Hypothesis Testing and Estimation under a Bayesian Approach

L.R. Pericchi* and M.E. Pérez [1]

[1]Department of Mathematics
Universidad de Puerto Rico, Río Piedras Campus
*Co-Leader of Biostatistics, Epidemiology and Bioinformatics (BEBiC) U54 MDA-UPR

Statistics Day in Puerto Rico, October 2015. Be part of ASA-PR!

The Essence of the Bayesian Approach

Observations X are random variables

Are Parameters Θ random variables?

Oxford English Dictionary: "Parameter: a Constant Variable"

For Bayes, parameters are random variables, and then is able to respond the scientific question:

$$Prob(Theory|Data) \text{ T}$$

as oppposed to the Mathematical question:

$$Prob(Data|Theory) \text{ D}$$

P-Values are Type D, but science is about Type T.

The Essence of the Bayesian Approach

Observations X are random variables
Are Parameters Θ random variables?
Oxford English Dictionary: "Parameter: a Constant Variable"
For Bayes, parameters are random variables, and then is able to respond the scientific question:

$$Prob(Theory|Data) \quad \text{T}$$

as oppposed to the Mathematical question:

$$Prob(Data|Theory) \quad \text{D}$$

P-Values are Type D, but science is about Type T.

The Essence of the Bayesian Approach

Observations X are random variables
Are Parameters Θ random variables?
Oxford English Dictionary: "Parameter: a Constant Variable"
For Bayes, parameters are random variables, and then is able to respond the scientific question:

$$Prob(Theory|Data) \quad \text{T}$$

as oppposed to the Mathematical question:

$$Prob(Data|Theory) \quad \text{D}$$

P-Values are Type D, but science is about Type T.

The Essence of the Bayesian Approach

Observations X are random variables
Are Parameters Θ random variables?
Oxford English Dictionary: "Parameter: a Constant Variable"
For Bayes, parameters are random variables, and then is able to respond the scientific question:

$$Prob(Theory|Data) \text{ T}$$

as oppposed to the Mathematical question:

$$Prob(Data|Theory) \text{ D}$$

P-Values are Type D, but science is about Type T.

Parameter Random Variables in Estimation and Testing

Testing:

$$H_0 : \theta = \theta_0 \text{ VS } H_1 : \theta = \theta_1$$

What is the $Prob(H_0|Data)$? So H_0 is a Random Variable!
Estimation: Likelihood: $Prob(X|\theta_1)$
Level I: $Prob(\theta_1|\theta_2)$
Level II: $Prob(\theta_2|\theta_3)$, So θ_1 and θ_2 are random variables.

Parameter Random Variables in Estimation and Testing

Testing:

$$H_0 : \theta = \theta_0 \text{ VS } H_1 : \theta = \theta_1$$

What is the $Prob(H_0|Data)$? So H_0 is a Random Variable!
Estimation: Likelihood: $Prob(X|\theta_1)$
Level I: $Prob(\theta_1|\theta_2)$
Level II: $Prob(\theta_2|\theta_3)$, So θ_1 and θ_2 are random variables.

The Evidence, The Bayes Factor, The Posterior Probability

$$H_i : X \text{ has density } f_i(x|\theta_i), i = 0, \ldots, I.$$

The Evidence: The marginal Likelihood

$$m_i(x) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i.$$

The Bayes Factor of M_i to M_j:

$$B_{ji} = \frac{m_j(x)}{m_i(x)}$$

Posterior Model Probabilities:

$$P(M_i|x) = \frac{P(M_i)m_i(x)}{\sum_{j=0}^{q} P(M_j)m_j(x)}$$

Testing: Bayes VS Non-Bayes: The difference is NOT about Mathematics

Neyman-Pearson Lemma: Optimal Test To Minimize $a * TypeIError + b * TypeIIerror$ is

$$\text{Reject} H_0 : \text{if: LikelihoodRatio}_{0,1} < b/a$$

The problem is how to choose b/a.

$$Pval = Prob(LikelihoodRatio_{0,1} < ObservedLikRatio)$$

b/a assigned indirectly.

$$\frac{\text{Posterior Probabilities}}{\text{Prior Probabilities}} = \text{Bayes Factor} = LikRatio_{0,1} < r$$

$b/a = r$, say $r = 1/20$ assigned directly, so BOTH type I error and Type II error go to zero as the sample size grows.
Pericchi and Pereira (2015) Brazilian Jour of Prob and Statistics, for generalizations.

