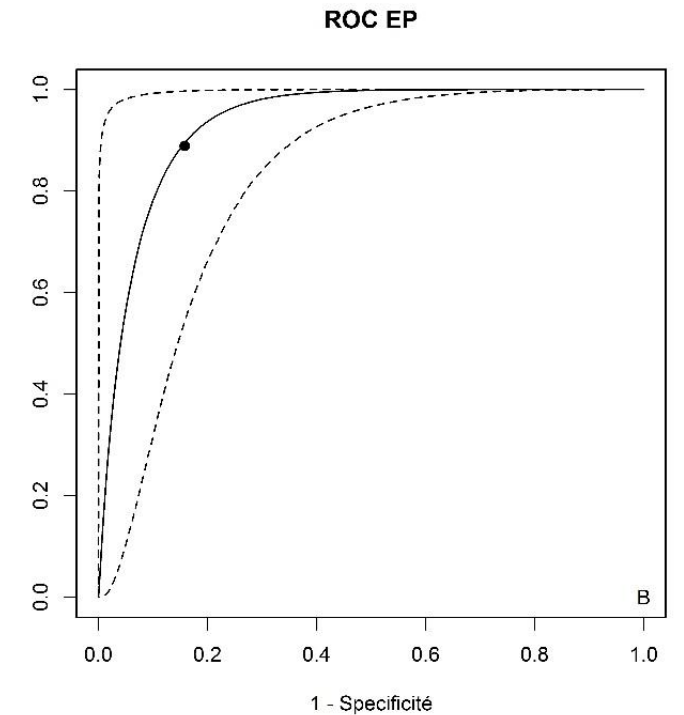
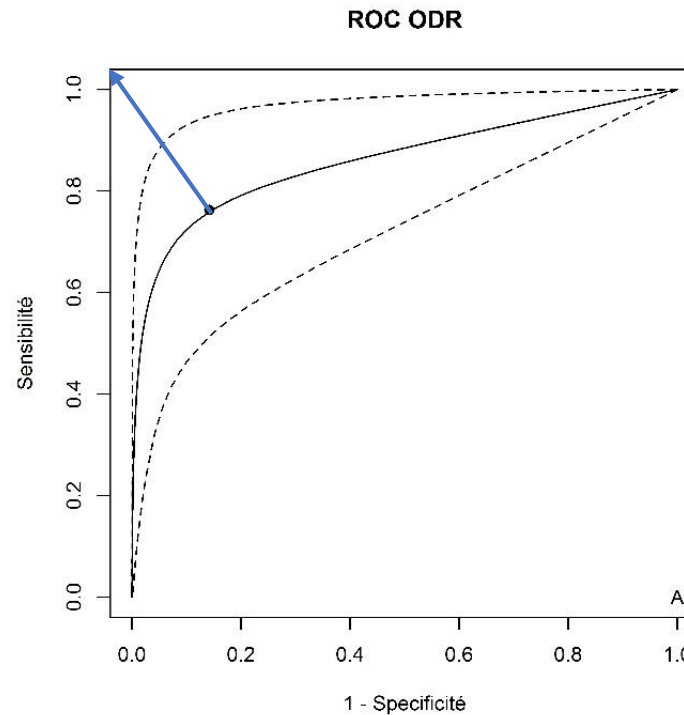
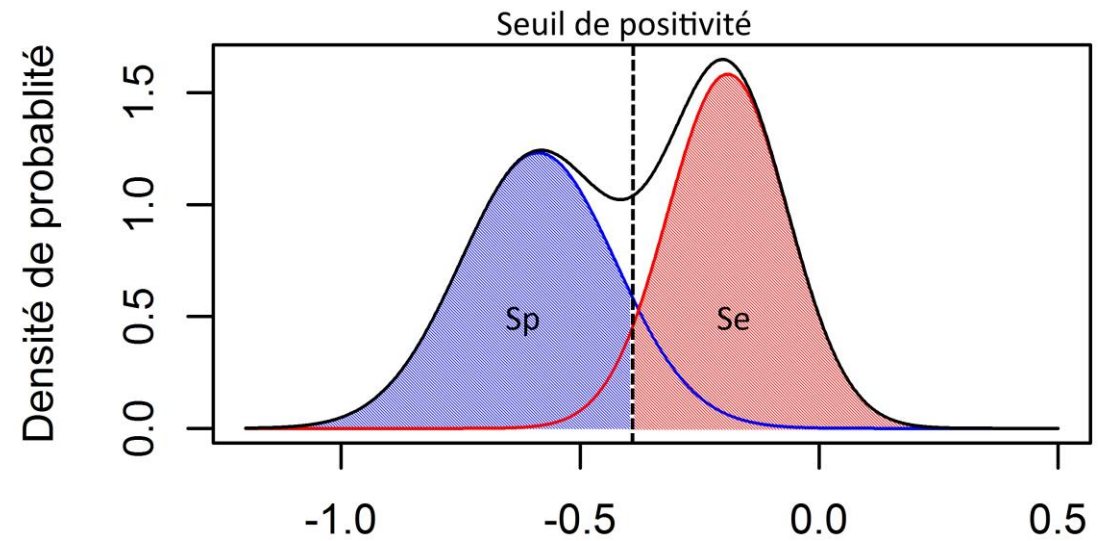
	Paramètres	Estimation <i>a posteriori</i> Médiane [IC à 95%]
Sensibilités des tests	Se^{BS}	0.063 [0.012; 0.18]
	Se^{BAL}	0.241 [0.113; 0.435]
Paramètres du mélange de distribution de ODR	μ_{non_patent}	-0.588 [-0.642; -0.527]
	θ	0.397 [0.301; 0.469]
	σ_{non_patent}	0.162 [0.128; 0.212]
	σ_{patent}	0.126 [0.088; 0.214]
Paramètres du mélange de distribution de EP	μ_{non_inf}	-4.852 [-5.871; -3.954]
	θ'	3.200 [2.264; 4.138]
	σ_{inf}	1.806 [0.870; 2.589]
	σ_{non_inf}	1.090 [0.8401; 1.470]



