Hierarchical models

Dr. Jarad Niemi

STAT 544 - Iowa State University

February 7, 2018

Outline

- Motivating example
  - Independent vs pooled estimates
- Hierarchical models
  - General structure
  - Posterior distribution
- Binomial hierarchical model
  - Posterior distribution
  - Prior distributions
- Stan for binomial hierarchical model
  - informative prior
  - default prior
  - integrating out θ
  - across seasons

Andre Dawkin's three-point percentage

Suppose Y_i are the number 3-pointers Andre Dawkin's makes in season i, and assume

$$Y_i \stackrel{ind}{\sim} Bin(n_i, \theta_i)$$

where

- n_i are the number of 3-pointers attempted and
- θ_i is the probability of making a 3-pointer in season$i$.

Do these models make sense?

- The 3-point percentage every season is the same, i.e. $\theta_i = \theta$.
- The 3-point percentage every season is independent of other seasons.
- The 3-point percentage every season should be similar to other seasons.

Andre Dawkin's three-point percentage

Suppose Y_i are the number of 3-pointers Andre Dawkin's makes in game i, and assume

$$Y_i \stackrel{ind}{\sim} Bin(n_i, \theta_i)$$

where

- n_i are the number of 3-pointers attempted in game i and
- θ_i is the probability of making a 3-pointer in game i.

Do these models make sense?

- The 3-point percentage every game is the same, i.e. $\theta_i = \theta$.
- The 3-point percentage every game is independent of other games.
- The 3-point percentage every game should be similar to other games.

```
Error in rbind(d, data.frame(date = NA, opponent = "Total", made = sum(d$made), :  could not find function
"%>%"
```

Andre Dawkin's 3-point percentage

```
Error in FUN(X[[i]], ...):  object 'lcl' not found
```

Andre Dawkin's 3-point percentage

	date	opponent	made	attempts
1	11/8/13	davidson	0	0
2	11/12/13	kansas	0	0
3	11/15/13	florida atlantic	5	8
4	11/18/13	unc asheville	3	6
5	11/19/13	east carolina	0	1
6	11/24/13	vermont	3	9
7	11/27/13	alabama	0	2
8	11/29/13	arizona	1	1
9	12/3/13	michigan	2	2
10	12/16/13	gardner-webb	4	8
11	12/19/13	ucla	1	5
12	12/28/13	eastern michigan	6	10
13	12/31/13	elon	5	7
14	1/4/14	notre dame	1	4
15	1/7/14	georgia tech	1	5
16	1/11/14	clemson	0	4
17	1/13/14	virginia	1	1
18	1/18/14	nc state	3	7
19	1/22/14	miami	2	6
20	1/25/14	florida state	3	6
21	1/27/14	pitt	6	7
22	2/1/14	syracuse	4	9
23	2/4/14	wake forest	4	7
24	2/8/14	boston college	0	1

```
Error in d %>% filter(opponent != "Total"):  could not find function "%>%"
```

Hierarchical models

Consider the following model

$$
\begin{array}{ll}
y_i & \stackrel{ind}{\sim} p(y|\theta_i) \\
\theta_i & \stackrel{ind}{\sim} p(\theta|\phi) \\
\phi & \sim p(\phi)
\end{array}
$$

where

- y_i is observed,
- $\theta = (\theta_1, \ldots, \theta_n)$ and ϕ are parameters, and
- only ϕ has a prior that is set.

This is a hierarchical or multilevel model.

Posterior distribution for hierarchical models

The joint posterior distribution of interest in hierarchical models is

$$p(\theta, \phi|y) \propto p(y|\theta, \phi)p(\theta, \phi) = p(y|\theta)p(\theta|\phi)p(\phi) = \left[\prod_{i=1}^{n} p(y_i|\theta_i)p(\theta_i|\phi)\right]p(\phi).$$