Testing: Bayes VS Non-Bayes: The difference is NOT about Mathematics

Neyman-Pearson Lemma: Optimal Test To Minimize $a * TypeIError + b * TypeIIerror$ is

$$\text{Reject} H_0 : \text{if: LikelihoodRatio}_{0,1} < b/a$$

The problem is how to choose b/a.

$$Pval = Prob(LikelihoodRatio_{0,1} < ObservedLikRatio)$$

b/a assigned indirectly.

$$\frac{\text{Posterior Probabilities}}{\text{Prior Probabilities}} = \text{Bayes Factor} = LikRatio_{0,1} < r$$

$b/a = r$, say $r = 1/20$ assigned directly, so BOTH type I error and Type II error go to zero as the sample size grows.
Pericchi and Pereira (2015) Brazilian Jour of Prob and Statistics, for generalizations.

The crisis of P-Values: Non Reproducible Findings

For many years there has been an important discussion on the validity of methods for Null Hypothesis Significance Testing (NHST).

As a worrying consequence of this controversy, statistical inference methods are losing the trust of sectors of the scientific community, as it is reflected by the recent editorial of *Basic and Applied Social Psychology* (Trafimow and Marks, 2015) banning the use of procedures as *p*-values, confidence intervals and related methods from the papers published in BASP.
As the editors remark, "In the NHSTP, the problem is in traversing the distance from the probability of the finding, given the null hypothesis, to the probability of the null hypothesis, given the finding". Increasingly large sections of the scientific community are speaking load and clear: p-values should no longer be the deciding balance of science.

The crisis of P-Values: Non Reproducible Findings

For many years there has been an important discussion on the validity of methods for Null Hypothesis Significance Testing (NHST).

As a worrying consequence of this controversy, statistical inference methods are losing the trust of sectors of the scientific community, as it is reflected by the recent editorial of *Basic and Applied Social Psychology* (Trafimow and Marks, 2015) banning the use of procedures as *p*-values, confidence intervals and related methods from the papers published in BASP.
As the editors remark, "In the NHSTP, the problem is in traversing the distance from the probability of the finding, given the null hypothesis, to the probability of the null hypothesis, given the finding". Increasingly large sections of the scientific community are speaking load and clear: p-values should no longer be the deciding balance of science.

How to convert P-Values into Bayes Factors to try to reduce non-reproducible findings? Calibrating the "Robust Lower Bound".

In Sellke, Bayarri and Berger (2001) (Infimum over Unimodal and Symmetric Priors), a lower bound is proposed for calibrating p-values when $p_{val} < e^{-1}$,

$$B_{01} \geq B_L(p_{val}) = -e p_{val} \log_e(p_{val})$$

It is very simple and it can be easily calculated, but becomes less informative when n increases.

Can we find a way of using the lower bound in the calibration of C_α for the adaptive α levels?
Can we calibrate this lower bound to make it closer to the actual value of a Bayes factor?

How to convert P-Values into Bayes Factors to try to reduce non-reproducible findings? Calibrating the "Robust Lower Bound".

In Sellke, Bayarri and Berger (2001) (Infimum over Unimodal and Symmetric Priors), a lower bound is proposed for calibrating p-values when $p_{val} < e^{-1}$,

$$B_{01} \geq B_L(p_{val}) = -e p_{val} \log_e(p_{val})$$

It is very simple and it can be easily calculated, but becomes less informative when n increases.

Can we find a way of using the lower bound in the calibration of C_α for the adaptive α levels?
Can we calibrate this lower bound to make it closer to the actual value of a Bayes factor?

How to convert P-Values into Bayes Factors to try to reduce non-reproducible findings? Calibrating the "Robust Lower Bound".