Estimation des prévalences très variable d'un élevage à l'autre (de 4 à 95%) et imprécise (intervalles de crédibilité à 95% très larges)

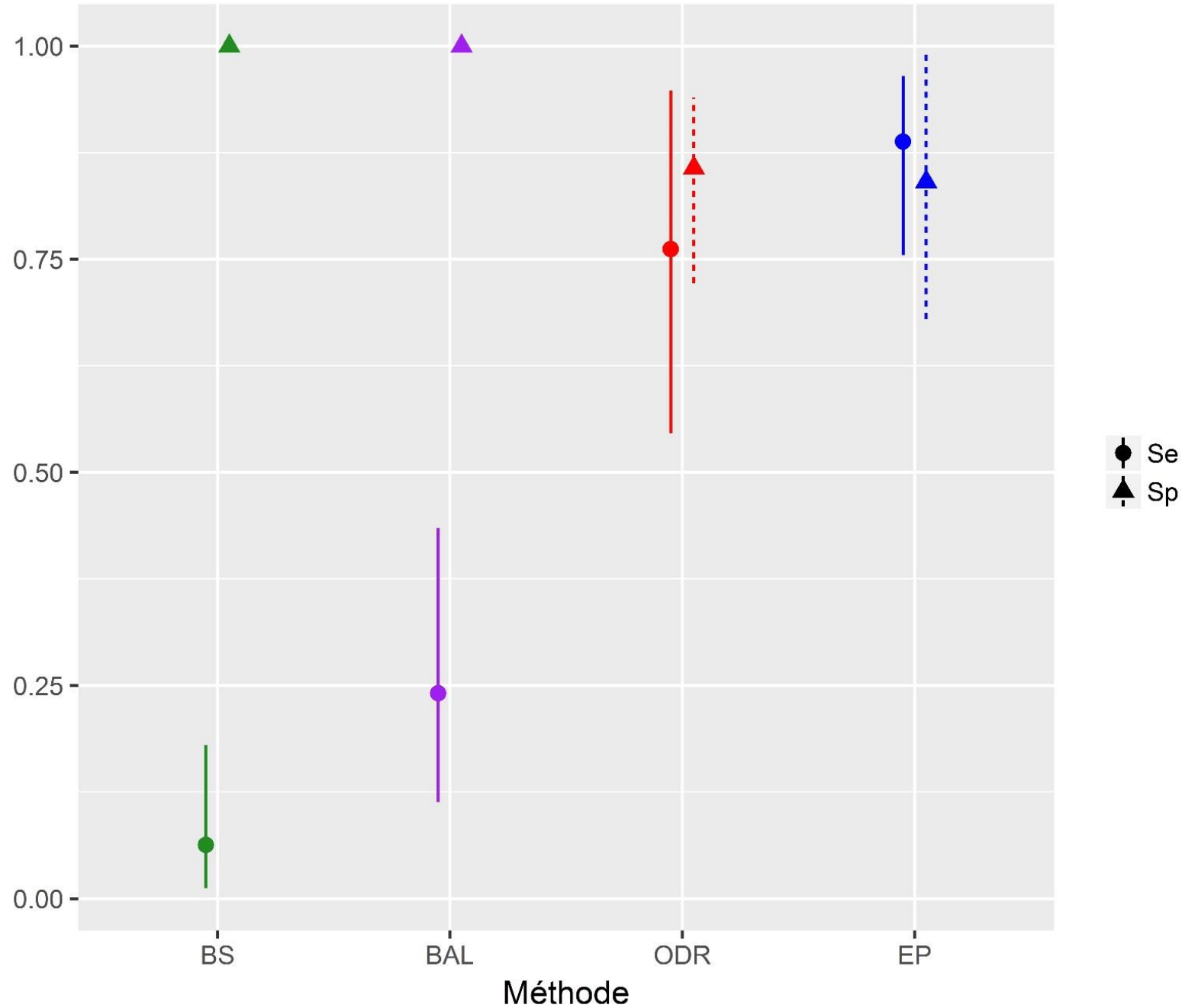
Courbe ROC - ODR et EP

- Estimation des courbes ROC
- Seuil optimal
- ODR
 - Seuil = 0.389 ODR
 - Se=76.2% [54.6; 94.8]
 - Sp = 85.7% [72.2; 94.0]
- EP
 - Seuil = 4.78%
 - Se = 88.8% [75.5; 96.5]
 - Sp = 84.1% [68.0; 99.7]



Résultats

- Sensibilité et spécificité des tests



Application

- Calcul des valeurs prédictives positives et négatives ainsi que de leur **intervalle de crédibilité !**

- D'après le théorème de Bayes $VPP = P(M^+ | T^+) = \frac{P(T^+ | M^+) \times P(M^+)}{P(T^+)}$

$$VPP = P(M^+ | T^+) = \frac{Se \times P}{Se \times P + (1 - P) \times (1 - Sp)} \quad VPN = P(M^- | T^-) = \frac{Sp \times (1 - P)}{(1 - Se) \times P + (1 - P) \times Sp}$$

- Calcul de la **sensibilité troupeau**, si on applique n fois le test dans un troupeau :

$$Se^T = 1 - (1 - Se)^n \quad Sp^T = Sp^n$$

- Et beaucoup d'autres indicateurs **directement calculables à partir de distribution a posteriori**

Travaux en cours

- Données sérologiques sur la fièvre Q de la plateforme ESA, en partenariat le LNR fièvre Q (ANSES Sophia Antipolis)
 - Maladie abortive des ruminants transmissible à l'homme
 - 3 kits diagnostiques sérologiques appliqués sur 1500 ovins, 1500 caprins et 1500 bovins
