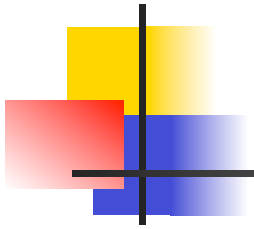
Bayesian modeling of animal and herd-level prevalence

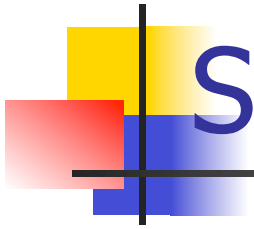
Adam Branscum
Wes Johnson
Dept. of Statistics, U.C. Davis

Ian Gardner
Vet Med: Medicine and Epidemiology, U.C. Davis



Overview

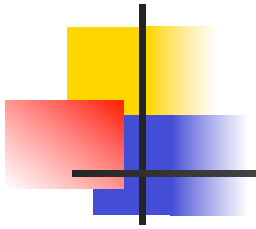
- Single herd
  - Binomial sampling
  - Hypergeometric sampling
- Multiple herds
- Importance of using a mixture model
- Conclusions



Single Herd, Binomial Sampling

GOALS:

- Estimate the prevalence π of disease in the herd
- Account for uncertainty and incorporate prior information about Se, Sp, and π
- Allow π to equal 0 with positive probability



Data

- Obtain random sample of size n assuming binomial sampling
- Each sampled animal subjected to an imperfect diagnostic test with sensitivity Se and specificity Sp.
- Let T+ denote the number of reactors out of the n tested

Model

T+ $|\pi$,Se,Sp ~ Bin[n, πSe + (1- π)(1-Sp)]

$\pi = \pi^*$ with probability τ

$= 0$ with probability 1- τ

$$\pi^* \sim Beta(a_\pi, b_\pi)$$

$$Se \sim Beta(a_{Se}, b_{Se}),\ Sp \sim Beta(a_{Sp}, b_{Sp})$$

τ = constant (or τ ~ Beta)

Example: Prevalence of Johne's Disease in a California Herd

- Of n=60 cows tested with an ELISA, T+=1
- Prior information for Se, Sp, τ and π^* elicited from experts
- Prior modes: Se=0.25, Sp=0.98, π^*=0.12 and τ =0.55
- Beta priors obtained using our free BetaBuster program.

