Multiparameter models

Dr. Jarad Niemi

STAT 544 - Iowa State University

February 7, 2019

Outline

- Independent beta-binomial
  - Independent posteriors
  - Comparison of parameters
  - JAGS
- Probability theory results
  - Scaled Inv-χ^2 distribution
  - t-distribution
  - Normal-Inv-χ^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?

Season	Year	Made	Attempts
1	2009-2010	36	95
2	2010-2011	64	150
3	2011-2012	67	171
4	2013-2014	64	152

Binomial model

Assume an independent binomial model,

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

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
- θ_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:

$$p(\theta) = \prod_{s=1}^{S} p(\theta_s) = \prod_{s=1}^{S} \frac{\theta_s^{a_s-1}(1-\theta_s)^{b_s-1}}{Beta(a_s, b_s)} \mathrm{I}(0 < \theta_s < 1).$$

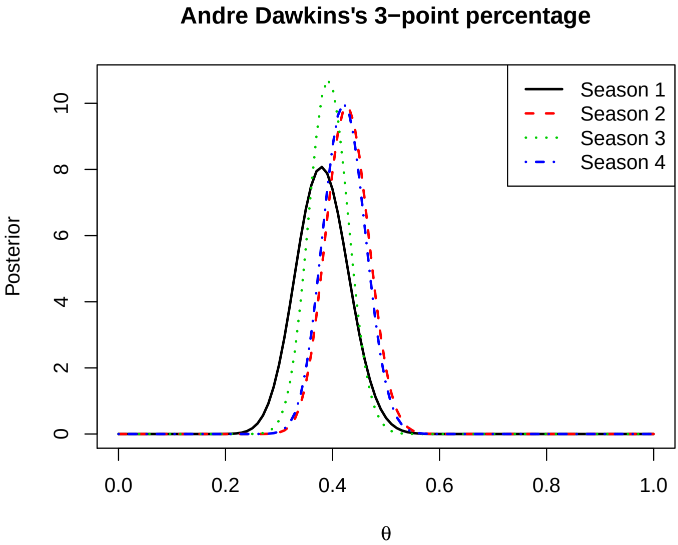
Joint posterior

Derive the posterior according to Bayes rule:

$$
\begin{aligned}
p(\theta|y) &\propto p(y|\theta)p(\theta) \\
&= \prod_{s=1}^{S} p(y_s|\theta_s) \prod_{s=1}^{S} p(\theta_s) \\
&= \prod_{s=1}^{S} p(y_s|\theta_s)p(\theta_s) \\
&\propto \prod_{s=1}^{S} \text{Beta}(\theta_s|a_s + y_s, b_s + n_s - y_s)
\end{aligned}
$$

So the posterior for each θ_s is exactly the same as if we treated each season independently.

Joint posterior



Monte Carlo estimates

Estimated means, medians, and quantiles.

```
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])
```

```
  year      mean    median       ciL       ciU      hpdL      hpdU
1    1 0.3828454 0.3816672 0.2893217 0.4821211 0.2851402 0.4803823
2    2 0.4283304 0.4297132 0.3498912 0.5050538 0.3509289 0.5054018
3    3 0.3951943 0.3958465 0.3235839 0.4694850 0.3208512 0.4662410
4    4 0.4228666 0.4235223 0.3464835 0.4982144 0.3465337 0.4981711
```

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

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

where

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

Estimated probabilities

```
# 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
```

Using JAGS

```
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:

           Mean      SD Naive SE Time-series SE
theta[1] 0.3777 0.04704 0.001487       0.001813
theta[2] 0.4278 0.04037 0.001277       0.001771
theta[3] 0.3943 0.03576 0.001131       0.001285
theta[4] 0.4223 0.03859 0.001220       0.001503

2. Quantiles for each variable:

           2.5%    25%    50%    75%  97.5%
theta[1] 0.2873 0.3438 0.3779 0.4100 0.4703
theta[2] 0.3546 0.3984 0.4272 0.4545 0.5111
theta[3] 0.3217 0.3707 0.3944 0.4177 0.4639
theta[4] 0.3492 0.3954 0.4216 0.4475 0.4982
```

```r
# 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
```

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

```
[1] 0.465
```

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

```
[1] 0.702
```

Background probability theory

- Scaled Inv-χ^2 distribution
- Location-scale t-distribution
- Normal-Inv-χ^2 distribution

