Assgn3: Students' goals

Salvador Garcia, s1655274

10 February 2016

1 Description of the problem

229 students aged 7-13 from 9 schools were asked whether popularity or sporting ability was most important to them. The aim is to create an appropriate model that takes into account variability between the schools (if necessary). **Statistical analysis.** Introduce the following random variables. Consider the ith student, $i = 1, 2, ..., N = 229$, and set $Zi = 1$ if popularity is more important to the student than sporting ability, and $Zi = 0$ otherwise.

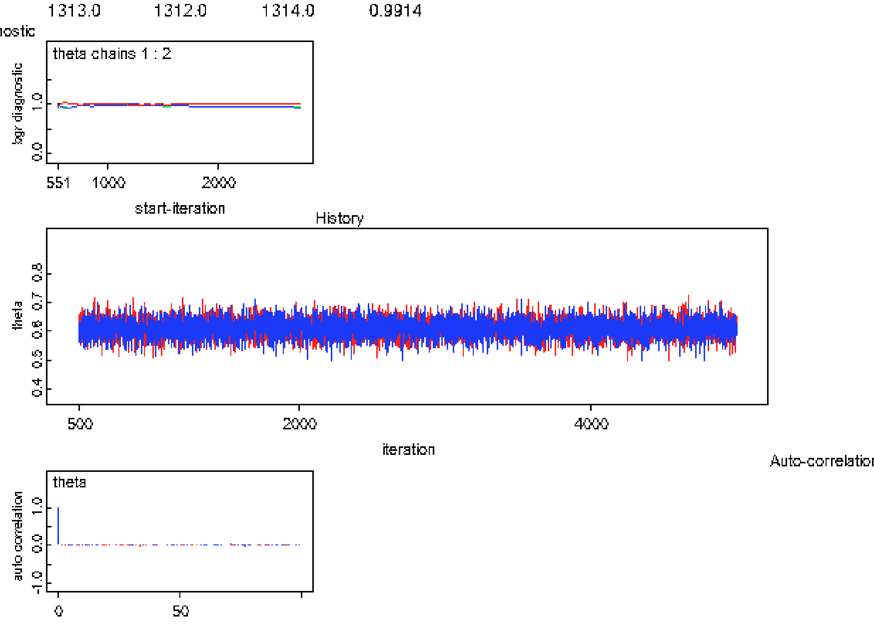
2 Likelihood

For this project, openbugs was used $Z_i|\theta_k \sim Bern(\theta_k)$, where k is the number of the school the ith student is from, independently given θ_k, $\theta_k \in (0, 1)$.

2.1 Prior case 1: $\theta_1 = \theta_2 = ... = \theta_9$

In this case, each parameter corresponding to the 9 schools are assumed identical, lets call this parameter θ and its non-informative prior distribution follows a $beta(1, 1)$. The MCMC was run with two chains, the first with initial point 0.1 and the second 0.9. As the autocorrelation graph shows, it is not necessary to change the thin parameter. Also, the chain converges relatively fast (Just 501 iterations were used as burning). Finally, the Gelman-Rubin gives us a good conclusion about the convergence.

	Dbar	Dhat	DIC	pD
Goals	307.0	306.0	308.0	0.9914
School	1006.0	1006.0	1006.0	-1.22E-10
total	1313.0	1312.0	1314.0	0.9914



Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
theta	0.6099	0.03199	3.415E-4	0.5455	0.6102	0.6717	501	9002

Figure 1: DIC, BGR diagnosis, history and autocorrelation of θ

2.2 Prior case 2: with independent non-informative priors

Now, we have 9 parameters $(\theta_1, \theta_2, ..., \theta_9)$, each one corresponding to different schools. The prior of each one of these parameters is assumed to be distributed $beta(1,1)$ (non-informative). The MCMC was run with two chains, the first with initial value $\theta_i = 0.1$ for all i and the second $\theta_i = 0.9$ for all i. As the autocorrelation graph shows, it was not necessary to change the thin parameter. Again, the chain converge relatively fast (just 1001 iterations were used as burning). Finally, the Gelman-Rubin gives a good conclusion about the convergence of the chains (although for θ_2 there is a very small gap.)

