HARMONY

Novel tools for test evaluation and disease prevalence estimation

Identifiability Issues in Bayesian Latent Class Analysis of Diagnostic Test Data

Geoff Jones

Massey University, NZ

cost

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

COST is supported by the EU Framework Programme Horizon 2020

Toxoplasmosis – Pig Data

- From a study of the accuracy of serological tests for detection of infected pigs, published by Dubey et al. (1995).
- Paired serum and heart samples collected from sows slaughtered at a single abattoir in Iowa;
- Three tests on each of 998 animals:
  - modified agglutination test (MAT);
  - enzyme-linked immunoassay (ELISA);
  - mouse bioassay (MB);
- Sampling and testing done in two batches (two populations?)

	MAT	ELISA	MB
Test+	222	241	107
Test–	776	757	891
Total	998	998	998

	MB+		MB–	
	ELISA+	ELISA–	ELISA+	ELISA–
MAT+	73	17	91	41
MAT–	4	13	73	686

Dubey, J.P., Thulliez, P.,Weigel, R.M., Andrews, C.D., Lind, P., Powell, E.C., 1995. Sensitivity and specificity of various serologic tests for detection of *Toxoplasma gondii* infection in naturally infected sows. *Am. J. Vet. Res*. 56, 1030-1036.

MAT Only

Data (x)

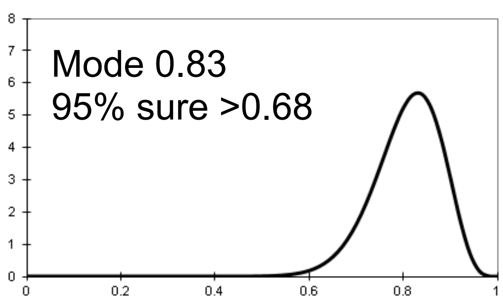
	MAT	
Test+	222	x1
Test−	776	
Total	998	n

x1 ~ binomial(n,p1)

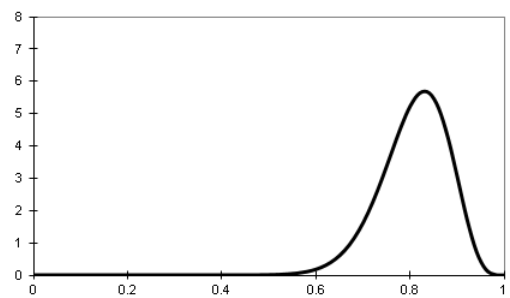
where p1 = π.Se + (1-π).(1-Sp)

1 df   3 parameters $p = F(\theta)$

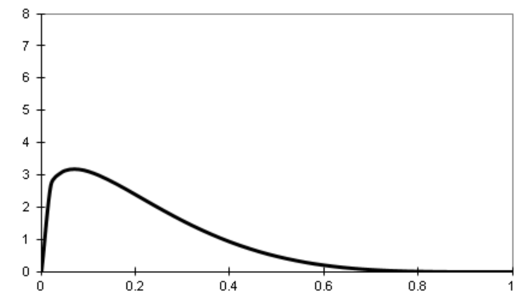
Priors (θ)



Se



Sp



π

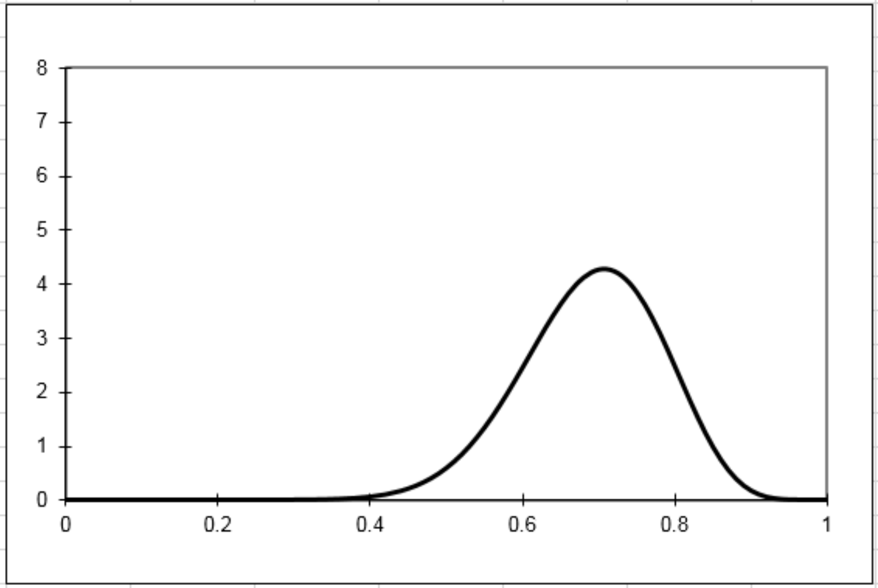
$$\boldsymbol{p(\theta|x)} \propto Se^{\alpha 1-1}(1-Se)^{\beta 1-1}Sp^{\alpha 2-1}(1-Sp)^{\beta 2-1}\pi^{\alpha 3-1}(1-\pi)^{\beta 3-1}p1^{222}(1-p1)^{776}$$

Marios P. Georgiadis MP, Johnson WO, Gardner IA, Singh R (2003) Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests, *JRSSC* 52, 63-76.

Setting Priors

alpha =	16.86547	beta =	7.577238	Beta =	3362095

x	f(x)	F(x)
0	0	0
0.02	3.27E-21	3.9E-24
0.04	1.7E-16	4.1E-19
0.06	9.22E-14	3.36E-16
0.08	7.68E-12	3.76E-14
0.1	2.29E-10	1.42E-12
0.12	3.57E-09	2.67E-11
0.14	3.54E-08	3.12E-10
0.16	2.52E-07	2.57E-09
0.18	1.39E-06	1.62E-08
0.2	6.31E-06	8.22E-08
0.22	2.42E-05	3.52E-07
0.24	8.12E-05	1.3E-06
0.26	0.000243	4.28E-06
0.28	0.000656	1.27E-05
0.3	0.00163	3.42E-05
0.32	0.003749	8.54E-05
0.34	0.008061	0.000199
0.36	0.016306	0.000434
0.38	0.031203	0.000895
0.4	0.056749	0.001753
0.42	0.098468	0.003272
0.44	0.163529	0.005847
0.46	0.260613	0.010027
0.48	0.39943	0.016549
0.5	0.589771	0.026348