 - Sous échantillon d'une étude épidémiologique plus large sur 20 000 individus testés avec un seul kit
- ⇒ Estimer les Se et Sp de chaque kit pour chaque espèce
- ⇒ Utiliser les résultats pour approfondir l'étude épidémiologique

Données

- Sérologie mesurée sur plaque avec 96 sérums analysés comprenant
 - 2 contrôles positifs et négatifs propres au kit
 - 2 sérums de référence communs à toutes les plaques

plaque 1

	1	2	3	4	5	6	7	8	9	10	11	12
A		Bovin 1	Bovin 9	Bovin 17	Bovin 25	Bovin 33	Bovin 41	Bovin 49	Bovin 57	Bovin 65	Bovin 73	Bovin 81
B	MRE 1/1	Bovin 2	Bovin 10	Bovin 18	Bovin 26	Bovin 34	Bovin 42	Bovin 50	Bovin 58	Bovin 66	Bovin 74	Bovin 82
C	MRE 1/2	Bovin 3	Bovin 11	Bovin 19	Bovin 27	Bovin 35	Bovin 43	Bovin 51	Bovin 59	Bovin 67	Bovin 75	Bovin 83
D		Bovin 4	Bovin 12	Bovin 20	Bovin 28	Bovin 36	Bovin 44	Bovin 52	Bovin 60	Bovin 68	Bovin 76	Bovin 84
E	Neg kit	Bovin 5	Bovin 13	Bovin 21	Bovin 29	Bovin 37	Bovin 45	Bovin 53	Bovin 61	Bovin 69	Bovin 77	Bovin 85
F	Neg kit	Bovin 6	Bovin 14	Bovin 22	Bovin 30	Bovin 38	Bovin 46	Bovin 54	Bovin 62	Bovin 70	Bovin 78	Bovin 86
G	POS kit	Bovin 7	Bovin 15	Bovin 23	Bovin 31	Bovin 39	Bovin 47	Bovin 55	Bovin 63	Bovin 71	Bovin 79	Bovin 87
H	POS kit	Bovin 8	Bovin 16	Bovin 24	Bovin 32	Bovin 40	Bovin 48	Bovin 56	Bovin 64	Bovin 72	Bovin 80	Bovin 88