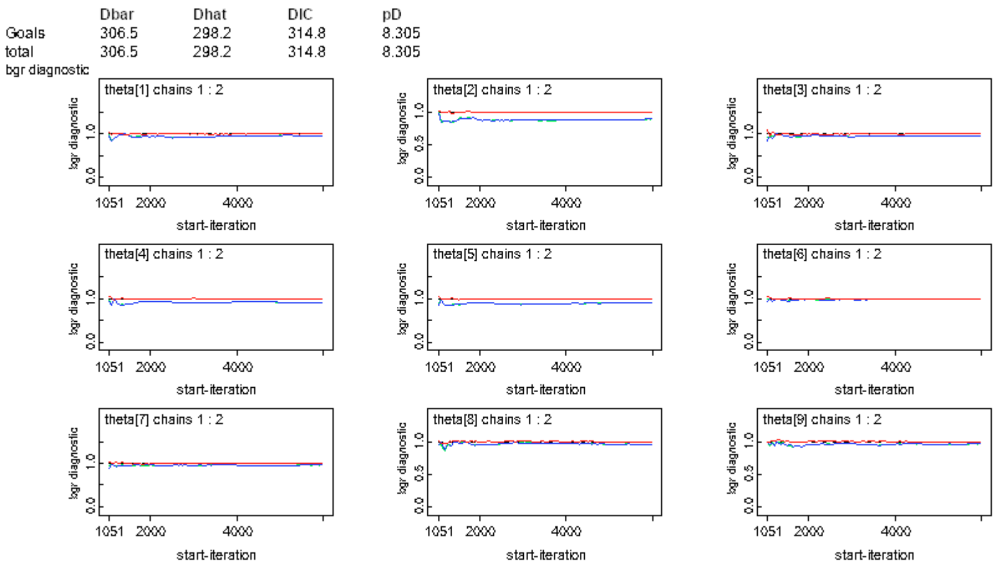


Figure 2: DIC and BGR diagnosis for each θ_i, a and b.

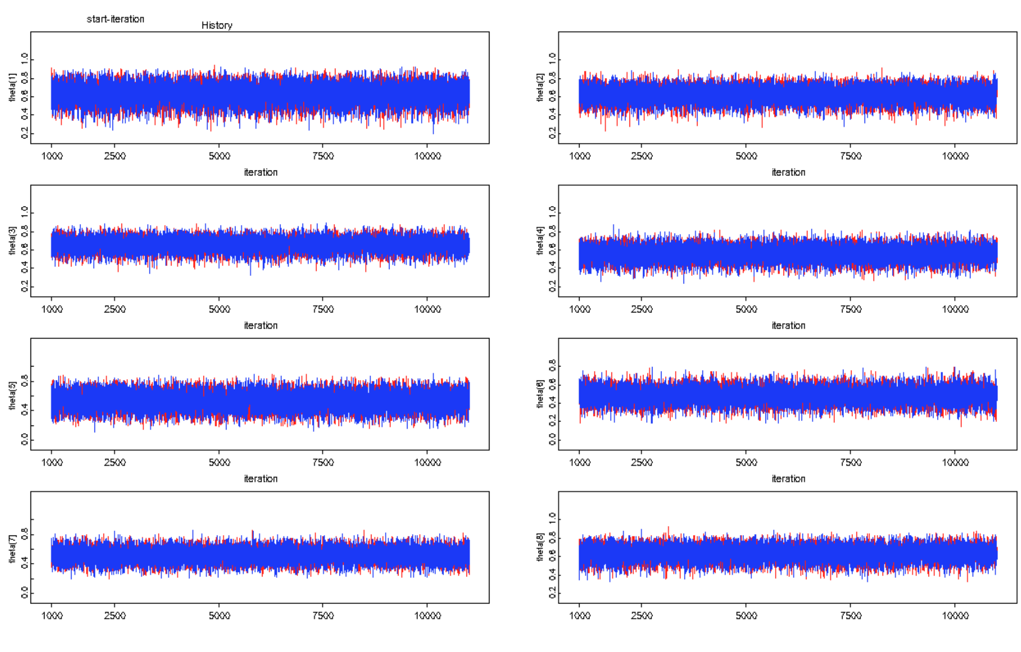


Figure 3: History for each θ_i, a and b.