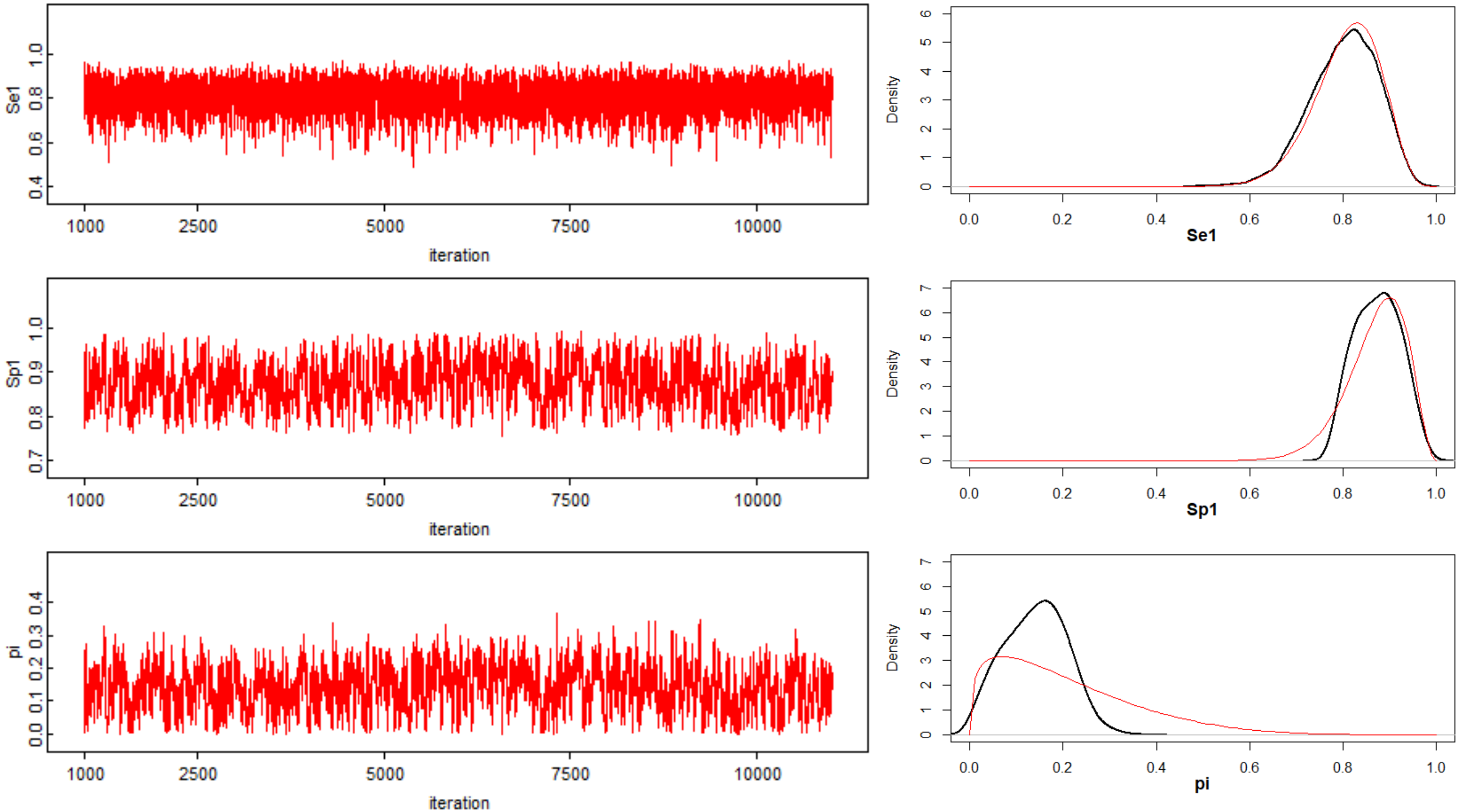
mean **spread**

	Mode =	0.706932
	Mean =	0.69
	SD =	0.09169
0.5	M =	0.695256
0.025	Q1 =	0.497665
0.975	Q3 =	0.852925

http://shiny.massey.ac.nz/kgovinda/PriorApp/

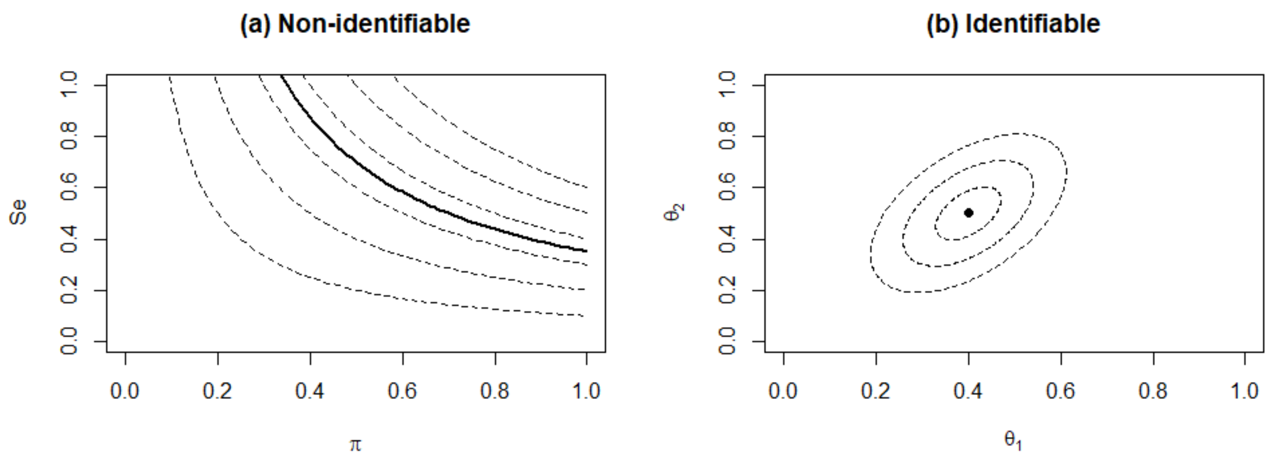
Jones G. and Johnson W.O. (2014) Prior elicitation: interactive spreadsheet graphics with sliders can be fun, and informative. *The American Statistician* 68(1): 42-51.

MAT Only - Results

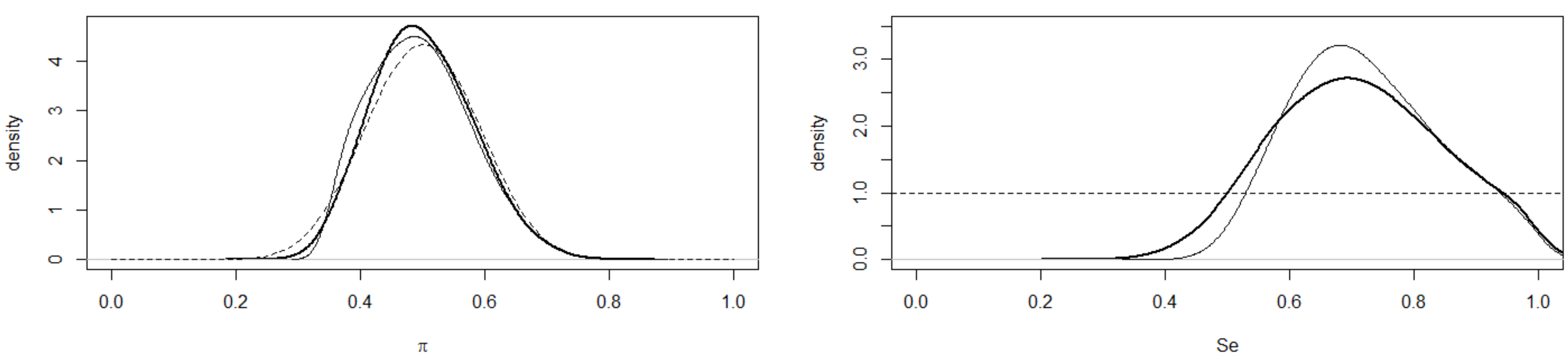


A Simpler Example: Sp = 1

Likelihood: X ~ Binomial(n, p=πSe)   Given X=35, n=100:

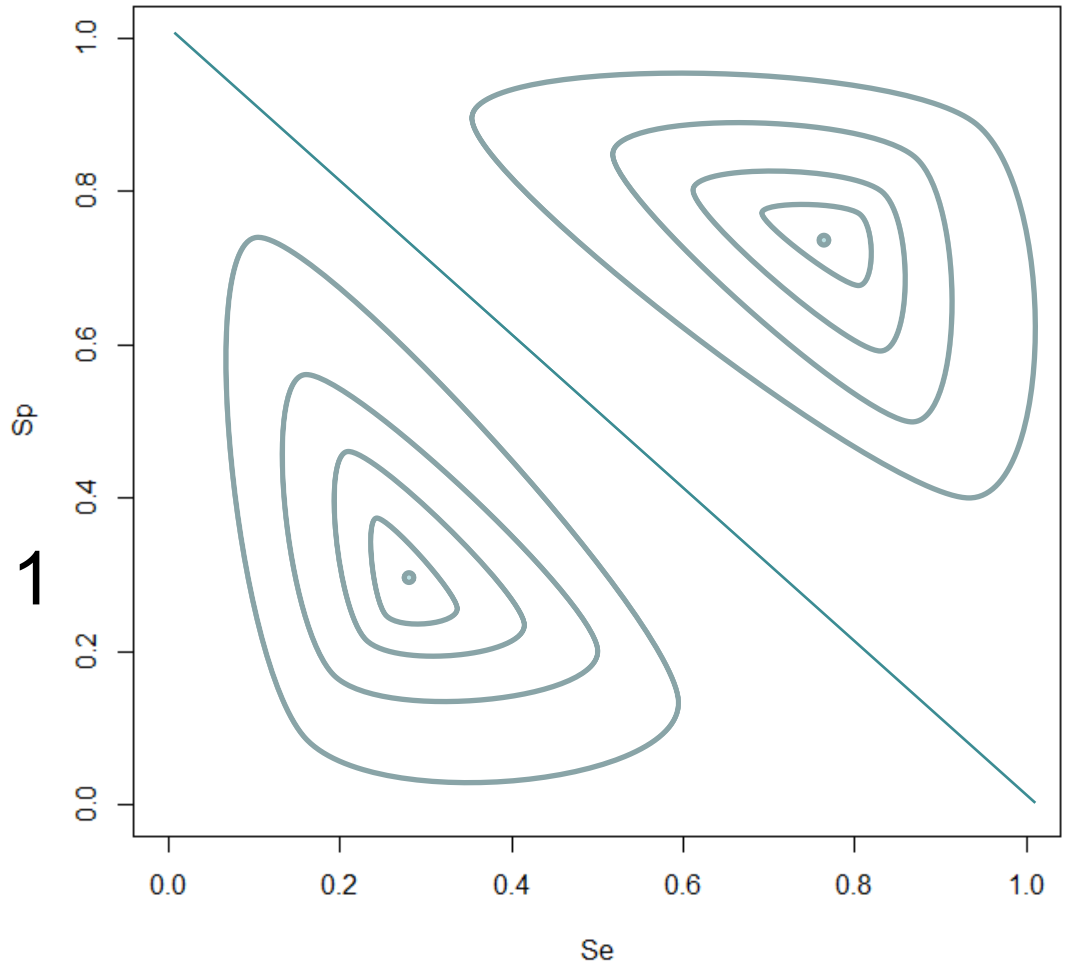


Priors: π ~ beta(15, 15); Se ~ beta(1,1)



