Multiparameter models

Dr. Jarad Niemi

STAT 544 - Iowa State University

February 12, 2019
Outline

- Independent beta-binomial
  - Independent posteriors
  - Comparison of parameters
  - JAGS
- Probability theory results
  - Scaled Inv-$\chi^2$ distribution
  - $t$-distribution
  - Normal-Inv-$\chi^2$ distribution
- Normal model with unknown mean and variance
  - Jeffreys prior
  - Natural conjugate prior
Motivating example

Is Andre Dawkins 3-point percentage higher in 2013-2014 than each of the past years?

<table>
<thead>
<tr>
<th>Season</th>
<th>Year</th>
<th>Made</th>
<th>Attempts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2009-2010</td>
<td>36</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>2010-2011</td>
<td>64</td>
<td>150</td>
</tr>
<tr>
<td>3</td>
<td>2011-2012</td>
<td>67</td>
<td>171</td>
</tr>
<tr>
<td>4</td>
<td>2013-2014</td>
<td>64</td>
<td>152</td>
</tr>
</tbody>
</table>
Binomial model

Assume an independent binomial model,

\[ Y_s \sim \text{Bin}(n_s, \theta_s), \text{ i.e. } p(y|\theta) = \prod_{s=1}^{S} p(y_s|\theta_s) = \prod_{s=1}^{S} \left( \binom{n_s}{y_s} \theta_s^{y_s} (1-\theta_s)^{n_s-y_s} \right) \]

where

- \( y_s \) is the number of 3-pointers made in season \( s \)
- \( n_s \) is the number of 3-pointers attempted in season \( s \)
- \( \theta_s \) is the unknown 3-pointer success probability in season \( s \)
- \( S \) is the number of seasons
- \( \theta = (\theta_1, \theta_2, \theta_3, \theta_4)' \) and \( y = (y_1, y_2, y_3, y_4) \)

and assume independent beta priors distribution:

\[ p(\theta) = \prod_{s=1}^{S} p(\theta_s) = \prod_{s=1}^{S} \frac{\theta_s^{a_s-1} (1-\theta_s)^{b_s-1}}{\text{Beta}(a_s, b_s)} I(0 < \theta_s < 1). \]
Derive the posterior according to Bayes rule:

\[
p(\theta | y) \propto p(y | \theta) p(\theta)
= \prod_{s=1}^{S} p(y_s | \theta_s) \prod_{s=1}^{S} p(\theta_s)
= \prod_{s=1}^{S} p(y_s | \theta_s) p(\theta_s)
\propto \prod_{s=1}^{S} \text{Beta}(\theta_s | a_s + y_s, b_s + n_s - y_s)
\]

So the posterior for each \( \theta_s \) is exactly the same as if we treated each season independently.
Joint posterior

Andre Dawkins's 3-point percentage

![Graph showing Andre Dawkins's 3-point percentage for different seasons.](image-url)
Monte Carlo estimates

Estimated means, medians, and quantiles.

```r
sim = ddply(d, .(year),
  function(x) data.frame(theta=rbeta(1e3, x$a, x$b),
    a = x$a, b = x$b))

# hpd
hpd = function(theta,a,b,p=.95) {
  h = dbeta((a-1)/(a+b-2),a,b)
  ftheta = dbeta(theta,a,b)
  r = uniroot(function(x) mean(ftheta>x)-p,c(0,h))
  range(theta[which(ftheta>r$root)])
}

# expectations
ddply(sim, .(year), summarize,
  mean = mean(theta),
  median = median(theta),
  ciL = quantile(theta, c(.025,.975))[1],
  ciU = quantile(theta, c(.025,.975))[2],
  hpdL = hpd(theta,a[1],b[1])[1],
  hpdU = hpd(theta,a[1],b[1])[2])
```

