STAT_205_HW3

HONGLEI REN

02/03/2020

Problem 1

Use BetaBuster to find the Beta(a,b) priors for mode 0.75 and 5th percentile 0.60, and for mode 0.01 and 99th percentile 0.02. What is the Beta prior when the mode is 1 and the first percentile is 0.80?

```
S = 100000 #Sample size
dist1 <- epi.betabuster(mode = 0.75, conf = 0.05, greaterthan = F, x = 0.6)
dist1$shape1;dist1$shape2
```

```
## [1] 23.567

## [1] 8.522333
```

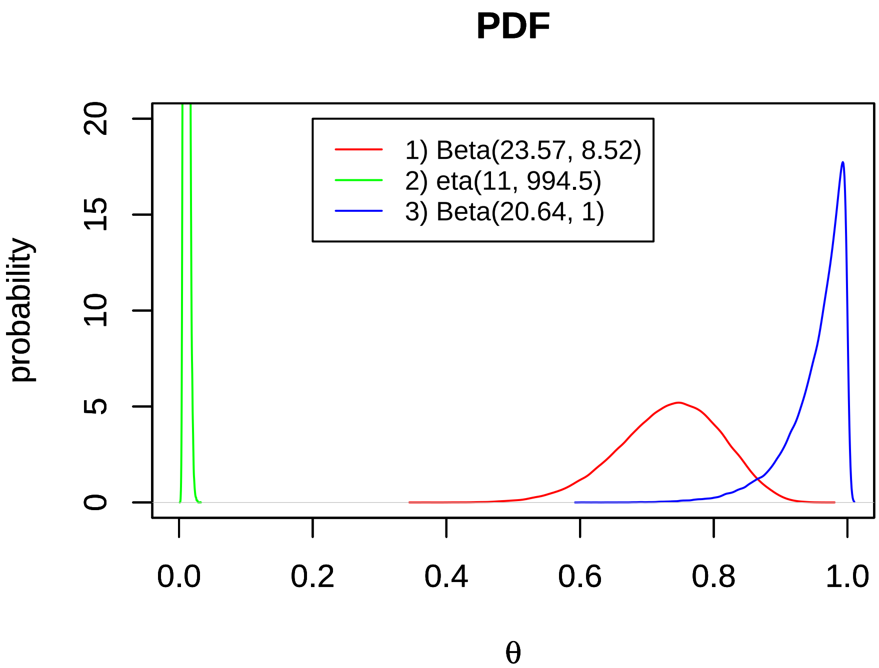
```
s1 <- rbeta(S, dist1$shape1, dist1$shape2)
dist2 <- epi.betabuster(mode = 0.01, conf = 0.99, greaterthan = F, x = 0.02)
dist2$shape1;dist2$shape2
```

```
## [1] 11.035

## [1] 994.465
```

```
s2 <- rbeta(S, dist2$shape1, dist2$shape2)
dist3 <- epi.betabuster(mode = 1, conf = 0.01, greaterthan = F, x = 0.8)
dist3$shape1;dist3$shape2
```

```
## [1] 20.638

## [1] 1
```

```
s3 <- rbeta(S, dist3$shape1, dist3$shape2)

# Plotting
plot(density(s1),col="red", type="l", xlab = expression(theta), xlim=range(c(0, 1)), ylab = "probability
par(new=TRUE)
plot(density(s2),col="green", type="l", xlab = expression(theta), xlim=range(c(0, 1)), ylab = "probabili
par(new=TRUE)
plot(density(s3),col="blue", type="l", xlab = expression(theta), xlim=range(c(0, 1)), ylab = "probabilit
legend(0.2, 20, legend=c("1) Beta(23.57, 8.52)", "2) eta(11, 994.5)", "3) Beta(20.64, 1)"),
       col=c("red", "green", "blue"), lty=1:1, cex=0.8)
```



Problem 2

2.1 Propose a model to conduct a meta-analysis

The model I considered for this study is as follows:

$$Y_i \overset{\text{ind}}{\sim} Binomial(n_i, \theta_i)$$

$$\theta_i \overset{\text{iid}}{\sim} Beta(\alpha, \beta)$$

$$\alpha, \beta \sim p(\alpha, \beta)$$

, which describes the probability of "hit" is from a prior distribution $Beta(\alpha, \beta)$, and it describes the variablity across trails.