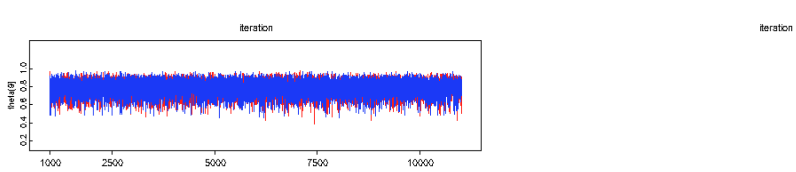


Figure 4: History for each θ_i, a and b.

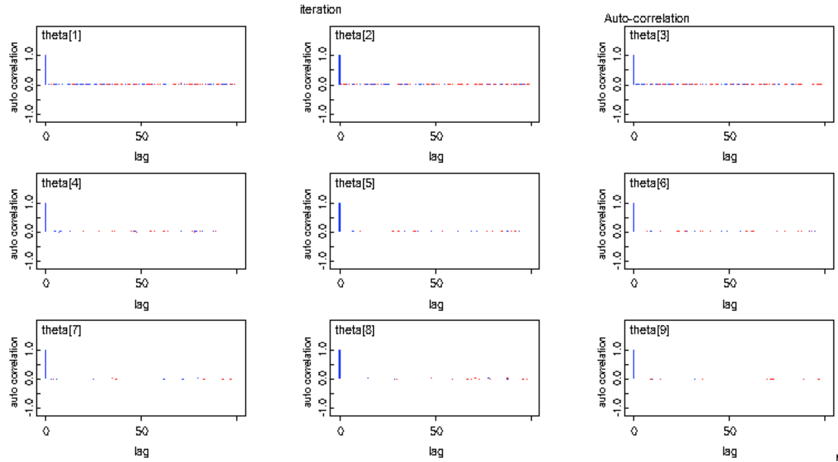


Figure 5: Autocorrelation for each θ_i, a and b.

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
theta[1]	0.6317	0.1076	7.58E-4	0.4101	0.637	0.827	1001	20000
theta[2]	0.6212	0.08832	6.397E-4	0.4425	0.6238	0.7839	1001	20000
theta[3]	0.6584	0.07588	5.159E-4	0.5038	0.6603	0.7997	1001	20000
theta[4]	0.5592	0.08392	6.207E-4	0.3905	0.5603	0.7194	1001	20000
theta[5]	0.5333	0.1238	8.563E-4	0.29	0.5349	0.7682	1001	20000
theta[6]	0.4839	0.08832	6.267E-4	0.3117	0.4844	0.6543	1001	20000
theta[7]	0.5198	0.09826	7.176E-4	0.3275	0.5207	0.7113	1001	20000
theta[8]	0.6358	0.08259	5.82E-4	0.4691	0.6383	0.7896	1001	20000
theta[9]	0.7828	0.08378	6.045E-4	0.5984	0.7911	0.9215	1001	20000

Figure 6: Statistics for each θ_i, a and b.

2.3 Prior case 3: with a hierarchical exchangeable priors

In the last example, two different prior distributions will be tested. The first will consider that $a \sim gamma(0.01, 0.01)$ and $b \sim gamma(0.01, 0.01)$ (independently). The second will consider $a \sim |t_3|$ and $b \sim |t_3|$ (independently).

2.3.1 a,b $\sim gamma(0.01, 0.01)$

Now, we have 9 parameters $(\theta_1, \theta_2, ..., \theta_9)$ and two additional hyper parameters a, b. The prior of each θ_i is assumed to be distributed $beta(a, b)$ (non-informative) and a, b are assumed to be distributed $gamma(0.01, 0.01)$. The MCMC was run with two chains. For the first chain, the initial values were $a = 10$, $b = 1$, and each $\theta_i = 0.1$ for all i. For the second chain, the initial values were $a = 1$, $b = .1$ and each $\theta_i = 0.9$ for all i. For this case, the model gives a strong autocorrelation for the parameters a and b (even when thin = 100). Then the prior distribution was changed for a $gamma(0.1, 0.1)$ (same mean, but less variance). Afther this, with a thin = 30 the autocorrelation in a and b was fixed. The chain converges relatively fast (just 501 iterations were used burned). Finally, the Gelman-Rubin gives a good conclusion about the convergence of the chains (although for a, b andθ_4 there is a very small gap.)

