# Certifying Countrywide Disease Freedom in Animal Livestock Populations:

# A Bayesian Approach

Eric A. Suess
California State University, Hayward

Ian Gardner and Wes Johnson
University of California, Davis

**The Goal:**

As part of free-trade agreements, countries might try to provide evidence as to their freedom from infectious agents in animal populations that potentially impact their ability to have unrestricted trade in animals and animal products with other countries.

**Criteria:**

Evidence of disease freedom might be based on a criteria such as a lack of clinical disease for a certain length of time, cessation of use of vaccines which might disguise the condition, no diagnoses at local diagnostic laboratories, and often some test based data (survey or ongoing routine surveillance).

**Risk:**

Some countries that attempt to demonstrate disease freedom might be of risk because neighboring countries have the infectious agent. For example, countries in Europe or South America may have higher risk of disease. Or others might be of substantially lower risk because of geographic isolation, e.g., Australia, New Zealand, etc..

**Data:**

Usually, most countries will try to conduct a national survey using internationally-recognized diagnostic tests on a large sample of animals. These surveys could be slaughterhouse-based or based on the testing of live animals in herds. If the latter occurs, usually the testing will be performed using a two-stage cluster sampling scheme with the selection of $k$ herds and then a random sample of $n$ animals.

The diagnostic tests will usually be serologic in nature because they are cheap and easy to use. However, such tests always will have imperfect sensitivity and specificity.

**The Problem:**

- Certification of a country as "free" from animal disease is modeled with an indicator variable:

$$\begin{aligned} Z = & \quad 1 \qquad \text{country is infected} \\ = & \quad 0 \qquad \text{not} \end{aligned}$$

- The data available are the diagnostic test results from a survey of individual live animals within herds.
  - Cluster Sampling.
  - Imperfect diagnostic test, i.e., sensitivity $P(T^+|D)$ and specificity $P(T^-|D^c)$ are below 1.
- We incorporate prior expert opinion, i.e., Bayesian Approach.

**Model:**

| | | | |
|---|---|---|---|
| $Z =$ | 1 | country is diseased | w.p. $\alpha$ |
| $=$ | 0 | not | w.p $1-\alpha$ |

The data results from cluster sampling. We assume two populations exist from which the $i^{th}$ herd can be selected, either diseased or non-diseased.

Define

$\lambda_i$ = prevalence of the disease in the population from which the $i^{th}$ herd was sampled.

$\pi_i$ = prevalence of the disease in a diseased herd.

**Model: (Continued)**

We also assume that the $i^{th}$ herd, if diseased, may be sampled from a population with varying prevalence. So

$$\pi_i \sim beta(a, b)$$

where $(a, b)$ is random.

Thus

$$\begin{aligned} \lambda_i|\pi_i = &\quad \pi_i \quad \text{w.p. } \tau \\ = &\quad 0 \quad \text{w.p. } 1-\tau \end{aligned}$$

So

$$\lambda_i|\pi_i, \tau \sim \pi_i Bernoulli(\tau)$$

**Latent Data for the herd level.**

Define $t_i$ = status of herd $i$.

$$\begin{aligned} t_i = &\quad 1 \quad \text{w.p. } \tau \\ = &\quad 0 \quad \text{w.p. } 1-\tau \end{aligned}$$

So now we have $\lambda_i = \pi_i t_i$.

**Latent Data for the individual animals.**

Define $v_{ij}$ = "true" status of animal $j$ within herd $i$.

$$\begin{aligned} v_{ij} = &\quad 1 \quad j^{th} \text{ animal in } i^{th} \text{ herd is diseased} \\ = &\quad 0 \quad \text{not} \end{aligned}$$

So

$$v_{ij}|Z = 1, t_i = 1 \sim Bernoulli(\pi_i)$$

**The Test Results Data:**

The data that collected from a cluster sample of herds are the individual test results $X_{ij}$, where

$$\begin{aligned} X_{ij} = & \quad 1 \quad \text{if the } j^{th} \text{ animal in the } i^{th} \text{ herd tests positive} \\ = & \quad 0 \quad \text{otherwise} \end{aligned}$$

