Outline
- Introduction
- Bayesian regression models
- Correlated data
- MCMC
- Meta-analysis
- Mixed treatment comparisons
- Diagnostic testing with missing gold standard

Overview of the Bayesian approach
- Goal: learning about an unknown parameter $\theta$ (possibly a vector)
  - $\theta_0$ = true disease status
  - $\theta_0$ = hazard ratio
  - $\theta_0$ = probability that experimental treatment is better
  - $\theta_0$ = vector of regression coefficients
  - $\theta_0$ = missing data
  - etc...
- Data: $y$ (e.g. test result)
- Input to analysis:
  - Prior distribution: $P(\theta)$
  - Likelihood Function: $L(\theta|y) \propto P(y|\theta)$
The likelihood function

- A likelihood function is a function of the parameters of a probability model given the outcomes.
  - The likelihood of $\theta$, given outcome $y$, is equal to the probability of that observed outcome given $\theta$.
- For example, a Bernoulli random variable $Y$ takes on two possible values: 0 or 1
  - Likelihood function based on a Bernoulli observation:
    - Given that $y=1$, the likelihood function of $\theta$ is:
      \[ L(\theta|y=1) = P(Y=1|\theta) = \theta \]
    - Given that $y=0$, the likelihood function of $\theta$ is:
      \[ L(\theta|y=0) = P(Y=0|\theta) = 1-\theta \]

Prior Distributions

- Quantifiable prior belief about the parameter of interest
- Choice of priors is based on judgments and a degree of subjectivity cannot be avoided
  - Prior is not unique!
  - Sensitivity analysis is crucial in assessing the impact of particular distributions on the conclusions
- Construct prior, $P(\theta)$, based on prior belief
  - Conjugate priors result in posterior in same family
  - Non-informative or flat prior

Posterior distribution

- Output from analysis:
  - Posterior distribution:
    \[ P(\theta|Y) = \frac{P(\theta)L(\theta|Y)}{\int P(\theta)L(\theta|Y) d\theta} \]
Inference based on posterior distribution

Inferences based on summaries of the posterior distribution

- Point estimates:
  - Mean/Median/Mode
- Interval estimates:
  - One-sided credible intervals
  - Two-sided credible intervals
  - Equi-tail area
  - Narrowest interval
[HPD: highest posterior density intervals]

Choices of summary measures justified with loss functions [decision theory].

---

Primary packages we will use

- INLA
  - Download at: http://www.r-inla.org/download
- rjags (alternative choices R2jags, runjags)

---

Bayesian Regression Models
Generalized Linear Regression Models

- Mean: $E[Y_i | X_{i1}, X_{i2}, \ldots, X_{ip}] = \mu_i = g^{-1}(\eta_i)$ where $g$ is a link function
- Regression Model: $g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$
  - Linear regression model
  - Logistic regression model $g(\mu_i) = \log \left( \frac{\mu_i}{1 - \mu_i} \right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$
  - Probit regression model $g(\mu_i) = \Phi^{-1}(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$
  - Poisson regression model $g(\mu_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$

Bayesian GLM

- Mean: $E[Y_i | X_{i1}, X_{i2}, \ldots, X_{ip}] = \mu_i = g^{-1}(\eta_i)$ where $g$ is a link function
- Regression Model: $g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}$
- Priors:
  - Regression parameters: $(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$
  - “Nuisance” parameters (e.g. in linear regression $\sigma^2$)
- Note:
  - Regression coefficients have the same interpretation (e.g. difference in means; log-odds ratio; etc.)
  - Interpretation of inferential results are different (e.g. posterior mean; probability that the regression parameter lies in some interval; etc.)

Bayesian GLM in R

- We will use the INLA package
- Different approaches to estimation of GLMs
  - Approximate posterior inference (Bayesian CLT)
- Advantages:
  - Syntax very similar to traditional GLMs
  - No need for heavy programming (e.g. MCMC methods)
- Disadvantages:
  - Approximate method under small samples
  - Constrained by model formulations handled by the packages
Bayesian GLM in R: INLA package

- Integrated Nested Laplace Approximations (INLA)
  - Alternative to MCMC in (latent) Gaussian models

- Regression Model:
  \[ g(\mu_i) = \eta_i = \beta_0 + \sum \beta_j X_{ij} + \sum f_k(\hat{x}_k) + \epsilon_i \]

  \( g(\cdot) \) = unknown functions of covariates \( \hat{X} \)
  \( \beta_j \) = linear effects of covariates \( X \)
  \( \epsilon_i \) = unstructured terms

- Assumption in latent Gaussian models:
  Gaussian Prior for: \( \beta_0, \{ \beta_j \}, \{ f_k(\cdot) \}, \{ \epsilon_i \} \)

---

Bayesian GLM in R: INLA package

- Latent model:
  Let \( z \) represent the collection of all Gaussian variables:
  \( \beta_0, \{ \beta_j \}, \{ f_k(\cdot) \}, \{ \epsilon_i \} \)
  with distribution \( \pi(z|\theta) \) with mean 0 precision matrix \( Q(\theta) \).

- Model: \( \pi(y|z, \theta) \)
- Prior: Let \( \theta = (\theta_1, \theta_2) \) with prior \( \pi(\theta) \).
- Via Gaussian & Laplace approximations:
  \[ \tilde{\theta}(\theta|y) \propto \frac{\pi(z|\theta, y)}{\mathcal{L}_C(z|\theta, y)} \]

  \( \tilde{\theta}(\theta) \): mode of \( \pi(\theta|y, y) \)

  \( \mathcal{L}_C \): Gaussian approximation of \( \pi(\theta|y, y) \)

---

Latent Models in INLA
For more info: [http://www.r-inla.org](http://www.r-inla.org)
Likelihoods

Priors on hyperparameters

Multiple Linear Regression in R

FEV dataset: contains data on 654 children.

- seqnbr: case number (the numbers 1 to 654)
- subjid: subject identification number (unique for each different child)
- age: subject age at time of measurement (years)
- fev: measured FEV (liters per second)
- height: subject height at time of measurement (inches)
- sex: subject sex (1 = male, 2 = female)
- smoke: smoking habits (1 = yes, 2 = no)

Our goal is to assess the association between FEV and smoking status adjusting for age.
Multiple Linear Regression in R

```r
# Read FEV data set

# Examine a few entries of the data set
head(data)

# Summarize the variables
summary(data)

# Scatter plot of log(fev) by age
plot(log(fev) ~ age, data=data)

# Scatter plot of log(fev) by age, but stratified by smoking status
plot(log(fev) ~ age, type="n", data=data)
points(log(fev) ~ age, col='red', pch=15, data=data[data$smoke==1,])
points(log(fev) ~ age, col='blue', pch=16, data=data[data$smoke==2,])
legend("topleft", c("Smokers", "No Smokers"), col=c("red", "blue"), pch=c(15, 16))
```

Bayesian GLM in R: INLA package
Bayesian linear regression: FEV data

```r
# Bayesian GLM in R: INLA package
# Bayesian linear regression: FEV data

library(INLA)
fit = inla(log(fev) ~ smoke + age, data=data)

fit$summary.fix

fit$summary.hy

fit.prior1 = inla(log(fev) ~ smoke + age, data=data, control.family=list(hyper=list(prec=list(prior="normal", param=c(0, 10)))))

fit.prior1$summary.fix

fit.prior1$summary.hy
```

Making prior assumptions explicit
Survival Models: Notation

- Let $T$ be a continuous non-negative random variable representing survival times of individuals in some population.
  - Density function (pdf): $f(t)$
  - Distribution function (cdf): $F(t)$
    - Fraction of people dying by time $t$
  - Survival function: $S(t)$
    - Fraction of people surviving at time $t$
  - Hazard function: $h(t)$
    - Instantaneous risk of death
  - Cumulative Hazard: $H(t)$

Survival Models: Relationships

$h(t) = \frac{f(t)}{S(t)}$

$H(t) = \int_0^t h(u)du$

$F(t) = \int_0^t f(u)du$

$S(t) = 1 - F(t) = \exp(-H(t))$

$f(t) = h(t)S(t) = h(t)\exp(-H(t))$

- Likelihood contribution for a subject who dies
  - $f(t) = h(t)S(t)$
- Likelihood contribution for a subject who is censored
  - $S(t)$
- Thus, if $d$ is the indicator of death, we can write:
  - $[h(t)]^d S(t)$

Survival Models: Proportional Hazards

- Proportional Hazards (PH) Model:
  $$h(t) = h_0(t)\exp\left(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p\right)$$
- Parametric vs Semi-parametric PH model?
  - What is the form of the baseline hazard ($h_0(t)$) function?
PH regression models in R

- Data from the German Breast Cancer Study Group 2 contains the observations of 686 women:
  - horTh: hormonal therapy, a factor at two levels no and yes.
  - age: age of the patients in years.
  - menostat: menopausal status, a factor at two levels pre (premenopausal) and post (postmenopausal).
  - tsize: tumor size (in mm).
  - tgrade: tumor grade, an ordered factor at levels I < II < III.
  - pnodes: number of positive nodes.
  - prog: progesterone receptor (in fmol).
  - estrec: estrogen receptor (in fmol).
  - time: recurrence free survival time (in days).
  - cens: censoring indicator (0 = censored, 1 = event).