Scaled-inverse χ^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 χ^2 has

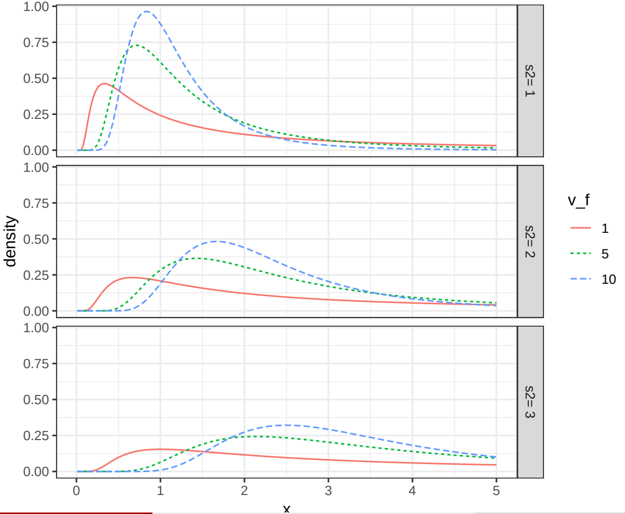
- Mean: $vz^2/(v-2)$ for $v > 2$
- Mode: $vz^2/(v+2)$
- Variance: $2v^2(z^2)^2/[(v-2)^2(v-4)]$ for $v > 4$

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 χ^2-distribution

```
dinvgamma = function(x, a, b, ...) dgamma(1/x, a, b,...)/x^2
dsichisq = function(x, v, s2, ...) dinvgamma(x, v/2, v*s2/2, ...)
```



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

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

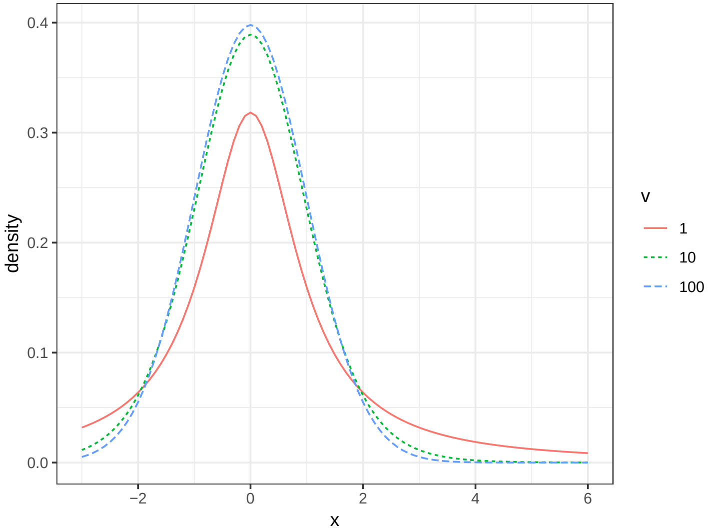
then $X = m + zT_v$ has pdf

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

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,

$$t_v(m, z^2) \stackrel{v\to\infty}{\longrightarrow} N(m, z^2).$$

t distribution as v changes



Normal-Inv-χ^2 distribution

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

$$\begin{aligned} p(\mu, \sigma^2) &= p(\mu|\sigma^2)p(\sigma^2) \\ &\propto (\sigma^2)^{-1/2} e^{-\frac{1}{2\sigma^2/k}(\mu-m)^2} (\sigma^2)^{-\frac{v}{2}-1} e^{-\frac{vz^2}{2\sigma^2}} \\ &= (\sigma^2)^{-(v+3)/2} e^{-\frac{1}{2\sigma^2}\left[k(\mu-m)^2+vz^2\right]} \end{aligned}$$

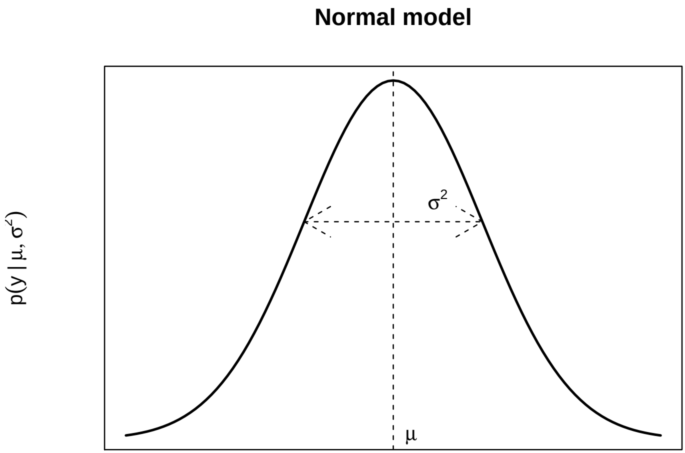
In addition, the marginal distribution for μ is