The conditional distribution of $X_i$ given that the country is diseased, $Z = 1$, the herd level prevalence $\lambda_i$ and the test parameters $\eta$ and $\theta$, is binomial. That is

$$X_i|Z = 1, \lambda_i, \eta, \theta \sim Bin(n_i, \lambda_i\eta + (1 - \lambda_i)(1 - \theta))$$

and for $Z = 0$

$$X_i|Z = 0 \sim Bin(n_i, (1 - \theta)).$$

Important: This model generates correlation in disease status between animals within each herd, but leaves the disease status independent between herds.

**Our Approach:**

Our approach to parameter estimation is from a Bayesian perspective and we implement the Gibbs sampler to produce posterior estimates or the parameters in the model.

**Bayesian Approach using the Gibbs sampler:**

The Bayesian approach to estimating the parameters in this model assumes that the unknown quantities are realizations of random variables having certain prior distributions. Inferences about the unknowns are made by investigating the joint posterior distributions that reflect the gain of knowledge about the unknown parameters given the observed data. Instead of directly computing the Bayesian estimates, the Gibbs Sampler is used to iteratively generate random samples from the joint posterior distribution of the unknowns and an estimate, such as the posterior mean, is used to estimate the parameters. Inferences about the unknown parameters are computed using the marginal conditional posterior distributions of these unknowns.

**Bayesian Statistics:**

Here $\mathbf{X}$ is the data and $\Theta$ is the parameter set.

Model:

$$p(\mathbf{X}|\mathbf{\Theta})$$

Prior distribution:

$$p(\mathbf{\Theta})$$

Joint distribution:

$$p(\mathbf{X}, \mathbf{\Theta}) = p(\mathbf{X}|\mathbf{\Theta})p(\mathbf{\Theta})$$

Posterior distribution:

$$\begin{aligned} p(\mathbf{\Theta}|\mathbf{X}) &= \frac{p(\mathbf{X}|\mathbf{\Theta})p(\mathbf{\Theta})}{\int p(\mathbf{X}|\mathbf{\Theta})p(\mathbf{\Theta})d\mathbf{\Theta}} \\ &\propto p(\mathbf{X}|\mathbf{\Theta})p(\mathbf{\Theta}) \end{aligned}$$

**Gibbs Sampler:** (ref. Gelfand and Smith 1990)

$\mathbf{\Theta} = (\theta_1, \theta_2)$

$p(\mathbf{\Theta}|\mathbf{X}) = p(\theta_1, \theta_2|\mathbf{X})$

Given the initial value $\theta_2^{(0)}$ for $h = 1, ..., Reps$

1. Sample $\theta_1^{(h)}$ from $p(\theta_1|\mathbf{X}, \theta_2^{(h-1)})$.
2. Sample $\theta_2^{(h)}$ from $p(\theta_2|\mathbf{X}, \theta_1^{(h)})$.
3. Set $h = h + 1$ and go to 1.

The our inferences are based on the simulated data after the chain has stabalized.

$$(\theta_1^{(BurnIn+1)}, \theta_2^{(BurnIn+1)}), ..., (\theta_1^{(Reps)}, \theta_2^{(Reps)})$$

The theory behind the Gibbs sampler says, $\mathbf{\Theta}^{(1)}, ..., \mathbf{\Theta}^{(Reps)}$ are realizations of a stationary Markov Chain, with transition probability from $\mathbf{\Theta}^{(h-1)}$ to $\mathbf{\Theta}^{(h)}$,

$$T(\mathbf{\Theta}^{(h-1)}, \mathbf{\Theta}^{(h)}) = p(\theta_1|\mathbf{X}, \theta_2^{(h-1)})p(\theta_2|\mathbf{X}, \theta_1^{(h)})$$

By Ergodic theory, we can calculate estimates of, for example, $\theta_1$ by

$$\frac{1}{Reps}\sum_h \theta_1^{(i)} \xrightarrow{a.s.} E[\theta_1|\mathbf{X}],$$

$$Reps \to \infty.$$

**Priors:**

The uncertainty in all of the parameters are modeled with probability distributions, where the hyperparameters are chosen to reflect expert opinion.

