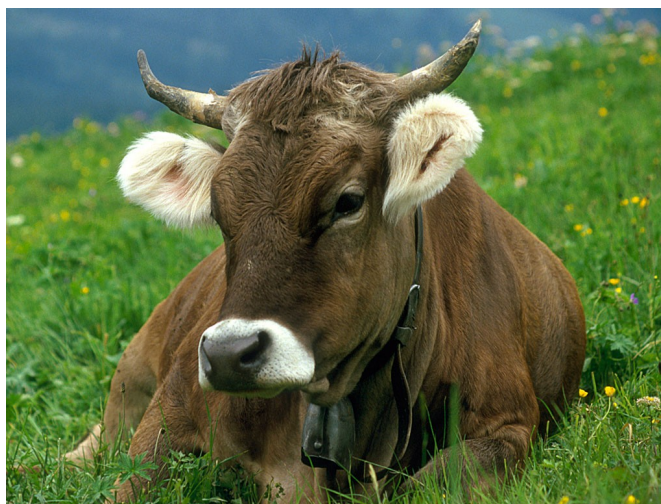


Mastitis in dairy production: Estimation of sensitivity, specificity and disease prevalence using MCMC

Caroline Uhler (cuhler@stat.berkeley.edu)



Aim: Determine which mastitis diagnostic test provides the best results and at what costs.

Mastitis

- Any inflammatory process in the mammary gland is called mastitis.
- Mastitis usually does not involve clinical signs and, therefore, can only be detected by diagnostic tests.
- Mastitis is an important cause of loss of efficiency in milk production: In Switzerland the average damage caused by mastitis is USD 300 per cow and year. The bulk of these costs accrue from decreased milk production.
- Early medical treatment substantially reduces the loss in milk production.

⇒ **Good tests for detecting mastitis are crucial!**

Mastitis diagnostic tests

The following four major mastitis diagnostic tests are analyzed in this study:

- Bacteriological test (Bact): USD 10 per test
- Somatic cell count (SCC): USD 1 per test
- Measurement of mammary associated amyloid A (MAA): USD 1 per test
- Measurement of electrical conductivity (MEC): USD 0.10 per test

None of these tests is a perfect test. So what is a ‘good’ test?

- a test with a high **sensitivity**:

$$\text{sensitivity} = \mathbb{P}(\text{positive test result} \mid \text{disease is present})$$

- a test with a high **specificity**:

$$\text{specificity} = \mathbb{P}(\text{negative test result} \mid \text{disease is not present})$$

Data

For this study 25 cows were tested at all four quarters of the udder with the four mastitis diagnostic tests (resulting in 100 independent samples). The test results are summarized in the following $2 \times 2 \times 2 \times 2$ table n :

$n_{0000} = 41$	$n_{1000} = 7$	$n_{0100} = 1$	$n_{1100} = 1$
$n_{0010} = 2$	$n_{1010} = 2$	$n_{0110} = 2$	$n_{1110} = 3$
$n_{0001} = 26$	$n_{1001} = 8$	$n_{0101} = 0$	$n_{1101} = 0$
$n_{0011} = 0$	$n_{1011} = 0$	$n_{0111} = 1$	$n_{1111} = 6$.

Notation: First index: Bact Second index: SCC
Third index: MAA Fourth index: MEC
Negative test result: 0 Positive test result: 1

Goal: Estimate the **disease prevalence** (π) as well as the **sensitivity** (**Se**) and the **specificity** (**Sp**) of the 4 tests.