- Scientific question of interest: does receipt of hormone therapy affect the length of time to breast cancer recurrence?

> data publicly available in an R-package
> data("GBSG2", package="TH.data")

> summary(GBSG2)

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>horTh</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>age</td>
<td>0.00</td>
<td>46.00</td>
<td>53.00</td>
<td>53.05</td>
<td>61.00</td>
<td>80.00</td>
</tr>
<tr>
<td>menostat</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>tsize</td>
<td>3.00</td>
<td>20.00</td>
<td>25.00</td>
<td>29.34</td>
<td>35.00</td>
<td>120.00</td>
</tr>
<tr>
<td>tgrade</td>
<td>1.00</td>
<td>1.00</td>
<td>3.00</td>
<td>5.01</td>
<td>7.00</td>
<td>51.00</td>
</tr>
<tr>
<td>pnodes</td>
<td>81.0</td>
<td>444.0</td>
<td>596.0</td>
<td>110.0</td>
<td>131.8</td>
<td>2380.0</td>
</tr>
<tr>
<td>prog</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.44</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>estrec</td>
<td>0.00</td>
<td>56.8</td>
<td>1124.5</td>
<td>96.25</td>
<td>1144.0</td>
<td>2659.0</td>
</tr>
<tr>
<td>time</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

> ## (Semi-parametric) Cox PH model
> fit1 <- coxph(Surv(time, cens) ~ horTh, data=GBSG2)
> summary(fit1)

> ## Parametric survival (Weibull regression)
> library(eha)
> fit3 <- phreg(Surv(time, cens) ~ horTh, data=GBSG2, dist="weibull")
> summary(fit3)
Bayesian PH regression models in R:

Non-parametric

```r
> library(INLA)

## Bayesian non-parametric PH model

> fit <- inla(inla.surv(time, cens) ~ horTh, family="coxph", data=GBSG2)

> summary(fit)

Call:

Time used:
Pre-processing Running inla Post-processing Total

0.0948          0.5117          0.0574          0.6639

Fixed effects:

mean   sd   0.025quant  0.5quant  0.975quant    mode
(Intercept) -7.7078 0.1403 -7.9948   -7.7039   -7.4426  -7.6965   0
horThyes -0.3660 0.1249 -0.6145   -0.3650   -0.1237  -0.3628   0

Random effects:

Name   Model
baseline.hazard RW1 model

Model hyperparameters:

mean   sd   0.025quant  0.5quant  0.975quant    mode
Precision for baseline.hazard 1451.61 943.88  363.52   1221.72  3902.23     849.37

Expected number of effective parameters (std dev): 9.484 (1.091)
Number of equivalent replicates: 497.48
Marginal Likelihood: -1379.80

Bayesian PH regression models in R:

Parametric

```r

## Bayesian parametric PH model

> fit <- inla(inla.surv(time, cens) ~ horTh, family="weibull", data=GBSG2)

> summary(fit)

Call:

Time used:
Pre-processing Running inla Post-processing Total

0.0698          1.2193          0.0485          1.3376

Fixed effects:

mean   sd   0.025quant  0.5quant  0.975quant    mode
(Intercept) -9.5518 0.4442 -10.2282   -9.3908   -8.9047  -9.3598
horThyes -0.3891 0.1248 -0.6373   -0.3880   -0.1470  -0.3859

The model has no random effects

Model hyperparameters:

mean   sd   0.025quant  0.5quant  0.975quant    mode
alpha parameter for weibull 1.2651 0.0749 1.1438    1.2557   1.4339     1.2290

Expected number of effective parameters (std dev): 2.005 (0.00)
Number of equivalent replicates: 342.15
Marginal Likelihood: -2641.95
Analysis of Correlated Data

Example: Longitudinal data

Covariation in the socioeconomic determinants of self rated health and happiness: a multivariate multilevel analysis of individuals and communities in the USA

J V Subramanian, Daniel Kim, Ichiro Kawachi
Example: Spatial data

Example: Time series data

Modeling of Correlated Data:

**Motivation**

- Degree of "similarity" may help with prediction!
  - Lyme disease incidence rates more similar in closer neighborhoods
  - Incidence rates of flu more similar within "short" time periods
  - Incidence rates of flu with similar seasonal patterns (e.g. Winter) across years
  - Happiness rates more similar from individuals within the same communities
  - ...
Hierarchical Model Example:

- **Goal:**
  - Study the effectiveness of cardiac treatments

Survival probability

\[ y_1 \quad y_2 \quad \ldots \quad y_J \]

Responses in hospitals

Independent Data
(Separate analysis using data from each study)

It may be reasonable to expect that estimates of \( q_j \)'s, which represent a sample of hospitals, should be related to each other:

\[ q_j \sim p_f, j=1, \ldots, J. \]

This implies, marginally, correlation between observations!
Hierarchical Model:

**Example:**

- **Goal:**
  - Study the effectiveness of cardiac treatments
  - \( q_j \): survival probability for patients in hospital \( j \)
  - \( \phi \): overall survival probability

- **Inference:**
  - Estimate \( q_j \)'s borrowing strength of information from all other hospitals
  - Estimate \( \phi \) taking into account the variability among hospitals

**Hierarchical Model:**

**Exchangeability**

- **Definition:** \( Y_1, \ldots, Y_n \) are judged **exchangeable** if the probability \( P(Y_1, \ldots, Y_n) \) is unaffected by permutations of the labels attached to the variables.

- **Example:**
  
  \[
  \begin{align*}
  P(Y_1, Y_2, Y_3) &= P(Y_2, Y_1, Y_3) = P(Y_2, Y_3, Y_1) = \\
  &= P(Y_3, Y_1, Y_2) = P(Y_3, Y_2, Y_1) = \ \\
  \end{align*}
  \]

  we would judge \( Y_1, Y_2, Y_3 \) exchangeable!

**Note:**

- An infinite sequence of random variables \( Y_1, Y_2, \ldots \) is exchangeable if any finite subsequence is exchangeable.

- Independence implies exchangeability, but not conversely! That is, independence is a stronger assumption than exchangeability.
Hierarchical Model:

Exchangeability

\[ P(Y_1 = 0; Y_2 = 1) = P(Y_1 = 1; Y_2 = 0) = 0.09 \]

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.01</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>1</td>
<td>0.09</td>
<td>0.81</td>
<td>0.90</td>
</tr>
<tr>
<td>Total</td>
<td>0.10</td>
<td>0.90</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Hierarchical Model:

Exchangeability

\[ P(Y_1 = 0; Y_2 = 1) = P(Y_1 = 1; Y_2 = 0) = 0.05 \]

If two random variables \( Y_1 \) and \( Y_2 \) are independent then they are exchangeable, but exchangeability does not imply independence...

Hierarchical Model:

Exchangeability

Checking exchangeability could be difficult if we had to assess the probabilities of all permutations

We can bypass this with a nice result...

Hierarchical Model:

Exchangeability: De Finetti's theorem

For all infinite sequences of exchangeable random binary variables \( (Y_1, Y_2, \ldots) \), there corresponds a distribution function \( F \) on \((0,1)\) such that for all \( n \) and \( k \leq n \),

\[ F(k, n-k) = \int_0^1 \theta^k (1-\theta)^{n-k} dF(\theta) \]

What is "cool" about this?