<table>
<thead>
<tr>
<th>year</th>
<th>mean</th>
<th>median</th>
<th>ciL</th>
<th>ciU</th>
<th>hpdL</th>
<th>hpdU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3828</td>
<td>0.38</td>
<td>0.29</td>
<td>0.48</td>
<td>0.29</td>
<td>0.48</td>
</tr>
<tr>
<td>2</td>
<td>0.43</td>
<td>0.43</td>
<td>0.35</td>
<td>0.5</td>
<td>0.35</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>0.39</td>
<td>0.39</td>
<td>0.32</td>
<td>0.47</td>
<td>0.32</td>
<td>0.47</td>
</tr>
<tr>
<td>4</td>
<td>0.42</td>
<td>0.42</td>
<td>0.34</td>
<td>0.49</td>
<td>0.34</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Jarad Niemi (STAT544@ISU)
Comparing probabilities across years

The scientific question of interest here is whether Dawkins’s 3-point percentage is higher in 2013-2014 than in each of the previous years. Using probability notation, this is

\[ P(\theta_4 > \theta_s | y) \text{ for } s = 1, 2, 3. \]

which can be approximated via Monte Carlo as

\[ P(\theta_4 > \theta_s | y) = E_{\theta|y}[I(\theta_4 > \theta_s)] \approx \frac{1}{M} \sum_{m=1}^{M} I\left(\theta_4^{(m)} > \theta_s^{(m)}\right) \]

where

- \( \theta_s^{(m)} \) \( \text{ind} \) \( Be(a_s + y_s, b_s + n_s - y_s) \)
- \( I(A) \) is in indicator function that is 1 if \( A \) is true and zero otherwise.
Estimated probabilities

```r
# Should be able to use dcast
d = data.frame(theta_1 = sim$theta[sim$year==1],
               theta_2 = sim$theta[sim$year==2],
               theta_3 = sim$theta[sim$year==3],
               theta_4 = sim$theta[sim$year==4])

# Probabilities that season 4 percentage is higher than other seasons
mean(d$theta_4 > d$theta_1)

[1] 0.758

mean(d$theta_4 > d$theta_2)

[1] 0.454

mean(d$theta_4 > d$theta_3)

[1] 0.697
```
library(rjags)
independent_binomials = "model {
  for (i in 1:N) {
    y[i] ~ dbin(theta[i],n[i])
    theta[i] ~ dbeta(1,1)
  }
}
"

d = list(y=c(36,64,67,64), n=c(95,150,171,152), N=4)
m = jags.model(textConnection(independent_binomials), d)

Compiling model graph
  Resolving undeclared variables
  Allocating nodes
Graph information:
  Observed stochastic nodes: 4
  Unobserved stochastic nodes: 4
  Total graph size: 14

Initializing model

res = coda.samples(m, "theta", 1000)
summary(res)

Iterations = 1001:2000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1000

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

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Naive SE</th>
<th>Time-series SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>theta[1]</td>
<td>0.3777</td>
<td>0.04704</td>
<td>0.001487</td>
<td>0.001813</td>
</tr>
<tr>
<td>theta[2]</td>
<td>0.4278</td>
<td>0.04037</td>
<td>0.001277</td>
<td>0.001771</td>
</tr>
<tr>
<td>theta[3]</td>
<td>0.3943</td>
<td>0.03576</td>
<td>0.001131</td>
<td>0.001285</td>
</tr>
<tr>
<td>theta[4]</td>
<td>0.4223</td>
<td>0.03859</td>
<td>0.001220</td>
<td>0.001503</td>
</tr>
</tbody>
</table>

2. Quantiles for each variable:

<table>
<thead>
<tr>
<th></th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>theta[1]</td>
<td>0.2873</td>
<td>0.3438</td>
<td>0.3779</td>
<td>0.4100</td>
<td>0.4703</td>
</tr>
<tr>
<td>theta[2]</td>
<td>0.3546</td>
<td>0.3984</td>
<td>0.4272</td>
<td>0.4545</td>
<td>0.5111</td>
</tr>
<tr>
<td>theta[3]</td>
<td>0.3217</td>
<td>0.3707</td>
<td>0.3944</td>
<td>0.4177</td>
<td>0.4639</td>
</tr>
<tr>
<td>theta[4]</td>
<td>0.3492</td>
<td>0.3954</td>
<td>0.4216</td>
<td>0.4475</td>
<td>0.4982</td>
</tr>
</tbody>
</table>
# Extract sampled theta values
theta = as.matrix(res[[1]])  # with only 1 chain, all values are in the first list element