In Sellke, Bayarri and Berger (2001) (Infimum over Unimodal and Symmetric Priors), a lower bound is proposed for calibrating p-values when $p_{val} < e^{-1}$,

$$B_{01} \geq B_L(p_{val}) = -e p_{val} \log_e(p_{val})$$

It is very simple and it can be easily calculated, but becomes less informative when n increases.

Can we find a way of using the lower bound in the calibration of C_α for the adaptive α levels?
Can we calibrate this lower bound to make it closer to the actual value of a Bayes factor?

Motivation

Lets go back to the approximation

$$
\begin{aligned}
-2log(B_{01}) &= -2\log(\frac{f_0(\mathbf{x}|\hat{\theta}_0)}{f_1(\mathbf{x}|\hat{\theta}_1)}) - q\log(n^*) + C^* \\
&\approx \chi^2_\alpha(q) - q\log(n^*) + C^*
\end{aligned}
$$

Idea: Assuming α fixed, select C^* such that our approximation for B_{01} equals $B_L(\alpha)$ for fixed (typically low) value of n^* (say n_L).

A first obvious selection:

$$C^* = -2logB_L(\alpha) + q\log(n_L) + \chi^2_\alpha(q)$$

where $B_L(\alpha) = -e\alpha\log\alpha$.

Motivation

Lets go back to the approximation

$$
\begin{aligned}
-2log(B_{01}) &= -2\log(\frac{f_0(\mathbf{x}|\hat{\theta}_0)}{f_1(\mathbf{x}|\hat{\theta}_1)}) - q\log(n^*) + C^* \\
&\approx \chi^2_\alpha(q) - q\log(n^*) + C^*
\end{aligned}
$$

Idea: Assuming α fixed, select C^* such that our approximation for B_{01} equals $B_L(\alpha)$ for fixed (typically low) value of n^* (say n_L).

A first obvious selection:

$$
C^* = -2logB_L(\alpha) + q\log(n_L) + \chi^2_\alpha(q)
$$

where $B_L(\alpha) = -e\alpha\log\alpha$.

This leads to

$$B_{01} \approx B_L(\alpha) \left(\frac{n^*}{n_L} \right)^{\frac{q}{2}}$$

First reaction:

Too good to be true!

Lets compare the behavior of this approximation with the behavior of Bayes factors based on proper objective priors (like intrinsic priors and Berger's robust priors) for several examples.

Two initial proposals for n_L are $n_L = m + 1$ and $n_L = 2m$, where m is the minimal training sample size.

This leads to

$$B_{01} \approx B_L(\alpha) \left(\frac{n^*}{n_L} \right)^{\frac{q}{2}}$$

First reaction:

Too good to be true!

Lets compare the behavior of this approximation with the behavior of Bayes factors based on proper objective priors (like intrinsic priors and Berger's robust priors) for several examples.

Two initial proposals for n_L are $n_L = m + 1$ and $n_L = 2m$, where m is the minimal training sample size.

This leads to

$$B_{01} \approx B_L(\alpha) \left(\frac{n^*}{n_L} \right)^{\frac{q}{2}}$$

First reaction:

Too good to be true!

Lets compare the behavior of this approximation with the behavior of Bayes factors based on proper objective priors (like intrinsic priors and Berger's robust priors) for several examples.

Two initial proposals for n_L are $n_L = m + 1$ and $n_L = 2m$, where m is the minimal training sample size.

This leads to

$$B_{01} \approx B_L(\alpha) \left(\frac{n^*}{n_L} \right)^{\frac{q}{2}}$$

First reaction:

Too good to be true!

Lets compare the behavior of this approximation with the behavior of Bayes factors based on proper objective priors (like intrinsic priors and Berger's robust priors) for several examples.

Two initial proposals for n_L are $n_L = m + 1$ and $n_L = 2m$, where m is the minimal training sample size.

Example 1: Normal distribution, σ known

(Berger, J.O.and Pericchi L.R 2015. Bayes Factors. Encyclopedia of Statistical Sciences.)