- Justifies the Bayesian approach:
  - If one is willing to assume that a collection of 0-1 variables is exchangeable, then one is prepared to re-frame the model into a sampling Bernoulli model with success probability \( \theta \) that is itself random with probability distribution \( F \) (the prior).
  - The theorem does not tell us anything about what the distribution \( F \) should be!
Hierarchical Models

Definition:
- A Bayesian Hierarchical model is a statistical model \( f(x|\theta) \) where the prior distribution \( \pi(\theta) \) is decomposed into conditional distributions
  \[ \pi_1(\theta_1), \pi_2(\theta_2|\theta_1), \ldots, \pi_n(\theta_n|\theta_{n-1}) \]
  and a marginal distribution \( \pi_{n+1}(\theta_n) \) such that
  \[ \pi(\theta) = \pi_1(\theta_1), \pi_2(\theta_2|\theta_1), \ldots, \pi_n(\theta_n|\theta_{n-1}) \pi_{n+1}(\theta_n) \text{ d}\theta_1 \ldots \text{ d}\theta_n \]
- Parameters \( \theta_i \) are called hyperparameters of level i

Higher level of hierarchy assumes known hyperparameters.
- Difficult to check propriety of posteriors with improper priors
  - Proper distributions which are almost vague can also approach impropriety with undesirable modeling results
- Sensitivity analysis is very important in hierarchical modeling
Hierarchical Models

- Approach to building complex models by specifying a series of conditional distributions
- Parameters in the model can be regarded as related or connected in some way by the structure of the problem
- Typically data have multi-level/hierarchical structure (observational units grouped into larger units)
  - Example: students are grouped into classes, which are grouped into schools, which are grouped by districts...
- Levels of inference dependent on scientific questions of interest
  - Example: Multi-center clinical trial
    - Magnitude of an “average” treatment effect?
    - Magnitude of treatment effect in each center?
    - Amount of variation of the effect across centers?
  - ...

Motivating example: HARVEST study

- We will use data from the HARVEST study to explore the use of INLA to estimate parameters in a hierarchical regression model
- Data are from 214 participants in a study of ambulatory blood pressure monitoring. Participants with suspected hypertension wore an ambulatory monitoring device for 24 hours to assess blood pressure on three occasions over the course of 5 months: https://raw.githubusercontent.com/rhubb/SISCR2017/master/data/harvest.csv
- The data set includes the following variables
  - ID: study id
  - Month: 1 = baseline, 3 = 1 month assessment, 5 = 5 month assessment
  - Smoke: number cigarettes smoked per day (0 = non-smoker, 1 = 1-5, 2 = 6-10, 3 = 11-20)
  - Sport: 0 = waterway, 1 = light activity, 2 = non-competitive sports, 3 = competitive sports
  - SBP: systolic blood pressure (continuous)
  - DBP: diastolic blood pressure (continuous)
  - HR: heart rate (continuous)
  - Age: age in years at baseline (continuous)
  - BMI: body mass index at baseline (continuous)
  - Male: male = 1, female = 0

- We would like to study the association between age, BMI, activity level, and diastolic blood pressure
- However, since multiple measurements are available for each subject it is necessary to account for correlation among DBP measures made for the same participant
Hierarchical Model for HAREST study

Let $y_{k,t} = \text{DBP for patient } k \text{ at time } t$

Population mean DBP

Mean DBP for patient $k$

DBP for patient $k$ at time $t$

Motivating example: HARVEST study

```r
# Load data

# Summary statistics
summary(harvest[,c("SBP", "DBP", "Age", "BMI", "Male")])

# Model specification
mod <- inla(DBP ~ factor(Sport) + Age + BMI + f(ID, model = "iid"), family = "gaussian", data = harvest)

# Model summary
summary(mod)
```

Motivating example: HARVEST study

- Fixed effects:
  - mean: 0.025quant 0.5quant 0.975quant mode
  - factor(Sport)1: 1.9735 1.4128 -4.7531 -1.9725 0.7982 -1.9705 0
  - factor(Sport)2: 2.9389 1.5057 -5.8948 -2.9402 0.0208 -2.9426 0
  - factor(Sport)3: 4.5822 1.8989 -8.3137 -4.5825 -0.8530 -4.5829 0
  - Age: 0.1617 0.0607 0.0424 0.1617 0.2809 0.1617 0
  - BMI: -0.0850 0.0700 -0.2226 -0.0850 0.0525 -0.0850 0

- Random effects:
  - Precision for the Gaussian observations: 0.0421 0.0053 0.0322 0.0419 0.0529 0.0418
  - Precision for ID: 0.0255 0.0039 0.0190 0.0251 0.0342 0.0243

Precision of DBP across participants
Markov Chain Monte Carlo (MCMC)
(Implementation via JAGS)

Markov Chains

Definition:
- A **Markov Chain** is a sequence of random variables $X_1, X_2, X_3, \ldots$ with the **Markovian property**, namely that, given the present state, the future and past states are independent. Formally,

\[
P(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n)
\]

Definition:
- A Markov Chain is **homogeneous** if

\[
P(X_{n+1} = y | X_n = x) = P(X_n = y | X_{n-1} = x) = P(x, y)
\]

Example:
- State Space: $S = \{0, 1\}$
- Transition Matrix: (conditional probs. in rows)

\[
P = \begin{bmatrix}
0.7 & 0.3 \\
0.4 & 0.6
\end{bmatrix}
\]

How does it behave?
Markov Chains

Transition matrix in n steps?

\[ P^n = S \Lambda^n S^{-1} \]

- In our example, the eigenvalues of \( P \) are 1 and 0.3 with corresponding eigenvectors \((1,1)\)' and \((0.3,-0.4)\)'.
- Thus:

\[ \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 0.3 \end{bmatrix}, \quad S^{-1} = \begin{bmatrix} 4/7 & 3/7 \\ 10/7 & -10/7 \end{bmatrix} \]

\[ P^n = \begin{bmatrix} 4/7(0.3)^n & 10/7 \end{bmatrix} \begin{bmatrix} 4/7 & 3/7 \\ 10/7 & -10/7 \end{bmatrix} \begin{bmatrix} 4/7 & 3/7 \\ 4/7 & 3/7 \end{bmatrix} \]

Markov Chains

Limiting distribution:

\[ \lim_{n \to \infty} P^n = \begin{bmatrix} 4/7 & 3/7 \\ 4/7 & 3/7 \end{bmatrix} \]

- Note that:
  - Largest eigenvalue is 1 (this gives the stationary distribution)
  - Rate of convergence is given by the second eigenvalue
  - Convergence describe "state" after many iterations
  - Stationary distribution does not depend on initial state

  "Subliminal" message:
  - If we want to generate an observation from \( \pi \), we can start anywhere and generate values from the transition probability matrix. After a length of time (burn-in), we can pick \( X_m \) whose distribution is \( \pi \).

Markov Chains

Conditions for convergence:

- Aperiodic
  - Avoids the chain from oscillating between different sets in a regular movement
- Irreducible
  - Starting from any point, the MC can reach any set with positive probability

Diagram:

- Periodic chain
- Reducible chain

Healthy

Disease

Death

Reducible chain
Markov Chains and MCMC

Q: How do we construct a Markov Chain whose stationary distribution is our target (posterior) distribution?

A: Markov Chain Monte Carlo (MCMC)

Luckily, for most models, you can use existing software. Bugs/Winbugs/Jags are very popular. However, some models are more complex and you would need to implement your own MCMC (beyond the scope of this module).

MCMC algorithms

- There are many MCMC algorithms
  - Gibbs Sampler
  - Metropolis Hastings
  - Hybrid MCMC
  - Adaptive MCMC
  - Slice Sampler
  - Reversible Jump MCMC (RJMCMC)
  - Particle filters
  - ...