2.2 Write model in Jags

```
ESP.data=read.csv("./GanzStudiesUsed-56.csv", header=T)
head(ESP.data)
```

```
##    n hits
## 1 32   14
## 2  7    6
## 3 30   13
## 4 30    7
## 5 20    2
## 6 10    9
```

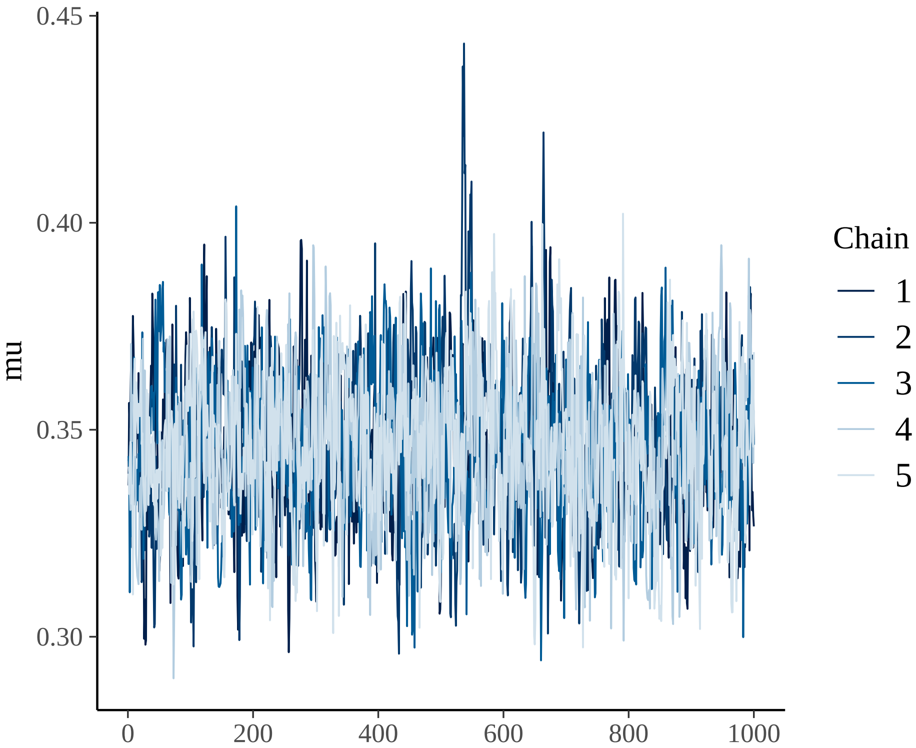
```
jags_model = "model{
  for (i in 1 : N){
  Y[i] ~ dbin(theta[i], n[i])
  theta[i] ~ dbeta(alpha, beta)
  }
```

```
  alpha = eta * mu
  beta = eta * (1-mu)
  eta ~ dlnorm(m, 1/C)
  mu ~ dbeta(a, b)
}"

jags.data = list(Y = ESP.data$hits, n = ESP.data$n, N = dim(ESP.data)[1], a = 20, b = 20, m=0, C=3)
jags.param <- c("theta", "alpha", "beta", "mu", "eta")
jags_fit <- jags(data = jags.data, n.chains = 5, inits = NULL, parameters.to.save = jags.param,n.iter=2(
```

```
## module glm loaded
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 56
##    Unobserved stochastic nodes: 58
##    Total graph size: 180
##
## Initializing model
```

```
jags.mcmc = as.mcmc(jags_fit)
mcmc_trace(jags.mcmc, pars = c("mu"))
```



2.3 Choice of Priors

```
open_mind_prior <- epi.betabuster(mode = 0.25, conf = 0.95, greaterthan = F, x = 0.3)
a1 = open_mind_prior$shape1; b1 = open_mind_prior$shape2;
```

```r
psi_believer_prior <- epi.betabuster(mode = 0.33, conf = 0.95, greaterthan = F, x = 0.36)
a2 = psi_believer_prior$shape1; b2 = psi_believer_prior$shape2;

psi_skeptic_prior <- epi.betabuster(mode = 0.25, conf = 0.95, greaterthan = F, x = 0.255)
a3 = psi_skeptic_prior$shape1; b3 = psi_skeptic_prior$shape2
```

The open-minded prior estimated is: Beta(58.825, 174.475)

The psi believer prior estimated is: Beta(100, 202)

The psi skepticr prior estimated is: Beta(100, 298)