$$\begin{aligned} p(\mu) &= \int p(\mu|\sigma^2)p(\sigma^2)d\sigma^2 = \cdots \\ &= \frac{\Gamma([v+1]/2)}{\Gamma(v/2)\sqrt{v\pi}z/\sqrt{k}} \left(1 + \frac{1}{v}\left[\frac{\mu-m}{z/\sqrt{k}}\right]^2\right)^{-(v+1)/2}. \end{aligned}$$

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

Univariate normal model

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



Confidence interval for μ

Let

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

Then,

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

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

$$\begin{aligned} 1-\alpha &= P\left(-t_{n-1,1-\alpha/2} \le T_{n-1} \le t_{n-1,1-\alpha/2}\right) \\ &= P\left(\overline{Y} - \frac{t_{n-1,1-\alpha/2}S}{\sqrt{n}} \le \mu \le \overline{Y} + \frac{t_{n-1,1-\alpha/2}S}{\sqrt{n}}\right) \end{aligned}$$

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

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

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

$$\begin{aligned} p(\mu, \sigma^2|y) &\propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right) \times \frac{1}{\sigma^2} \\ &= (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \overline{y} + \overline{y} - \mu)^2\right) \times \frac{1}{\sigma^2} \\ &\vdots \\ &= (\sigma^2)^{-(n-1+3)/2} \exp\left(-\frac{1}{2\sigma^2} \left[n(\mu - \overline{y})^2 + (n-1)s^2\right]\right) \end{aligned}$$

Thus

$$\mu|\sigma^2, y \sim N(\overline{y}, \sigma^2/n) \qquad \sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2).$$

Marginal posterior for μ

The marginal posterior for μ is

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

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

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

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

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

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

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

independent of each other. Thus

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

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

$$\tilde{y}|y \sim t_{n-1}(\overline{y}, s^2[1 + 1/n]).$$

Conjugate prior for μ and σ^2

The joint conjugate prior for μ and σ^2 is

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

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

The posterior under this prior is

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

where

$$\begin{aligned} k' &= k + n \\ m' &= [km + n\overline{y}]/k' \\ v' &= v + n \\ v'(z')^2 &= vz^2 + (n-1)S^2 + \tfrac{kn}{k'}(\overline{y} - m)^2 \end{aligned}$$

Marginal posterior for μ

The marginal posterior for μ is

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

An equal-tailed $100(1-\alpha)\%$ credible inteval 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 μ is to simulate from the distribution. Recall that

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

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

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

Not only does this provide a sample from the joint distribution for μ, σ but it also (therefore) provides a sample from the marginal distribution for μ. The integral was suggestive:

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

Predictive distribution via simulation

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 σ^2, simulate $\mu \sim N(m', \sigma^2/k')$, and
3. using these μ and σ^2, simulate $\tilde{y} \sim N(\mu, \sigma^2)$.

Summary of normal inference

- Default analysis
  - Prior: (think $\mu \sim N(0,\infty)$ and $\sigma^2 \sim \text{Inv-}\chi^2(0,0)$)
    $$p(\mu,\sigma^2) \propto 1/\sigma^2$$
  - Posterior:
    $$\mu|\sigma^2,y \sim N(\overline{y},\sigma^2/n),\ \sigma^2|y \sim \text{Inv-}\chi^2(n-1,S^2),\ \mu|y \sim t_{n-1}(\overline{y},S^2/n)$$
- Conjugate analysis
  - Prior:
    $$\mu|\sigma^2 \sim N(m,\sigma^2/k),\ \sigma^2 \sim \text{Inv-}\chi^2(v,z^2),\ \mu \sim t_v(m,z^2/k)$$
  - Posterior:
    $$\mu|\sigma^2,y \sim N(m',\sigma^2/k'),\ \sigma^2|y \sim \text{Inv-}\chi^2(v',(z')^2),\ \mu|y \sim t_{v'}(m',(z')^2/k')$$
    with
    $$k' = k+n, m' = [km+n\overline{y}]/k', v' = v+n,$$
    $$v'(z')^2 = vz^2 + (n-1)S^2 + \frac{kn}{k'}(\overline{y}-m)^2$$