BetaBuster

Input Information

95 % sure that x greater than 0.5 and Mode at : 0.95

Density: Beta

a : 4.7719

b : 1.1985

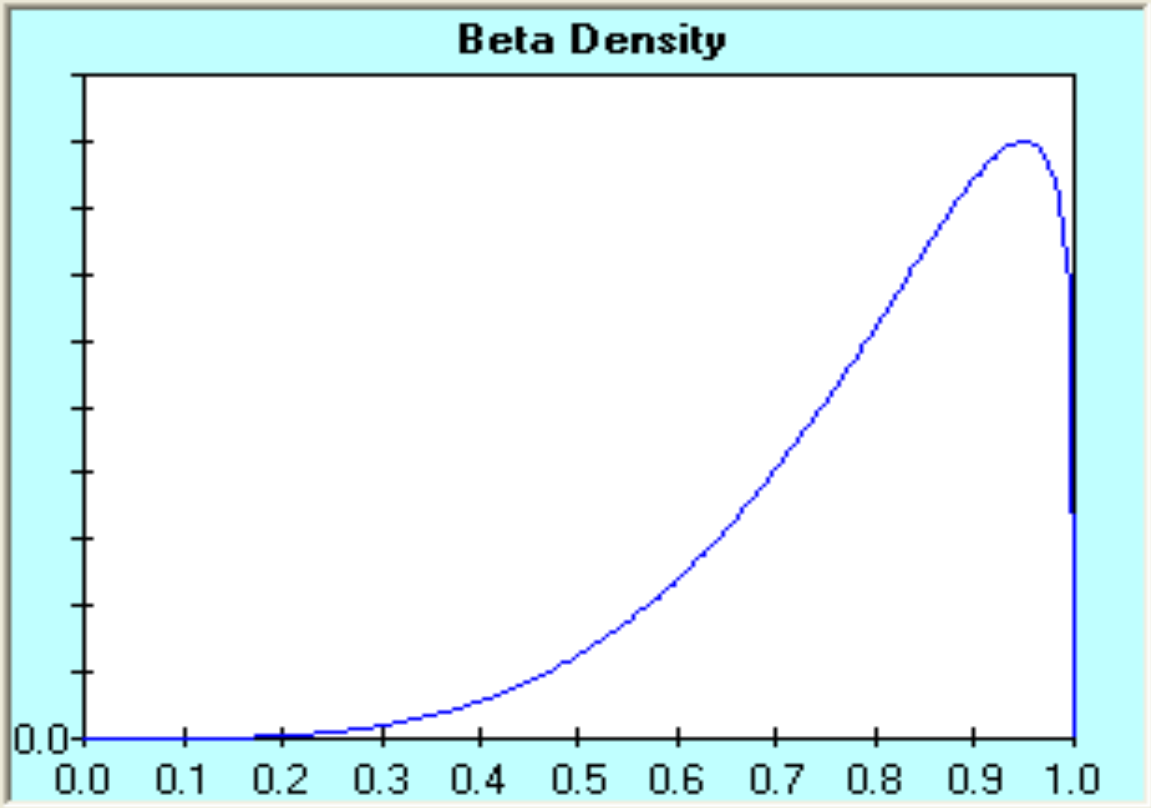
Mean : 0.7993

Variance : 0.023017868

Median : 0.8334

2.5% : 0.4307

97.5% : 0.9896



Set Priors Help Exit

Inferences

- Model fitted in freeware program WinBUGS

  www.mrc-bsu.cam.ac.uk/bugs

- Inferences:

  P(π =0 | T+=1) = 0.61

  Post. median for π^* = 0.11

  95% PI (0.05, 0.19)

Single Herd, Hypergeometric Sampling

- If n~herd size, distribution of T+ is very complicated
- Implemented in the free program BDFree (also a sample size calculator)
- Available at www.epi.ucdavis.edu/diagnostictests/

Multiple Herd Setting

GOALS:

- Estimate proportion of infected herds, τ
- A simple extension of the single herd model

$$T_i + | \pi_i, Se, Sp \sim Bin(n_i, p_i)$$
$$p_i = \pi_i Se + (1 - \pi_i)(1 - Sp)$$

- π_i modeled as a mixture of point mass at 0 and a beta distribution
- τ also modeled as a mixture to allow $\tau = 0$

Example: Herd-level prevalence estimation for Johne's Disease

- 29 California dairy herds with 60 cows sampled per herd

No. T+	0	1	2	3	>3
No. herds	4	4	8	3	10

- Inferences

  - Posterior median for $\tau = 0.58$

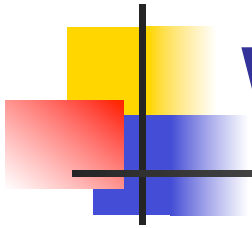
    95% PI:(0.30, 0.85)

Importance of allowing π =0

- Sample n=50 animals from a single herd with $\pi_T = 0, Se_T = 0.8, Sp_T = 1$
  Hence T+=0

Inf. herds τ	0.25	0.50	0.75	1
P(π=0|T+=0)	0.88	0.70	0.44	0

- When τ = 1, PM for π =0.01 (0.0004,0.05)

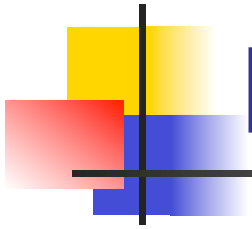


WinBUGS Code

```
model
{  for (i in 1:k)
y[i] ~ dbin(p[i], n[i])
p[i] <- pi[i]*Se + (1-pi[i])*(1-Sp)
z[i] ~ dbern(tau)
pistar[i] ~ dbeta(alpha,beta)
pi[i]  <- z[i]*pistar[i]
```

