Introduction to Bayesian Statistics

JAGS course, Part I

Inga Schwabe

September 6, 2014

Outline

- Why use Bayesian statistics?
- Bayes theorem: A simple example
- Bayes theorem $\rightarrow$ Bayesian statistics
- Calculating the posterior distribution: Simple example
- JAGS syntax example: Twin studies

Why Bayesian statistics?

- Incorporating prior knowledge (idealistic use)
- Bayesian methods can easily be extended to more complex problems (pragmatic use)
- Bayesian methods allow testing competing hypotheses
- Bayesian methods are very intuitive

Bayes' Theorem

- Bayesian statistics are based on Bayes' theorem:

Bayes theorem

If A and B are events in the same sample space, then Bayes' theorem states:

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$$

Bayes' theorem: A simple example

Cookie bowl 1

10 chocolate cookies,
30 vanilla cookies

Cookie bowl 2

20 chocolate cookies,
20 vanilla cookies

Fred picks one bowl at random and then picks one cookie at random.
Given that it is a vanilla cookie, what is P(bowl 1)?
$\rightarrow P(\text{bowl } 1|\text{Vanilla cookie})$

What is $P(\text{bowl 1}|\text{vanilla cookie})$?

Cookie bowl 1:
10 chocolate cookies,
30 vanilla cookies

Cookie bowl 2:
20 chocolate cookies,
20 vanilla cookies

Bayes' theorem: $P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$

In this case:
$P(\text{bowl 1}|\text{vanilla cookie}) = \ldots$

$$P(A) \quad = \quad ?$$

What is $P(\text{bowl } 1|\text{vanilla cookie})$?

Cookie bowl 1:
10 chocolate cookies,
30 vanilla cookies

Cookie bowl 2:
20 chocolate cookies,
20 vanilla cookies

Bayes' theorem: $P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$

In this case:
$P(\text{bowl } 1|\text{vanilla cookie}) = \ldots$

$$\begin{aligned} P(A) &= 0.5 \\ P(B|A) &= ? \end{aligned}$$

What is $P(\text{bowl 1}|\text{vanilla cookie})$?

Cookie bowl 1:
10 chocolate cookies,
30 vanilla cookies

Cookie bowl 2:
20 chocolate cookies,
20 vanilla cookies

Bayes' theorem: $P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$

In this case:
$P(\text{bowl 1}|\text{vanilla cookie}) = \ldots$

$$\begin{aligned} P(A) &= 0.5 \\ P(B|A) &= 0.75 \\ P(B) &= ? \end{aligned}$$

What is $P(\text{bowl 1}|\text{vanilla cookie})$?

Cookie bowl 1:
10 chocolate cookies,
30 vanilla cookies

Cookie bowl 2:
20 chocolate cookies,
20 vanilla cookies

Bayes' theorem: $P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$

In this case:
$P(\text{bowl 1}|\text{vanilla cookie}) = \ldots$

$$\begin{aligned} P(A) &= 0.5 \\ P(B|A) &= 0.75 \\ P(B) &= \frac{5}{8} \end{aligned}$$

What is $P(\text{bowl } 1|\text{vanilla cookie})$?

Cookie bowl 1:
10 chocolate cookies,
30 vanilla cookies

Cookie bowl 2:
20 chocolate cookies,
20 vanilla cookies

Bayes' theorem: $P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$

In this case:

$$
\begin{aligned}
P(\text{Bowl } 1|\text{vanilla cookie}) &= \\
\frac{P(\text{bowl } 1) \times P(\text{vanilla cookie}|\text{bowl } 1)}{P(\text{vanilla cookie})} &= \\
\frac{0.5 \times 0.75}{5/8} &= \\
0.6 &
\end{aligned}
$$

From Bayes' theorem to Bayesian Statistics

Bayesian statistics is a branch of statistics that applies Bayes' theorem to solve inferential questions of interest where A represents unknown parameters and B data:

$$P(\theta|y) = \frac{P(y|\theta) \times P(\theta)}{P(y)}$$

$$\begin{aligned} \theta &= \text{Unknown parameters} \\ y &= \text{Data} \\ P(\theta|y) &= \text{Posterior distribution} \\ P(y|\theta) &= \text{Likelihood of the data} \\ P(\theta) &= \text{Prior} \\ P(y) &= \text{Marginal likelihood} \end{aligned}$$

From Bayes' theorem to Bayesian Statistics

The marginal likelihood does not involve any unknown parameter θ and is just a normalizing constant.

Therefore, we can also write:

$$P(\theta|y) = \frac{P(y|\theta) \times P(\theta)}{P(y)} \propto P(y|\theta) \times P(\theta)$$

Often simply put as:

$$P(\theta|y) \propto \text{Likelihood} \times \text{Prior}$$

Calculating the posterior distribution: One short example

Normal distribution: Posterior distribution for μ
Prior for μ:

$$p(\mu) = (2\pi\sigma_0^2)^{-\frac{1}{2}}\exp(-\frac{1}{2}(\mu - \mu_0)^2/\sigma_0^2)$$

Likelihood for one data point:

$$p(y_i|\mu) = (2\pi\sigma^2)^{-\frac{1}{2}}\exp(-\frac{1}{2}(y_i - \mu)^2/\sigma^2)$$