- Only discussing the simplest algorithm

Gibbs Sampling

- Derive full conditional distributions (distribution of each parameter conditional on all other parameters and data): \( \{ \theta_j | \theta_{-j} \} \)

- Algorithm:
  1. Set starting values \( \{ \theta^0_1, ..., \theta^0_k \} \)
  2. Iteration \( i \)
     - Draw \( \theta^i_1 \) from \( \{ \theta^i_1 | \theta^i_{-1}, ..., \theta^i_{-k} \} \)
     - Draw \( \theta^i_2 \) from \( \{ \theta^i_2 | \theta^i_1, \theta^i_{-2}, ..., \theta^i_{-k} \} \)
     - Draw \( \theta^i_k \) from \( \{ \theta^i_k | \theta^i_1, ..., \theta^i_{k-1} \} \)
  3. Repeat step 2: After \( t \) iterations obtain

- Theorem:
  \[ \{ \theta^t_1, ..., \theta^t_k \} \rightarrow D \{ \theta_1, ..., \theta_k \} \text{ as } t \rightarrow \infty \]
Gibbs Sampling

Note:
- You should be able to sample from the full conditionals.
- If the full conditional is not in closed form, may need to use other algorithms (beyond the scope of today’s lecture).
- In some problems, full conditionals are also complex...

Example 3: Normal model with noninformative priors

Let $Y_i \sim N(\mu, \sigma^2)$ and $\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$.

We had:
\[
\pi(\mu, \sigma^2 | y) \propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \times \exp\left\{-\sum (y_i - \mu)^2 \right\}
\]

Let $\tau = 1/\sigma^2$. Easy to derive:
\[
\pi(\mu | \sigma^2, y) = N(\bar{y}, \sigma^2/n)
\]
\[
\pi(\sigma^2 | \mu, y) = \Gamma\left(\frac{n}{2}, \frac{1}{2} \sum (y_i - \mu)^2\right)
\]
Example 3:
Normal model with noninformative priors & Gibbs Sampling

First 100 iterations...

Example 3:
Normal model with noninformative priors & Gibbs Sampling

Example 3:
Normal model with noninformative priors & Gibbs Sampling

Gibbs Sampling
(with two different starting points)
Example 3:
Normal model with noninformative priors

Prediction of a new observation
- In this model, the posterior predictive distribution is a Student-t with (n-1) degrees of freedom, location at the sample mean, scale \( \frac{1}{\sqrt{\nu}} \)
- Analytical results not always available. What to do?
- Recall
  \[
  P(Y_{\text{new}} | \text{Data}) = \int P(Y_{\text{new}} | \text{Data}, \theta) P(\theta | \text{Data}) d\theta
  \]
  \[
  = \int P(Y_{\text{new}} | \theta) P(\theta | \text{Data}) d\theta
  \]
- Can get samples from predictive distribution:
  - Draw samples \( Y_{\text{new}} \) given a posterior sample of the parameters

Example 3:
Normal model with noninformative priors

Prediction of a new observation
- Using simulation draws

```
post = post[-c(1:100),]
nsamples = nrow(post)
ynew = rnorm(nsamples, post[,1], sqrt(post[,2]))
hist(ynew)
```

MCMC methods

- Implementing your own MCMC can be challenging!
- A large variety of models can be implemented in Bugs/Winbugs/Jags
  - "Black-Box"
    - You will not need to derive full conditionals
    - You will not need to decide on MCMC samplers
- Input:
  - Likelihood
  - Priors
  - [Define any quantity of interest (e.g. Odds Ratio, etc)]
- Output:
  - Posterior samples
Jags (Just Another Gibbs Sampler)

Website:
http://mcmc-jags.sourceforge.net

- Very similar to WinBUGS (with a few differences)

- Goals/features:
  - Cross-platform engine for the BUGS language
  - Extensible, allowing users to write their own functions, distributions and samplers.
  - Platform for experimentation with ideas in Bayesian modelling

- Packages:
  - rjags: Allows you to run Jags from within R
  - coda: Allows you to perform convergence diagnosis

---

**Using Jags**

<table>
<thead>
<tr>
<th>Name</th>
<th>Range</th>
<th>Density</th>
<th>Lower</th>
<th>Upper</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evian</td>
<td>$a, b, c, x, y$</td>
<td>$a, b, c, x, y$</td>
<td>$a, b, c, x, y$</td>
<td>$a, b, c, x, y$</td>
<td>$a, b, c, x, y$</td>
</tr>
<tr>
<td>Smooth</td>
<td>alphas($x$), betas($x$)</td>
<td>$a, b, c, x, y$</td>
<td>$a, b, c, x, y$</td>
<td>$a, b, c, x, y$</td>
<td>$a, b, c, x, y$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Gamma</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Logistic</td>
<td>$y &gt; 0$</td>
<td>$y &gt; 0$</td>
<td>$y &gt; 0$</td>
<td>$y &gt; 0$</td>
<td>$y &gt; 0$</td>
</tr>
<tr>
<td>Lognormal</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Multinomial</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Normal</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Poisson</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Rank1</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Uniform</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Wilcox</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
</tbody>
</table>

Table 3.1: Common and custom distributions in the Jags module

---

**Using Jags**

<table>
<thead>
<tr>
<th>Name</th>
<th>Range</th>
<th>Density</th>
<th>Lower</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>$x, y, z$</td>
<td>$x, y, z$</td>
<td>$x, y, z$</td>
<td>$x, y, z$</td>
</tr>
<tr>
<td>Uniform</td>
<td>$x, y, z$</td>
<td>$x, y, z$</td>
<td>$x, y, z$</td>
<td>$x, y, z$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Logistic</td>
<td>$y &gt; 0$</td>
<td>$y &gt; 0$</td>
<td>$y &gt; 0$</td>
<td>$y &gt; 0$</td>
</tr>
<tr>
<td>Normal</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Poisson</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Rank1</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
<tr>
<td>Uniform</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
<td>$x &gt; 0$</td>
</tr>
</tbody>
</table>

Table 3.2: Custom and choice distributions in the Jags module
Example 1: using jags

```
model{
  # define likelihood of observations
  for (i in 1:n)
    y[i] ~ dnorm(mu, tausq)
  
  # define priors
  mu ~ dnorm(0.0, 0.0001)
  tausq <- 1/sigmasq
  sigmasq ~ dunif(0,100)
}
```

Code saved in a text file (in this case, example1.jag)
Example 1: using jags

```r
## define jags model within R
jags.m <- jags.model(file="example1.jag", data=data,
inits=inits, n.chains=2, n.adapt=500)
## specify parameters to be monitored
params <- c("mu","sigmasq")
## run jags and save posterior samples
samps <- coda.samples(jags.m, params, n.iter=10000)
## summarize posterior samples
summary(samps)
plot(samps)
```

Example 1: using jags

```r
> jags.m <- jags.model(file="example1.jag", data=data,
inits=inits, n.chains=2, n.adapt=500)
> params <- c("mu","sigmasq")
> samps <- coda.samples(jags.m, params, n.iter=10000)
> summary(samps)
> summary(window(samps, start=1000))

Iterations = 501:10500
Thinning interval = 1
Number of chains = 2
Sample size per chain = 10000
1. Empirical mean and standard deviation for each variable, plus standard error of the mean:
   Mean     SD Naive SE Time
          -
mu     0.2417 0.2057 0.001454       0.001454
sigmasq 4.2044 0.6123 0.004329       0.005847
2. Quantiles for each variable:
   2.5%    25%    50%    75%  97.5%
mu    -0.1595 0.1037 0.2430 0.3812 0.6408
sigmasq 3.1678 3.7724 4.1500 4.5722 5.5553
```

Example 1: using jags

```r
> summary(window(samps, start=1000))
Iterations = 1000:10500
Thinning interval = 1
Number of chains = 2
Sample size per chain = 9501
1. Empirical mean and standard deviation for each variable, plus standard error of the mean:
   Mean     SD Naive SE Time
          -
mu     0.2420 0.2057 0.001492       0.001492
sigmasq 4.2070 0.6138 0.004453       0.006198
2. Quantiles for each variable:
   2.5%    25%    50%    75%  97.5%
mu    -0.1594 0.1034 0.2434 0.3815 0.6401
sigmasq 3.1672 3.7730 4.1526 4.5749 5.5636
```
Convergence Diagnostics Methods