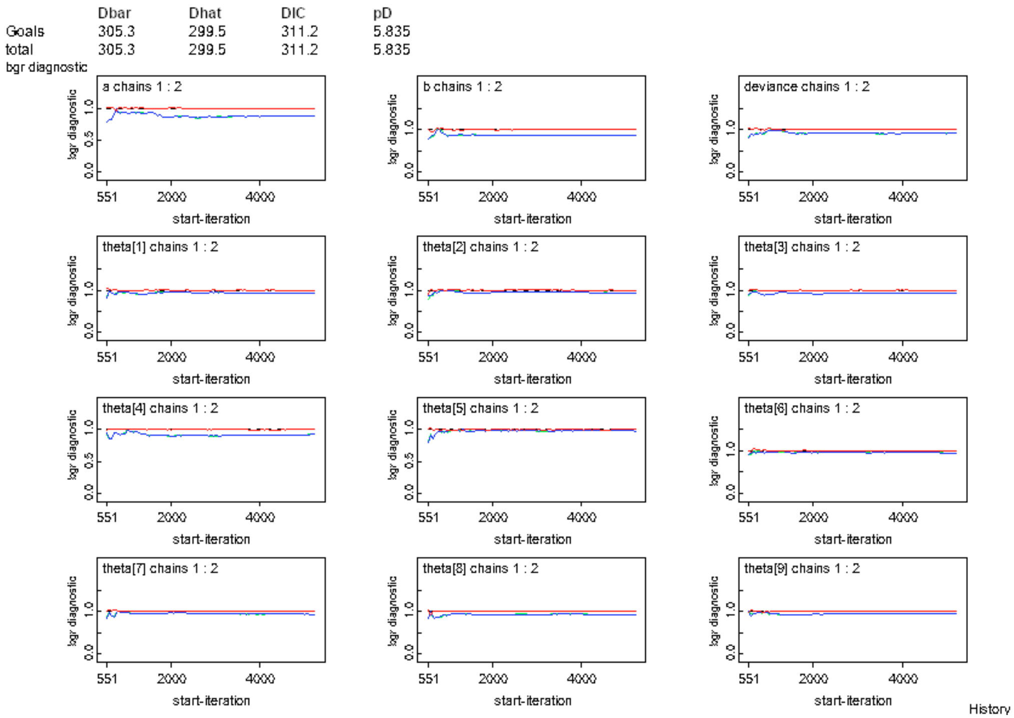


Figure 7: DIC and BGR diagnosis for each θ_i, a and b.

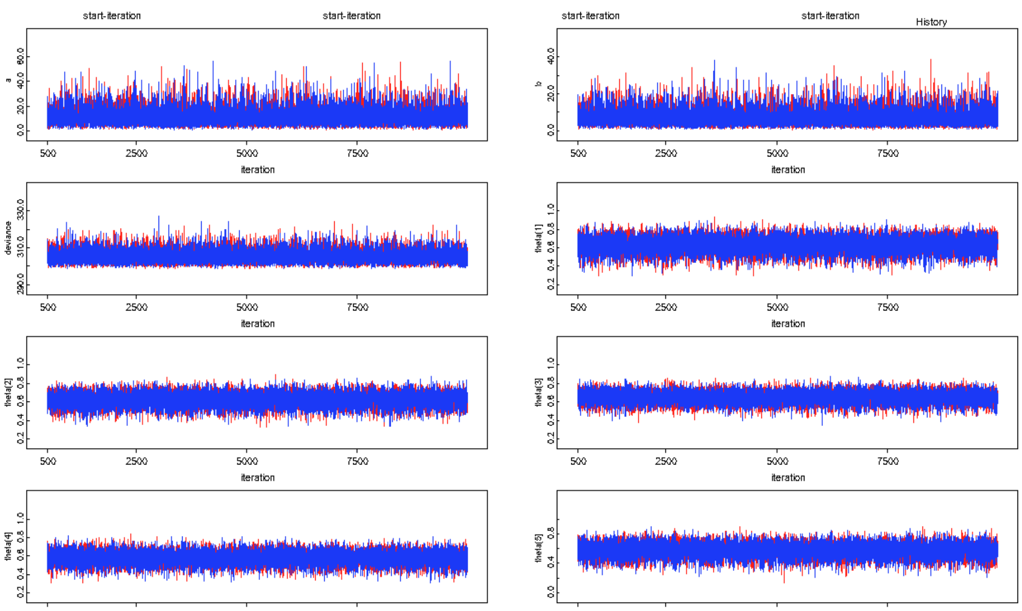


Figure 8: History for each θ_i, a and b.