# Calculate probabilities that season 4 percentage is higher than other seasons
mean(theta[,4] > theta[,1])

[1] 0.772

mean(theta[,4] > theta[,2])

[1] 0.465

mean(theta[,4] > theta[,3])

[1] 0.702
Background probability theory

- Scaled Inv-\( \chi^2 \) distribution
- Location-scale \( t \)-distribution
- Normal-Inv-\( \chi^2 \) distribution
Scaled-inverse $\chi^2$-distribution

If $\sigma^2 \sim IG(a, b)$ with shape $a$ and scale $b$, then $\sigma^2 \sim \text{Inv-}\chi^2(v, z^2)$ with degrees of freedom $v$ and scale $z^2$ have the following

- $a = v/2$ and $b = vz^2/2$, or, equivalently,
- $v = 2a$ and $z^2 = b/a$.

Deriving from the inverse gamma, the scaled-inverse $\chi^2$ has

- Mean: $vz^2/(v - 2)$ for $v > 2$
- Mode: $vz^2/(v + 2)$
- Variance: $2v^2(z^2)^2/[(v - 2)^2(v - 4)]$ for $v > 4$

So $z^2$ is a point estimate and $v \to \infty$ means the variance decreases, since, for large $v$,

$$
\frac{2v^2(z^2)^2}{(v - 2)^2(v - 4)} \approx \frac{2v^2(z^2)^2}{v^3} = \frac{2(z^2)^2}{v}.
$$
Scaled-inverse $\chi^2$-distribution

dinvgamma = function(x, shape, scale, ...) dgamma(1/x, shape = shape, rate = scale, ...) / x^2
dsichisq = function(x, dof, scale, ...) dinvgamma( x, shape = dof/2, scale = dof*scale/2, ...)

![Graph showing the scaled-inverse chi-square distribution for different values of $s^2$ and $v_f$.](image-url)
Location-scale $t$-distribution

The $t$-distribution is a location-scale family (Casella & Berger Thm 3.5.6), i.e. if $T_v$ has a standard $t$-distribution with $v$ degrees of freedom and pdf

$$f_t(t) = \frac{\Gamma([v + 1]/2)}{\Gamma(v/2)\sqrt{v\pi}} (1 + t^2/v)^{-(v+1)/2},$$

then $X = m + zT_v$ has pdf

$$f_X(x) = f_t([x - m]/z)/z = \frac{\Gamma([v + 1]/2)}{\Gamma(v/2)\sqrt{v\pi}z} \left(1 + \frac{1}{v} \left[\frac{x - m}{z}\right]^2\right)^{-(v+1)/2}.$$

This is referred to as a $t$ distribution with $v$ degrees of freedom, location $m$, and scale $z$; it is written as $t_v(m, z^2)$. Also,

$$t_v(m, z^2) \xrightarrow{v \to \infty} N(m, z^2).$$
$t$ distribution as $\nu$ changes
Normal-Inv-$\chi^2$ distribution

Let $\mu|\sigma^2 \sim N(m, \sigma^2/k)$ and $\sigma^2 \sim \text{Inv-}\chi^2(v, z^2)$, then the kernel of this joint density is

$$p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$$

$$\propto (\sigma^2)^{-1/2} e^{-\frac{1}{2}\sigma^2/k} (\mu-m)^2 (\sigma^2)^{-\frac{v}{2}} - 1 e^{-\frac{vz^2}{2\sigma^2}}$$

$$= (\sigma^2)^{-v+3/2} e^{-\frac{1}{2\sigma^2} [k(\mu-m)^2 + vz^2]}$$

In addition, the marginal distribution for $\mu$ is

$$p(\mu) = \int p(\mu|\sigma^2)p(\sigma^2) d\sigma^2 = \cdots$$

$$= \frac{\Gamma([v+1]/2)}{\Gamma(v/2)\sqrt{v\pi z}/\sqrt{k}} \left(1 + \frac{1}{v} \left[\frac{\mu-m}{z/\sqrt{k}}\right]^2\right)^{-(v+1)/2}$$

with $\mu \in \mathbb{R}$. Thus $\mu \sim t_v(m, z^2/k)$. 
Univariate normal model

Suppose $Y_i \overset{ind}{\sim} N(\mu, \sigma^2)$.