- Brooks, Gelman & Rubin
  - Two or more parallel chains (different starting values)
  - Comparison of within and between chain variance for each variable using the second half of chains
  - "Rule-of-thumb": Samples are considered to arise from the stationary distribution if estimates are approximately equal to 1 (0.975 quantile is less than or equal to 1.2)

- Geweke
  - Individual chain
  - Chain divided in two "windows" – comparison of the mean of sampled values in the first window to the mean in the second window
  - "Rule-of-thumb": Lack of convergence if p-values < 0.05

- Heidelberger and Welch
  - Individual chains
  - Based on Brownian bridge theory and uses Cramer-von-Mises statistic
  - Repeatedly discards 10% of iterations until the chain passes the test, or more than 50% of the iterations have been discarded
  - "Rule-of-Thumb": Failure of the chain to pass the test indicates that a longer run is needed

- Raftery and Lewis
  - Individual chains
  - "Rule-of-Thumb": Dependence factors greater than 5 indicate lack of convergence

Example 1: using jags

- autocor.plot(samps)
- gelman.diag(samps)
  - Potential scale reduction factors:
    - Point est. Upper C.I.
    - mu
    - sigmasq
    - Multivariate psrf
- gelman.plot(samps)
Example 1: using jags

```
geweke.diag(samps)
```

```
[1] Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

mu sigmaq
-0.4963 -0.6335
```

```
raftery.diag(samps)
```

```
[1] Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

mu sigmaq
-0.2554 -0.2781
```

```
heidel.diag(samps)
```

```
[1] Stationarity start passed
   test         iteration
   mu          passed    1         0.025
   sigmaq      passed    1         0.025

Halfwidth Mean
   test
   mu          passed    0.242 0.0040
   sigmaq      passed    4.210 0.0158
```

Example 1: using jags

```
## adding observation at last position for prediction (value is missing with NA)
y <- c(y, NA)
n <- length(y)
data <- list(y=y, n=n)
inits <- list(mu=0, sigmaq=1)
jags.m <- jags.model(file="example1.jag", data=data, inits=inits, n.chains=2, n.adapt=500)
params <- c("mu", "sigmaq", "y")
samps <- coda.samples(jags.m, params, n.iter=2000)
summary(samps)
```

Example 1: using jags

```
Posterior predictive distribution
```
Example 1: using jags

Posterior predictive distribution

> summary(samps)

Iterations = 501:2500
Thinning interval = 1
Number of chains = 2
Sample size per chain = 2000

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>SD</td>
<td>Naive SE</td>
<td>Time</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>mu</td>
<td>0.243185</td>
<td>0.2036</td>
<td>0.003219</td>
</tr>
<tr>
<td>sigmasq</td>
<td>4.199118</td>
<td>0.6070</td>
<td>0.009597</td>
</tr>
<tr>
<td>y[1]</td>
<td>-1.400793</td>
<td>0.0000</td>
<td>0.000000</td>
</tr>
<tr>
<td>y[2]</td>
<td>0.410639</td>
<td>0.0000</td>
<td>0.000000</td>
</tr>
<tr>
<td>y[3]</td>
<td>-1.868522</td>
<td>0.0000</td>
<td>0.000000</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>y[100]</td>
<td>1.058556</td>
<td>0.0000</td>
<td>0.000000</td>
</tr>
<tr>
<td>y[101]</td>
<td>0.297378</td>
<td>2.0684</td>
<td>0.032704</td>
</tr>
</tbody>
</table>

2. Quantiles for each variable:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5%</td>
<td>25%</td>
<td>50%</td>
<td>75%</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>mu</td>
<td>-0.153727</td>
<td>0.111128</td>
<td>0.242856</td>
</tr>
<tr>
<td>sigmasq</td>
<td>3.183236</td>
<td>3.764582</td>
<td>4.134646</td>
</tr>
<tr>
<td>y[1]</td>
<td>-1.400793</td>
<td>-1.400793</td>
<td>-1.400793</td>
</tr>
<tr>
<td>y[2]</td>
<td>-1.058556</td>
<td>-1.058556</td>
<td>-1.058556</td>
</tr>
<tr>
<td>y[3]</td>
<td>-3.825483</td>
<td>-1.086550</td>
<td>0.326903</td>
</tr>
</tbody>
</table>

Analyzing FEV Data Set with Jags

## Analysis of the FEV dataset using Jags

## read FEV data set


## now prepare data for Jags
datajag = list(n=length(data$fev), y=log(data$fev), smoke=1*(data$smoke==2), age=data$age-mean(data$age))
inits = list(beta=rep(0,3), sigmasq=1)

## define jags model within R
model = jags.model(file="fev.jag", data=datajag, inits=inits, n.chains=2, n.adapt=500)
params = c("beta", "sigmasq", "ratiogm")
fev.post = coda.samples(model, params, n.iter=1000)

## summarize posterior samples
summary(fev.post)
plot(fev.post)

## convergence diagnosis
autocorr.plot(fev.post)
gelman.plot(fev.post)
heidel.diag(fev.post)

model{
## define likelihood of observations
for (i in 1:n){
y[i] ~ dnorm(mu[i], tausq)
}

## define priors
for (i in 1:3){
beta[i] ~ dnorm(0, 0.0001)
}
tausq <- 1/sigmasq

## deriving quantities of interest (ratios of geometric means)
for (i in 1:2){
ratiogm[i] <- exp(beta[i+1])
}
}
Meta-analysis

Table 1: Results from Clinical Trials Providing Estimates of Efficacy of BCG Vaccine Against Cases of Tuberculosis (TB) and TB Death That Were Used in the Meta-analysis

<table>
<thead>
<tr>
<th>Study, Year</th>
<th>Population</th>
<th>Cases of TB</th>
<th>Cases of TB Death</th>
<th>OR</th>
<th>95% CI</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conti et al., 1995</td>
<td>1,000</td>
<td>10</td>
<td>2</td>
<td>1.0</td>
<td>(0.4-2.5)</td>
<td>0.99</td>
</tr>
<tr>
<td>Evans &amp; Guo, 1997</td>
<td>1,500</td>
<td>15</td>
<td>3</td>
<td>1.4</td>
<td>(0.7-2.7)</td>
<td>0.33</td>
</tr>
<tr>
<td>Khorram &amp; et al., 1999</td>
<td>2,000</td>
<td>20</td>
<td>4</td>
<td>1.0</td>
<td>(0.5-2.1)</td>
<td>0.99</td>
</tr>
<tr>
<td>Yamey &amp; et al., 2001</td>
<td>3,000</td>
<td>30</td>
<td>6</td>
<td>1.0</td>
<td>(0.5-2.0)</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Note: OR = Odds Ratio; 95% CI = 95% Confidence Interval; p-Value = Probability Value.
Motivation:
- Many individual clinical trials are not large enough to answer the questions of interest reliably

Solutions
- Advocacy for large trials
  - Not always feasible
- Informal evidence synthesis from different studies
  - Possibility of biased selection of evidence
- Formal systematic review

Goals of Systematic Reviews:
- To review systematically the available evidence from a particular research area
- To provide quantitative summaries of the results from each study
- To combine the results across studies if appropriate; such combination of results leads to greater statistical power in estimating treatment effects
- To assess the amount of variability between studies
- To estimate the degree of benefit associated with a particular study treatment
- To identify study characteristics associated with particularly effective treatments.