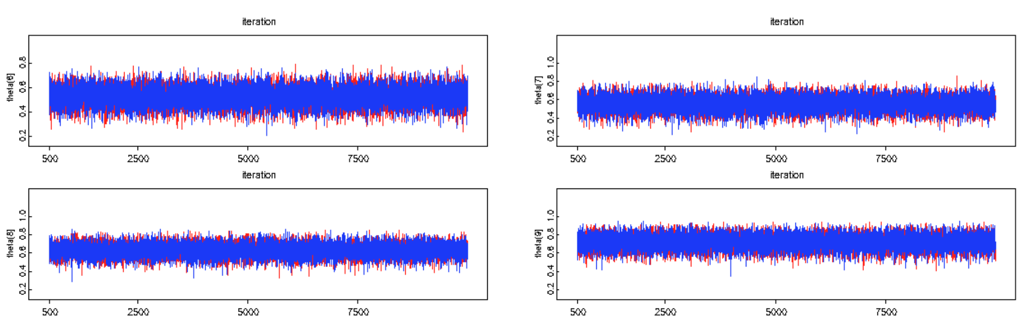


Figure 9: History for each θ_i, a and b.

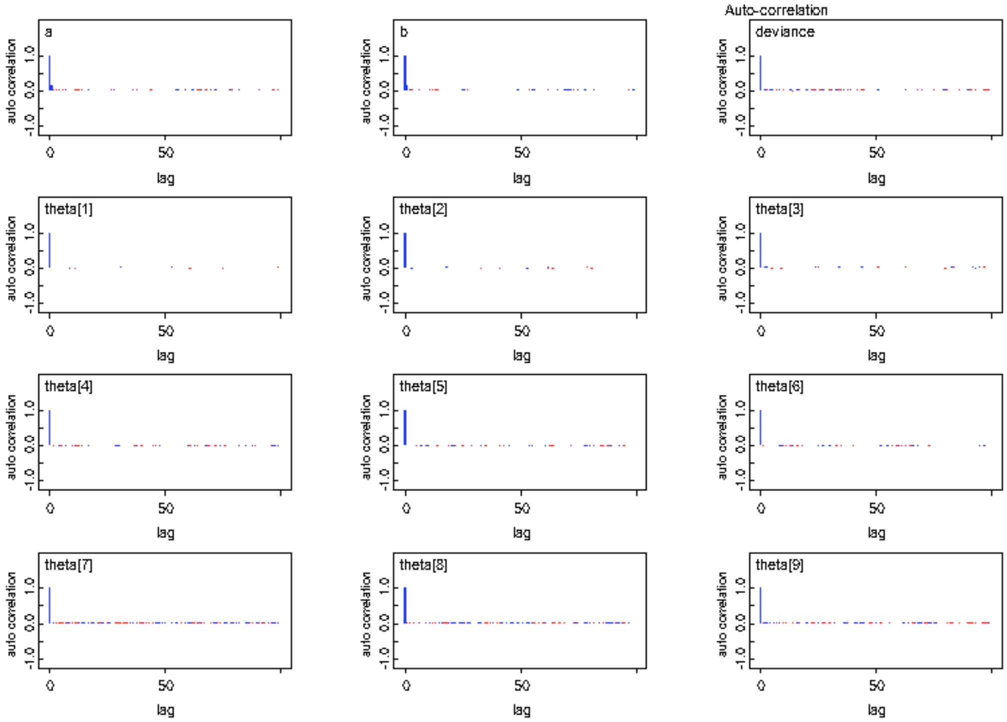


Figure 10: Autocorrelation for each θ_i, a and b.