The joint posterior distribution can be decomposed via

$$p(\theta, \phi|y) = p(\theta|\phi, y)p(\phi|y)$$

where

$$\begin{array}{rl}
p(\theta|\phi, y) & \propto p(y|\theta)p(\theta|\phi) = \prod_{i=1}^{n} p(y_i|\theta_i)p(\theta_i|\phi) \propto \prod_{i=1}^{n} p(\theta_i|\phi, y_i) \\
p(\phi|y) & \propto p(y|\phi)p(\phi) \\
p(y|\phi) & = \int p(y|\theta)p(\theta|\phi)d\theta \\
& = \int \cdots \int \prod_{i=1}^{n} \left[p(y_i|\theta_i)p(\theta_i|\phi)\right] d\theta_1 \cdots d\theta_n \\
& = \prod_{i=1}^{n} \int p(y_i|\theta_i)p(\theta_i|\phi)d\theta_i \\
& = \prod_{i=1}^{n} p(y_i|\phi)
\end{array}$$

Three-pointer example

Our statistical model

$$\begin{array}{ll} Y_i & \stackrel{ind}{\sim} Bin(n_i, \theta_i) \\ \theta_i & \stackrel{ind}{\sim} Be(\alpha, \beta) \\ \alpha, \beta & \sim p(\alpha, \beta) \end{array}$$

In this example,

- $\phi = (\alpha, \beta)$
- $Be(\alpha, \beta)$ describes the variability in 3-point percentage across games, and
- we are going to learn about this variability.

Decomposed posterior

$$Y_i \stackrel{ind}{\sim} Bin(n_i, \theta_i) \quad \theta_i \stackrel{ind}{\sim} Be(\alpha, \beta) \quad \alpha, \beta \sim p(\alpha, \beta)$$

Conditional posterior for θ:

$$p(\theta|\alpha, \beta, y) = \prod_{i=1}^{n} p(\theta_i|\alpha, \beta, y_i) = \prod_{i=1}^{n} Be(\theta_i|\alpha + y_i, \beta + n_i - y_i)$$

Marginal posterior for (α, β):

$$\begin{array}{ll}
p(\alpha, \beta|y) & \propto p(y|\alpha, \beta)p(\alpha, \beta) \\
p(y|\alpha, \beta) & = \prod_{i=1}^{n} p(y_i|\alpha, \beta) = \prod_{i=1}^{n} \int p(y_i|\theta_i)p(\theta_i|\alpha, \beta)d\theta_i \\
& = \prod_{i=1}^{n} \int Bin(y_i|n_i, \theta_i)Be(\theta_i|\alpha, \beta)d\theta_i \\
& = \prod_{i=1}^{n} \int_0^1 \binom{n_i}{y_i}\theta_i^{y_i}(1-\theta_i)^{n_i-y_i}\frac{\theta_i^{\alpha-1}(1-\theta_i)^{\beta-1}}{B(\alpha,\beta)}d\theta_i \\
& = \prod_{i=1}^{n} \binom{n_i}{y_i}\frac{1}{B(\alpha,\beta)}\int_0^1 \theta_i^{\alpha+y_i-1}(1-\theta_i)^{\beta+n_i-y_i-1}d\theta_i \\
& = \prod_{i=1}^{n} \binom{n_i}{y_i}\frac{B(\alpha+y_i,\beta+n_i-y_i)}{B(\alpha,\beta)}
\end{array}$$

Thus $y_i|\alpha, \beta \stackrel{ind}{\sim}$ Beta-binomial(n_i, α, β).

A prior distribution for α and β

Recall the interpretation:

- α: prior successes
- β: prior failures

A more natural parameterization is

- prior expectation: $\mu = \frac{\alpha}{\alpha+\beta}$
- prior sample size: $\eta = \alpha + \beta$

Place priors on these parameters or transformed to the real line:

- logit $\mu = \log(\mu/[1-\mu]) = \log(\alpha/\beta)$
- $\log \eta$

A prior distribution for α and β

It seems reasonable to assume the mean (μ) and size (η) are independent *a priori*:

$$p(\mu, \eta) = p(\mu)p(\eta)$$

```
a = 20
b = 30
m = 0
C = 3
```

Let's assume an informative prior for μ and η perhaps

- $\mu \sim Be(20, 30)$
- $\eta \sim LN(0, 3^2)$

where $LN(0, 3)$ is a log-normal distribution, i.e. $\log(\eta) \sim N(0, 3)$.