Components of Systematic Reviews:
- Qualitative:
  - Description of available trials in terms of relevance and methodological strengths and weaknesses
- Quantitative
  - Means of combining results from different studies
  - This is known as Meta-Analysis

Critical Step:
- Study selection
Statistical Methodology

- Fixed effects models
  - Each individual study used to estimate a common, unknown, overall pooled effect

- Random effects models
  - Each individual study has its own underlying effect, which in turn are used to estimated a common population effect.
  - Accounts for two sources of heterogeneity:
    - Within-study heterogeneity
    - Between-study heterogeneity

---

**Fixed-Effects (Mantel-Haenszel):**

- Pooled Effect: \( \hat{Y} = \frac{\sum W_i Y_i}{\sum W_i} \) with \( \text{Var}(\hat{Y}) = \frac{1}{\sum W_i} \)
- \( k \): number of studies
- \( Y_i \): effect size in the i-th study
- \( W_i \): weight (inverse of within-study variance for i-th study)

---

**Random-Effects (DerSimonian-Laird):**

- \( Y_i \sim \mu_i + \sigma_i \varepsilon_i \) for \( i = 1, \ldots, k \)
  - \( \mu_i \sim N(\mu, \tau^2) \); \( \varepsilon_i \sim N(0,1) \)
- Pooled Effect: \( \hat{Y} = \frac{\sum W_i Y_i}{\sum W_i} \) with \( \text{Weights}: W_i = \frac{1}{W_i^2 + \tau^2} \)
- \( \hat{Q} = \begin{cases} 
0, & \text{if } Q < k-1 \\
\frac{(Q-k+1)}{U}, & \text{if } Q > k-1 
\end{cases} \)
- \( Q = \sum W_i (Y_i - \hat{Y})^2 \); \( U = (k-1)\tau^2 - \sum X_i^2 / kW \)
Systematic Reviews and Meta-Analysis

- Heterogeneity is very likely in meta-analysis
  - Many possible sources of heterogeneity
  - Estimating how these various factors affect the effect size is often of considerable interest and importance
    - Meta-regression!

Efficacy of BCG Vaccine in the Prevention of Tuberculosis

- Bacille Calmette Guerin (BCG)
  - Most widely used vaccine against tuberculosis (TBC)

- Expanded Data: publicly available in R
  - 13 clinical trials of BCG investigating efficacy in the treatment of tuberculosis
    - Number of subjects with TB with or without BCG vaccination
  - Heterogeneity among trials may be explained by geographic location and year

- Efficacy measure: Odds Ratio (OR)

BCG Example

- Data:
  
<table>
<thead>
<tr>
<th>Trial</th>
<th>author</th>
<th>year</th>
<th>tpos</th>
<th>tneg</th>
<th>cpos</th>
<th>cneg</th>
<th>alloc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aronson 1948</td>
<td>4</td>
<td>119</td>
<td>11</td>
<td>128</td>
<td>44</td>
<td>random</td>
</tr>
<tr>
<td>2</td>
<td>Ferguson &amp; Jones 1949</td>
<td>4</td>
<td>300</td>
<td>274</td>
<td>55</td>
<td>random</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Ronesch et al 1960</td>
<td>3</td>
<td>228</td>
<td>220</td>
<td>42</td>
<td>random</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Hart &amp; Abbot 1977</td>
<td>62</td>
<td>13536</td>
<td>248</td>
<td>12619</td>
<td>52</td>
<td>random</td>
</tr>
<tr>
<td>5</td>
<td>Frimodt-Moller et al 1973</td>
<td>33</td>
<td>5036</td>
<td>47</td>
<td>5761</td>
<td>13</td>
<td>alternate</td>
</tr>
<tr>
<td>6</td>
<td>Stein &amp; Aronson 1953</td>
<td>62</td>
<td>13536</td>
<td>248</td>
<td>12619</td>
<td>13</td>
<td>random</td>
</tr>
<tr>
<td>7</td>
<td>Vandecruy et al 1960</td>
<td>4</td>
<td>300</td>
<td>274</td>
<td>55</td>
<td>random</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>TPT Madras 1980</td>
<td>505</td>
<td>87846</td>
<td>493</td>
<td>87822</td>
<td>13</td>
<td>random</td>
</tr>
<tr>
<td>9</td>
<td>Coetzee &amp; Berjak 1968</td>
<td>29</td>
<td>1699</td>
<td>65</td>
<td>1600</td>
<td>42</td>
<td>systematic</td>
</tr>
<tr>
<td>10</td>
<td>Rosenthal et al 1961</td>
<td>17</td>
<td>1699</td>
<td>65</td>
<td>1600</td>
<td>42</td>
<td>systematic</td>
</tr>
<tr>
<td>11</td>
<td>Comstock &amp; Webster 1963</td>
<td>5</td>
<td>2493</td>
<td>3</td>
<td>2338</td>
<td>33</td>
<td>systematic</td>
</tr>
<tr>
<td>12</td>
<td>Comstock et al 1974</td>
<td>146</td>
<td>52448</td>
<td>141</td>
<td>27197</td>
<td>18</td>
<td>systematic</td>
</tr>
<tr>
<td>13</td>
<td>Comstock et al 1974</td>
<td>27</td>
<td>14886</td>
<td>29</td>
<td>17825</td>
<td>33</td>
<td>systematic</td>
</tr>
</tbody>
</table>

- The 13 studies provide data in terms of 2x2 tables in the form:

<table>
<thead>
<tr>
<th>vaccinated group</th>
<th>TB positive TB negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>tpos</td>
<td>tneg</td>
</tr>
<tr>
<td>cpos</td>
<td>cneg</td>
</tr>
</tbody>
</table>
## Meta-Analysis

```r
library(metafor)

## load data

data(dat.bcg)

## Part A: Frequentist analysis

### meta-analysis of the log odds ratio using the Mantel-Haenszel method

res.fe <- rma.mh(measure="OR", ai=tpos, bi=tneg, ci=cpos, di=cneg, data=dat.bcg, slab=paste(author, year, sep=","))

### forest plot of the observed odds ratio with summary estimate

forest(res.fe, atransf=exp, xlim=c(-7,5), ylim=c(-2.5,16))

### meta-analysis of the log odds ratio using a random-effects model

res.re <- rma(measure="OR", ai=tpos, bi=tneg, ci=cpos, di=cneg, data=dat.bcg, slab=paste(author, year, sep=","))

### add summary estimate from the random-effects model to forest plot

cat("Forest plot of the observed odds ratio with summary estimate for random-effects model: OR = ", round(coef(res.re), 3), " (95\% CI: ", round(confint(res.re)[1], 3), ", ", round(confint(res.re)[2], 3), ")\n")

### forest plot of the observed odds ratio with summary estimate

forest(res.re, atransf=exp, xlim=c(-7,5), ylim=c(-2.5,16))
```

### BCG Example (A): Standard Meta-Analysis Mantel-Haenszel

#### Fixed-Effects Model (k = 13)

Test for Heterogeneity: 
\[ Q(df = 12) = 163.9426, \text{ p-val} < .0001 \]

Model Results (log scale):

<table>
<thead>
<tr>
<th>estimate</th>
<th>se</th>
<th>zval</th>
<th>pval</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4734</td>
<td>0.0410</td>
<td>-11.5444</td>
<td>&lt;.0001</td>
<td>0.5538</td>
<td>0.3930</td>
</tr>
</tbody>
</table>

Model Results (OR scale):

<table>
<thead>
<tr>
<th>estimate</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6229</td>
<td>0.5748</td>
<td>0.6750</td>
</tr>
</tbody>
</table>

Cochran Mantel-Haenszel Test: \[ CMH = 135.6889, \text{ df} = 1, \text{ p-val} < .0001 \]

Tarone’s Test for Heterogeneity: \[ X^2 = 171.7567, \text{ df} = 12, \text{ p-val} < .0001 \]
BCG Example (A): Standard Meta-Analysis
DerSimonian-Laird

> res.re
Random-effects Model (k = 13; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.3378 (SE = 0.1784)
tau (square root of estimated tau^2 value): 0.5812
I^2 (total heterogeneity / total variability): 92.07%
H^2 (total variability / sampling variability): 12.61
Test for Heterogeneity:
Q(df = 12) = 163.1649, p-val < .0001

Model Results:
<table>
<thead>
<tr>
<th>estimate</th>
<th>se</th>
<th>zval</th>
<th>pval</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7452</td>
<td>0.1860</td>
<td>-4.0057</td>
<td>&lt;.0001</td>
<td>-1.1098</td>
<td>-0.3806</td>
</tr>
</tbody>
</table>

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The heterogeneity test shows strong evidence of heterogeneity in the 13 trials.