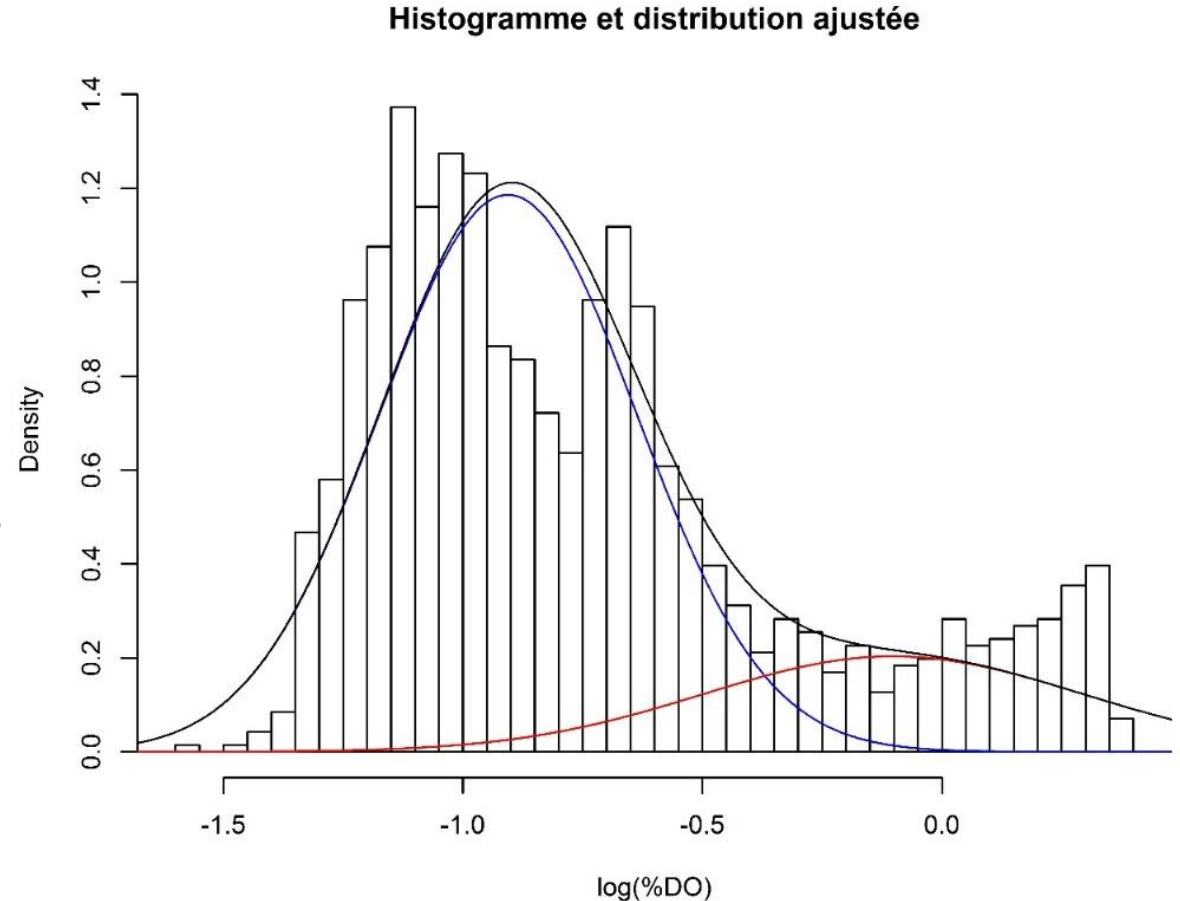
- Résultats en % de DO :

$$\% DO = \frac{DO_{ech} - \left(\frac{DO_{neg_1} + DO_{neg_2}}{2} \right)}{\left(\frac{DO_{pos_1} + DO_{pos_2}}{2} \right) - \left(\frac{DO_{neg_1} + DO_{neg_2}}{2} \right)}$$