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
a	11.19	6.631	0.05007	2.764	9.736	28.01	501	19000
b	7.22	4.274	0.03191	1.789	6.278	18.1	501	19000
deviance	305.4	3.364	0.02589	300.3	304.9	313.3	501	19000
theta[1]	0.6287	0.08621	6.041E-4	0.4553	0.6309	0.7902	501	19000
theta[2]	0.6216	0.07551	5.093E-4	0.4666	0.6236	0.7637	501	19000
theta[3]	0.6475	0.06733	5.292E-4	0.5123	0.6489	0.7755	501	19000
theta[4]	0.5785	0.07273	5.227E-4	0.4324	0.5806	0.7158	501	19000
theta[5]	0.5754	0.0956	7.521E-4	0.375	0.5784	0.7558	501	19000
theta[6]	0.5265	0.07725	5.595E-4	0.3717	0.5283	0.6747	501	19000
theta[7]	0.5569	0.08219	5.991E-4	0.3915	0.5586	0.7122	501	19000
theta[8]	0.6329	0.07146	5.355E-4	0.4874	0.6345	0.7679	501	19000
theta[9]	0.7211	0.07939	5.465E-4	0.5616	0.7235	0.8687	501	19000

Figure 11: Summary for each θ_i, a and b.

2.3.2 a,b $\sim |t_3|$

Again, we have 9 parameters $(\theta_1, \theta_2, ..., \theta_9)$ and two additional hyper parameters a, b. The prior of each θ_i is assumed to be distributed $beta(a, b)$ (non-informative) and a, b are assumed to be distributed $|t_3|$. The MCMC was run with two chains. For the first chain, the initial values were $a = 1$, $b = 10$, and each $\theta_i = 0.1$ for all i. For the second chain, the initial values were $a = 10$, $b = 1$ and each $\theta_i = 0.9$ for all i.For this experiment, there was a high autocorrelation in the parameters a and b, so there was a change the thin parameter (now, equal to 30). The chains converge slow (compared with the other examples, now 4000 iterations were used as burning). Finally, the Gelman-Rubin diagnosis gives us a good conclusion about the convergence of the chains of θ_i, but for a, b there is a considerable gap.)

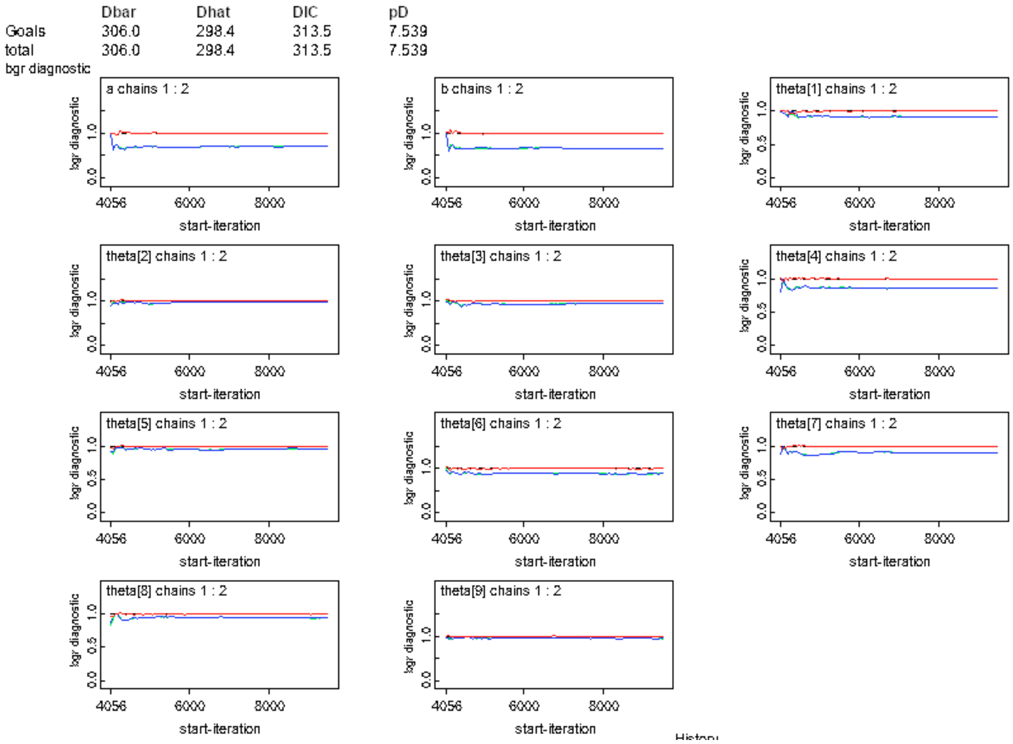


Figure 12: DIC and BGR diagnosis for each θ_i, a and b.

