

Bayesian Hierarchical true prevalence estimation

Eleftherios Meletis & Aurélien Madouasse

2022-03-23

Objectives of this Section

- ▶ Explain Bayes' theorem | Bayesian Inference
- ▶ Why apparent prevalence is different than true prevalence?
- ▶ Apply a simple apparent & true prevalence estimation model
- ▶ Understand the full hierarchical model

Bayes' theorem

Bayes' rule

Describes the probability of an event based on prior knowledge

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (1)$$

Components

- ▶ $P(A|B)$: Prob of event A occurring given that B is true - Posterior probability
- ▶ $P(B|A)$: Prob of event B occurring given that A is true - Likelihood ~ function of A
- ▶ $P(A)$: Prob of event A occurring - Prior probability
- ▶ $P(B)$: Prob of event B occurring - Marginal probability ~ sum over all possible values of A

What we usually see/use

Bayes' rule

θ : parameter of interest | y : observed data

$$P(\theta|y) = \frac{P(y|\theta) * P(\theta)}{P(y)} \rightarrow P(\theta|y) \propto P(y|\theta) * P(\theta) \quad (2)$$

Where:

- ▶ $P(\theta)$: Prior probability of parameter(s) of interest;
- ▶ $P(y|\theta)$: Likelihood of the data given the parameters value(s)
- ▶ $P(\theta|y)$: Posterior probability of parameter(s) of interest given the data and the priors

Bayesian Inference - Summary & Example

To estimate the posterior distribution $P(\theta|y)$ we need to:

- ▶ *Specify* the **Prior distribution**: $P(\theta)$
- ▶ *Define* the **Likelihood** of the data: $P(y|\theta)$

Example: Bayesian apparent prevalence (ap) estimation

y out of n individuals test positive. Estimate the apparent prevalence.

Parameter of interest: $ap \in [0,1]$

Data: n tested, y positive

- ▶ Prior distribution for ap: $ap \sim \text{Beta}(a,b)$
- ▶ Likelihood: $y \sim \text{Binomial}(n,ap)$

Let's write our first JAGS model

```
ap_model <-  
'model {  
  
  # Define likelihood distribution of the data  
  # JAGS Binomial distribution Arguments: p, n  
  
  y ~ dbin(ap,n)  
  
  # Specify prior distribution for par of interest  
  # Uniform (non-informative) prior distribution  
  ap ~ dbeta(1,1)  
  
  #data# n, y  
  #monitor# ap  
  #inits# ap  
}
```

Let's run our first JAGS model

```
# Call JAGS
```

```
library(runjags)
```

```
# Provide Data
```

```
n = 4072
```

```
y = 1210
```

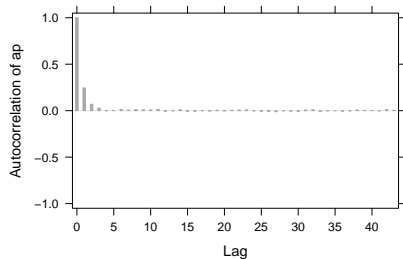
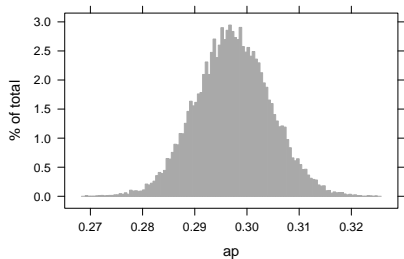
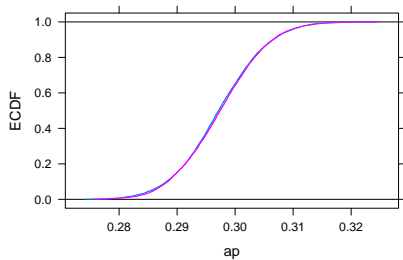
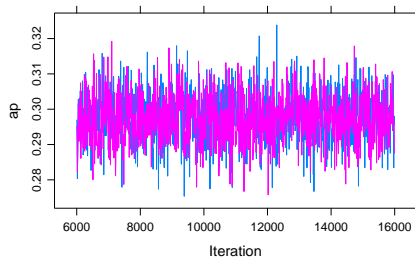
```
# Initial values for par of interest
```

```
ap <- list(chain1=0.05, chain2=0.95)
```

```
# Run the model
```

```
results <- run.jags(ap_model, n.chains=2,  
                    burnin=5000, sample=10000)
```

View results



Example: Bayesian true prevalence (tp) estimation

Assuming the absence of a perfect test we do not know how many individuals are truly positive/negative.

Instead we know that n individuals are tested with an imperfect test and y have a positive result.