WB Code

```
pieq0[i] <- equals(p[i],0)
prpistar ~ dbeta(alpha,beta)}
mu <- alpha/gamma
gamma <- alpha + beta
Se ~ dbeta(.,.)     Sp ~ dbeta(.,.)
mu ~ dbeta(.,.)     gamma ~ dgamma(.,.)
}
```



More WB Code

list(n=50, y=0, tau=..)

list(pistar=c(0.10,…,0.1),
Se=0.80, Sp=0.99,
z=c(0,1,0,1…), gamma = 1, mu = .5, prpistar = .5)

Prevalence Distribution

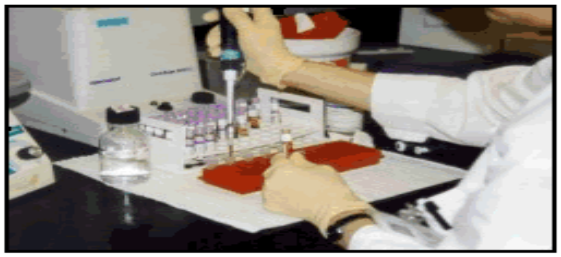
- Note the quantity prpistar
- The smoothed histogram of the MC interates of prpistar will give the Bayesian inference to the distribution of prevalences among the infected herds
- The mean of this distribution should be very close to the posterior mean for mu

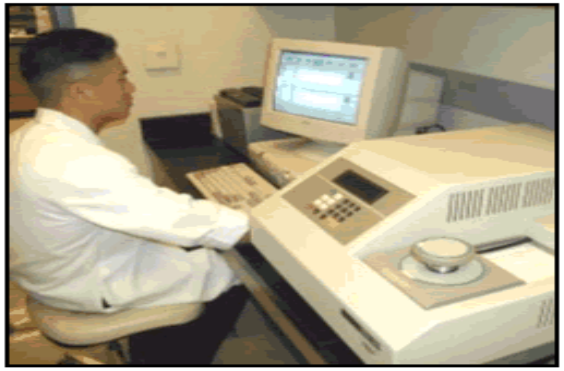
90th Percentile of the Prev Dist

- We could add a line to the code: prprob <- step(prpistar-.5), say
- The posterior mean of this would be the predictive probability that pistar is less than 0.5.
- If that turned out to be .9, say, then our inference would be that 90% of the herds have prevalences less than 0.5

Research Focus

The primarily **focus** of our research is the application of Bayesian methods to epidemiological problems. Bayesian methods are appealing because of their flexibility and analytic results have direct probabilistic interpretability, even in the absence of large sample sizes.





Research Goals

Our research **goals** are to develop:

- Innovative techniques that assess test accuracy in disease diagnosis at individual- and population-levels with or without a gold standard, facilitate interpretation of test results, and improve disease risk modeling based on diagnostic test results.

- Statistical methods to make improved inferences that a herd (or group of herds in a zone, state, or country) is free of important infectious agents based on herd-level test results.

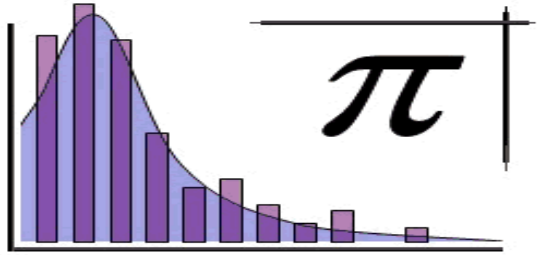
Website Synopsis

This web site includes:

- Software for a variety of diagnostic test and prevalence estimation problems.

- Reference papers and worked examples through a series of modules.

- A glossary of epidemiologic terms.



Funding for research described on this site was provided in part by the USDA NRI Competitive Grants Program