- The "initial" probability the country is diseased, $\alpha \sim beta(a_\alpha, b_\alpha)$.
- The probability of a diseased herd, $\pi_i \sim beta(a, b)$.
- The hyperparameters that give a different distribution for the prevalence, $(a, b)$, are assumed to be independently gamma distributed.
- The herd level prevalence, $\tau \sim beta(a_\tau, b_\tau)$.
- Sensitivity, $\eta \sim beta(a_\eta, b_\eta)$.
- Specificity, $\theta \sim beta(a_\theta, b_\theta)$.

**The Likelihood:** For $Z = 1$

$$X_{ij}|Z = 1, \{v_{ij}\}, \{t_i\}, \{\pi_i\}, \eta, \theta \sim Bernoulli[\eta^{v_{ij}}(1-\theta)^{1-v_{ij}}].$$

For $Z = 0$

$$X_{ij}|Z = 0 \sim Bernoulli[(1-\theta)].$$

**The parameter set:**

$\mathbf{\Theta} = \{Z, \{v_{ij}\}, \{t_i\}, \{\pi_i\}, (a, b), \alpha, \tau, \eta, \theta\}$

**Hyperparamters:**

$a_\alpha, b_\alpha, a_\tau, b_\tau, a_\eta, b_\eta, a_\theta, b_\theta$

**Fixed parameters:**

$k, n_i$

**Overall Prior:**

$$p(\{\pi_i\}, (a, b), \alpha, \tau, \eta, \theta)$$

**Joint Density:**

$$p(\{X_{ij}\}, \{v_{ij}\}, \{t_i\}, \{\pi_i\}, (a, b), \alpha, \tau, \eta, \theta | Z = 1)$$

**Joint Posterior:**

$$p(\mathbf{\Theta}|\mathbf{X}) \propto p(\mathbf{X}|\mathbf{\Theta})p(\mathbf{\Theta})$$

**Conditional Marginal Posterior Distributions:**

$$v_{ij}|Z=1, X_{ij}=1, t_i=1, \text{“}rest\text{”} \sim Bernoulli(P(D|X_{ij}=1))$$

$$P(D|X_{ij}=1) = \frac{\eta\pi_i}{\eta\pi_i + (1-\theta)(1-\pi_i)}$$

$$v_{ij}|Z=1, X_{ij}=0, t_i=1, \text{“}rest\text{”} \sim Bernoulli(P(D|X_{ij}=0))$$

$$P(\bar{D}|X_{ij}=1) = \frac{(1-\eta)\pi_i}{(1-\eta)\pi_i + \theta(1-\pi_i)}$$

$$v_{ij}|Z=1, X_{ij}=0, t_i=0, \text{“}rest\text{”} \sim Bernoulli(0)$$

For $t_i$, the indicator of herd $i$ being infected or not. If $v_{ij} = 1$ for any $j$ within herd $i$, then

$$t_i | \{X_{ij}\}, \text{“}rest\text{”} \sim Bernoulli(1).$$

If $v_{ij} = 0$ for all $j$ within herd $i$, then

$$t_i | \{X_{ij}\}, \text{“}rest\text{”} \sim Bernoulli\left(\frac{(1-\pi)^{n_i}\tau}{(1-\pi_i)^{n_i} + 1(1-\tau)}\right).$$

Or if $Z = 0$, then

$$t_i | Z = 0, \text{“}rest\text{”} \sim Bernoulli(0).$$

$(a, b)|\{X_{ij}\}$, "$rest$" are sampled using Adaptive Rejection.