The Label-switching Problem

Se + Sp > 1



Se + Sp < 1

Two Conditionally Independent Tests (and multiple populations)

Diseased

	+	–
+	η_{11} = Se_1Se_2	η_{10} = $Se_1(1\text{-}Se_2)$
–	η_{10} = $(1\text{-}Se_1)Se_2$	η_{00} = $(1\text{-}Se_1)(1\text{-}Se_2)$

Disease-free

	+	–
+	θ_{11} = $(1\text{-}Sp_1)(1\text{-}Sp_2)$	θ_{10} = $(1\text{-}Sp_1)Sp_2$
–	θ_{10} = $Sp_1(1\text{-}Sp_2)$	θ_{00} = Sp_1Sp_2

$\times\pi_m$

$\times(1\text{-}\pi_m)$

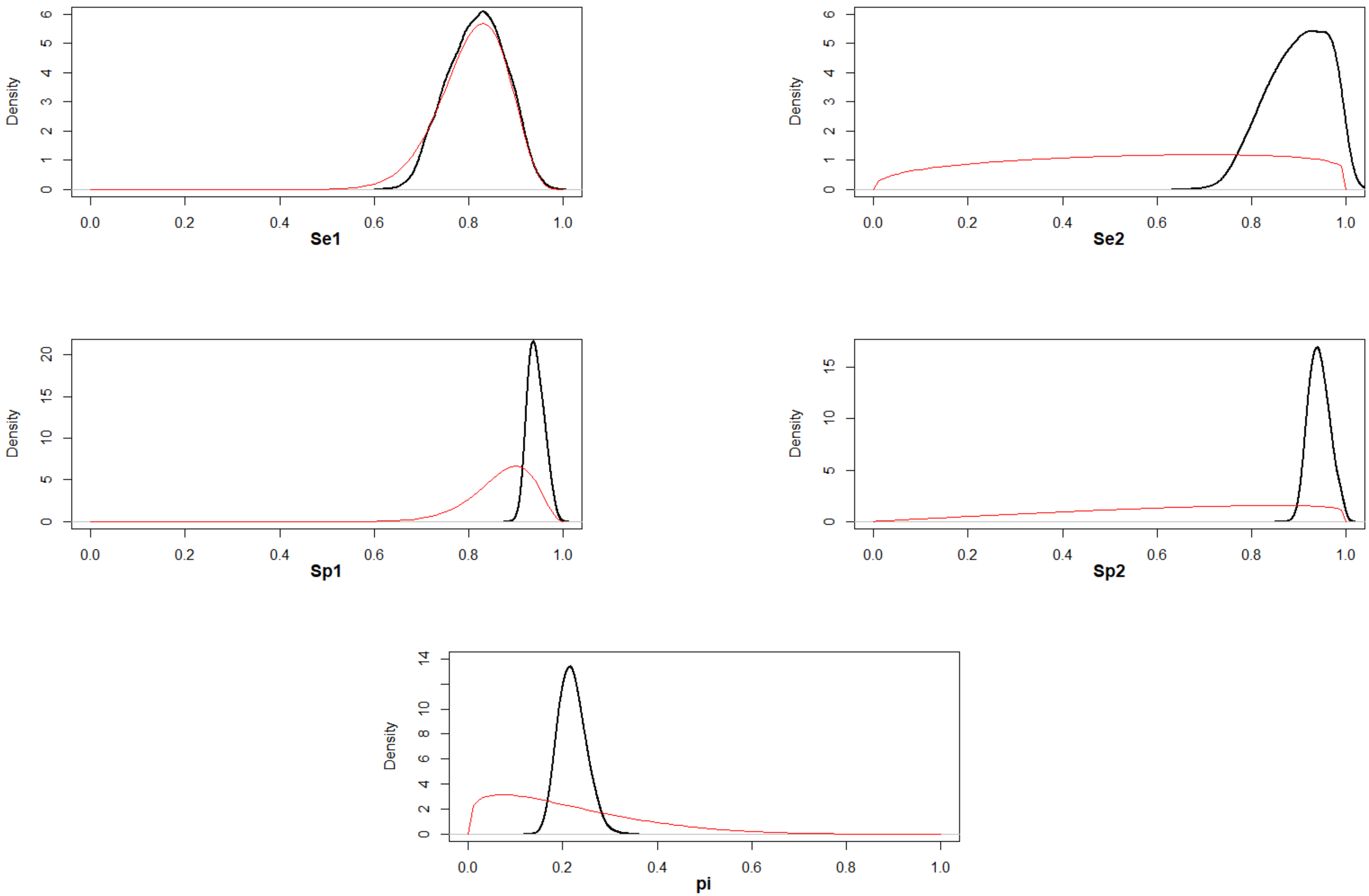
$p_{m,11}$	$p_{m,10}$
$p_{m,01}$	$p_{m,00}$

Population *m*

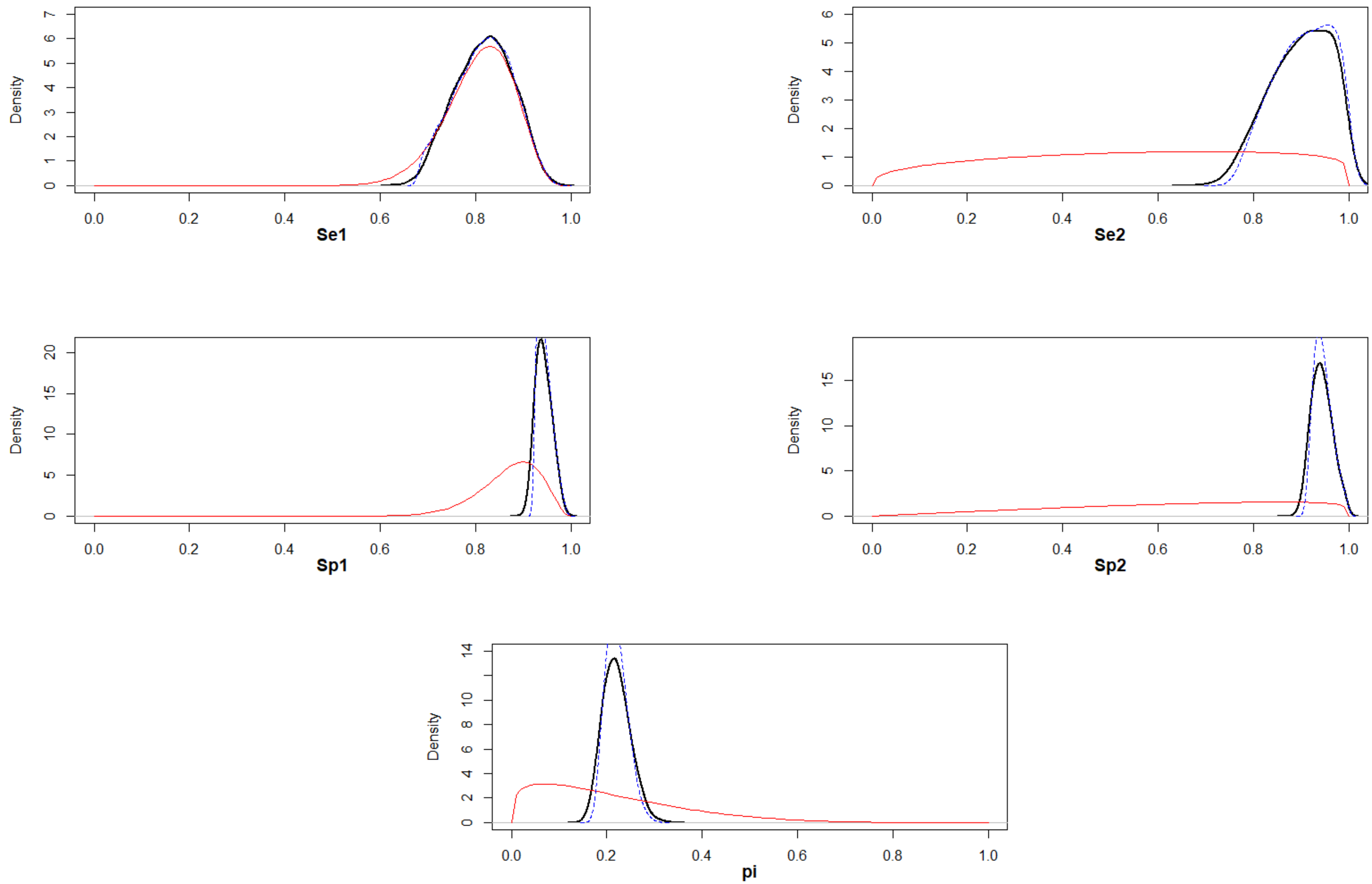
m=1
5 parameters
3 dof

m=2
6 parameters
6 dof

MAT + ELISA (ind, 1 popn) - Results



MAT + ELISA (ind) – with LPD



Some Theory (for those who enjoy it)

WIKIPEDIA

In statistics, **identifiability** is a property which a model must satisfy in order for precise inference to be possible. A model is **identifiable** if it is theoretically possible to learn the true values of this model's underlying parameters after obtaining an infinite number of observations from it. Mathematically, this is equivalent to saying that different values of the parameters must generate different probability distributions of the observable variables.