Reversible jump MCMC

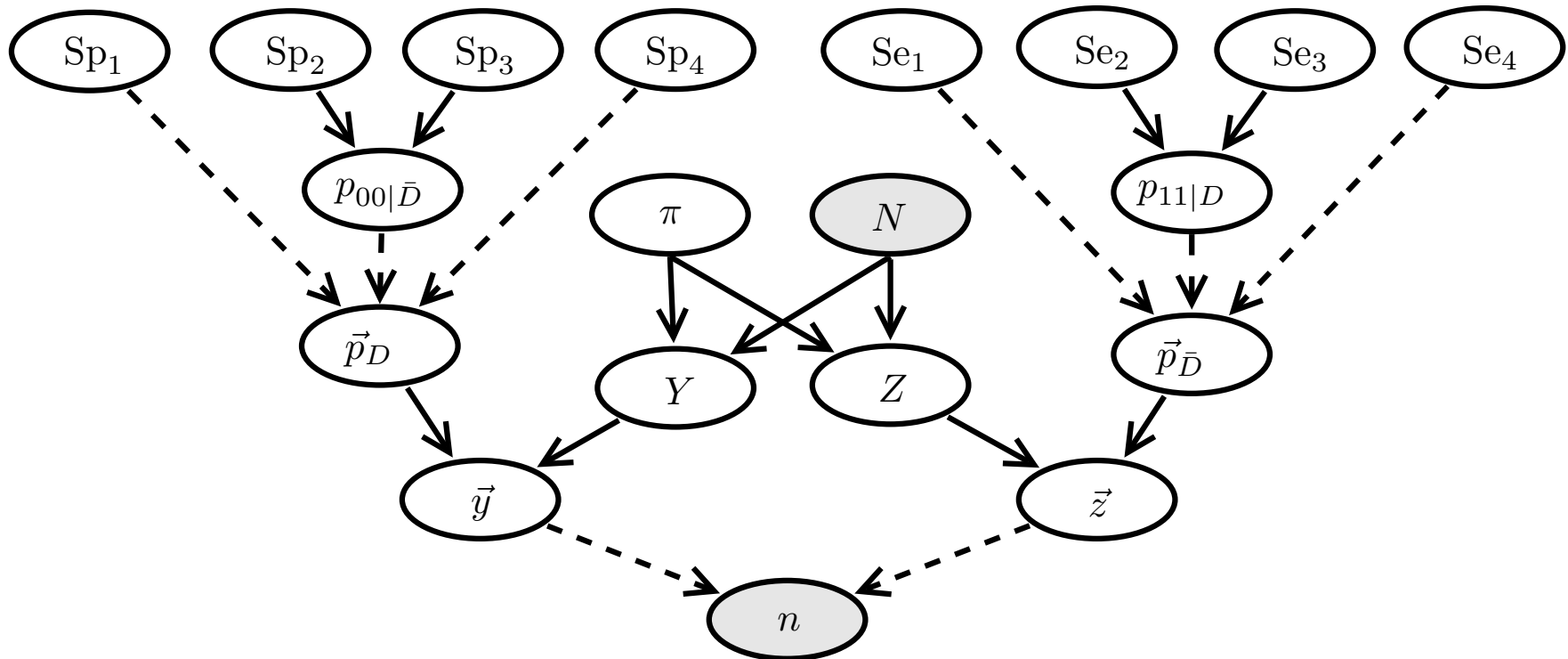
Because the SCC test and the MAA measure similar biological processes, they are likely to be **conditionally dependent**. So two more variables have to be introduced:

$$\begin{aligned} p_{00|\bar{D}} &:= P[T_{\text{SCC}} = 0, T_{\text{MAA}} = 0 \mid \text{non-diseased}] \geq \text{Sp}_{\text{SCC}} \cdot \text{Sp}_{\text{MAA}} \\ p_{11|D} &:= P[T_{\text{SCC}} = 1, T_{\text{MAA}} = 1 \mid \text{diseased}] \geq \text{Se}_{\text{SCC}} \cdot \text{Se}_{\text{MAA}} \end{aligned}$$