$X_1, X_2, \ldots, X_n$ i.i.d sample from $N(\theta, \sigma^2)$, σ^2 known.

It is desired to test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

Assume a prior for θ that is $N(\theta_0, \tau^2)$. It is also usual to select $\tau^2 = k\sigma^2$. In particular, $k = 2$ corresponds to the intrinsic prior for θ

The Bayes factor obtained using the intrinsic prior is

$$B_{01} = \sqrt{1 + 2n} \exp\left(\frac{-z^2}{2 + 1/n}\right)$$

where $z = \sqrt{n}(\bar{x} - \theta_0)/\sigma$

Example 1: Normal distribution, σ known

(Berger, J.O.and Pericchi L.R 2015. Bayes Factors. Encyclopedia of Statistical Sciences.)

$X_1, X_2, \ldots, X_n$ i.i.d sample from $N(\theta, \sigma^2)$, σ^2 known.

It is desired to test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

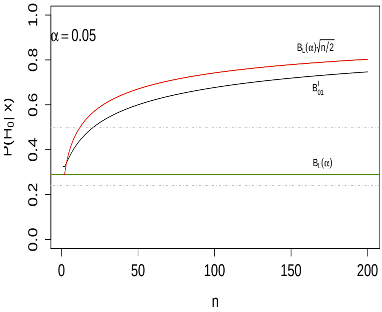
Assume a prior for θ that is $N(\theta_0, \tau^2)$. It is also usual to select $\tau^2 = k\sigma^2$. In particular, $k = 2$ corresponds to the intrinsic prior for θ

The Bayes factor obtained using the intrinsic prior is

$$B_{01} = \sqrt{1 + 2n} \exp\left(\frac{-z^2}{2 + 1/n}\right)$$

where $z = \sqrt{n}(\bar{x} - \theta_0)/\sigma$

Fixed α, n varying.



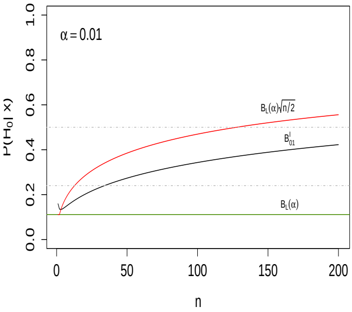


Table of α to Posterior Probabiities of H_0, $N_L = 4$

N	α	0.1	0.05	0.01	0.005	0.001	0.0005
4		0.38	0.29	0.1	0.07	0.02	0.01
20		0.58	0.48	0.22	0.14	0.04	0.02
50		0.69	0.59	0.31	0.20	0.06	0.03

Jeffreys Table of Evidence:
$P(H_0) > 0.5, H_0$ Supported, $0.5 > P(H_0) > 0.25$ Mild Evidence, $0.25 > P(H_0) > 0.1$ Substantial, $0.1 > P(H_0) > 0.03$ Strong $0.03 > P(H_0) > 0.01$ very strong, $0.01 > P(H_0)$ Decisive.

Part II
Bayesian estimation in practice: Using MCMC software.

Bayesian estimation (as hypothesis testing) is based on the posterior distribution.

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}$$

where $p(y) = \int p(\theta)p(y|\theta)d\theta$ (continuous case)

The calculation of the integral in the denominator can be very difficult (or even impossible analytically). This is specially true in the high dimension case.

Markov Chain Monte Carlo methods

Instead of solving the integral(s), the usual approach is using *Markov Chain Monte Carlo* methods to obtain samples from the posterior distribution. Here “Monte Carlo” implies random sampling, while “Markov Chains” refer to the method of simulation: iterative methods in which each iterate depends only on the previous one. MCMC algorithms are built to guarantee that the stationary distribution of the chain is the desired posterior distribution.

Correlation, convergence and "burn in"

As every sample depends on the previous one, contiguous samples from the Markov Chain can be correlated. This fact have some consequences:

- Selected initial values can impact on the simulation chain until a large number of samples has been obtained. For this reason, a certain number of initial observations is discarded. ("*burn in period*").
- Because of correlation, each observation gives only a fraction of the information that would be obtained when using non correlated iterates. so, if the correlation is high a large number of samples will be needed to obtain precise results.