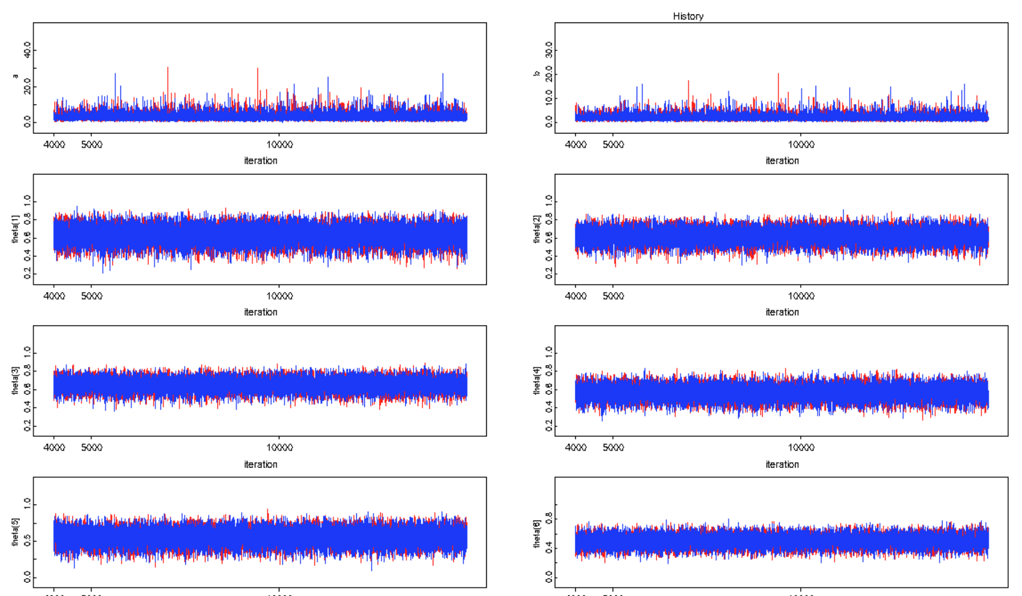


Figure 13: History for each θ_i, a and b.

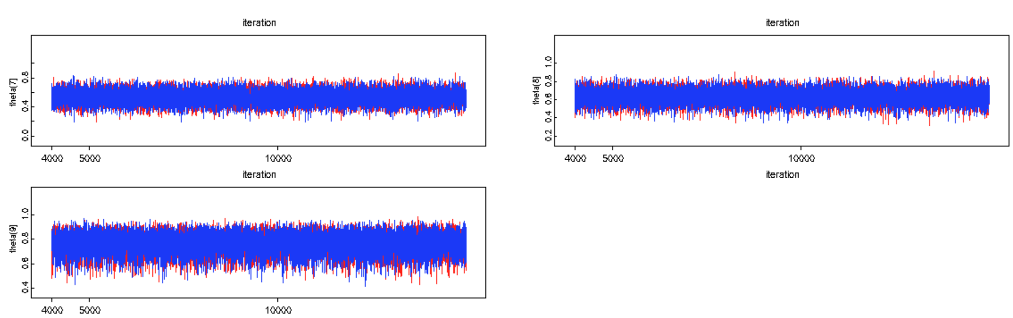


Figure 14: History for each θ_i, a and b.