Prior draws

```
n = 1e4

prior_draws = data.frame(mu  = rbeta(n, a, b),
                         eta = rlnorm(n, m, C)) %>%
  mutate(alpha = eta*   mu,
         beta  = eta*(1-mu))

prior_draws %>%
  tidyr::gather(parameter, value) %>%
  group_by(parameter) %>%
  summarize(lower95 = quantile(value, prob = 0.025),
            median  = quantile(value, prob = 0.5),
            upper95 = quantile(value, prob = 0.975))
```

```
      lower95    median  upper95
1 0.002736002 0.4238952 177.3007
```

```
cor(prior_draws$alpha, prior_draws$beta)
```

```
[1] 0.9451046
```

```
model_informative_prior = "
data {
  int<lower=0> N;    // data
  int<lower=0> n[N];
  int<lower=0> y[N];
  real<lower=0> a;   // prior
  real<lower=0> b;
  real<lower=0> C;
  real m;
}
parameters {
  real<lower=0,upper=1> mu;
  real<lower=0> eta;
  real<lower=0,upper=1> theta[N];
}
transformed parameters {
  real<lower=0> alpha;
  real<lower=0> beta;

  alpha = eta*   mu ;
  beta  = eta*(1-mu);
}
model {
  mu     ~ beta(a,b);
  eta    ~ lognormal(m,C);

  # implicit joint distributions
  theta ~ beta(alpha,beta);
  y     ~ binomial(n,theta);
}
"
```

Stan

```
dat = list(y=d$made, n=d$attempts, N=nrow(d),a=a, b=b, m=m, C=C)
m = stan_model(model_code=model_informative_prior)
r = sampling(m, dat, c("mu","eta","alpha","beta","theta"),
             iter = 10000,
             control = list(adapt_delta = 0.9))
```

```
Warning:  There were 19 divergent transitions after warmup.  Increasing adapt_delta above 0.9 may help.
See
http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
Warning:  There were 4 chains where the estimated Bayesian Fraction of Missing Information was low.  See
http://mc-stan.org/misc/warnings.html#bfmi-low
Warning:  Examine the pairs() plot to diagnose sampling problems
```

stan

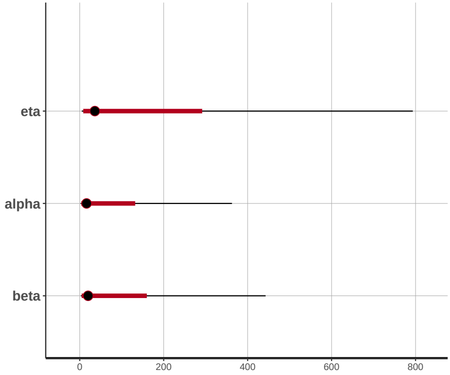
```
r
```

```
Inference for Stan model: 165e18050a3b1562e943b25f76c728ea.
4 chains, each with iter=10000; warmup=5000; thin=1;
post-warmup draws per chain=5000, total post-warmup draws=20000.

             mean se_mean     sd    2.5%     25%     50%     75%   97.5% n_eff Rhat
mu           0.45    0.00   0.04    0.37    0.42    0.45    0.47    0.53  3170 1.00
eta        117.21   10.84 229.80    4.83   15.43   35.97  106.51  793.43   450 1.01
alpha       52.54    4.78 102.44    2.06    6.77   16.26   47.85  362.59   460 1.01
beta        64.68    6.09 128.23    2.71    8.56   19.89   58.65  442.77   444 1.01
theta[1]     0.45    0.00   0.11    0.21    0.39    0.45    0.51    0.68 20000 1.00
theta[2]     0.45    0.00   0.11    0.21    0.38    0.45    0.51    0.68 20000 1.00
theta[3]     0.49    0.00   0.09    0.32    0.43    0.48    0.54    0.70  4425 1.00
theta[4]     0.46    0.00   0.09    0.28    0.40    0.45    0.51    0.65 20000 1.00
theta[5]     0.43    0.00   0.11    0.18    0.37    0.43    0.49    0.64 20000 1.00
theta[6]     0.42    0.00   0.09    0.22    0.37    0.43    0.48    0.59  4979 1.00
theta[7]     0.41    0.00   0.11    0.16    0.36    0.42    0.48    0.60  4448 1.00
theta[8]     0.47    0.00   0.11    0.26    0.40    0.46    0.52    0.73 20000 1.00
theta[9]     0.49    0.00   0.11    0.30    0.42    0.48    0.54    0.75  4130 1.00
theta[10]    0.46    0.00   0.09    0.29    0.40    0.46    0.51    0.65 20000 1.00
theta[11]    0.41    0.00   0.10    0.18    0.35    0.42    0.47    0.59  3726 1.00
theta[12]    0.49    0.00   0.09    0.33    0.43    0.48    0.54    0.68  4421 1.00
theta[13]    0.50    0.00   0.10    0.34    0.44    0.49    0.56    0.73  3175 1.00
theta[14]    0.42    0.00   0.10    0.20    0.37    0.43    0.48    0.60 20000 1.00
theta[15]    0.41    0.00   0.10    0.18    0.35    0.42    0.47    0.58  4197 1.00
theta[16]    0.39    0.00   0.11    0.13    0.33    0.40    0.46    0.56  1983 1.00
theta[17]    0.47    0.00   0.11    0.27    0.41    0.46    0.53    0.72 20000 1.00
theta[18]    0.44    0.00   0.09    0.26    0.39    0.44    0.50    0.63 20000 1.00
theta[19]    0.43    0.00   0.09    0.23    0.37    0.43    0.48    0.60 20000 1.00
```