I account for the uncertainty about the dependence structure of these two tests by running reversible jump MCMC on the four possible models:

$$\begin{aligned} M_1 : \quad p_{11|D} &= \text{Se}_{\text{SCC}} \cdot \text{Se}_{\text{MAA}}, & p_{00|\bar{D}} &= \text{Sp}_{\text{SCC}} \cdot \text{Sp}_{\text{MAA}} \\ M_2 : \quad p_{11|D} &> \text{Se}_{\text{SCC}} \cdot \text{Se}_{\text{MAA}}, & p_{00|\bar{D}} &= \text{Sp}_{\text{SCC}} \cdot \text{Sp}_{\text{MAA}} \\ M_3 : \quad p_{11|D} &= \text{Se}_{\text{SCC}} \cdot \text{Se}_{\text{MAA}}, & p_{00|\bar{D}} &> \text{Sp}_{\text{SCC}} \cdot \text{Sp}_{\text{MAA}} \\ M_4 : \quad p_{11|D} &> \text{Se}_{\text{SCC}} \cdot \text{Se}_{\text{MAA}}, & p_{00|\bar{D}} &> \text{Sp}_{\text{SCC}} \cdot \text{Sp}_{\text{MAA}} \end{aligned}$$

Graphical Model



Notation: Solid arrow: probabilistic dependency
 Dashed arrow: deterministic relationship

Notation:

N : Total number of cows	$\vec{n} = (n_{ijkl})$ # of cows with test results i, j, k, l
Z : # of diseased cows	$\vec{z} = (z_{ijkl})$ # of diseased cows with test results i, j, k, l
Y : # of healthy cows	$\vec{y} = (y_{ijkl})$ # of healthy cows with test results i, j, k, l
D : diseased	$\vec{p}_{\bar{D}} = (p_{ijkl D}) = (\mathbb{P}[T_1 = i, T_2 = j, T_3 = k, T_4 = l \mid D])$
\bar{D} : non-diseased	$\vec{p}_{\bar{D}} = (p_{ijkl \bar{D}}) = (\mathbb{P}[T_1 = i, T_2 = j, T_3 = k, T_4 = l \mid \bar{D}])$

Model specification:

$\text{Sp}_t \sim \text{Beta}(a_{\text{Sp}_t}, b_{\text{Sp}_t}), \quad t = 1, 2, 3, 4$	$\text{Se}_t \sim \text{Beta}(a_{\text{Se}_t}, b_{\text{Se}_t}), \quad t = 1, 2, 3, 4$
$\pi \sim \text{Beta}(a_\pi, b_\pi)$	$n = z + y$
$Y \mid N, \pi \sim \text{Binomial}(N, 1 - \pi)$	$y \mid Y, \vec{p}_{\bar{D}} \sim \text{Multinomial}(Y, \vec{p}_{\bar{D}})$
$Z \mid N, \pi \sim \text{Binomial}(N, \pi)$	$z \mid Z, \vec{p}_D \sim \text{Multinomial}(Z, \vec{p}_D)$
$p_{00 \bar{D}} \sim \text{Uniform}(\text{Sp}_1 \cdot \text{Sp}_2, \min(\text{Sp}_1, \text{Sp}_2))$	$p_{11 D} \sim \text{Uniform}(\text{Se}_1 \cdot \text{Se}_2, \min(\text{Se}_1, \text{Se}_2))$

$$p_{ijkl|\bar{D}} = \text{Sp}_1^{1-i} (1 - \text{Sp}_1)^i (1 - \text{Sp}_2 - \text{Sp}_3 + p_{00|\bar{D}})^{j \cdot k} (\text{Sp}_2 - p_{00|\bar{D}})^{(1-j) \cdot k} \cdot (\text{Sp}_3 - p_{00|\bar{D}})^{j \cdot (1-k)} p_{00|\bar{D}}^{(1-j)(1-k)} \text{Sp}_4^{1-l} (1 - \text{Sp}_4)^l$$

$$p_{ijkl|D} = \text{Se}_1^i (1 - \text{Se}_1)^{1-i} p_{11|D}^{j \cdot k} (\text{Se}_2 - p_{11|D})^{j \cdot (1-k)} (\text{Se}_3 - p_{11|D})^{(1-j) \cdot k} \cdot (1 - \text{Se}_2 - \text{Se}_3 + p_{11|D})^{(1-j)(1-k)} \text{Se}_4^l (1 - \text{Se}_4)^{1-l}$$