Convergence can be checked in several ways:

- Plot samples vs. iteration number. The behavior should appear random.
- Run several chains using different initial values (they should all converge to the same values)
- Formal convergence testing methods.

Some MCMC software

In many cases, existing software for MCMC methods can be used (some complex cases require the researcher to code his/hers own algorithms)

- **BUGS** (Bayes using Gibbs Sampler) Development of BUGS began in 1989. Currently, the most used "flavors" are *WinBUGS* (version 1.4.3, running over Windows) and *OpenBUGS* (runing natively on Windows and Linux).
- **JAGS** (Just another Gibbs Sampler) Developed independently, it runs natively on Windows, Mac, Linux and several other varieties of Unix. It uses essentially the same model description language than BUGS.
- **STAN** A more recent option, uses a similar model description language but is conceptually different.

The examples shown in this talk use WinBUGS.

Example: Efron and Morris Baseball Data

Efron and Morris (1975, 1977) obtained a sample of batting averages for 18 baseball player during the 1970 season. They used the average obtained during the first 45 at-bats for predicting the batting average for the rest of the season for each player.
The **Direct Evidence Estimator** is the individual MLE (inadmissable and bad here), the **Indirect Evidence Estimator** is the overall sample mean M=0.2654 (amazingly good here).

Player	Batting average for first 45 at bats	Batting average for remainder of season	At bats for remainder of season
Clemente (Pitts, NL)	0.400	0.346	367
F. Robinson (Balt, AL)	0.378	0.298	426
F. Howard (Wash,AL)	0.356	0.276	521
Johnstone (Cal, AL)	0.333	0.222	275
Berry (Chi, AL)	0.311	0.273	418
Spencer (Cal, AL)	0.311	0.270	466
Kessinger (Chi, NL)	0.289	0.263	586
Alvarado (Bos, AL)	0.267	0.210	138
Santo (Chi, NL)	0.244	0.269	510
Swoboda (NY, NL)	0.244	0.230	200
Unser (Wash, AL)	0.222	0.264	277
Williams (Chi, AL)	0.222	0.256	270
Scott (Bos, AL)	0.222	0.303	435
Petrocelli (Bos, AL)	0.222	0.264	538
E. Rodriguez (KC, AL)	0.222	0.226	186
Campaneris (Oak, AL)	0.200	0.285	558
Munson (NY, AL)	0.178	0.316	408
Alvis (Mil, NL)	0.156	0.200	70

Table: Original data: 1970 batting averages for 18 MBL players. Overall MEAN M=0.2654

How to Combine Direct and Indirect Evidence?

Efron and Morris assumption about the data is:

$$Y_i \sim \frac{1}{45}\text{Bin}(45, p_i)$$

where Y_i is the batting average for the first 45 at-bats, and p_i depends on each player's ability.

The batting average for the rest of the season, R_i can be modelled as

$$R_i \sim \frac{1}{n_i}\text{Bin}(n_i, p_i)$$

where n_i is the number of at bats for player i during the remainder of the season.

They applied a variance stabilizing transformation to Y_i,

$$X_i = \sqrt{45}\,\arcsin\left(2Y_i - 1\right)$$

In the sequel, we will use the transformed variable.

Model 1: Empirical Bayes analysis using conjugate model

This analysis is equivalent to Efron and Morris (1975)

$$
\begin{aligned}
X_i &\sim \text{Normal}(\mu_i, 1), i = 1, \ldots, 18 \\
\mu_i &\sim \text{Normal}(\text{M}, \sigma^2)
\end{aligned}
$$

Here $\text{M} = \bar{X} = -3.3166$ and σ^2 such that $\frac{1}{(1+\sigma^2)} = \frac{k-3}{\sum_{i=1}^{k}(X_i - \bar{X})^2}$,
so $\tau = (\sigma^2)^{-1} = 3.7853$.
The calculations for this model can be made in closed form.