stan

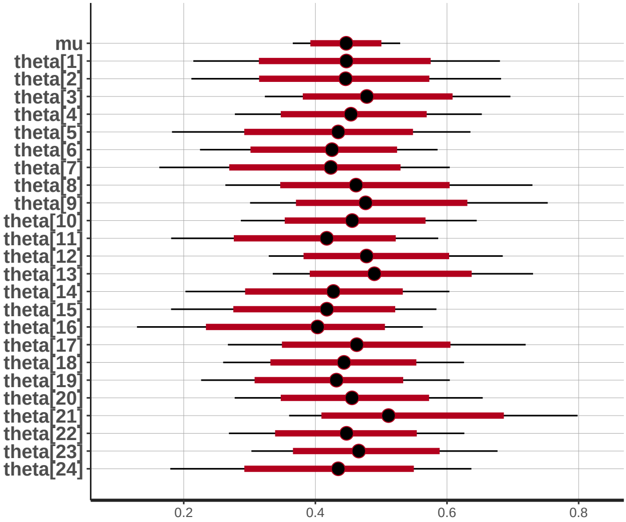
```r
plot(r, pars=c('eta','alpha','beta'))
```

```
ci_level: 0.8 (80% intervals)
outer_level: 0.95 (95% intervals)
```



stan

```
plot(r, pars=c('mu','theta'))
```



Comparing independent and hierarchical models

```
Error in rbind(deparse.level, ...):  numbers of columns of arguments do not match
```

A prior distribution for α and β

In Bayesian Data Analysis (3rd ed) page 110, several priors are discussed

- $(\log(\alpha/\beta), \log(\alpha+\beta)) \propto 1$ leads to an improper posterior.
- $(\log(\alpha/\beta), \log(\alpha+\beta)) \sim Unif([-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}])$ while proper and seemingly vague is a very informative prior.
- $(\log(\alpha/\beta), \log(\alpha+\beta)) \propto \alpha\beta(\alpha+\beta)^{-5/2}$ which leads to a proper posterior and is equivalent to $p(\alpha, \beta) \propto (\alpha+\beta)^{-5/2}$.

Stan - default prior

```
model_default_prior = "
data {
  int<lower=0> N;
  int<lower=0> n[N];
  int<lower=0> y[N];
}
parameters {
  real<lower=0> alpha;
  real<lower=0> beta;
  real<lower=0,upper=1> theta[N];
}

model {
  # default prior
  target += -5*log(alpha+beta)/2;

  # implicit joint distributions
  theta ~ beta(alpha,beta);
  y     ~ binomial(n,theta);
}
"

m2 = stan_model(model_code=model_default_prior)
r2 = sampling(m2, dat, c("alpha","beta","theta"), iter=10000,
              control = list(adapt_delta = 0.9))
```