Let $\mathcal{P} = \{P_\theta\colon\ \theta \in \Theta\}$ be a statistical model where the parameter space Θ is either finite- or infinite-dimensional. We say that $\mathcal{P}$ is **identifiable** if the mapping $\theta \mapsto P_\theta$ is one-to-one: if $\theta_1 \neq \theta_2$, then also $P_{\theta 1} \neq P_{\theta 2}$.

Multinomial Models $\quad p = F(\theta) \quad \mathbf{J} = \frac{\partial F}{\partial \theta}$ (Jacobian matrix)

The model is **locally identifiable** at θ if $\mathbf{J}(\theta)$ has no zero singular values. This means the model will be identifiable in a neighbourhood of θ.

Check if $|\mathbf{J}^t\mathbf{J}| = 0$

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.

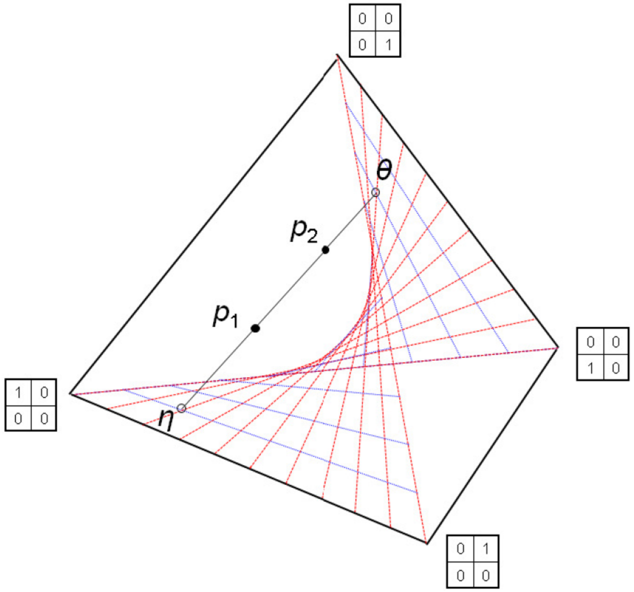
Hui-Walter Model: Two Tests, Two Popns

$$p_{1,11} = \pi_1 Se_1 Se_2 + (1-\pi_1)(1-Sp_1)(1-Sp_2)$$
$$p_{1,10} = \pi_1 Se_1 (1-Se_2) + (1-\pi_1)(1-Sp_1) Sp_2$$
$$p_{1,01} = \pi_1 (1-Se_1) Se_2 + (1-\pi_1) Sp_1 (1-Sp_2)$$
$$p_{2,11} = \pi_2 Se_1 Se_2 + (1-\pi_2)(1-Sp_1)(1-Sp_2)$$
$$p_{2,10} = \pi_2 Se_1 (1-Se_2) + (1-\pi_2)(1-Sp_1) Sp_2$$
$$p_{2,01} = \pi_2 (1-Se_1) Se_2 + (1-\pi_2) Sp_1 (1-Sp_2)$$

$$p = F(\theta)$$

$$\mathbf{J} = \frac{\partial F}{\partial \theta}$$

$$|\mathbf{J}| = (\pi_1 - \pi_2)^2 (Se_1 + Sp_1 - 1)^2 (Se_2 + Sp_2 - 1)^2$$



Assumptions:

1. $\pi_1 \neq \pi_2$
2. $Se_i + Sp_i > 1$
3. Conditional independence
4. Test homogeneity

Jones G., Johnson W.O., Hansen T.E. and Christensen R. (2010) Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics* 66, 855-863.

Three Independent Tests (one population)

$$p_{111} = \pi Se_1 Se_2 Se_3 + (1-\pi)(1-Sp_1)(1-Sp_2)(1-Sp_3)$$
$$p_{110} = \pi Se_1 Se_2 (1-Se_3) + (1-\pi)(1-Sp_1)(1-Sp_2)Sp_3$$
$$p_{101} = \pi Se_1 (1-Se_2) Se_3 + (1-\pi)(1-Sp_1)Sp_2(1-Sp_3)$$

etc.

Number of parameters = 7
Number of dof = 7

$$|J| = \pi^3(\pi-1)^3(Se_1+Sp_1-1)^2(Se_2+Sp_2-1)^2 \times (Se_3+Sp_3-1)^2.$$

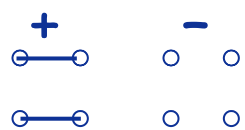
Four Dependent Tests (one population)

- Four tests for Kala-Azar (“black fever”) - a potentially fatal parasitic disease transmitted by sandflies.
  - –T1 freeze-dried direct agglutination
  - –T2 the rk30 dipstick test
  - –T3 Katex urine antigen test
  - –T4 direct microscopic examination of a tissue smear

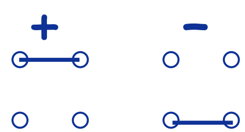
Number of dof =

- Expected pairwise correlations between T1&T2, T3&T4

Model 1



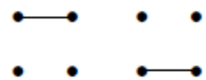
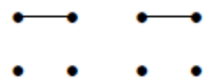
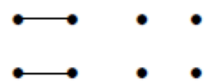
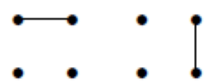
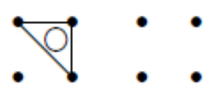
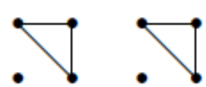
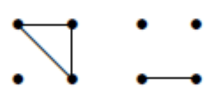
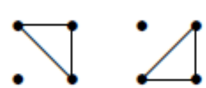
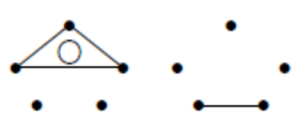
Model 2



Number of parameters =

Menton, J., Boelaert, M., and Lesaffre, E. (2008). Bayesian latent class models with conditionally dependent diagnostic tests: A case study. *Statistics in Medicine*, 27, 4469-4488.

Four or FiveTests (one population)

Model	Architecture	$D+$	$D-$	# Null Vectors
a		[12]	[34]	1
b		[12]	[12]	0
c		[12] [34]	–	0
d		[12]	[23]	0
e		[123]	–	0
f		[12] [23] [13]	[12] [23] [13]	0
g		[12] [23] [13]	[34]	1
h		[12] [23] [13]	[23] [24] [34]	2
i		[123]	[45]	1

HARMONY

Novel tools for test evaluation and disease prevalence estimation

Conditional Dependence in Bayesian Latent Class Analysis of Diagnostic Test Data

Geoff Jones
Massey University, NZ

cost

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

COST is supported by the EU Framework Programme Horizon 2020

Two Conditionally Independent Tests (and multiple populations)

Diseased

	+	–
+	$\eta_{11} = Se_1Se_2$	$\eta_{10} = Se_1(1\text{-}Se_2)$
–	$\eta_{10} = (1\text{-}Se_1)Se_2$	$\eta_{00} = (1\text{-}Se_1)(1\text{-}Se_2)$

Disease-free

	+	–
+	$\theta_{11} = (1\text{-}Sp_1)(1\text{-}Sp_2)$	$\theta_{10} = (1\text{-}Sp_1)Sp_2$
–	$\theta_{10} = Sp_1(1\text{-}Sp_2)$	$\theta_{00} = Sp_1Sp_2$

$\times\pi_m$

$\times(1\text{-}\pi_m)$

$p_{m,11}$	$p_{m,10}$
$p_{m,01}$	$p_{m,00}$

Population *m*

m=1
5 parameters
3 dof

m=2
6 parameters
6 dof

MAT and ELISA – Independent, 2 Popns

Data (x1,x2)