BCG Example (A): Standard Meta-Analysis
DerSimonian-Laird

### meta-regression
```r
### calculate log odds ratios and corresponding sampling variances
dat <- escalc(measure="OR", ai = tpos, bi = tneg, ci = cpos, di = cneg, data = dat.bcg)
head(dat)

### random-effects model (output is the same as seen for res.re)
res <- rma(yi, vi, data = dat)
res

### average relative risk with 95% CI (this will give you the OR from combined studies)
predict(res, transf = exp)
```

### meta-regression model with absolute latitude as moderator
```r
res.mr1 <- rma(yi, vi, mods = ~ ablat, data = dat)
res.mr1
```

### meta-regression model with absolute latitude as moderator and 95% CI
```r
res.mr2 <- rma(yi, vi, mods = ~ ablat, data = dat, transf = exp)
res.mr2
```
Some evidence that latitude is associated with observed effect size.

Model Results:

<table>
<thead>
<tr>
<th>estimate</th>
<th>se</th>
<th>zval</th>
<th>pval</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>intrcpt</td>
<td>0.3010</td>
<td>0.2146</td>
<td>1.4025</td>
<td>0.1608</td>
<td>-0.1197   0.7217</td>
</tr>
<tr>
<td>ablat</td>
<td>-0.0315</td>
<td>0.0063</td>
<td>-5.0242</td>
<td>&lt;.0001</td>
<td>-0.0438   -0.0192 ***</td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

We will consider several models and compare the results

First, we need to re-organize the data...

```r
# Some code to re-organize the data...
```

```r
dat <- NULL
dat$trial <- rep(seq(1,13),2)
dat$group <- c(rep(1, 13), rep(0,13))
dat$Y <- c(dat.bcg$tpos, dat.bcg$cpos)
dat$N <- rep(NA, 26)
dat$N[1:13] <- dat.bcg$tpos + dat.bcg$tneg
dat$N[14:26] <- dat.bcg$cpos + dat.bcg$cneg
dat$Latitude <- rep(dat.bcg$ablat,2)
dat$centeredLatitude <- dat.bcg$ablat - mean(dat.bcg$ablat)
dat$Year <- rep(dat.bcg$year, 2)
dat$centeredYear <- dat.bcg$year - mean(dat.bcg$year)
dat1 <- as.data.frame(dat)
```
**Model 1 (B):**

\[ \logit(p_i) = \delta_i + \beta_i \text{Group}, \]

\[ \delta_i = N(0, \sigma^2), \quad j = 1, \ldots, 13; \quad \beta_i = N(0, \sigma^2_i) \]

The overall posterior median OR=exp(-0.48)=0.62 (95% PCI= 0.57,0.67).

- Very similar results to those obtained using Mantel-Haenszel (fixed-effects).

**Model 2 (B):**

\[ \logit(p_i) = \delta_i + \beta_i \text{Group}, \]

\[ \delta_j = N(0, \sigma^2), \quad j = 1, \ldots, 13; \quad \nu^2 = \text{IG}(a, b); \quad \beta_i = N(0, \sigma^2_i) \]

The overall posterior median OR=exp(-0.46)=0.63 (95% PCI= 0.57,0.67).

Posterior median precision = 4.43 (posterior median variance = 1/4.43=0.23).

Estimated variance under frequentist is much smaller (since it doesn't account for uncertainty in random effects).
Model 3 (B): \[ \log(p_i) = \delta_i + \beta_{\text{Group}} + \beta_{\text{Latitude}} + \beta_{\text{Group} \times \text{Latitude}}, \]

\[ \delta_i \sim N(0, \sigma^2), \quad \beta_{\text{Group}} \sim N(0, \sigma^2_{\beta}) \]

- **Time used:**
  - Pre-processing: 0.0933
  - Running: 0.0473
  - Post-processing: 0.0649
  - Total: 0.2055

- **Fixed effects:**
  - \( k_{\text{ld}} \) (Intercept): -4.1401, sd: 0.3571
  - Factor (group): 1, mean: -0.7166, sd: 0.0480
  - Centered Latitude, mean: 0.0736, sd: 0.0256
  - Factor (group): 1: Centered Latitude, mean: -0.0334, sd: 0.0028

- **Random effects:**
  - Mean: 1.25 model
  - Hyperparameters: mean: 0.6693, sd: 0.2933

- **Marginal Likelihood:** -147.66

The posterior mean log-odds ratio (comparing the odds of TB among vaccinated versus not) decreases by approximately 0.03 for each unit difference from the average latitude.

**BCG Example: recap**

- With this example we illustrated a few ways in which we could combine the data from the different studies.
  - Random effects: model heterogeneity
    - (example: no trivial variation in the response rates across studies!)
  - Which model?
    - Model choice guided by scientific questions
    - Model choice guided by statistical criteria
Meta-Analysis:
Mixed and Indirect Treatment Comparisons

Suppose there are several trials

- Comparing treatment A to B (AB trials)
  - Trials AB provide “direct evidence” of the effect of treatment B relative to A.

- Comparing treatment A to C (AC trials)
  - Trials AC provide “direct evidence” of the effect of treatment C relative to A.

- Comparing treatment B to C (BC trials)
  - Trials BC provide “direct evidence” of the effect of treatment C relative to B.
Meta-Analysis: Mixed and Indirect Treatment Comparisons

- Suppose there are several trials
  - What if: NO LONGER TRIALS AB!!

- Comparing treatment A to C (AC trials)
  - Trials AC provide “direct evidence” of the effect of treatment C relative to A.

- Comparing treatment B to C (BC trials)
  - Trials BC provide “direct evidence” of the effect of treatment C relative to B.

Best evidence on the effect of treatment B relative to A is provided by head-to-head trials.

In the absence (or even sparsity) of such trials, there can be “indirect” evidence of the effect of B relative to A:

- \[ d_{AB}^{indirect} = d_{AC}^{direct} - d_{BC}^{direct} \]

The mixing of direct and indirect evidence is called “mixed treatment comparison” (MTC)

More generically:
- With K treatments, there are a total of \( K(K-1)/2 \) possible pairwise comparisons
  - E.g. \( K=6 \) means 15 potential comparisons of interest

- Direct evidence for a subset of pairwise comparisons

- Extending (pairwise) meta-analysis for MTD
  - Fixed effects model
  - Random effects model
Meta-Analysis: Mixed and Indirect Treatment Comparisons

Data from a Smoking Cessation Study

Randomized trials: 24 RCTs

Interventions:
- A: No Contact
- B: Self-Help
- C: Individual Counseling
- D: Group Counseling

Response:
- Number of patients ceasing smoking

Direct evidence for:
- Total number of contrasts: 6

Indirect evidence for:
- Consistency:
  - "Rationale": If \((b-a)=2, \ (c-a)=3\), then \((c-b)\) must be 1

Meta-Analysis: Mixed and Indirect Treatment Comparisons

Four Treatments:
- A (reference)
- B
- C
- D

Direct evidence for:
- \(d_{AB}, d_{AC}, d_{AD}\)
  - (basic parameters)

Total number of contrasts: 6

Meta-Analysis: Mixed and Indirect Treatment Comparisons

Fixed Effects
- \(r_j \sim \text{Binomial}(p_j, N_j)\)
- \(\logit(p_j) \sim N(\mu_j, \sigma^2)\)
- \(d_{AB} = d_{AC} - d_{AB}\)
- \(d_{AD} = d_{AC} - d_{AD}\)
- \(d_{BC} = d_{AC} - d_{ABC}\)
- \(d_{BD} = d_{AD} - d_{AB}\)
- \(d_{CD} = d_{AD} - d_{AC}\)
- \(\mu_A, d_{BC}, d_{BD}, d_{CD} \sim N(0, 100^2)\)
- \(\sigma \sim U(0, 2)\)

Random Effects
- \(r_j \sim \text{Binomial}(p_j, N_j)\)
- \(\logit(p_j) \sim N(\mu_j, \sigma^2)\)
- \(d_{AB} = d_{AC} - d_{AB}\)
- \(d_{AD} = d_{AC} - d_{AD}\)
- \(d_{BC} = d_{AC} - d_{ABC}\)
- \(d_{BD} = d_{AD} - d_{AB}\)
- \(d_{CD} = d_{AD} - d_{AC}\)
- \(\mu_A, d_{BC}, d_{BD}, d_{CD} \sim N(0, 100^2)\)
- \(\sigma \sim U(0, 2)\)
### Meta-Analysis: Mixed and Indirect Treatment Comparisons