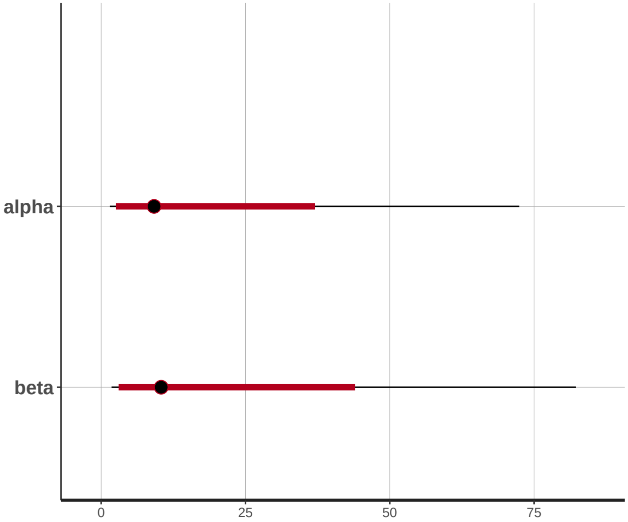
```
Warning:  There were 1700 divergent transitions after warmup.  Increasing adapt_delta above 0.9 may help.
See
http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
Warning:  There were 4 chains where the estimated Bayesian Fraction of Missing Information was low.  See
```

Stan - default prior

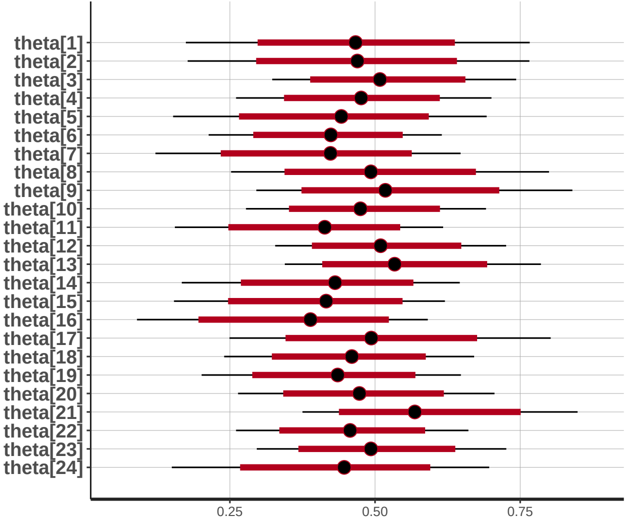
```
r2


Inference for Stan model: 5846b411d351181d50f65dae5af5de43.
4 chains, each with iter=10000; warmup=5000; thin=1;
post-warmup draws per chain=5000, total post-warmup draws=20000.

           mean se_mean    sd  2.5%   25%   50%   75% 97.5% n_eff Rhat
alpha     16.28    1.74 20.89  1.52  4.55  9.17 19.33 72.42   145 1.05
beta      18.40    1.94 23.31  1.79  5.20 10.40 21.59 82.22   144 1.05
theta[1]   0.47    0.00  0.14  0.17  0.39  0.47  0.55  0.77  8162 1.00
theta[2]   0.47    0.00  0.14  0.18  0.39  0.47  0.55  0.77 12071 1.00
theta[3]   0.52    0.00  0.11  0.32  0.44  0.51  0.58  0.74  3380 1.00
theta[4]   0.48    0.00  0.11  0.26  0.41  0.48  0.54  0.70 12595 1.00
theta[5]   0.44    0.00  0.13  0.15  0.36  0.44  0.52  0.69  7005 1.00
theta[6]   0.42    0.00  0.10  0.21  0.36  0.42  0.49  0.61  5082 1.00
theta[7]   0.41    0.00  0.13  0.12  0.33  0.42  0.50  0.65  3914 1.00
theta[8]   0.50    0.00  0.13  0.25  0.42  0.49  0.58  0.80  6900 1.00
theta[9]   0.53    0.00  0.14  0.30  0.44  0.52  0.61  0.84  2789 1.00
theta[10]  0.48    0.00  0.10  0.28  0.41  0.47  0.54  0.69 20000 1.00
theta[11]  0.40    0.00  0.12  0.16  0.33  0.41  0.48  0.62  3357 1.00
theta[12]  0.52    0.00  0.10  0.33  0.45  0.51  0.58  0.73  3377 1.00
theta[13]  0.54    0.00  0.11  0.34  0.47  0.53  0.61  0.79  3497 1.00
theta[14]  0.42    0.00  0.12  0.17  0.35  0.43  0.50  0.65  5756 1.00
theta[15]  0.41    0.00  0.12  0.15  0.33  0.42  0.49  0.62  3393 1.00
theta[16]  0.37    0.00  0.13  0.09  0.30  0.39  0.46  0.59  2314 1.00
theta[17]  0.50    0.00  0.13  0.25  0.42  0.49  0.58  0.80  7366 1.00
theta[18]  0.46    0.00  0.11  0.24  0.39  0.46  0.52  0.67  8298 1.00
theta[19]  0.43    0.00  0.11  0.20  0.36  0.44  0.50  0.65  4583 1.00
theta[20]  0.48    0.00  0.11  0.26  0.41  0.47  0.55  0.71  6131 1.00
theta[21]  0.58    0.00  0.12  0.38  0.50  0.57  0.66  0.85  1636 1.01
```