Apparent and True prevalence

Apparent/True prevalence: ap/tp | Sensitivity: Se | Specificity: Sp

$$\begin{aligned} ap &= P(T^+) = P(T^+ \cap D^+) + P(T^+ \cap D^-) = \\ &P(D^+) * P(T^+|D^+) + P(D^-) * P(T^+|D^-) \\ &\rightarrow ap = tp * Se + (1 - tp) * (1 - Sp) \end{aligned} \tag{3}$$

Create a JAGS model for true prevalence estimation

Parameters of interest:

- ▶ $tp \in [0,1]$
- ▶ $Se \in [0,1]$
- ▶ $Sp \in [0,1]$

Prior distributions

- ▶ $tp \sim \text{dbeta}(1,1)$
- ▶ $Se \sim \text{dbeta}(25.4, 3.4)$
- ▶ $Sp \sim \text{dbeta}(95, 5)$

Data: n tested, y positive

- ▶ Likelihood: $y \sim \text{Binomial}(n, ap)$, $ap = tp * Se + (1 - tp) * (1 - Sp)$

Write JAGS model

```
tp_model <-  
'model {  
  
  y ~ dbin(ap,n)  
  ap <- tp*Se + (1-tp)*(1-Sp)  
  
  # Uniform (non-informative) prior distribution  
  tp ~ dbeta(1,1)  
  # Informative priors for Se and Sp  
  Se ~ dbeta(25.4, 3.4)  
  Sp ~ dbeta(95, 5)  
  
  #data# n, y  
  #monitor# tp, Se, Sp  
  #inits# tp, Se, Sp  
}
```

Let's run our JAGS model

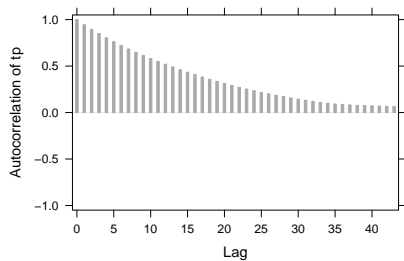
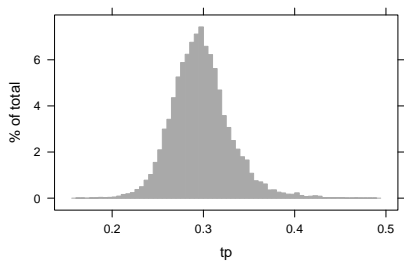
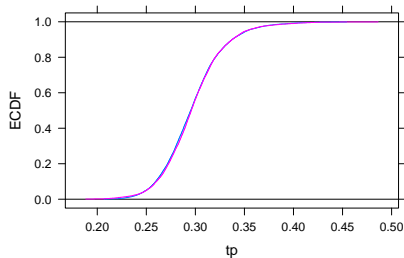
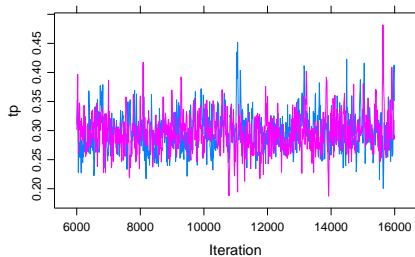
```
# Call JAGS
library(runjags)

# Provide Data
n = 4072
y = 1210

# Initial values for pars of interest
tp <- list(chain1=0.05, chain2=0.95)
Se <- list(chain1=0.05, chain2=0.95)
Sp <- list(chain1=0.05, chain2=0.95)

# Run the model
results <- run.jags(tp_model, n.chains=2,
                    burnin=5000, sample=10000)
```

View results



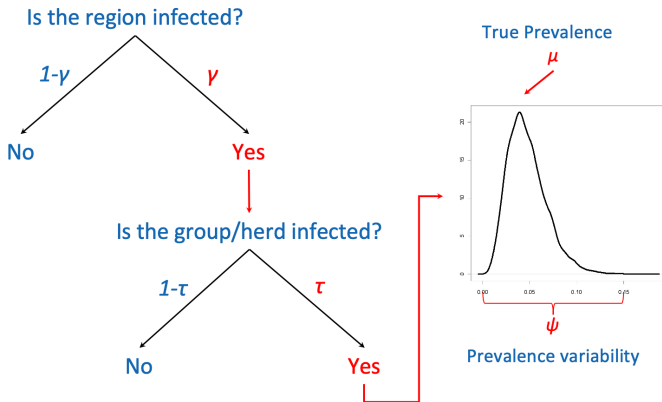
Bayesian hierarchical true prevalence estimation

Model breakdown

Captures the population structure and is divided in three parts that describe distribution of disease:

- ▶ Within herds which are
- ▶ Within regions in the
- ▶ Country under consideration

Full Model Figure



Branscum et al., 2005

Model outputs

- ▶ Animal-level true prevalence of infection within infected herds
- ▶ Herd-level prevalence of infection in the region
- ▶ Region-level prevalence of infection
- ▶ Probability that the whole country is free from infection
- ▶ Probability of freedom from infection at each level of hierarchy
- ▶ Probability that infection does not exceed a pre-specified critical level (e.g. 5%) in a specific region

Run the model with JAGS

The model and the data are already saved in the Section's folder.
Let's load and view the data.

```
##      region y    n
## 1         1 1 107
## 2         1 1  92
## 3         1 1 114
## 4         1 1 121
## 5         1 1 105
## 6         1 1  92
```

Run the model

```
hier_model <- run.jags("hier_model.txt",data=hier_data)

## Compiling rjags model...
## Calling the simulation using the rjags method...
## Adapting the model for 1000 iterations...
## Burning in the model for 4000 iterations...
## Running the model for 10000 iterations...
## Simulation complete
## Note: Summary statistics were not produced as there are
## variables
## [To override this behaviour see ?add.summary and ?runjags]
## FALSEFinished running the simulation

res <- summary(hier_model)

## Calculating summary statistics...
## Note: The monitored variables 'a[2,1]', 'a[2,2]', 'a[2,3]
```

Visualise results