2.4

2.4.1 Posterior Mean is 0.319 and 95% CI is (0.292, 0.348) for Open Mind Prior.

```r
jags.param <- c("mu", "alpha", "beta")
data_open_mind_prior = list(Y = ESP.data$hits, n = ESP.data$n, N = dim(ESP.data)[1], a = a1, b = b1, m=(
```

```r
print(jags_fit_open_mind_prior)
```

```
## Inference for Bugs model at "5", fit using jags,
##  5 chains, each with 2000 iterations (first 1000 discarded)
##  n.sims = 5000 iterations saved
##          mu.vect sd.vect    2.5%     25%     50%     75%   97.5%  Rhat
## alpha      9.532   4.126   4.314   6.707   8.610  11.295  20.685 1.010
## beta      20.304   8.673   9.325  14.376  18.373  24.024  43.313 1.011
## mu         0.319   0.015   0.290   0.309   0.319   0.329   0.348 1.004
## deviance 274.549  11.145 254.857 266.692 274.018 281.563 298.789 1.006
##          n.eff
## alpha      460
## beta       410
## mu         760
## deviance   590
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 61.7 and DIC = 336.3
## DIC is an estimate of expected predictive error (lower deviance is better).
```

2.4.2 Posterior Mean is 0.335 and 95% CI is (0.308, 0.364) for Psi Believer Prior.

```r
data_psi_believer_prior = list(Y = ESP.data$hits, n = ESP.data$n, N = dim(ESP.data)[1], a = a2, b = b2,
```

```r
print(jags_fit_psi_believer_prior)
```

```
## Inference for Bugs model at "6", fit using jags,
##  5 chains, each with 2000 iterations (first 1000 discarded)
##  n.sims = 5000 iterations saved
##          mu.vect sd.vect    2.5%     25%     50%     75%   97.5%  Rhat
## alpha      9.382   3.719   4.572   6.881   8.677  11.005  18.307 1.006
## beta      18.558   7.436   8.988  13.482  17.071  21.822  36.869 1.006
## mu         0.336   0.014   0.309   0.327   0.336   0.346   0.366 1.002
```

```
## deviance 273.225  10.889 253.416 265.785 272.587 280.100 296.485 1.003
##          n.eff
## alpha      820
## beta       870
## mu        1800
## deviance  1200
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 59.1 and DIC = 332.3
## DIC is an estimate of expected predictive error (lower deviance is better).
```

2.4.3 Posterior Mean is 0.310 and 95% CI is (0.282, 0.336) for Psi Skeptic Prior.

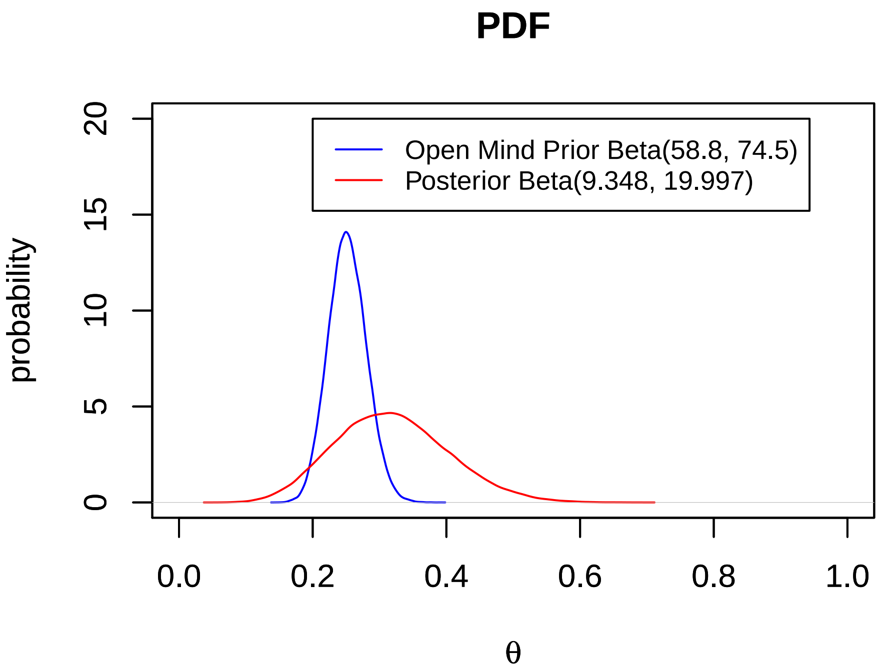
```
data_psi_skeptic_prior = list(Y = ESP.data$hits, n = ESP.data$n, N = dim(ESP.data)[1], a = a3, b = b3, 
```

```
print(jags_fit_psi_skeptic_prior)
```

```
## Inference for Bugs model at "7", fit using jags,
##  5 chains, each with 2000 iterations (first 1000 discarded)
##  n.sims = 5000 iterations saved
##          mu.vect sd.vect    2.5%     25%     50%     75%   97.5%  Rhat
## alpha      9.077   4.283   4.138   6.475   8.237  10.535  18.333 1.015
## beta      20.212   9.480   9.415  14.590  18.333  23.393  40.785 1.016
## mu         0.309   0.014   0.283   0.300   0.310   0.319   0.336 1.002
## deviance 274.773  11.614 253.741 266.758 274.190 281.795 299.988 1.005
##          n.eff
## alpha      520
## beta       480
## mu        2200
## deviance  1100
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 67.2 and DIC = 342.0
## DIC is an estimate of expected predictive error (lower deviance is better).
```