Stan - default prior

```
plot(r2, pars=c('alpha','beta'))
```



Stan - default prior



Comparing all models

```
Error in rbind(deparse.level, ...):  numbers of columns of arguments do not match
Error in factor(bind_d$prior, c("informative", "default", "independent")):  object 'bind_d' not found
Error in ggplot(bind_d, aes(x = lcl, xend = ucl, y = game + e * (prior == :  object 'bind_d' not found
```

Marginal posterior for α, β

An alternative to jointly sampling θ, α, β is to

1. sample $\alpha, \beta \sim p(\alpha, \beta|y)$, and then
2. sample $\theta_i \stackrel{ind}{\sim} p(\theta_i|\alpha, \beta, y_i) \stackrel{d}{=} Be(\alpha + y_i, \beta + n_i - y_i)$.

The maginal posterior for α, β is

$$p(\alpha, \beta|y) \propto p(y|\alpha, \beta)p(\alpha, \beta) = \left[\prod_{i=1}^{n} \text{Beta-binomial}(y_i|n_i, \alpha, \beta)\right] p(\alpha, \beta)$$

Stan - beta-binomial

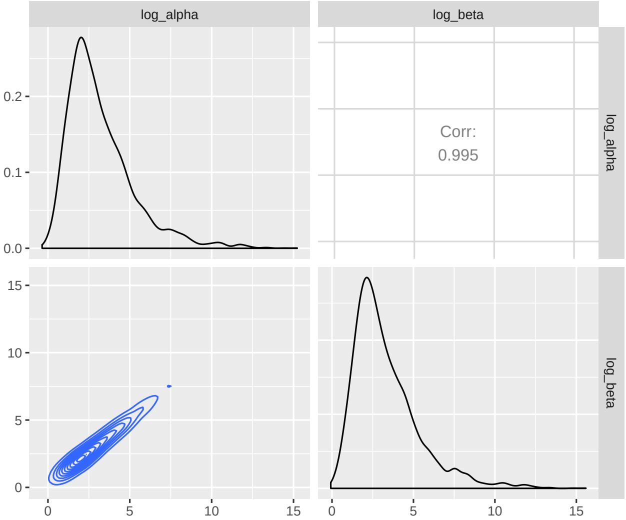
```
# Marginalized (integrated) theta out of the model
model_marginalized = "
data {
  int<lower=0> N;
  int<lower=0> n[N];
  int<lower=0> y[N];
}
parameters {
  real<lower=0> alpha;
  real<lower=0> beta;
}
model {
  target += -5*log(alpha+beta)/2;
  y     ~ beta_binomial(n,alpha,beta);
}
"

m3 = stan_model(model_code=model_marginalized)
r3 = sampling(m3, dat, c("alpha","beta"))
```

Stan - beta-binomial

```
Inference for Stan model: e43d085e5efc74fdcaa9b1ceb76cdc65.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

         mean se_mean        sd   2.5%    25%    50%    75%   97.5% n_eff Rhat
alpha 3222.94 1872.28  73940.36   1.86   6.05  15.23  63.20 4590.35  1560    1
beta  4020.31 2519.13 104083.61   2.09   6.85  16.76  68.59 4807.47  1707    1
lp__   -84.61    0.04      1.09 -87.61 -85.09 -84.26 -83.81  -83.50   732    1

Samples were drawn using NUTS(diag_e) at Thu Feb  9 09:42:20 2017.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

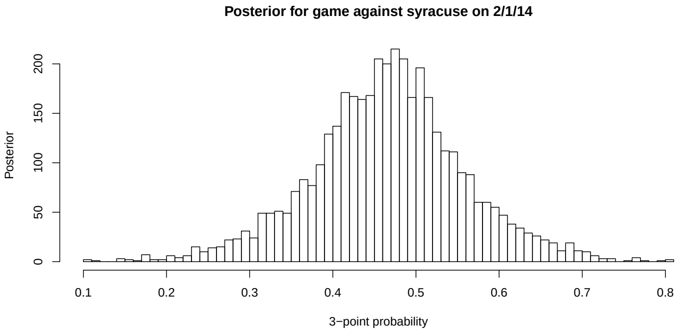
Posterior samples for α and β