Popn 1	ELISA+	ELISA−	
MAT+	x111 67	x112 25	92
MAT−	x121 41	x122 328	369
	108	353	461

n1

Popn 2	ELISA+	ELISA−	
MAT+	97	33	130
MAT−	36	371	407
	133	404	537

etc.

x1 ~ multinomial(n1,p1)
x2 ~ multinomial(n2,p2)

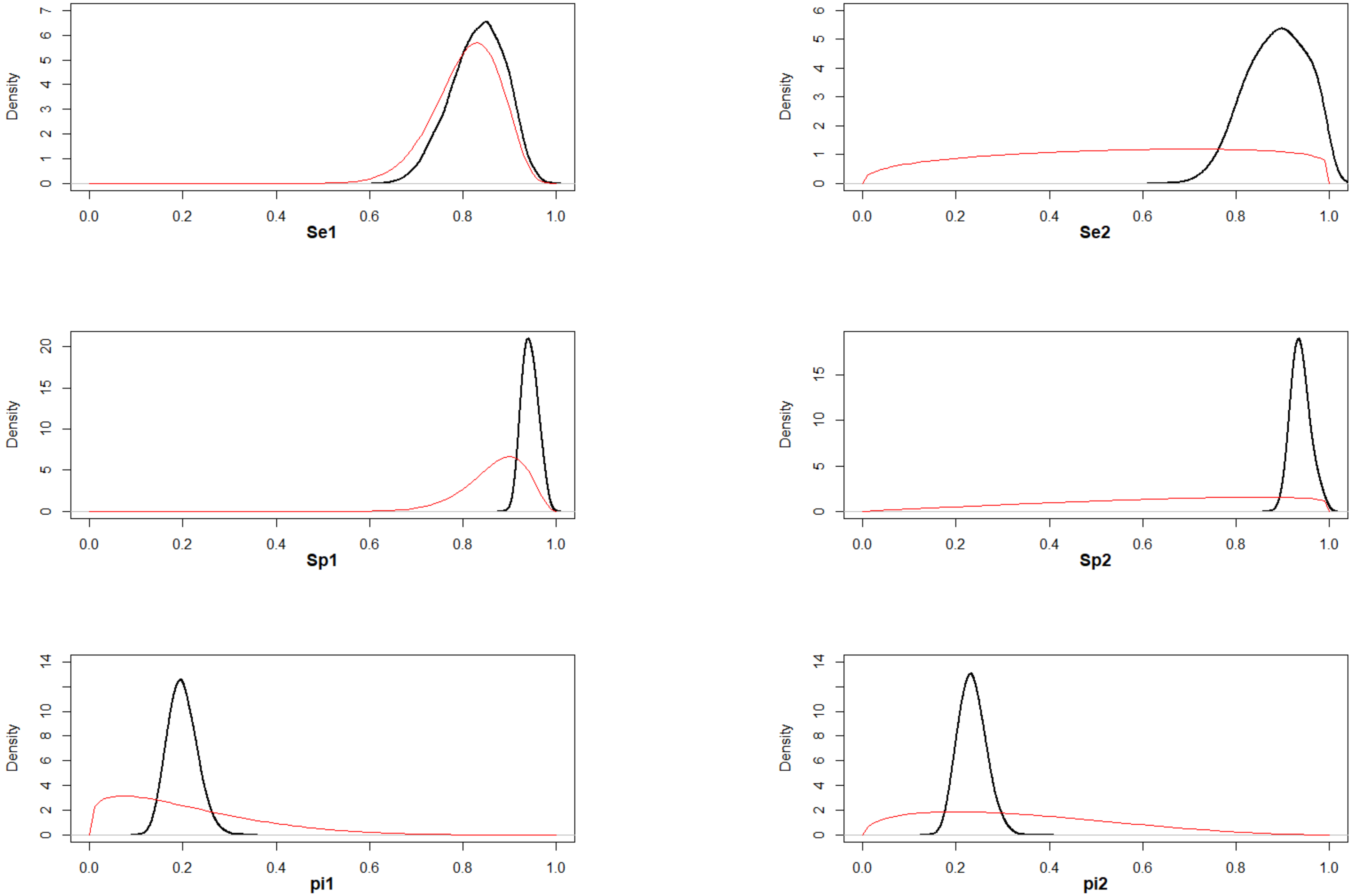
where $p111 = \pi1Se1Se2 + (1-\pi1)(1-Sp1)(1-Sp2)$

$p112 = \pi1Se1(1-Se2) + (1-\pi1)(1-Sp1)Sp2$

$p121 = \pi1(1-Se1)Se2 + (1-\pi1)Sp1(1-Sp2)$

$p122 = \pi(1-Se1)(1-Se2) + (1-\pi1)Sp1Sp2$

MAT + ELISA (ind, 2 popns) - Results



Two Correlated Tests (and multiple populations)

Diseased

	+	–
+	$\eta_{11} =$ Se_1Se_2 $+Cp_{12}$	$\eta_{10} =$ $Se_1(1\text{-}Se_2)$ $-Cp_{12}$
–	$\eta_{10} =$ $(1\text{-}Se_1)Se_2$ $-Cp_{12}$	$\eta_{00} =$ $(1\text{-}Se_1)(1\text{-}Se_2)$ $+Cp_{12}$

Disease-free

	+	–
+	$\theta_{11} =$ $(1\text{-}Sp_1)(1\text{-}Sp_2)$ $+Cn_{12}$	$\theta_{10} =$ $(1\text{-}Sp_1)Sp_2$ $-Cn_{12}$
–	$\theta_{10} =$ $Sp_1(1\text{-}Sp_2)$ $-Cn_{12}$	$\theta_{00} =$ Sp_1Sp_2 $-Cn_{12}$

$\times\pi_m$ $\qquad$ $\times(1\text{-}\pi_m)$

$p_{m,11}$	$p_{m,10}$
$p_{m,01}$	$p_{m,00}$

Population *m*

Not identifiable!

m=3

9 parameters

9 dof

Two Correlated Tests (three populations)

Null singular vectors for three populations, two correlated tests

Parameter	1	2
π_1	$1-\pi_1$	π_1
π_2	$1-\pi_2$	π_2
π_3	$1-\pi_3$	π_3
Se_1	0	$-Q_1$
Se_2	0	$-Q_2$
Sp_1	Q_1^*	0
Sp_2	Q_2	0
C^+	0	$C^- - C^+ - Q_1Q_2$
C^-	$C^- - C^+ + Q_1Q_2$	0

$^*Q_i = Se_i + Sp_i - 1.$

MAT and ELISA – Test Dependence I

- Some probability theory: if A and B are binary (0/1) then:
  $$P(A \text{ and } B) = E(AB) = E(A)E(B) + Cov(A, B)$$
- If A and B are independent then $Cov(A, B) = 0$;

For a diseased individual:

- $P(\text{MAT+}, \text{ELISA +}) = Se1Se2 + Cp$;

For a disease-free individual:

- $P(\text{MAT+}, \text{ELISA +}) = (1 - Sp1)(1 - Sp2) + Cn$;

So p11 = $\pi 1$(Se1Se2+Cp) + (1-$\pi 1$)((1-Sp1)(1-Sp2)+Cn)
p12 = $\pi 1$(Se1(1-Se2)–Cp) + (1-$\pi 1$)((1-Sp1)Sp2–Cn)
p21 = $\pi 1$((1–Se1)Se2–Cp) + (1-$\pi 1$)(Sp1(1-Sp2)–Cn)
p22 = π ((1–Se1)(1-Se2)+Cp) + (1-$\pi 1$)(Sp1Sp2)+Cn)

Dendukuri, N. and Joseph, L. (2001). Bayesian approaches to modeling conditional dependence between multiple diagnostic tests. *Biometrics* 57, 158–167.

MAT and ELISA – Test Dependence I

- The new parameters *Cp*, *Cn* have to be restricted to keep all probabilities in [0,1];
- "A little algebra" gives:

$$(Se1-1)(1-Se2) \le Cp \le \min(Se1, Se2) - Se1Se2$$

$$(Sp1-1)(1-Sp2) \le Cn \le \min(Sp1, Sp2) - Sp1Sp2$$

- These restrictions can be enforced using the priors

- How do we get the conditional correlations between the tests?
- $Corr(X,Y) = Cov(X,Y)/\sqrt{V[X]V[Y]}$;

So for a diseased individual:

- $Corr(MAT, ELISA) = \frac{Cp}{\sqrt{Se1(1-Se1)Se2(1-Se2)}}$

MAT and ELISA – Test Dependence II

- Some probability theory: $P(A \text{ and } B) = P(A) \times P(B|A)$;
- $P(B|A)$ is the conditional probability of B given A;
- If A and B are independent then $P(B|A) = P(B)$;

For a diseased individual:

- $P(\text{MAT}+, \text{ELISA}+) = P(\text{MAT}+)P(\text{ELISA}+|\text{MAT}+)$;
- If independent, this is $Se1 \times Se2$;
- If not, $Se1 \times cSe2p$ – the conditional probability that Test 2 is positive given that test 1 is positive.