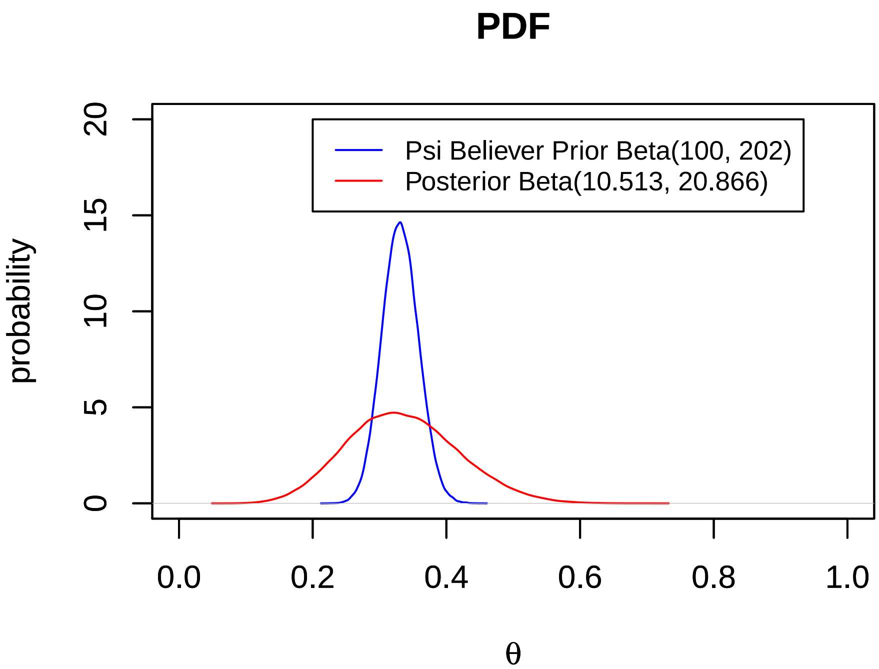
#2.5 Plotting ## 2.5.1 Plot for Open Mind Prior

```
s_open_mind_prior <- rbeta(S, a1, b1)
s_post_open_mind_prior <- rbeta(S, 9.348, 19.997)

plot(density(s_open_mind_prior),col="blue", type="l", xlab = expression(theta), xlim=range(c(0, 1)), yla
par(new=TRUE)
plot(density(s_post_open_mind_prior),col="red", type="l", xlab = expression(theta), xlim=range(c(0, 1))
legend(0.2, 20, legend=c("Open Mind Prior Beta(58.8, 74.5)", "Posterior Beta(9.348, 19.997)"),
       col=c("blue",  "red"), lty=1:1, cex=0.8)
```



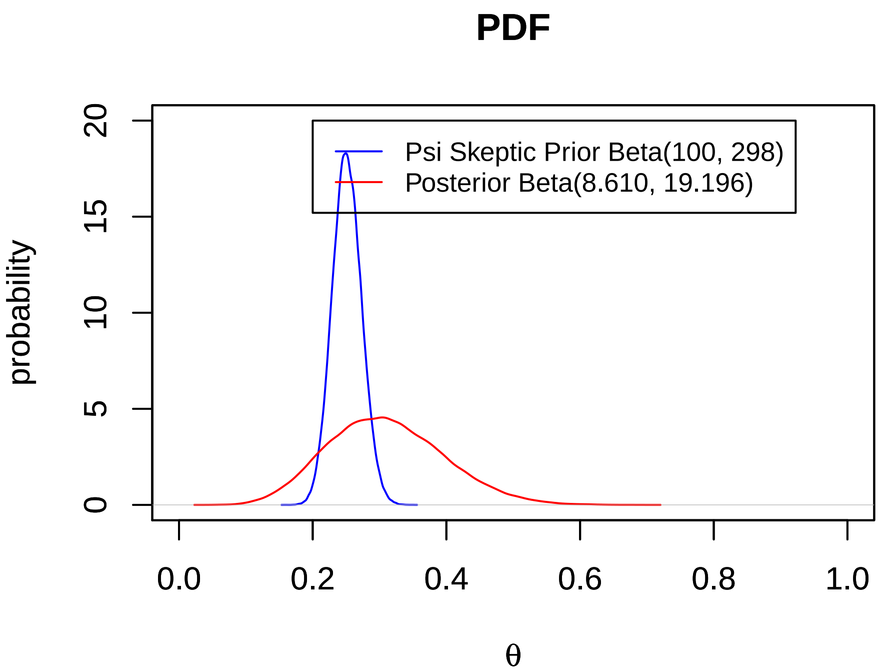
2.5.2 Plot for Psi Believer Prior

```
s_psi_believer_prior <- rbeta(S, a2, b2)
s_post_psi_believer_prior <- rbeta(S, 10.513, 20.866)

plot(density(s_psi_believer_prior),col="blue", type="l", xlab = expression(theta), xlim=range(c(0, 1)),
par(new=TRUE)
plot(density(s_post_psi_believer_prior),col="red", type="l", xlab = expression(theta), xlim=range(c(0, 
legend(0.2, 20, legend=c("Psi Believer Prior Beta(100, 202)", "Posterior Beta(10.513, 20.866)"),
       col=c("blue",  "red"), lty=1:1, cex=0.8)
```