Plan de plaque ELISA

Modélisation des données

- Difficulté pour « normaliser » les DO relatives rendues par l'appareil (%DO)
 - Phénomène de saturation lié à la réaction ELISA
 - Nécessité de prendre en compte la dépendance conditionnelle entre les 3 kits
- ⇒ Modèle multi-normal



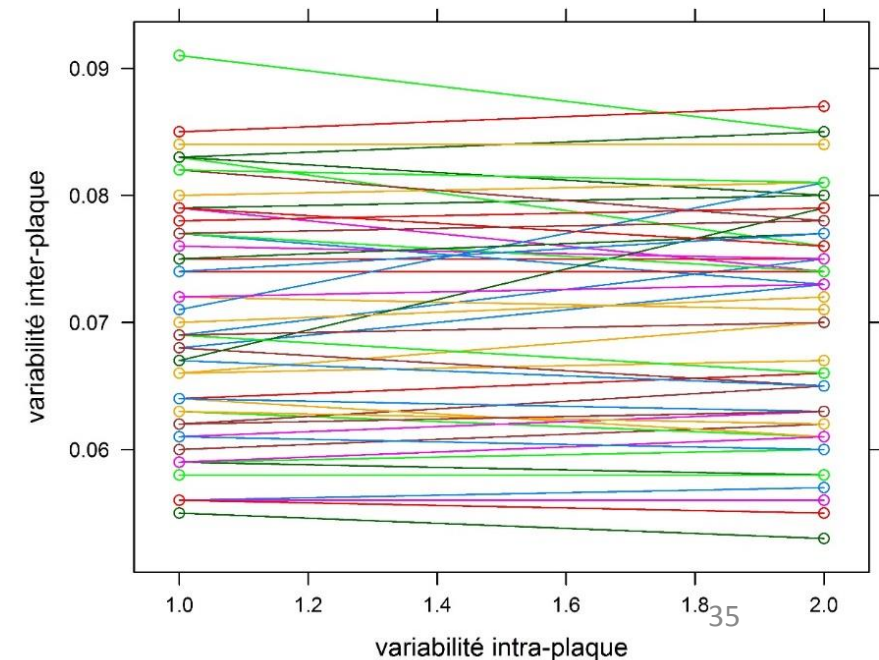
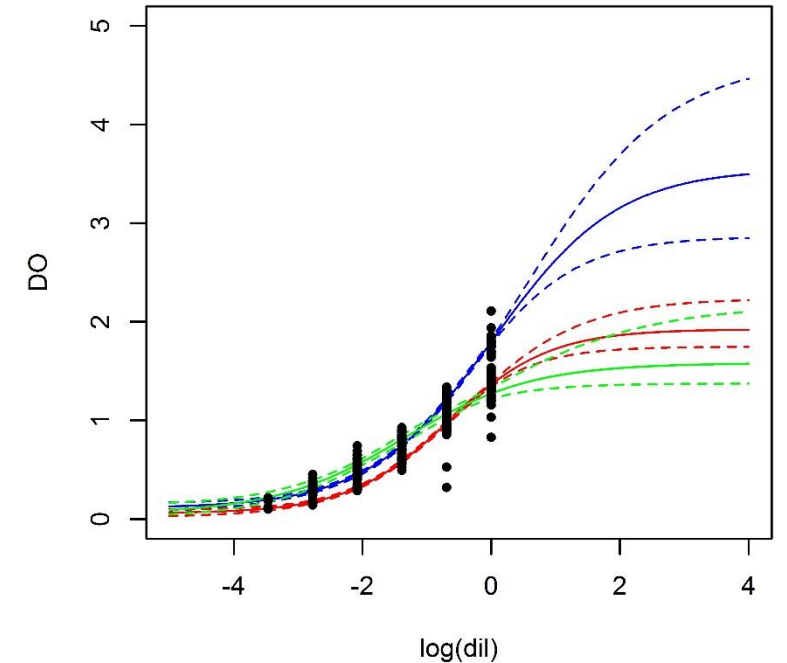
Transformation des données