**Preparing for Coding in Jags:**

<table>
<thead>
<tr>
<th>Treatment Contrast</th>
<th>Treatment[i]</th>
<th>Baseline[i]</th>
<th>$d_{\text{Treatment}[i]} - d_{\text{Baseline}[i]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>2</td>
<td>1</td>
<td>$d_{2} - d_{1}$</td>
</tr>
<tr>
<td>1,3</td>
<td>3</td>
<td>1</td>
<td>$d_{3} - d_{1}$</td>
</tr>
<tr>
<td>1,4</td>
<td>4</td>
<td>1</td>
<td>$d_{4} - d_{1}$</td>
</tr>
<tr>
<td>2,3</td>
<td>3</td>
<td>2</td>
<td>$d_{3} - d_{2}$</td>
</tr>
<tr>
<td>2,4</td>
<td>4</td>
<td>2</td>
<td>$d_{4} - d_{2}$</td>
</tr>
<tr>
<td>3,4</td>
<td>4</td>
<td>3</td>
<td>$d_{4} - d_{3}$</td>
</tr>
</tbody>
</table>

### Meta-Analysis: Mixed and Indirect Treatment Comparisons

**Sometimes it is useful to have the absolute risk difference instead of odds ratios...**

- Can get this from (log-) odds ratios but need information about the "baseline" probability of the outcome:
  - What is the probability of smoking cessation in the "no treatment" group?
  - Can get this information from cohort studies, trials, etc.
  - Assume, for example, that for "no treatment", the log-odds of smoking cessation has $N(-2.6, 0.38^2)$ distribution.

- Absolute effects for other treatments are:
  - $\text{Logit}(T_{k}) = A + d_{k}$

### Fixed Effects model

```jags
model {
  # loop over 50 observations
  for (i in 1:50) {
    # likelihood
    Response[i] ~ dbin(p[i], N[i])
    logit(p[i]) <- mu[Study[i]] + delta[i]*(1 - equals(Treatment[i], Baseline[i]))
    delta[i] <- d[Treatment[i]] - d[Baseline[i]]
  }
  # vague priors for intercepts (effect for baseline comparison group)
  for (j in 1:24) { mu[j] ~ dnorm(0,.0001) }
  # set effect of Treatment 1 as 0 (effects of other Treatments is relative to this Treatment 1)
  d[1] <- 0
  # flat priors for 3 basic treatment effect parameters
  for (k in 2:4) { d[k] ~ dnorm(0,.001) }
  # Absolute treatment effects
  # prior precision for Treatment 1, $sd = 0.38$
  precA <- pow(0.38, -2)
  # external info on A.
  A ~ dnorm(-2.6, precA)
  for (i in 1:4) { logit(T[i]) <- A + d[i] }
  # Rank the treatment effects (with 1=best) & record the best treatment
  rk <- 5 - rank(d)
  best <- equals(rk, 1)
  # All pairwise log odds ratios and odds ratios (some of these calculations are redundant, but needed to run)
  for (c in 1:4) {
    for (k in 1:4) {
      lor[c,k] <- d[k] - d[c]
    }
  }
}
```
Meta-Analysis: Mixed and Indirect Treatment Comparisons

- Prior to analyzing the data, it is important to define the JAGS model within R. This involves specifying hierarchical priors for the parameters of interest, which are used to update the posterior distributions.

For (j in 1:24) {
# vague priors for intercepts (baseline group)
lor <- datajag$lor

# external info on A.
precA <- datajag$precA

# flat priors for 3 basic treatment parameters
# set effect of treatment 1 as 0 (all other treatment effects are relative to this one)
lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lor

lore
Meta-Analysis: Mixed and Indirect Treatment Comparisons

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF WASHINGTON
School of Public Health

Diagnostic testing with missing gold standard

Meta-Analysis: Mixed and Indirect Treatment Comparisons

TABLE 1. Results of serological and stool testing for epimastigote infection on 182 Cambodian refugees arriving in Montreal, Canada, between July 1982 and February 1983

<table>
<thead>
<tr>
<th>Serumology</th>
<th>+</th>
<th>−</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stool examination</td>
<td>35</td>
<td>67</td>
<td>125</td>
</tr>
<tr>
<td>+</td>
<td>2</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>−</td>
<td>40</td>
<td>122</td>
<td>162</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>90% CI</th>
<th>95% CI</th>
<th>99% CI</th>
<th>Median</th>
<th>90% CI</th>
<th>95% CI</th>
<th>99% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.84</td>
<td>0.80–0.89</td>
<td>0.83</td>
<td>0.80</td>
<td>0.84</td>
<td>0.80</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>0.87</td>
<td>0.80–0.91</td>
<td>0.86</td>
<td>0.80</td>
<td>0.87</td>
<td>0.80</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>0.61</td>
<td>0.49–0.74</td>
<td>0.57</td>
<td>0.48</td>
<td>0.61</td>
<td>0.49</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>0.88</td>
<td>0.84–0.92</td>
<td>0.85</td>
<td>0.83</td>
<td>0.88</td>
<td>0.84</td>
<td>0.85</td>
<td></td>
</tr>
</tbody>
</table>

+ CI credible interval.
Reproducing analyses:
Using only one diagnostic test

Recall: In the absence of 'gold standard' we only observe totals

<table>
<thead>
<tr>
<th>Truth</th>
<th>+</th>
<th>-</th>
<th>N-(Y1+Y2)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>Y1</td>
<td>a-Y1</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y2</td>
<td>b-Y2</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>Y1+Y2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Y1 and Y2 are latent/unobserved data

Probability model for positive test result?
- a ~ Binomial(N, PPT)
  - Where N is the total sample size (i.e. a+b)
  - PPT is the probability of a positive test

\[ PPT = \frac{P(T+)}{N} \]
\[ = \frac{P(T+|D)P(D) + P(T+|D')P(D')}{1-P(T-)} \]
\[ = S\pi + (1-C)(1-\pi) \]

(recall: S is sensitivity and C is the specificity)

Jags Code

```r
model{
  ## model
  a ~ dbin(PPT, N)  # a ~ Binomial(N, PPT)
  ## definition of probability of positive test
  PPT <- a/PPT + (1-PPT)
  ## priors
  S ~ dbeta(aS, bS) # prior for sensitivity
  C ~ dbeta(aC, bC) # prior for specificity
  pi ~ dbeta(api, bpi) # prior for prevalence
  ## computing probability of disease given test results
  pY1 <- pi*S/PPT
  pY2 <- pi*(1-S)/(1-PPT)
  ## simulating Y1, Y2
  Y1 ~ dbin(pY1, a)
  Y2 ~ dbin(pY2, N-a)
}
```

Note: original paper derived full conditionals that allows one to implement full MCMC (Gibbs Sampler) – but that is out of the scope of this introductory course.
Reproducing analyses: Using only one diagnostic test

```r
# Initial values
data <- list(N = 162, a = 40, Y1 = NA, Y2 = NA,
api = 1, bpi = 1,
aS = 4.4, bS = 13.31,
aC = 71.25, bC = 3.75)

## Initial values
inits = function(){list(pi = 0.5, S = 0.9, C = 0.8, Y1 = 10, Y2 = 10)}

## Model specification
jags.m = jags.model(file = diagnostic.jag, data = data,
n.chains = 2,
inits = inits())

## Parameters to be monitored
params = c("pi", "S", "C", "Y1", "Y2")

## Sampling
samps = coda.samples(jags.m, params, n.iter = 5000, thin = 5)

## Summarize posterior samples and save output results
aux = summary(samps)

par(mfrow = c(3, 2))
plot(samps)
output = cbind(aux[[1]][, c(1, 2)], aux[[2]][, c(1, 3, 5)])

Mean          SD       2.5%        50%       97.5%  
C   0.9469572  0.02796301  0.8771444  0.9516872  0.9862617
S   0.3128120  0.06713805  0.2093188  0.3023151  0.4756297
Y1 37.4385000  3.10580759 29.0000000 38.0000000 40.0000000
Y2 82.8770000 24.44454211 35.9750000 85.0000000 120.0000000
pi 0.7401364  0.16335664  0.4075330  0.7581660  0.9855292
```

Final Comments

- There is ‘art’ in Bayesian Analysis

- Achieving ‘mastery’ requires practice!