```
samples = extract(r3, c("alpha","beta"))
ggpairs(data.frame(log_alpha = log(as.numeric(samples$alpha)), log_beta  = log(as.numeric(samples$beta))),
        lower  = list(continuous='density'))
```

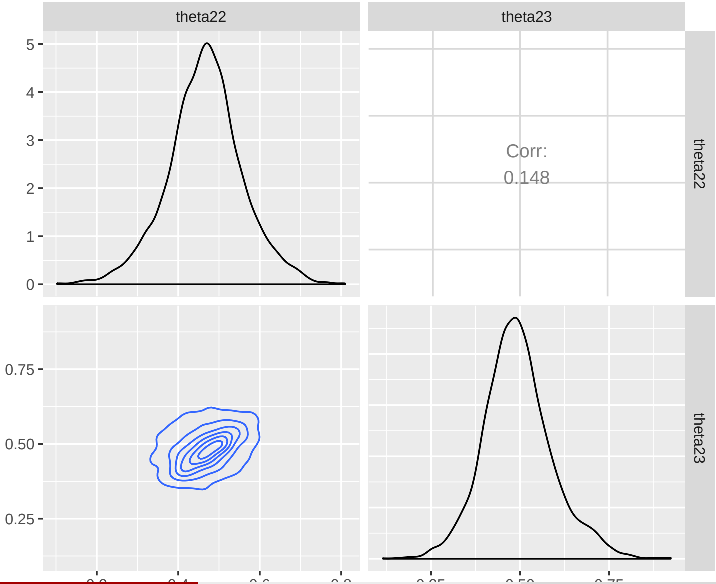


Posterior sample for θ_{22}

```
samples = extract(r3, c("alpha","beta"))
game = 22
theta22 = rbeta(length(samples$alpha),
                samples$alpha + d$made[game],
                samples$beta  + d$attempts[game] - d$made[game])
hist(theta22, 100,
     main=paste("Posterior for game against", d$opponent[game], "on", d$date[game]),
     xlab="3-point probability",
     ylab="Posterior")
```



θs are not independent in the posterior



3-point percentage across seasons

An alternative to modeling game-specific 3-point percentage is to model 3-point percentage in a season. The model is exactly the same, but the data changes.

	season	y	n
1	1	36	95
2	2	64	150
3	3	67	171
4	4	64	152

Due to the low number of seasons (observations), we will use an informative prior for α and β.

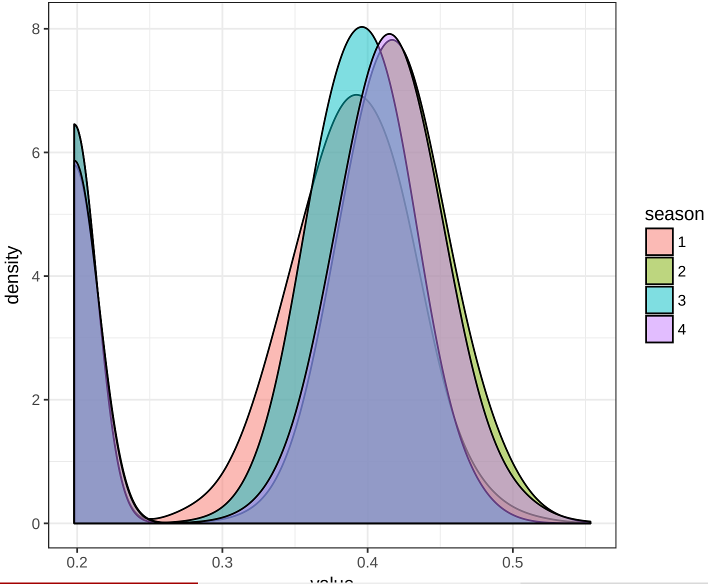
Stan - beta-binomial