2.5.3 Plot for Psi Skeptic Prior

```r
s_psi_skeptic_prior <- rbeta(S, a3, b3)
s_post_psi_skeptic_prior <- rbeta(S, 8.610, 19.196)

plot(density(s_psi_skeptic_prior),col="blue", type="l", xlab = expression(theta), xlim=range(c(0, 1)), 
par(new=TRUE)
plot(density(s_post_psi_skeptic_prior),col="red", type="l", xlab = expression(theta), xlim=range(c(0, 1
legend(0.2, 20, legend=c("Psi Skeptic Prior Beta(100, 298)", "Posterior Beta(8.610, 19.196)"),
       col=c("blue",  "red"), lty=1:1, cex=0.8)
```



#2.6 Comments Although the data given is the same for three senarios, the posteriors we got are different, they somewhat refects different prior bias.

For Psi Believer Prior which has higher expectation for hit probability, the posterior we got for believer prior also shows the highest posterior mean and 95% posterior credible interval.

For the Psi Skeptic Prior, we have and got the opposite, which shows lowest mean and 95% CI for both pior and posterior distribution.

For the Open Mind Prior, it holds a mild estimation for the hit ability, and the posterior shows a moderate mean and 95% CI.

Problem 4

4.1 Model

The model I assumed is as follows:

$$z_s^{c_s} \overset{\text{ind}}{\sim} Binomial(N_s^{c_s}, \theta_s^{c_s})$$

$$\theta_s^{c_s} \overset{\text{iid}}{\sim} Beta(\alpha^c, \beta^c)$$

$$\mu^c = \frac{\alpha^c}{\alpha^c + \beta^c}, \eta^c = \alpha^c + \beta^c$$

$$\mu^c \sim Beta(a, b), \eta^c \sim logNormal(m, C)$$

where N_s is the number of opportunities at bat, z_s is the number of hits, c_s is the primary position of player s. c is the position indicator.

The upper model means the number of hit for each player follows a Binomial distribution, and has different parameters between each other. The batting average for each player follows a Beta distribution with a position specific parameter. Finally, all position specific parameters come from the same distribution.