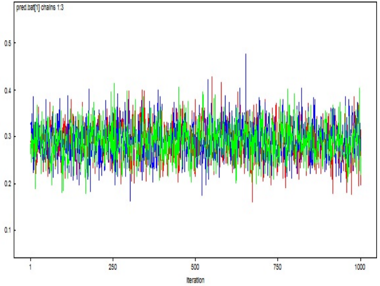
WinBUGS model:

```
model
{
     for (i in 1: nplayers)
     {
        #
        # Likelihood for X[i]= sqrt(45)*arcsin(2Y[i]-1)
        #
        X[i]~ dnorm(mu[i], 1)
        mu[i]~ dnorm(Mu,tau)
        pbat[i]<-0.5*(sin(mu[i]/sqrt(45.))+1)
        }
        #
        # Predicted average for the rest of the season
        #
        for(i in 1:nplayers)
        {
            theta[i]<-mu[i]/sqrt(45)
            R[i]~ dnorm(theta[i],atbat[i])
            pred.bat[i]<-0.5*(sin(R[i])+1)
        }
}
```

Data for the model

```
list(nplayers=18,X=c(-1.35074999608559, -1.65349439900760, -1.95971921115431,
-2.28443930236967, -2.60033503716442, -2.60033503716442, -2.92243068710524,
-3.25189908299015, -3.60573680589079, -3.60573680589079, -3.95492619729744,
-3.95492619729744, -3.95492619729744, -3.95492619729744, -3.95492619729744,
-4.31673666857481, -4.69383371133901, -5.08971235251556) ,
atbat=c(367, 426, 521, 275, 418, 466, 586, 138, 510, 200, 277, 270,435, 538,
186, 558, 408, 70),
tau=3.78527897894347, Mu=-3.31656313614355)
```

Some results for model 1



Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
pred.bat[1]	0.29	0.03829	3.336E-4	0.2177	0.2894	0.3674	1000	12003
pred.bat[2]	0.2862	0.03779	3.792E-4	0.2151	0.2857	0.3624	1000	12003
pred.bat[3]	0.2822	0.03684	2.949E-4	0.2118	0.2819	0.3566	1000	12003
pred.bat[4]	0.2774	0.04028	3.407E-4	0.2019	0.2771	0.3577	1000	12003
pred.bat[5]	0.2734	0.03685	3.181E-4	0.2036	0.2726	0.3472	1000	12003
pred.bat[6]	0.2737	0.03676	3.188E-4	0.2038	0.2727	0.3466	1000	12003
pred.bat[7]	0.2689	0.03534	3.363E-4	0.2023	0.268	0.3411	1000	12003
pred.bat[8]	0.2654	0.04765	3.968E-4	0.1762	0.2637	0.3628	1000	12003
pred.bat[9]	0.2597	0.03503	3.092E-4	0.194	0.2588	0.3312	1000	12003
pred.bat[10]	0.2595	0.04258	3.829E-4	0.1793	0.2586	0.3467	1000	12003
pred.bat[11]	0.2545	0.03913	3.642E-4	0.1805	0.2531	0.3333	1000	12003
pred.bat[12]	0.255	0.03995	3.806E-4	0.1805	0.2542	0.3356	1000	12003
pred.bat[13]	0.2548	0.03608	3.366E-4	0.1861	0.2537	0.3271	1000	12003
pred.bat[14]	0.2553	0.03486	3.192E-4	0.1889	0.255	0.3243	1000	12003
pred.bat[15]	0.2557	0.04397	3.795E-4	0.1737	0.2543	0.3471	1000	12003
pred.bat[16]	0.2499	0.03444	3.444E-4	0.1843	0.2491	0.3199	1000	12003
pred.bat[17]	0.2448	0.03576	3.137E-4	0.1774	0.2439	0.3182	1000	12003
pred.bat[18]	0.2415	0.05781	5.395E-4	0.1356	0.2393	0.3615	1000	12003

Model 2: Full Bayes hierarchical analysis with high tail prior for the precision

Instead of assigning fixed values to M and σ^2, we will assign *hyperpriors* to them. In this example, we assign a "vague" normal (with large variance) to M and a high tail distribution to σ^2 (a Beta2 with parameters (1,1), which has polynomial tails)