So p111 = π1Se1cSe2p + (1-π1)(1-Sp1)(1-cSp2p)
p112 = π1Se1(1-cSe2p) + (1-π1)(1-Sp1)cSp2p
p121 = π1(1–Se1)cSe2n + (1-π1)Sp1(1-cSp2n)
p122 = π(1–Se1)(1-cSe2n) + (1-π1)Sp1cSp2n

Georgiadis,M. P., Johnson,W. O., Gardner, I. A., and Singh, R. (2003). Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Applied Statistics* 52, 63–76.

MAT and ELISA – Test Dependence II

- Given the conditional sensitivities, how do we get the unconditional sensitivity?
- More probability theory: $P(B) = P(A)P(B|A) + P(A')\,P(B|A')$;
- So $Se2 = Se1 \times cSe2p + (1 - Se1) \times cSe2n$

- How do we get the conditional correlations between the tests?
- $Corr(X,Y) = (E[XY] - E[X]E[Y])/\sqrt{V[X]V[Y]}$;

So for a diseased individual:

- $Corr(MAT, LAT) = \frac{Se1\,(cSe2p - Se2)}{\sqrt{Se1(1-Se1)Se2(1-Se2)}}$

HARMONY
Novel tools for test evaluation and disease prevalence estimation

Exploring Identifiability in Bayesian Latent Class Analysis of Diagnostic Test Data

Geoff Jones
Massey University, NZ

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

COST is supported by the EU Framework Programme Horizon 2020

Exploring Identifiability – Plan

- **Label-switching** eg H-W model
- **Effect of large *n*** eg H-W model vs 2 dependent tests
- **Limiting posterior distribution (LPD)** eg 2 dep tests, I popn
- **Adding a popn vs adding a test** to 2 dep tests, 2 popns
- **Checking identifiability** eg four tests, I popn

Two Conditionally Independent Tests (and multiple populations)

Diseased

	+	−
+	$\eta_{11} =$ $Se_1 Se_2$	$\eta_{10} =$ $Se_1(1-Se_2)$
−	$\eta_{10} =$ $(1-Se_1)Se_2$	$\eta_{00} =$ $(1-Se_1)(1-Se_2)$

Disease-free

	+	−
+	$\theta_{11} =$ $(1-Sp_1)(1-Sp_2)$	$\theta_{10} =$ $(1-Sp_1)Sp_2$
−	$\theta_{10} =$ $Sp_1(1-Sp_2)$	$\theta_{00} =$ Sp_1Sp_2

$\times \pi_m$

$\times(1-\pi_m)$

$p_{m,11}$	$p_{m,10}$
$p_{m,01}$	$p_{m,00}$

Population *m*

m=1
5 parameters
3 dof

m=2
6 parameters
6 dof

MAT and ELISA - Independent

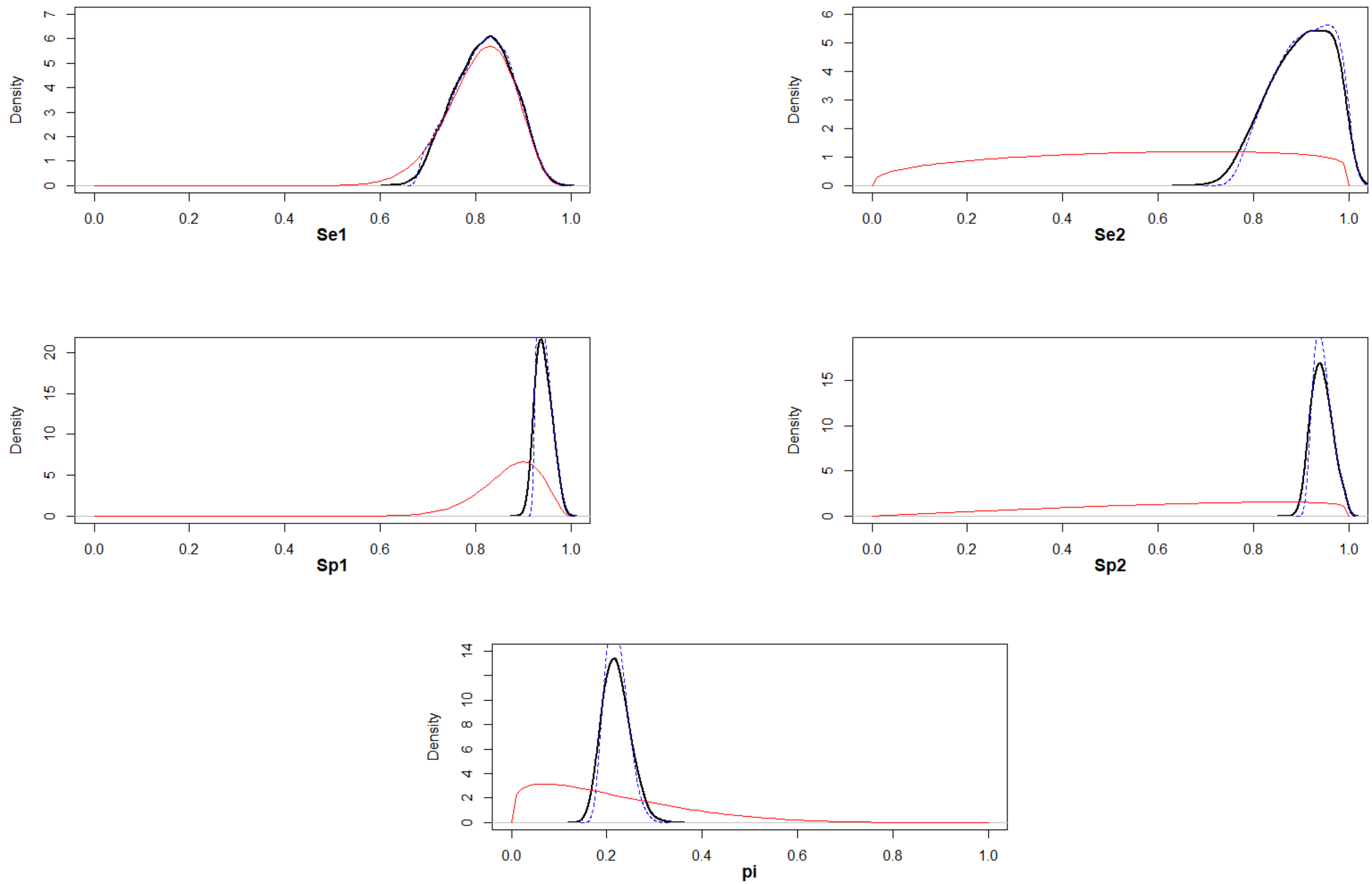
Data (x)

	ELISA+	ELISA−	
MAT+	x11 164	x12 58	222
MAT−	x21 77	x22 699	776
	241	757	n 998

$$x \sim \text{multinomial}(n,p)$$

where
$$p11 = \pi Se1Se2 + (1-\pi)(1-Sp1)(1-Sp2)$$
$$p12 = \pi Se1(1-Se2) + (1-\pi)(1-Sp1)Sp2$$
$$p21 = \pi(1-Se1)Se2 + (1-\pi)Sp1(1-Sp2)$$
$$p22 = \pi(1-Se1)(1-Se2) + (1-\pi)Sp1Sp2$$

MAT + ELISA (ind) – with LPD



Calculating the LPD

Take the original parameter vector as $\theta = (\pi, Se1, Sp1, Se2, Sp2)$

and the transparent parametrisation as $\Phi = (p_{11}, p_{10}, p_{01}, Se1, Sp1)$

The transformation linking them is:

$$\begin{aligned}
p_{11} &= \pi Se1Se2 + (1-\pi)(1-Sp1)(1-Sp2) \\
p_{10} &= \pi Se1(1-Se2) + (1-\pi)(1-Sp1)Sp2 \\
p_{01} &= \pi(1-Se1)Se2 + (1-\pi)Sp1(1-Sp2) \\
Se1 &= Se1 \\
Sp1 &= Sp1
\end{aligned}$$