4.2 RJags Implementation

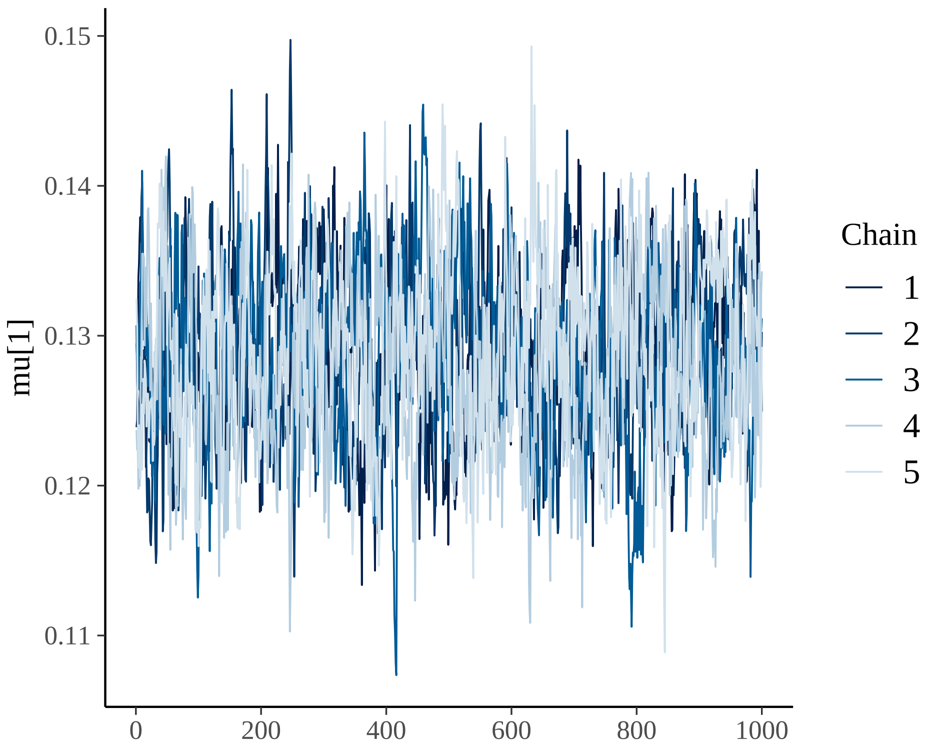
```
BA.data=read.csv("./BattingAverage.csv", header=T)
head(BA.data)
```

```
##           Player     PriPos Hits AtBats PlayerNumber PriPosNumber
## 1 Fernando Abad    Pitcher    1      7            1            1
## 2   Bobby Abreu Left Field   53    219            2            7
## 3    Tony Abreu   2nd Base   18     70            3            4
## 4 Dustin Ackley   2nd Base  137    607            4            4
## 5    Matt Adams   1st Base   21     86            5            3
## 6 Nathan Adcock    Pitcher    0      1            6            1
```

```
jags.data=list(Z = BA.data$Hits, n = BA.data$AtBats, c=BA.data$PriPosNumber,N =nrow(BA.data), m=0, C=3,

jags_model = "model{
  for (i in 1 : N){
    Z[i] ~ dbin(theta[i], n[i])
    theta[i] ~ dbeta(alpha[c[i]], beta[c[i]])
  }
  for(j in 1: 9)
  {
    alpha[j] =eta[j]*mu[j]
    beta[j] = eta[j]*(1-mu[j])
    eta[j] ~ dlnorm(m, 1/C)
    mu[j] ~ dbeta(a, b)
  }
}"

jags.param <- c("theta", "alpha", "beta", "mu", "eta")
jags_fit <- jags(data = jags.data, n.chains = 5, inits = NULL, parameters.to.save = jags.param,n.iter=4
```

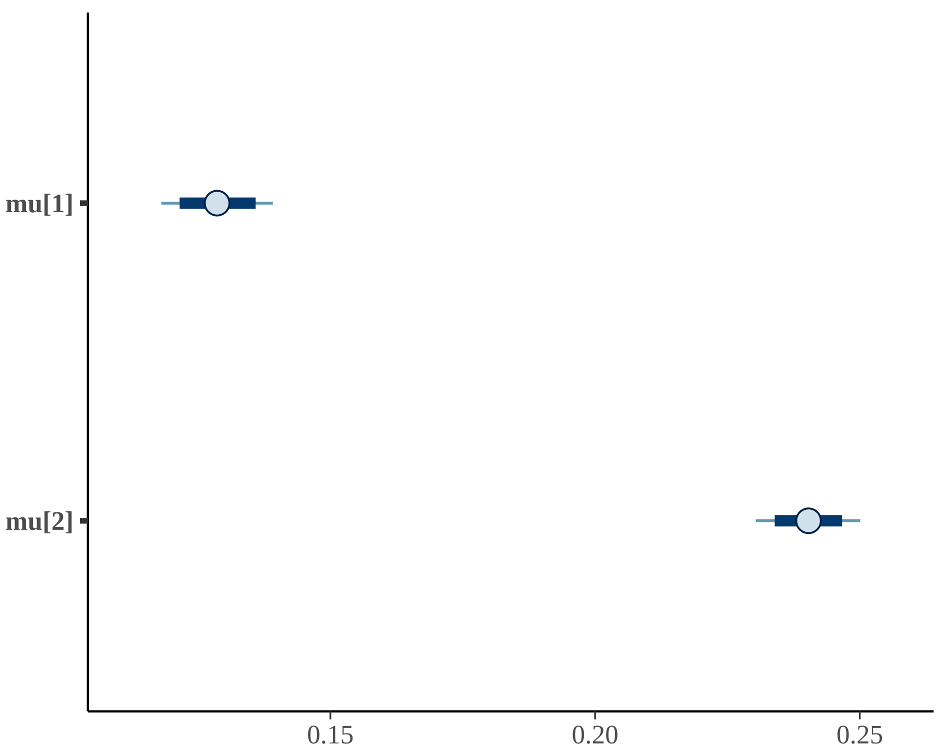
```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 948
##    Unobserved stochastic nodes: 966
##    Total graph size: 3844
##
## Initializing model
```

```
jags.mcmc = as.mcmc(jags_fit)
mcmc_trace(jags.mcmc, pars = c("mu[1]"))
```