$$\pi_i|\{X_{ij}\}, \text{“}rest\text{”} \sim beta\left(a + \sum_{j=1}^{n_i} v_{ij}, b + n_i - \sum_{j=1}^{n_i} v_{ij}\right)$$

$$\alpha|\{X_{ij}\}, \text{“}rest\text{”} \sim beta\left[a_\alpha + Z, b_\alpha + (1 - Z)\right]$$

$$\tau|\{X_{ij}\}, \text{“}rest\text{”} \sim beta\left(a_\tau + \sum_{i=1}^{k} t_i, b_\tau + k - \sum_{i=1}^{k} t_i\right)$$

$$\eta|\{X_{ij}\}, \text{“}rest\text{”} \sim beta\left(a_\eta + \sum x_{ij}v_{ij}, b_\eta + \sum (1 - x_{ij})v_{ij}\right)$$

$$\theta|\{X_{ij}\}, \text{“}rest\text{”} \sim beta(a_\theta + \sum (1 - x_{ij})(1 - v_{ij}), b_\theta + \sum x_{ij}(1 - v_{ij}))$$

$$Z|\{X_{ij}\}, \text{“}rest\text{”} \sim Bernoulli\left(\frac{\alpha(1-\tau)^k}{\alpha(1-\tau)^k + (1-\alpha)}\right)$$

**Steps To Perform The Gibbs Sampler:**

Given the initial values:

1. For $i = 1, ..., k, j = 1, ..., n_i$ sample $v_{ij}^{(h)} \sim Bernoulli.$
2. For $i = 1, ..., k$ sample $t_i^{(h)} \sim Bernoulli.$
3. Sample $(a, b)$ using Adaptive Rejection.
4. For $i = 1, ..., k$ sample $\pi_i^{(h)} \sim beta.$
5. Sample $\alpha^{(h)} \sim beta.$
6. Sample $\tau^{(h)} \sim beta.$
7. Sample $\eta^{(h)} \sim beta.$
8. Sample $\theta^{(h)} \sim beta.$
9. Sample $Z^{(h)} \sim Bernoulli.$

**Simulation Results:**

Presented are the results for a simulated data set. The data set contains $k = 50$ herds with $n = 20$ animals tested with in each herd. The herd level prevalence is $\tau = 0.02$ and the within herd level prevalence is $\pi_i = 0.10$. The testing parameters are $\eta = 0.95$ and $\theta = 0.90$.

**Priors used:**

Each prior is selected from the following information solicited from an Epidemiologist with knowledge of the subject.

1. $\alpha \sim beta$ with mode 0.02 and a 95 percentile of 0.04.
2. $\tau \sim beta$ with mode 0.02 and a 95% percentile 0.05.
3. $\pi_i \sim beta$ with mode 0.05 and 95% percentile 0.10.
4. $\eta \sim beta$ with mode 0.95 and a 5% percentile 0.90.
5. $\theta \sim beta$ with mode 0.90 and a 5% percentile 0.85.
6. Note that $(a, b)$ is held fixed in these results. The A.R. part is still under construction.

**Estimation:**

The posterior estimates are produced from the means of the Markov chains produced by the Gibbs sampler.

1. Posterior estimate that this country has the disease is 0.3602.
2. Posterior estimate of the herd level prevalence is 0.242.

**Conclusions:**

1. This is a *Bayesian approach* to the problem.
2. The model extensively uses *latent data*.
3. Incorporates expert knowledge.
4. Gibbs sampling is used with an Adaptive Rejection step.
5. Ultimately we will be able to make estimates of the probability that a country has a disease (or not) and put them onto existing scales of disease freedom.

**References:**


Audige, L., Doherr, M.G., and Salman, M. (1999) A Quantitative Approach in Declaring a Country as Free of a Disease. Accepted for publication.

Cowling, David, Johnson, Wesley O. and Gardner, Ian (1998) Effect of Dependence on Prevalence Inferences using Pooled Samples. Accepted for publication.

Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-Based Approaches to Calculating Marginal Densities *Journal of the American Statistical Association* **85**, 398-409.

Gilks, W.R. and Wild, P. (1992)Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics* **41**, 337-348.

Vose, D.J. (1998) Risk Analysis in Relation to the Imporation and Exportation of Animal Products. *Rev. sci. tech. Off. int. Epiz* **16**, 17-29.