![Normal model graph]
Confidence interval for $\mu$

Let

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \quad \text{and} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

Then,

$$T_{n-1} = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

and an equal-tail $100(1 - \alpha)$% confidence interval can be constructed via

$$1 - \alpha = P\left(-t_{n-1,1-\alpha/2} \leq T_{n-1} \leq t_{n-1,1-\alpha/2}\right) = P\left(\bar{Y} - \frac{t_{n-1,1-\alpha/2}S}{\sqrt{n}} \leq \mu \leq \bar{Y} + \frac{t_{n-1,1-\alpha/2}S}{\sqrt{n}}\right)$$

where $t_{n-1,1-\alpha/2}$ is the t-critical value, i.e. $P(T_{n-1} > t_{n-1,1-\alpha/2}) = \alpha/2$.

Thus

$$\bar{y} \pm t_{n-1,1-\alpha/2}s/\sqrt{n}$$

is an equal-tail $100(1 - \alpha)$% confidence interval with $\bar{y}$ and $s$ the observed values of $\bar{Y}$ and $S$. 
Default priors

Jeffreys prior can be shown to be $p(\mu, \sigma^2) \propto (1/\sigma^2)^{3/2}$. But alternative methods, e.g. reference prior, find that $p(\mu, \sigma^2) \propto 1/\sigma^2$ is a more appropriate prior.

The posterior under the reference prior is

$$p(\mu, \sigma^2 | y) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right) \times \frac{1}{\sigma^2}$$

$$= (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \bar{y} + \overline{y} - \mu)^2\right) \times \frac{1}{\sigma^2}$$

$$\vdots$$

$$= (\sigma^2)^{-(n-1+3)/2} \exp\left(-\frac{1}{2\sigma^2} \left[n(\mu - \bar{y})^2 + (n - 1)s^2\right]\right)$$

Thus

$$\mu | \sigma^2, y \sim N(\bar{y}, \sigma^2/n) \quad \sigma^2 | y \sim \text{Inv-}\chi^2(n - 1, s^2).$$
Marginal posterior for $\mu$

The marginal posterior for $\mu$ is

$$\mu|y \sim t_{n-1}(\bar{y}, s^2/n).$$

An equal-tailed $100(1 - \alpha)%$ credible interval can be obtained via

$$\bar{y} \pm t_{n-1,1-\alpha/2}s/\sqrt{n}.$$ 

This formula is exactly the same as the formula for a $100(1 - \alpha/2)%$ confidence interval. But the interpretation of this credible interval is a statement about your belief when your prior belief is represented by the prior $p(\mu, \sigma^2) \propto 1/\sigma^2$. 
Predictive distribution

Let $\tilde{y} \sim N(\mu, \sigma^2)$. The predictive distribution is

$$p(\tilde{y}|y) = \int \int p(\tilde{y}|\mu, \sigma^2)p(\mu|\sigma^2, y)p(\sigma^2|y)d\mu d\sigma^2$$

The easiest way to derive this is to write $\tilde{y} = \mu + \epsilon$ with

$$\mu|\sigma^2, y \sim N(\bar{y}, \sigma^2/n) \quad \epsilon|\sigma^2, y \sim N(0, \sigma^2)$$

independent of each other. Thus

$$\tilde{y}|\sigma^2, y \sim N(\bar{y}, \sigma^2[1 + 1/n]).$$

with $\sigma^2|y \sim \text{Inv-}\chi^2(n - 1, s^2)$. Now, we can use the Normal-Inv-\chi^2 theory, to find that

$$\tilde{y}|y \sim t_{n-1}(\bar{y}, s^2[1 + 1/n]).$$
Conjugate prior for $\mu$ and $\sigma^2$

The joint conjugate prior for $\mu$ and $\sigma^2$ is

$$
\mu | \sigma^2 \sim N(m, \sigma^2/k) \quad \sigma^2 \sim \text{Inv-}\chi^2(v, z^2)
$$

where $z^2$ serves as a prior guess about $\sigma^2$ and $v$ controls how certain we are about that guess.