4.3 Pitcher vs Catcher

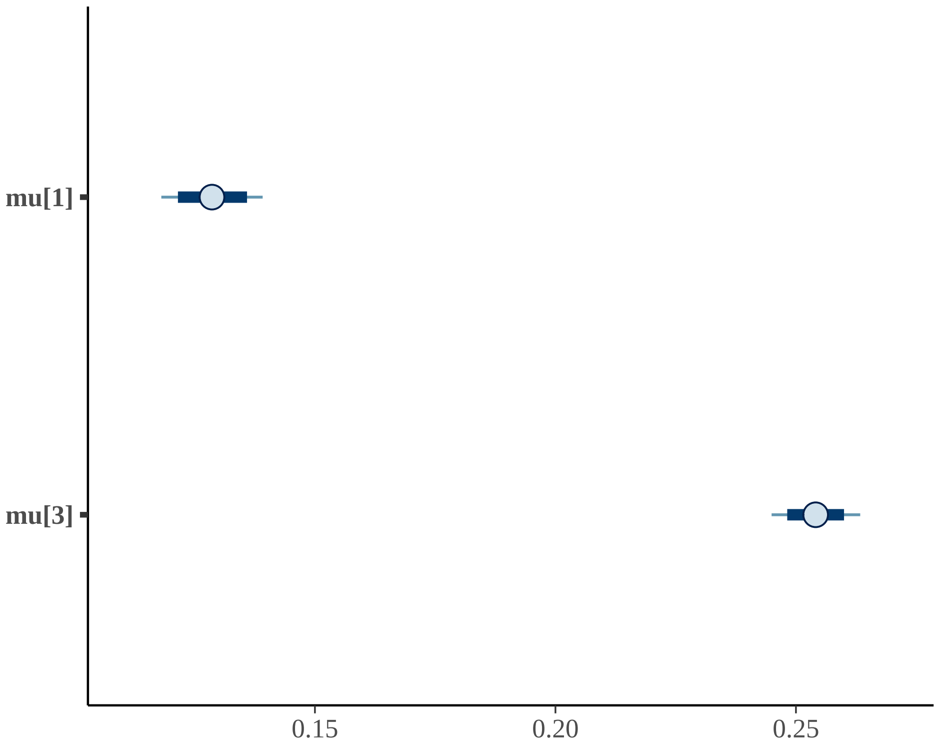
```
mcmc_intervals(jags.mcmc, pars=c("mu[1]", "mu[2]"), prob = 0.8, # 80% intervals - inner
    prob_outer = 0.95, # 95% - outer
    point_est = "mean")
```



The batting average of pitcher and catcher are estimated by mu[1] and mu[2], respectively. From the plot above, we can see catcher has a hihger batting ability (around 0.24) overall, whereas the pitcher only shows around 0.13.

4.4 Pitcher vs First Base Player

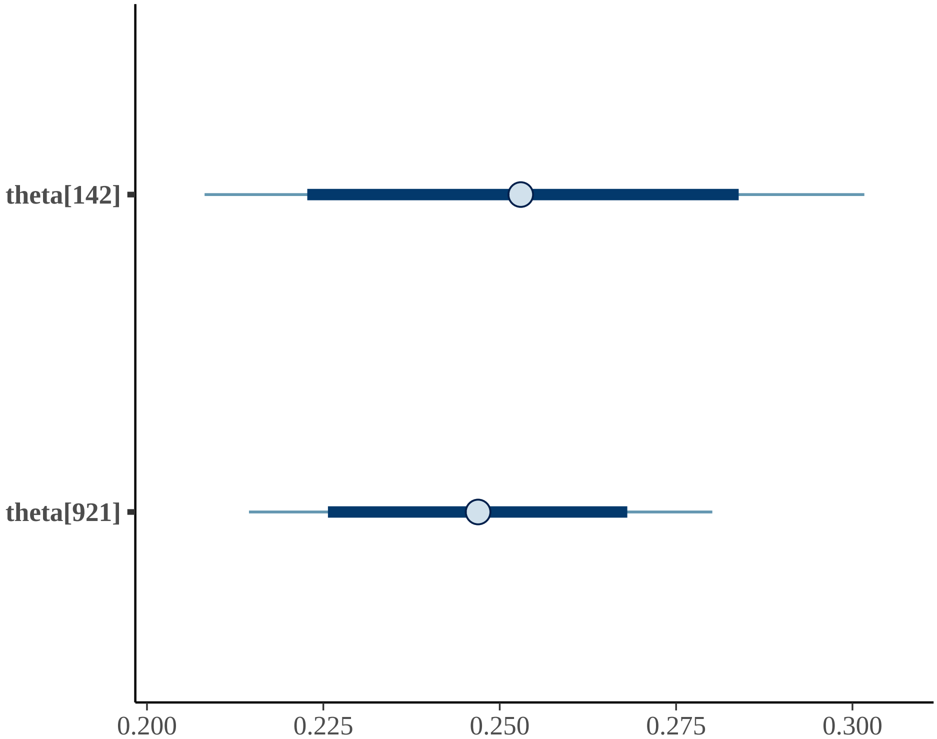
```
mcmc_intervals(jags.mcmc, pars=c("mu[1]", "mu[3]"), prob = 0.8, # 80% intervals - inner
    prob_outer = 0.95, # 95% - outer
    point_est = "mean")
```



