Hierarchical Bayesian model for certification of a country as “free” from an animal disease

Eric A. Suess
Department of Statistics
California State University, Hayward

Ian Gardner
Department of Medicine and Epidemiolgy
School of Veterinary Medicine
University of California, Davis

Wes Johnson
Division of Statistics
University of California, Davis

Outline:

- Background
- Disease Freedom
- Current Approaches
  - Qualitative
  - Quantitative
- Hierarchical Model
- Bayesian Approach: Gibbs Sampling
- Example: Newcastle Disease

Background:

- Risk analysis for trade in animal products.
- Countries are interested in new trading opportunities and maintaining current trade.
- Is the country disease free?
- Needed for risk analysis to make policy decisions.

Disease Freedom:

- “Disease freedom”
  - requires a *perfectly* sensitive test
  - *All* animals
  - *All* negative tests
- Level of Disease Freedom may not mean total freedom in all situations.
  - prevalence $<$ threshold

Current Approaches:

- Qualitative
- Quantitative
  - Monte Carlo Simulation

Data:

- Two-stage cluster sample.
- Random selection of k herds and a random sample of n animals within each herd.

Bayesian Model:

Country level:

$Z =$ "true" status of the country.

$$\begin{array}{rccl} Z = & 1 & \alpha & \text{diseased} \\ = & 0 & 1-\alpha & \end{array}$$

Herd level: (latent data)

$t_i =$ "true" status of herd i.

$$\begin{aligned} t_i = & \quad 1 \quad \tau \qquad\qquad i^{th} \text{ herd is diseased} \\ = & \quad 0 \quad 1 - \tau \end{aligned}$$

Within herd level:

Assume two populations exist from which the i^{th} herd can be selected, either diseased or non-diseased.

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

π_i = prevalence of the disease in a diseased herd.

Prevalence Within Diseased Herds:

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

where (a, b) is random.

Prevalence Within Herds:

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

So

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

Within herd level: (latent data)

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

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

So

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

Data:

$$X_{ij} = 1 \quad \text{if the } j^{th} \text{ animal in the } i^{th} \text{ herd tests positive}$$
$$= 0 \quad \text{otherwise test negative}$$

Imperfect diagnostic test:

1. sensitivity $\eta = P(T^+|D) < 1$
2. specificity $\theta = P(T^-|\bar{D}) < 1$

$X_i = \sum_j X_{ij}$ = number of test positive animals in the i^{th} herd.

If $Z = 1$

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

If $Z = 0$

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

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

Bayesian Statistics:

$\mathbf{X}$ is the data vector and Θ is the parameter set.

Model: (or likelihood)

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

Prior distribution:

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

Joint distribution:

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

Posterior distribution: (Bayes Theorem)

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

Gibbs Sampler: (ref. Gelfand and Smith 1990)

Two parameter case: $\mathbf{\Theta} = (\theta_1, \theta_2)$ $p(\mathbf{\Theta}|\mathbf{X}) = p(\theta_1, \theta_2|\mathbf{X})$

Want the marginals, $p(\theta_1|\mathbf{X})$ and $p(\theta_2|\mathbf{X})$

Can get the conditional marginals, $p(\theta_1|\mathbf{X}, \theta_2)$ and $p(\theta_2|\mathbf{X}, \theta_1)$

Initial value $(\theta_1^{(0)}, \theta_2^{(0)})$ 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.

Inferences are calculated using the simulated data after the chain has stabilized.

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

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 $\mathbf{\Theta}^{(h)}$,

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

Justified by Ergodic theory, we can calculate estimates of the parameter using means. For example,

$$\hat{\theta}_1 = \frac{1}{Reps} \sum_h \theta_1^{(i)} \xrightarrow{a.s.} E[\theta_1 | \mathbf{X}],$$

$$Reps \to \infty.$$

Priors:

- Country level, $\alpha \sim beta(a_\alpha, b_\alpha)$.
- Herd level, $\tau \sim beta(a_\tau, b_\tau)$.
- Within herd level, $\pi_i \sim beta(a, b)$.
- The distribution from which the prevalence a diseased herd results from is assumed to be a beta distribution with unknown parameters (a, b). Adaptive Rejection Sampling.
- Sensitivity, $\eta \sim beta(a_\eta, b_\eta)$.
- Specificity, $\theta \sim beta(a_\theta, b_\theta)$.

The Likelihood:

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

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

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

Joint Posterior:

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

Conditional Marginal Posterior Distributions:

- "true" status of each animal.

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

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

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

$$P(D|X_{ij}=0) = \frac{(1-\eta)\pi_i}{(1-\eta)\pi_i + \theta(1-\pi_i)}$$

$$v_{ij}|Z=1, t_i=0, \text{``}rest\text{''} \sim Bernoulli(0)$$

- "true" status of each herd

  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

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

  If $Z = 0$, then

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

- prevalence of a diseased herd.

$$\pi_i | \{X_{ij}\}, \text{“}rest\text{”} \sim beta\left(a + \sum_j v_{ij}, b + n - \sum_j v_{ij}\right)$$

- hyperparameters for the distribution on the π_i.

$$(a, b) | \{X_{ij}\}, \text{“}rest\text{” sampled using Adaptive Rejection.}$$

- probability the country is diseased.

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

- probability a herd is diseased.

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

- sensitivity.

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

- specificity.

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

- country status.

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

Adaptive Rejection Sampling: (ref. Gilks and Wild 1992)

$$\prod_{i=1}^{k} \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi_i^{a-1} (1-\pi_i)^{b-1} p(a) \right]^{t_i}$$

log-concavity in a given b.

Steps To Perform The Gibbs Sampler:

Given the initial values:

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

Conclusions:

- Bayesian approach.
  - Uses prior knowledge.
  - Latent data.
  - Imperfect test.
- Variable within herd-level prevalence.
- Country/Herd/Animal inference.
- Priors for other risk analysis.

Further Work:

- Variable sample size within herds.
- Application to continuous surveillance.

References:

Audigé, L., Beckett (1999) A Quantitative Assessment of the Validity of Animal-health Surveys Using Stochastic Modeling *Journal of Preventitive Veterinary Medicine* **38**, 259-276.

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.

Gohm, Thür, Audigé, and Hoffmann (1999) A Survey of Newcastle Disease in Swiss Lay-hen Flocks Using Serological Testing and Simulation Modeling. *Journal of Preventitive Veterinary Medicine* **38**, 277-288.

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