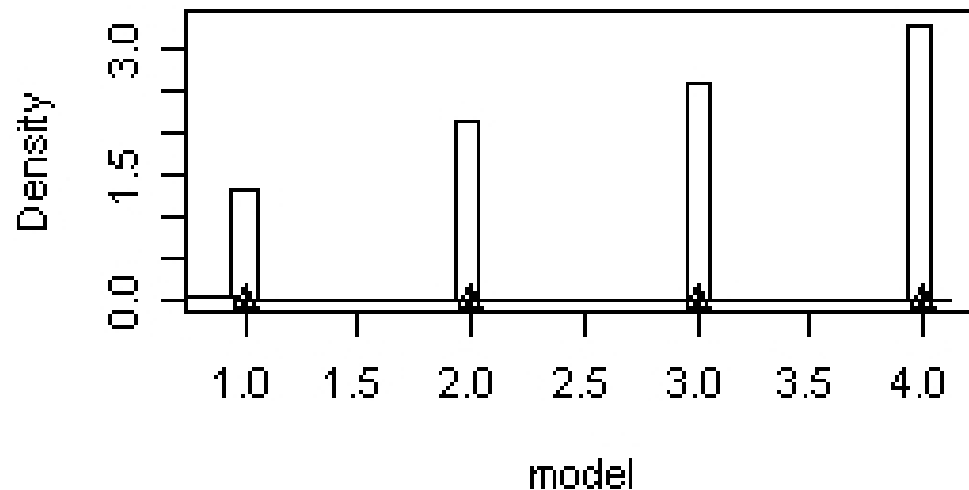
Summary statistic

Because no prior information on the model parameters was available, I used Beta(1,1) priors. One iteration of the MCMC algorithm consists of first updating the parameters and then jumping between the models. The resulting summary statistic is given below.

	mean	sd	2.5%	25%	50%	75%	97.5%
π	0.17	0.07	0.07	0.13	0.16	0.20	0.33
Se_{Bact}	0.75	0.13	0.49	0.67	0.76	0.86	0.98
Se_{SCC}	0.76	0.17	0.34	0.59	0.73	0.84	0.97
Se_{MAA}	0.77	0.16	0.39	0.67	0.80	0.90	0.99
Se_{MEC}	0.51	0.14	0.25	0.41	0.51	0.60	0.79
Sp_{Bact}	0.82	0.05	0.71	0.79	0.82	0.85	0.94
Sp_{SCC}	0.96	0.03	0.89	0.95	0.97	0.98	1.00
Sp_{MAA}	0.94	0.03	0.87	0.93	0.95	0.97	1.00
Sp_{MEC}	0.61	0.05	0.59	0.57	0.61	0.64	0.71
$p_{00 \bar{D}}$	0.92	0.04	0.83	0.90	0.93	0.95	0.98
$p_{11 D}$	0.61	0.17	0.24	0.49	0.63	0.74	0.91

Model selection

The posterior density of the parameter ‘model’ shows that the chains tend to remain longer in the full conditional dependence model. This suggests that the SCC test and the MAA test are conditionally dependent.



However, further analysis has shown that this tendency is not significant. So I can conclude that all tests are independent.

Discussion

- In Europe the **prevalence** of mastitis was estimated to be around 20%. So this farm lies near the average.
- The **MEC test** is very cheap to perform, but it has also a very bad performance. This is due to the fact that the electrical conductivity is influenced by mastitis infection but also by factors such as fat content and temperature of the milk, milking intervals and food types.
- The **bact. test** achieves moderately good results and is very expensive.
- The **MAA test** and the **SCC test** have the same moderate costs and both show a very good test accuracy.
- All **sensitivities** are lower than the **specificities**. This suggests that the tests could be improved by choosing a lower discrimination threshold for a positive test result.
- Further research should include computing the **ROC curves** (plot of Se versus $(1 - Sp)$ as the threshold is varied) to compare the overall test accuracy of the mastitis diagnostic tests and find the optimal threshold for each test.