Hence:

$$p(\mu|y) = (2\pi\sigma_0^2)^{-\frac{1}{2}}\exp(-\frac{1}{2}(\mu - \mu_0)^2/\sigma_0^2) \times \prod_{i=1}^{N}(2\pi\sigma^2)^{-\frac{1}{2}}\exp(-\frac{1}{2}(y_i - \mu)^2/\sigma^2)$$

Can be shown that this has the form of a normal distribution (e.g. Box & Tiao)

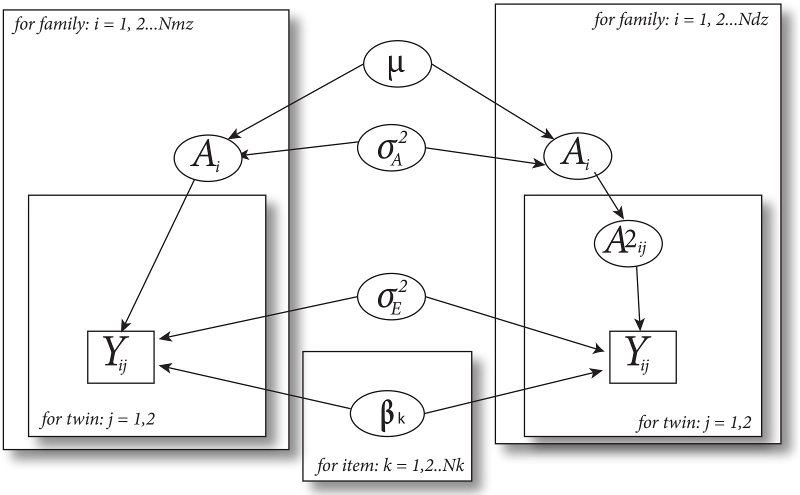
My own research: Twin studies

- Twin studies: Monozygotic (MZ) and dizygotic (DZ) twins
- Decomposition of variance into
  - Genetic influences (A)
  - Shared-environmental influences (C)
  - Unique environmental influences (E)

Most simple case: Data of only MZ twins

- Decomposition into genetic & environmental variances (A+E)
- Aim: Integrate biological model with IRT measurement model

Integrate AE decomposition and IRT model



How can we estimate this in a Bayesian way?

Do it yourself:

- Joint posterior distribution for all parameters
- Gibbs sampling

Use JAGS:

- Build hierarchical model
- Choose priors

The model

For MZ twins, we have:

$$\begin{aligned}
A_i &\sim N(\mu, \sigma_A^2) \\
\theta_{ij} &\sim N(A_i, \sigma_E^2) \\
\ln(P_{ijk}/(1 - P_{ijk})) &= \theta_{ij} - \beta_k \\
Y_{ijk} &\sim \text{Bernoulli}(P_{ijk})
\end{aligned}$$

For DZ twins, we have:

$$\begin{aligned}
A_i &\sim N(\mu, \sigma_A^2/2) \\
A2_{ij} &\sim N(A_i, \sigma 2_A/2) \\
\theta_{ij} &\sim N(A2_{ij}, \sigma_E^2) \\
\ln(P_{ijk}/(1 - P_{ijk})) &= \theta_{ij} - \beta_k \\
Y_{ijk} &\sim \text{Bernoulli}(P_{ijk})
\end{aligned}$$

Model in JAGS: MZ twins

```
for (fam in 1:NMZ){
  aMZ[fam] ~ dnorm(mu, tauA)

  for (twin in 1:2){
    phenoMZ[fam,twin] ~ dnorm(aMZ[fam], tauE)
  }

  #1pl model for twin 1
    for (k in 1:n.items){
         logit(p2[fam,k]) <- phenoMZ[fam,1] - beta[k]
         Ymz[fam,k] ~ dbern(p2[fam,k])
       }

     #1pl model for twin 2
       for (k in (n.items +1) : (2*n.items)){
         logit(p[fam,k]) <- phenoMZ[fam, 2] - beta[k-n.items]
         Ymz[fam,k] ~ dbern(p[fam,k])
       }
  }
```

Model in JAGS: 2: DZ twins

```
for (fam in 1:NDZ){
  aDZ[fam] ~ dnorm(mu[fam], doubletauA)

for (twin in 1:2){
  a2DZ[fam,twin] ~ dnorm(aDZ[fam], doubletauA)
  phenoDZ[fam,twin] ~ dnorm(a2DZ[fam,twin], tauE)
  }

#1pl model twin1 (DZ twins)
  for (k in 1:n.items){
       logit(p2[fam,k]) <- phenoDZ[fam,1] - beta[k]
       Ydz[fam,k] ~ dbern(p2[fam,k])
     }

  #1pl model for twin 2 (DZ twins)
    for (k in (n.items +1) : (2*n.items)){
      logit(p2[fam,k]) <- phenoDZ[fam, 2] - beta[k-n.items]
          Ydz[fam,k] ~ dbern(p2[fam,k])
     }
   }
```

Model in JAGS: 3 Priors

Item parameters assumed known:

```
mu ~ dnorm(0, .1)
tauA ~dgamma(1,1)
tauE ~ dgamma(1,1)
```

Estimate item parameters as well:

```
for (i in 1: n.items){
 beta[i] ~ dnorm(0, .1)
}
tauA ~dgamma(1,1)
tauE ~ dgamma(1,1)
```

But . . .

- What is JAGS doing? (Gibbs sampling)
- How do we know we are sampling from the joint posterior distribution? (Convergence issues)
- How can we call JAGS from R?