```
model_seasons = "
data {
  int<lower=0> N; int<lower=0> n[N]; int<lower=0> y[N];
  real<lower=0> a; real<lower=0> b; real<lower=0> C; real m;
}
parameters {
  real<lower=0,upper=1> mu;
  real<lower=0> eta;
}
transformed parameters {
  real<lower=0> alpha;
  real<lower=0> beta;
  alpha = eta *    mu;
  beta  = eta * (1-mu);
}
model {
  mu  ~ beta(a,b);
  eta ~ lognormal(m,C);
  y   ~ beta_binomial(n,alpha,beta);
}
generated quantities {
  real<lower=0,upper=1> theta[N];
  for (i in 1:N) theta[i] = beta_rng(alpha+y[i], beta+n[i]-y[i]);
}
"

dat  = list(N=nrow(d), y=d$y, n=d$n, a=20, b=30, m=0, C=2)
m4 = stan_model(model_code=model_seasons)
r_seasons = sampling(m4, dat, c("alpha","beta","mu","eta","theta"))
```

Stan - hierarchical model for seasons

```
Inference for Stan model: 84e424f232c33e2cee8bc8e53273d456.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

                 mean      se_mean           sd    2.5%     25%     50%           75%        97.5% n_eff
alpha    1.465388e+17 1.794502e+17 2.538443e+17    2.37   14.59   42.63 1.465388e+17 5.861551e+17     2 2.63285
beta     5.936712e+17 7.270048e+17 1.028397e+18    3.52   21.54   61.73 5.936712e+17 2.374685e+18     2 8.75384
mu       3.500000e-01 6.000000e-02 9.000000e-02    0.20    0.27    0.39 4.200000e-01 4.700000e-01     2 3.28000
eta      7.402099e+17 9.064550e+17 1.282242e+18    6.02   36.04  103.88 7.402099e+17 2.960840e+18     2 1.11403
theta[1] 3.400000e-01 6.000000e-02 9.000000e-02    0.20    0.22    0.37 4.100000e-01 4.600000e-01     2 2.72000
theta[2] 3.600000e-01 7.000000e-02 1.000000e-01    0.20    0.26    0.40 4.300000e-01 4.900000e-01     2 3.54000
theta[3] 3.500000e-01 6.000000e-02 9.000000e-02    0.20    0.26    0.38 4.100000e-01 4.600000e-01     2 3.34000
theta[4] 3.600000e-01 7.000000e-02 1.000000e-01    0.20    0.27    0.40 4.300000e-01 4.800000e-01     2 3.48000
lp__     1.190471e+04 1.509601e+04 2.135433e+04 -425.17 -422.84 -422.02 1.190553e+04 4.888687e+04     2 2.54520

Samples were drawn using NUTS(diag_e) at Thu Feb  9 09:43:04 2017.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

Stan - hierarchical model for seasons



Stan - hierarchical model for seasons

Probabilities that 3-point percentage is greater in season 4 than in the other seasons:

```
theta = extract(r_seasons, "theta")[[1]]
mean(theta[,4] > theta[,1])

[1] 0.651

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

[1] 0.48225

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

[1] 0.62575
```

Summary - hierarchical models

Two-level hierarchical model:

$$y_i \stackrel{ind}{\sim} p(y|\theta) \qquad \theta_i \stackrel{ind}{\sim} p(\theta|\phi) \qquad \phi \sim p(\phi)$$

Conditional independencies:

- $y_i \perp\!\!\!\perp y_j|\theta$ for $i \neq j$
- $\theta_i \perp\!\!\!\perp \theta_j|\phi$ for $i \neq j$
- $y \perp\!\!\!\perp \phi|\theta$
- $y_i \perp\!\!\!\perp y_j|\phi$ for $i \neq j$
- $\theta_i \perp\!\!\!\perp \theta_j|\phi, y$ for $i \neq j$

Summary - extension to more levels

Three-level hierarchical model:

$$y \sim p(y|\theta) \qquad \theta \sim p(\theta|\phi) \qquad \phi \sim p(\phi|\psi) \qquad \psi \sim p(\psi)$$

When deriving posteriors, remember the conditional independence structure, e.g.

$$p(\theta, \phi, \psi|y) \propto p(y|\theta)p(\theta|\phi)p(\phi|\psi)p(\psi)$$