The posterior under this prior is

$$
\mu | \sigma^2, y \sim N(m', \sigma^2/k') \quad \sigma^2 | y \sim \text{Inv-}\chi^2(v', (z')^2)
$$

where

$$
k' = k + n \quad m' = [km + n\bar{y}] / k' \quad v' = v + n \quad v'(z')^2 = vz^2 + (n - 1)S^2 + \frac{kn}{k'}(\bar{y} - m)^2
$$
The marginal posterior for $\mu$ is

$$\mu | y \sim t_{v'}(m', (z')^2 / k').$$

An equal-tailed $100(1 - \alpha)\%$ credible interval can be obtained via

$$m' \pm t_{v', 1-\alpha/2} z' / \sqrt{k'}.$$
Marginal posterior via simulation

An alternative to deriving the closed form posterior for $\mu$ is to simulate from the distribution. Recall that

$$
\mu | \sigma^2, y \sim N(m', \sigma^2 / k') \quad \sigma^2 | y \sim \text{Inv-}\chi^2(v', (z')^2)
$$

To obtain a simulation from the posterior distribution $p(\mu, \sigma^2 | y)$, calculate $m', k', v', \text{ and } z'$ and then

1. simulate $\sigma^2 \sim \text{Inv-}\chi^2(v', (z')^2)$ and
2. using the simulated $\sigma^2$, simulate $\mu \sim N(m', \sigma^2 / k')$.

Not only does this provide a sample from the joint distribution for $\mu, \sigma$ but it also (therefore) provides a sample from the marginal distribution for $\mu$.

The integral was suggestive:

$$
p(\mu | y) = \int p(\mu | \sigma^2, y)p(\sigma^2 | y) d\sigma^2
$$
Similarly, we can obtain the predictive distribution via simulation. Recall that

\[ p(\tilde{y}|y) = \int \int p(\tilde{y}|\mu, \sigma^2)p(\mu|\sigma^2, y)p(\sigma^2|y)\,d\mu\,d\sigma^2 \]

To obtain a simulation from the predictive distribution \( p(\tilde{y}|y) \), calculate \( m', k', v', \) and \( z' \)

1. simulate \( \sigma^2 \sim \text{Inv-\chi}^2(v', (z')^2) \),
2. using this \( \sigma^2 \), simulate \( \mu \sim N(m', \sigma^2/k') \), and
3. using these \( \mu \) and \( \sigma^2 \), simulate \( \tilde{y} \sim N(\mu, \sigma^2) \).
Summary of normal inference

• Default analysis
  • Prior: (think $\mu \sim N(0, \infty)$ and $\sigma^2 \sim Inv-\chi^2(0, 0)$)
    $$p(\mu, \sigma^2) \propto 1/\sigma^2$$
  • Posterior:
    $$\mu|\sigma^2, y \sim N(\overline{y}, \sigma^2/n), \sigma^2|y \sim Inv-\chi^2(n - 1, S^2), \mu|y \sim t_{n-1}(\overline{y}, S^2/n)$$

• Conjugate analysis
  • Prior:
    $$\mu|\sigma^2 \sim N(m, \sigma^2/k), \sigma^2 \sim Inv-\chi^2(v, z^2), \mu \sim t_v(m, z^2/k)$$
  • Posterior:
    $$\mu|\sigma^2, y \sim N(m', \sigma^2/k'), \sigma^2|y \sim Inv-\chi^2(v', (z')^2), \mu|y \sim t_{v'}(m', (z')^2/k')$$
    with
    $$k' = k + n, m' = [km + n\overline{y}]/k', v' = v + n,$$
    $$v'(z')^2 = vz^2 + (n - 1)S^2 + \frac{kn}{k'}(\overline{y} - m)^2$$