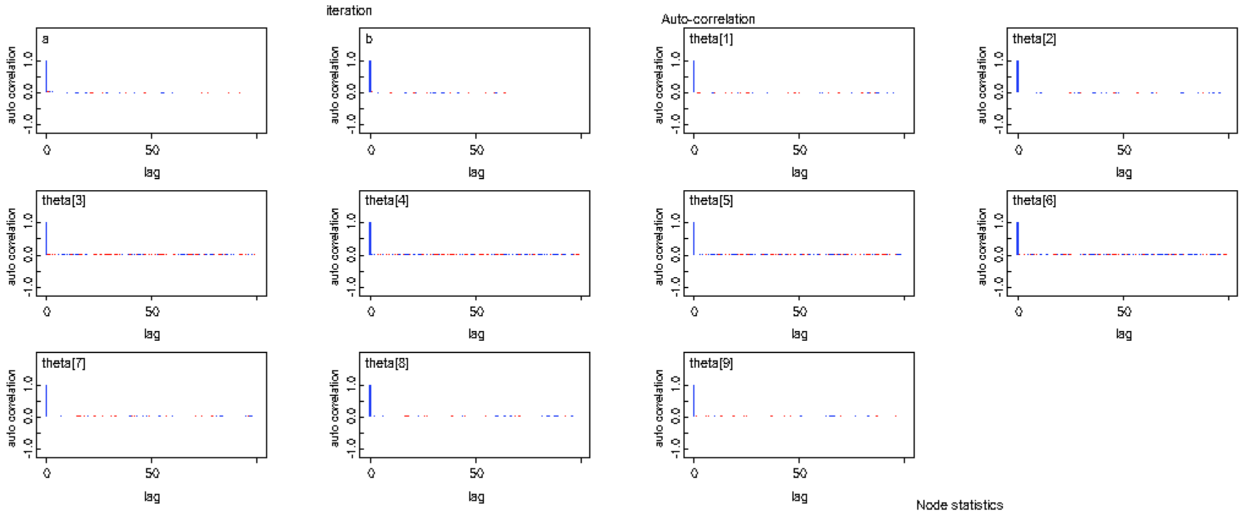


Figure 15: Autocorrelation for each θ_i, a and b.

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
a	3.464	1.998	0.01467	1.107	2.997	8.521	4001	22000
b	2.324	1.276	0.009579	0.777	2.043	5.598	4001	22000
theta[1]	0.635	0.1003	6.322E-4	0.4306	0.6383	0.8218	4001	22000
theta[2]	0.624	0.08451	6.075E-4	0.4532	0.6265	0.7812	4001	22000
theta[3]	0.6572	0.07361	5.046E-4	0.5069	0.6595	0.7929	4001	22000
theta[4]	0.5681	0.08031	5.872E-4	0.4047	0.5693	0.7208	4001	22000
theta[5]	0.5556	0.1152	7.811E-4	0.3257	0.5576	0.7724	4001	22000
theta[6]	0.5007	0.08535	5.597E-4	0.3346	0.5016	0.6652	4001	22000
theta[7]	0.5357	0.09346	5.994E-4	0.3504	0.5378	0.7154	4001	22000
theta[8]	0.6367	0.07924	5.49E-4	0.4756	0.639	0.7852	4001	22000
theta[9]	0.765	0.0835	5.702E-4	0.5832	0.7708	0.9076	4001	22000

Figure 16: Summary for each θ_i, a and b.

3 Conclusions

For this experiment, the best option was to have a pooled θ. Some experiments that can be done are the following:

- Try different priors for the hyper parameters
- Consider another criteria to group the schools

The first point is very important. We can try different prior in order to consider important or not so important the differences between the schools. The second point is an idea to group some schools with the same parameter (for example, the school 1,2,3,4,5 in one group and the school 6,7,8,9 in another). This grouping can be done geographically or with other relevant information. This way, we can reduce the p_D factor of the DIC.

Is important to remark that it is necessary to consider all the values written in the tables of DIC of each model. This way, if we only are worried about fitting, we shoud consider the value *Dhat*. Then, under this assumption, we should select another model as the best one. But as seen in class, we should penalize the number of parameters (this is because of the overfitting, we can have a better *Dhat*, but in prediction this model will have a worse predictions). Again, in the DIC table we can see the value p_D that, according with the theory, is an approximation of the number of parameters. ($p_D \approx p$ with p the number of parameters). In fact, this number is used as a penalization parameter for the DIC (DIC = Dhat + 2pD).

Table 1: My caption

Case	DIC	pD
θ_i are equal with non informative prior (pooled)	308.0	0.9914
Independent non informative priors	314.8	8.305
Hierarchical exchangeable prior (Gamma)	311.2	5.835
Hierarchical exchangeable prior ($\|t_3\|$)	313.5	7.538

Now, interpreting the results according to the problem, when we select the first model (θ_i are equal with non informative prior (pooled)), we arrive at a conclussion that $\theta = .6099$, i.e. that 60% of the students will prefer popularity over sporting ability. This can be given with the credibility interval of $(0.5455, 0.6717)$.

Another point is that, in the models of exchangeable prior, the sd deviation of each θ_i is less than the second model (Independence). Also, the mean of θ_i in the the hierarchical model is closer to the mean of θ the first model (partially pooling in direction to a common mean). This is directly related with the assumption that in the hierarchical models the have a dependency between them.