The Jacobian of the transformation $|J| = \partial\Phi/\partial\theta$ is then

$$\begin{vmatrix}
Se1Se2 - (1-Sp1)(1-Sp2) & \pi Se2 & -(1-\pi)(1-Sp2) & \pi Se1 & -(1-\pi)(1-Sp1) \\
Se1(1-Se2) - (1-Sp1)Sp2 & \pi(1-Se2) & -(1-\pi)Sp2 & -\pi Se1 & (1-\pi)(1-Sp1) \\
(1-Se1)Se2 - Sp1(1-Sp2) & -\pi Se2 & (1-\pi)(1-Sp2) & \pi(1-Se1) & -(1-\pi)Sp1 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0
\end{vmatrix}$$

$$= \pi(1-\pi)(Se1 + Sp1 - 1)^2$$

Calculating the LPD

Now consider a "convenience prior" on Φ that is Dirichlet(1,1,1,1) on the observable proportions and independent betas on $Se1, Sp1$. With infinite data, (p_{11}, p_{10}, p_{01}) are known. Set them to their observed values. The posterior distribution of $(Se1, Sp1)$ is

$$p^*(Se1, Se2 \mid p_{11}, p_{10}, p_{01}) \propto Se1^{\alpha_1}(1 - Se1)^{\beta_1} Sp1^{\alpha_2}(1 - Sp1)^{\beta_2} I(A)$$

The actual prior on Φ induced by the prior on θ is

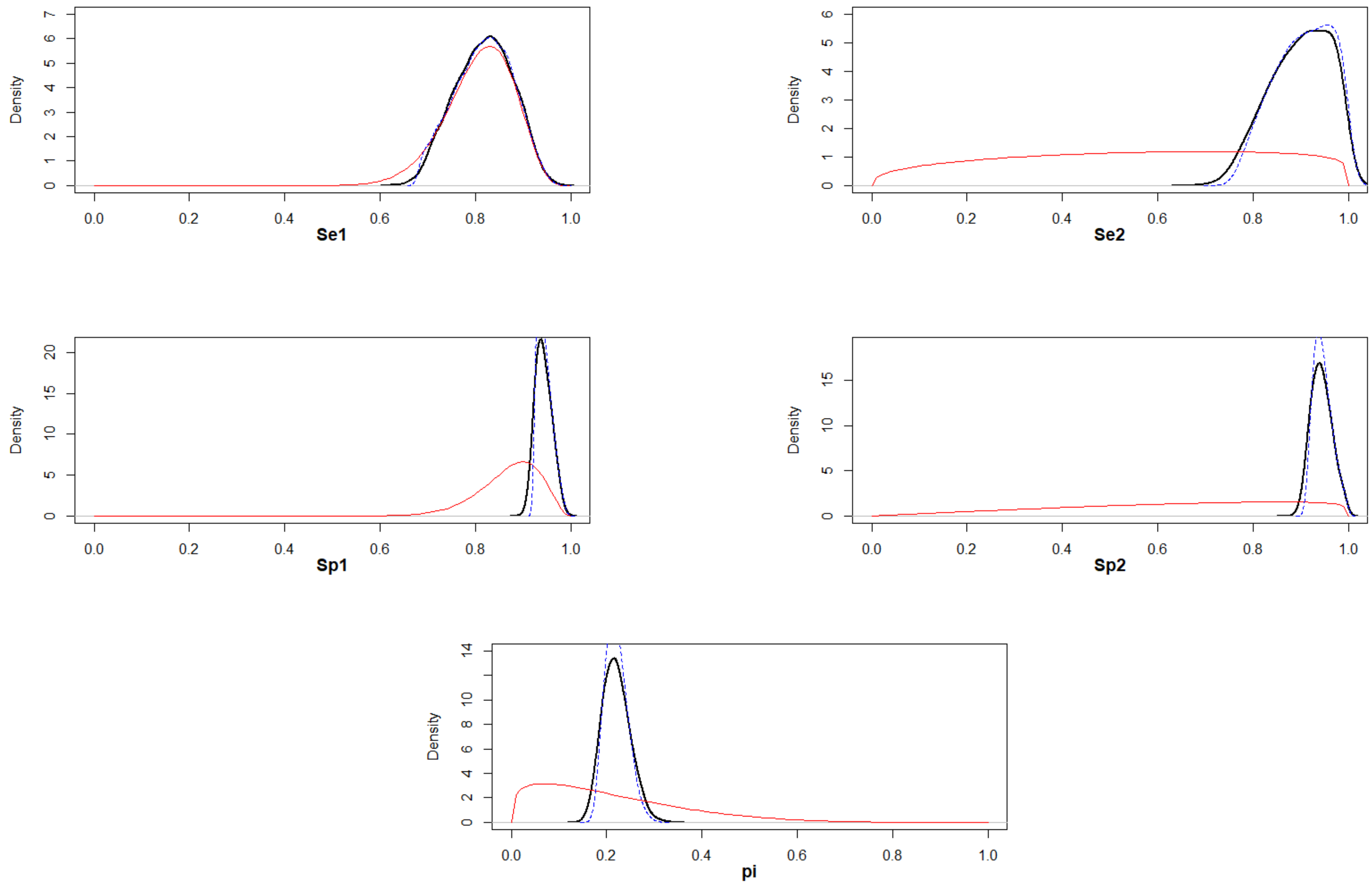
$$p(\Phi) \propto Se1^{\alpha_1}(1 - Se1)^{\beta_1} Sp1^{\alpha_2}(1 - Sp1)^{\beta_2} Se2^{\alpha_3}(1 - Se2)^{\beta_3} Sp2^{\alpha_4}(1 - Sp2)^{\beta_4} \pi^{\alpha_5}(1 - \pi)^{\beta_5} |J|^{-1} I(A)$$

To sample from $p(\theta \mid p_{11}, p_{10}, p_{01})$

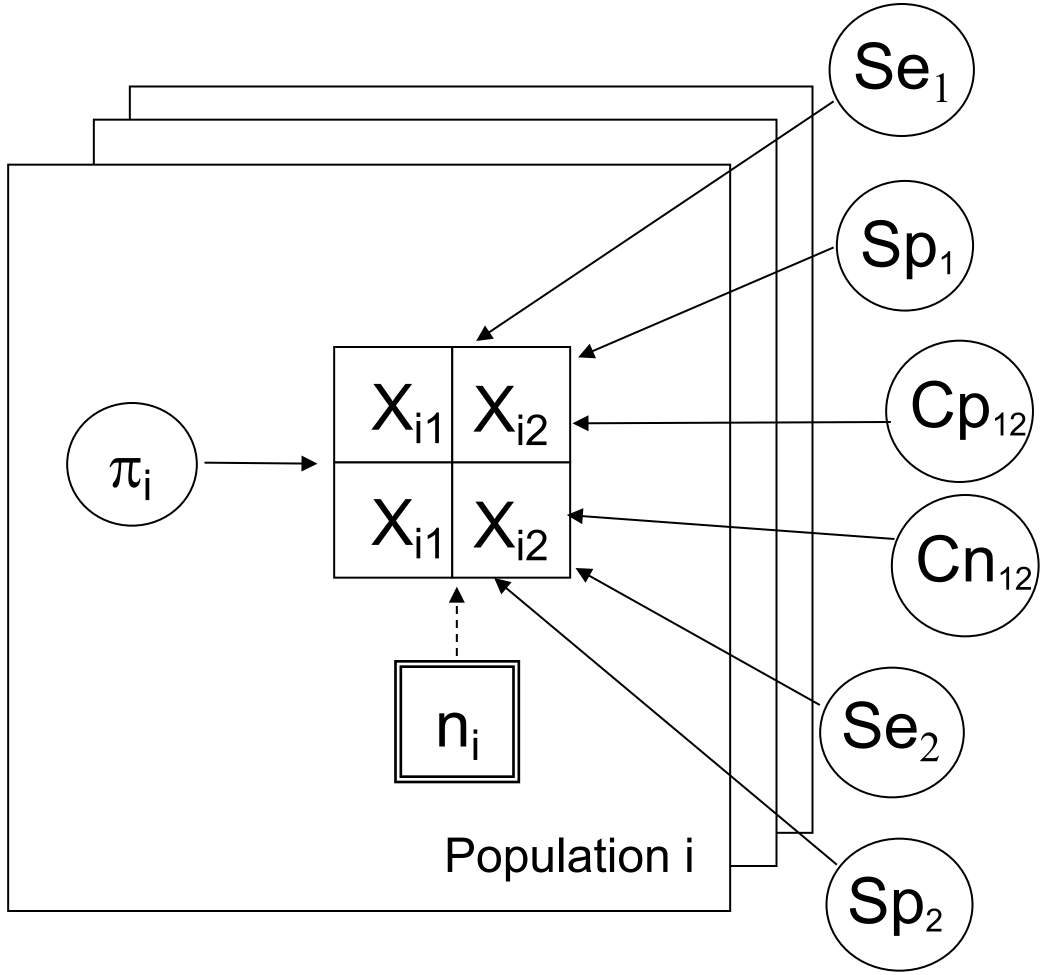
1. draw $Se1$ from beta(α_1, β_1) and $Sp1$ from beta(α_2, β_2),
2. transform Φ to θ,
3. accept if $\theta \in A$,
4. calculate importance weights

$$w = \frac{p(\Phi)}{p^*(\Phi)} = Se2^{\alpha_3}(1 - Se2)^{\beta_3} Sp2^{\alpha_4}(1 - Sp2)^{\beta_4} \pi^{\alpha_5}(1 - \pi)^{\beta_5} |J|^{-1}$$

MAT + ELISA (ind) – with LPD



Two Tests, Three Populations



Likelihood: (X_{i1}, X_{i1}, X_{i1}, X_{i1}) ~ Multinomial (n_i; …..)