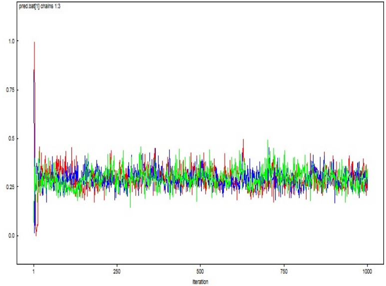
$$
\begin{aligned}
X_i &\sim \text{Normal}(\mu_i, 1), i = 1, \ldots, 18 \\
\mu_i &\sim \text{Normal}(\text{M}, \sigma^2) \\
M &\sim N(0, 10^5) \\
\sigma^2 &\sim \text{Beta2}(1, 1)
\end{aligned}
$$

This model cannot be calculated in closed form!

WinBUGS code:

```
model
{
     for (i in 1: nplayers)
     {
        # Likelihood for X[i]= arcsin(2Y[i]-1)
        X[i]~dnorm(mu[i], 1)
        mu[i]~dnorm(Mu,tau)
        pbat[i]<-0.5*(sin(mu[i]/sqrt(45.))+1)
        }
        # Prior for the common mean
        Mu ~ dnorm(0,0.00001)
        # Prior for the precision tau
        tau<- 1/sigma2
        P ~ dbeta(1,1)
        sigma2 <- P/(1-P)
        # Predicted average for the rest of the season
        for(i in 1:nplayers)
        {
            theta[i]<-mu[i]/sqrt(45)
            R[i]~dnorm(theta[i],atbat[i])
            pred.bat[i]<-0.5*(sin(R[i])+1)
        }
}
```

Data can be written in a similar way



Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
pred.bat[1]	0.2959	0.04835	9.851E-4	0.2117	0.2915	0.4039	1000	12003
pred.bat[2]	0.2907	0.04603	8.956E-4	0.2108	0.2869	0.394	1000	12003
pred.bat[3]	0.2857	0.04341	7.448E-4	0.2112	0.2823	0.3838	1000	12003
pred.bat[4]	0.2799	0.04615	7.104E-4	0.1952	0.2771	0.3801	1000	12003
pred.bat[5]	0.2755	0.04193	5.761E-4	0.1983	0.2735	0.3645	1000	12003
pred.bat[6]	0.2748	0.04154	5.855E-4	0.1995	0.2725	0.3632	1000	12003
pred.bat[7]	0.2695	0.03955	5.214E-4	0.1956	0.268	0.3542	1000	12003
pred.bat[8]	0.2647	0.05088	6.649E-4	0.1681	0.2632	0.3685	1000	12003
pred.bat[9]	0.2579	0.03992	5.991E-4	0.1805	0.2573	0.3393	1000	12003
pred.bat[10]	0.2589	0.04629	5.797E-4	0.1717	0.2575	0.355	1000	12003
pred.bat[11]	0.2533	0.0437	6.162E-4	0.1675	0.2529	0.3398	1000	12003
pred.bat[12]	0.2531	0.04337	5.806E-4	0.1687	0.2528	0.3409	1000	12003
pred.bat[13]	0.2529	0.04005	5.569E-4	0.1735	0.2531	0.333	1000	12003
pred.bat[14]	0.2535	0.03992	6.283E-4	0.1741	0.2534	0.3327	1000	12003
pred.bat[15]	0.2533	0.04727	6.328E-4	0.1615	0.2524	0.3469	1000	12003
pred.bat[16]	0.2475	0.03949	6.825E-4	0.1655	0.2486	0.3229	1000	12003
pred.bat[17]	0.2414	0.0417	7.459E-4	0.1568	0.2424	0.322	1000	12003
pred.bat[18]	0.2379	0.06213	9.578E-4	0.1211	0.2363	0.3629	1000	12003

In this graph we can see an initial oscilation for the chains, but convergence is very fast anyway.
The results for this model show less "shrinkage" towards the general mean.