The batting average of pitcher and first base player are estimated by mu[1] and mu[3], respectively. From the estimation above, first base player shows stronger batting ability (> 0.25) than pitcher (around 0.13). Also first base player shows more stable batting ability indicated by shorter posterior credible interval.

4.5 Welington Castillo(Catcher) vs Matt Wieters(Catcher)

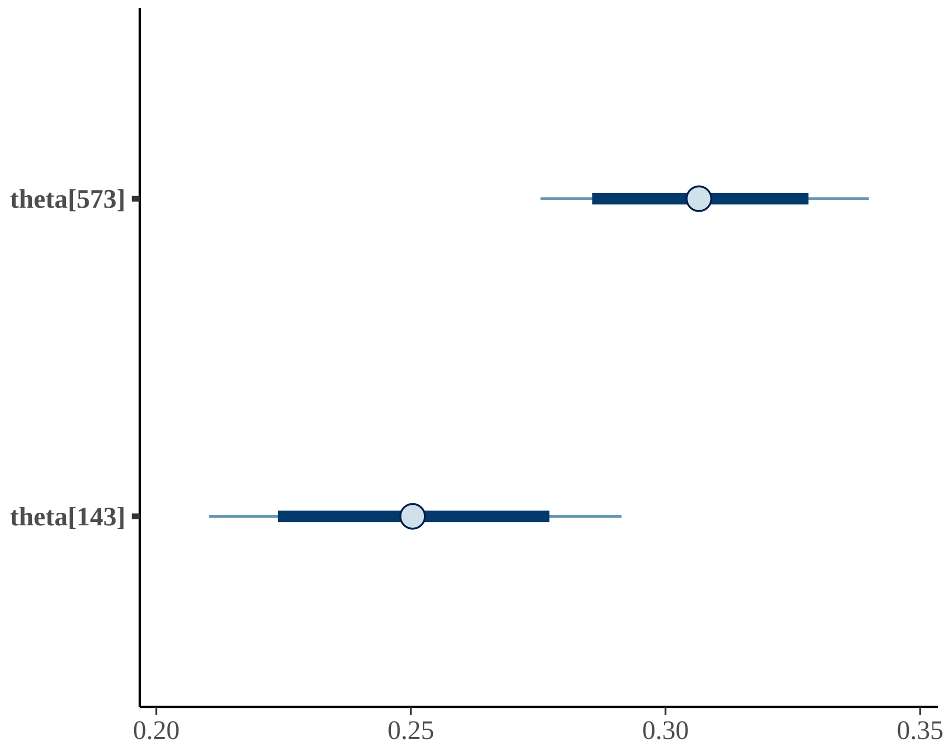
```
mcmc_intervals(jags.mcmc, pars=c("theta[142]", "theta[921]"), prob = 0.8, # 80% intervals - inner
    prob_outer = 0.95, # 95% - outer
    point_est = "mean")
```



The estimated batting averge of Welington Castillo(Catcher) and Matt Wieters(Catcher) are shown as theta[142] and theta[921] in the upper figure. Welington Castillo was estimated as higher batting averge but less stable in performance, which is indicated by larger credible interval. On contarary, although Matt Wieters shows slightly lower batting ability, but have more stable performance, which we can tell from the shrinkage in the estimate.

4.6 Andrew McCutchen(Center Field) vs Jason Castro(Catcher)

```
mcmc_intervals(jags.mcmc, pars=c("theta[573]", "theta[143]"), prob = 0.8, # 80% intervals - inner
    prob_outer = 0.95, # 95% - outer
    point_est = "mean")
```



The estimated batting averge of Andrew McCutchen(Center Field) and Jason Castro(Catcher) are shown as theta[573] and theta[143] in the upper figure. Andrew McCutchen was estimated with higher batting averge(around 0.32) and more stable in performance, since we can see the the shrinkage in the estimate. On contarary, Jason Castro shows much lower batting ability(0.25), and less stable performance.