Two Tests, Three Populations: Example

Data

	+	-		+	-		+	-
+	43	20		78	15		97	16
-	20	117		21	86		20	67

Priors

$Se_1 \sim \text{beta}(42, 5)$, $Sp_1 \sim \text{beta}(45, 2.5)$

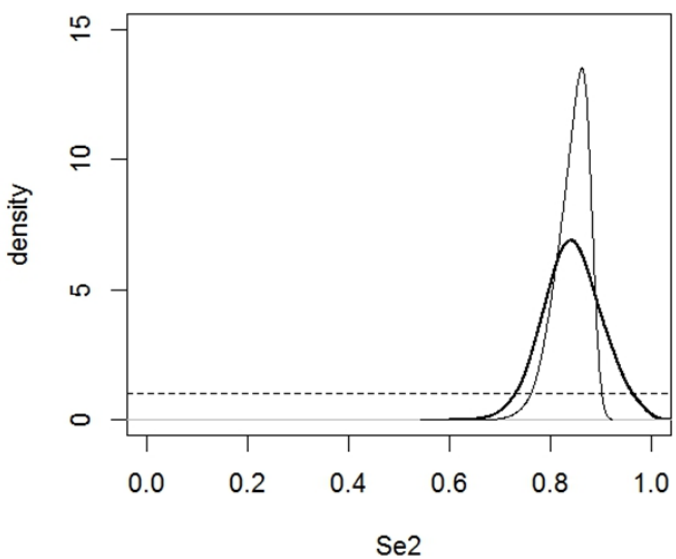
$Se_2 \sim \text{beta}(1, 1)$, $Sp_2 \sim \text{beta}(1, 1)$

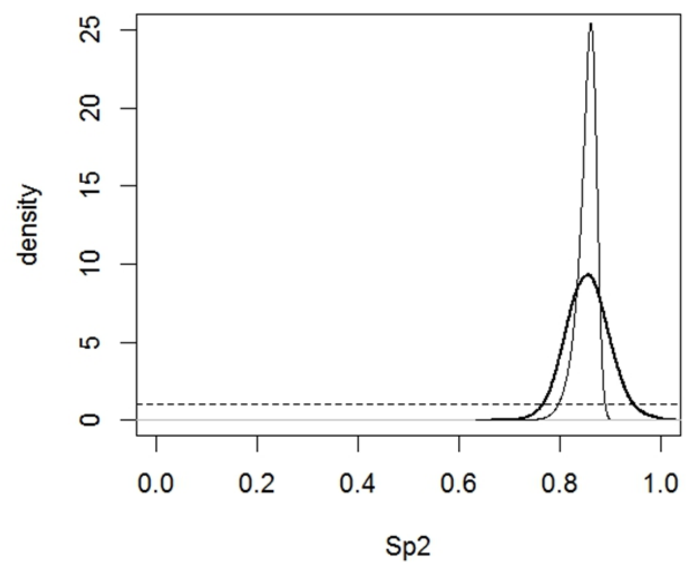
$\pi_1 \sim \text{beta}(4.57, 10.7)$, $\pi_2 \sim \text{beta}(8.26, 8.26)$, $\pi_3 \sim \text{beta}(10.7, 4.57)$

Gustafson's *transparent parametrization*:

$$\boldsymbol{\phi} = (p_{1,11}, p_{1,10}, p_{1,01}, p_{3,11}, p_{3,10}, p_{3,01}, (\pi_2 - \pi_1)/\,(\pi_3 - \pi_1))'$$

$$\boldsymbol{\lambda} = (\pi_1, \pi_3)'$$





MAT, ELISA and MB

Data (x1,x2)

Popn 1	MB+		MB−	
	ELISA+	ELISA−	ELISA+	ELISA−
MAT+	x1111 22	x1121 6	45	19
MAT−	2	1	39	327
Popn 2	**MB+**		**MB−**	
	ELISA+	ELISA−	ELISA+	ELISA−
MAT+	51	11	46	22
MAT−	2	12	34	359

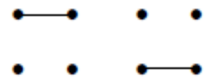
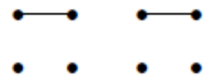
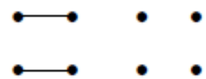
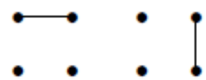
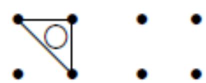
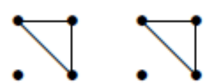
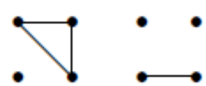
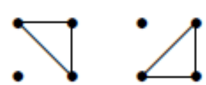
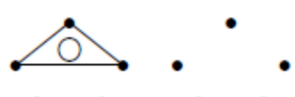
x1 ~ multinomial(n1,p1)
x2 ~ multinomial(n2,p2)

Assume MB conditionally independent of (MAT,ELISA)

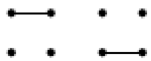
$p111 = \pi 1 Se1cSe2pSe3 + (1-\pi 1)(1-Sp1)(1-cSp2p)(1-Se3)$

etc

Four or FiveTests (one population)

Model	Architecture	$D+$	$D-$	# Null Vectors
a		[12]	[34]	1
b		[12]	[12]	0
c		[12] [34]	–	0
d		[12]	[23]	0
e		[123]	–	0
f		[12] [23] [13]	[12] [23] [13]	0
g		[12] [23] [13]	[34]	1
h		[12] [23] [13]	[23] [24] [34]	2
i		[123]	[45]	1

Four Correlated Tests Model *a*

Model	Architecture	$D+$	$D-$	# Null Vectors
a		[12]	[34]	1

- q=15, p=11 so J is 15×11 but has rank 10

- In the order $\pi, Se_1, Se_2, Se_3, Se_4, Sp_1, Sp_2, Sp_3, Sp_4, C^+_{12}, C^-_{34}$

  the null vector is

$$\left(1,\ -\frac{Q_1}{\pi},\ -\frac{Q_2}{\pi},\ 0,\ 0,\ 0,\ 0,\ \frac{Q_3}{(1-\pi)},\ \frac{Q_4}{(1-\pi)},\ \frac{(Q_1Q_2 - C^+_{12})}{\pi},\ -\frac{(Q_3Q_4 - C^+_{34})}{(1-\pi)}\right)$$

Jones G., Johnson W.O., Hansen T.E. and Christensen R. (2010) Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics* 66, 855-863.

Simulation Results

Parameter	Model a	Model a'	Model b
$Se_1 = 0.70$	0.72 (0.63,0.81)	0.70 (0.65,0.76)	0.70 (0.67,0.73)
$Se_2 = 0.90$	0.91 (0.84,0.99)	0.90 (0.86,0.94)	0.90 (0.88,0.92)
$Se_3 = 0.80$	0.80 (0.78,0.82)	0.80 (0.78,0.82)	0.80 (0.78,0.82)
$Se_4 = 0.98$	0.98 (0.96,1.00)	0.98 (0.96,1.00)	0.98 (0.95,1.00)
$Sp_1 = 0.85$	0.85 (0.83,0.87)	0.85 (0.84,0.86)	0.85 (0.83,0.87)
$Sp_2 = 0.55$	0.55 (0.53,0.57)	0.55 (0.53,0.57)	0.55 (0.53,0.57)
$Sp_3 = 0.70$	0.69 (0.64,0.76)	0.70 (0.67,0.73)	0.70 (0.68,0.72)
$Sp_4 = 0.90$	0.88 (0.81,0.99)	0.90 (0.86,0.94)	0.90 (0.88,0.92)
$C_{12}^{+} = 0.05$	0.04 (0.00,0.07)	0.05 (0.03,0.07)	0.05 (0.04,0.06)
$C_{12}^{-} = 0.05$	NA	NA	0.05 (0.04,0.06)
$C_{34}^{-} = 0.05$	0.06 (0.00,0.09)	0.05 (0.03,0.07)	NA
$\pi_1 = 0.40$	0.39 (0.33,0.46)	0.40 (0.36, 0.44)	0.40 (0.38,0.42)
$\pi_2 = 0.20$	NA	0.20 (0.17, 0.23)	NA

Jones G., Johnson W.O., Hansen T.E. and Christensen R. (2010) Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics* 66, 855-863.