- Récupération des DO mesurées brutes pour transformation log-logit (*Schick et Schwack 2017, Huet et al. 1988*)

$$DO = \theta_1 + \frac{\theta_2}{1 + \exp(\theta_3 \times (\ln(x) - \theta_4))}$$

- Estimation des paramètres avec un sérum analysé à 6 dilutions successives
- Pas de sérum « hyper-concentré » pour estimer le plateau supérieur (θ_2) mais utilisation de priors possible à partir des données tout venant
- Variabilité inter-plaque importante à prendre en compte

⇒ Modèle hiérarchique



Merci de votre attention

Et aux différents partenaires des projets présentés

Gilles Bourgoïn et Marie Anne Arcangioli pour l'étude sur la
Dictyocaulose

Elodie Rousset (LNR fièvre Q) et plateforme ESA



- Branscum, A.J., Johnson, W.O., Hanson, T.E., Gardner, I.A., 2008. Bayesian semiparametric ROC curve estimation and disease diagnosis. *Stat. Med.* 27, 2474–2496.
- Brier, G.W., 1950. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* 78, 1–3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2)
- Choi, Y.-K., Johnson, W.O., Collins, M.T., Gardner, I.A., 2006. Bayesian inferences for receiver operating characteristic curves in the absence of a gold standard. *J. Agric. Biol. Environ. Stat.* 11, 210–229.
- Collins, J., Huynh, M., 2014. Estimation of diagnostic test accuracy without full verification: a review of latent class methods. *Stat. Med.* 33, 4141–4169.
- Dendukuri, N., Hadgu, A., Wang, L., 2009. Modeling conditional dependence between diagnostic tests: A multiple latent variable model. *Stat. Med.* 28, 441–461.
- Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. *Stat. Comput.* 24, 997–1016.
- Gneiting, T., Raftery, A.E., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Stat. Assoc.* 102, 359–378.
- Gschlößl, S., Czado, C., 2007. Spatial modelling of claim frequency and claim size in non-life insurance. *Scand. Actuar. J.* 2007, 202–225.
- Hanley, J.A., 1996. The use of the “binormal” model for parametric ROC analysis of quantitative diagnostic tests. *Stat. Med.* 15, 1575–1585.
- Hui, S.L., Walter, S.D., 1980. Estimating the error rates of diagnostic tests. *Biometrics* 36, 167–171.
- Jafarzadeh, S.R., Johnson, W.O., Utts, J.M., Gardner, I.A., 2010. Bayesian estimation of the receiver operating characteristic curve for a diagnostic test with a limit of detection in the absence of a gold standard. *Stat. Med.* 29, 2090–2106.
- Ladouceur, M., Rahme, E., Bélisle, P., Scott, A.N., Schwartzman, K., Joseph, L., 2011. Modeling continuous diagnostic test data using approximate Dirichlet process distributions. *Stat. Med.* 30, 2648–2662.
- Marin, J.-M., Mengersen, K., Robert, C.P., 2005. Bayesian Modelling and Inference on Mixtures of Distributions, in: *Handbook of Statistics*. Elsevier, pp. 459–507.
- Matheson, J.E., Winkler, R.L., 1976. Scoring Rules for Continuous Probability Distributions. *Manag. Sci.* 22, 1087–1096.
- Nielsen, S.S., Toft, N., Jørgensen, E., Bibby, B.M., 2007. Bayesian mixture models for within-herd prevalence estimates of bovine paratuberculosis based on a continuous ELISA response. *Prev. Vet. Med.* 81, 290–305.
- Plummer, M., Stukalov, A., Denwood, M., 2016. *rjags: Bayesian graphical models using MCMC*.
- Raftery, A.E., Lewis, S.M., 1992. [Practical Markov Chain Monte Carlo]: Comment: One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo. *Stat. Sci.* 7, 493–497.
- Scott, A.N., Joseph, L., Bélisle, P., Behr, M.A., Schwartzman, K., 2008. Bayesian modelling of tuberculosis clustering from DNA fingerprint data. *Stat. Med.* 27, 140–156.
- Wang, C., Turnbull, B.W., Gröhn, Y.T., Nielsen, S.S., 2006. Estimating receiver operating characteristic curves with covariates when there is no perfect reference test for diagnosis of Johne’s disease. *J. Dairy Sci.* 89, 3038–3046.

