Universiteit Utrecht

Juan Claramunt (5876060)
Methodology and Statistics

Predict NBA points using a Bayesian approach

Data & competing models

NBA database. Season 2016-2017 (March 2017). 474 players.
Outcome variable: Points (PTS).
Predictors: Free throw attempts (FTA), Field goal attempts (FGA). Both centered.
Our goal is to predict points by FTA and FGA. To do so, we propose the following linear regression models.

Model 1:

$$\boldsymbol{PTS_i = a_0 + b_1 FTA_i + b_2 FGA_i + e_i \,, \quad e_i \sim N(0, \sigma^2)}$$

Model 2:

$$PTS_i = a_0 + b_2 FGA_i + e_i \;\;, \quad e_i \sim N(0, \sigma^2)$$

Priors

Prior distributions:
Based on historical data from the 2015-2016 NBA season:
$p(a_0)$ = N(466.676, 3.694).
$p(b_1)$: t-distribution. Mean = 0.971 , standard deviation = 0.041 ,df = 1.
$p(b_2)$ = N(1.022, 0.012).
Uninformative:
$p(\sigma^2)$ =Γ(0.001, 0.001)

Posteriors and sampling

We sample from the joint posterior using the Gibbs sampler with a MH step to draw values from the conditional posterior of b_1.

Gibbs sampler: *To sample from well-known posterior distributions.*
Given that we cannot sample from the joint posterior distribution, we do it from the conditional posterior distributions in order.
First, we sample from the conditional posterior of a_0:

$$p(a_0|X, Y, b_{1,(t-1)}, b_{2,(t-1)}, \sigma^2_{t-1}) = N(\mu_{01}, \tau^2_{01})^*$$

Metropolis-Hastings step: *To sample from any distribution.*
To sample from the conditional posterior of b_1 we perform a random walk MH:
1) Sample from a proposal distribution, β*:

$$q(\beta^*|b_{1,(t-1)}) = N(\beta^* - b_{1,(t-1)}, 0.1)$$

2) Compute the acceptance rate, *r*, (around 18%):

$r = \frac{p(\beta^*|X,Y,a_{0,t},b_{2,(t-1)},\sigma^2_{t-1})}{p(b_{1,(t-1)}|X,Y,a_{0,t},b_{2,(t-1)},\sigma^2_{t-1})}$ where $p(b|X, Y, a_{0,t}, b_{2,(t-1)}, \sigma^2_{t-1})$ is the conditional posterior of b_1.
3) Sample a value *u* from a uniform(0,1) distribution.
4) If *r* > *u*, $b_{1,t}$=β*. Otherwise, $b_{1,t}=b_{1,(t-1)}$

Third, we sample from the conditional posterior of b_2:

$$p(b_2|X, Y, a_{0,t}, b_{1,t}, \sigma^2_{t-1}) = N(\mu_{21}, \tau^2_{21})^*$$

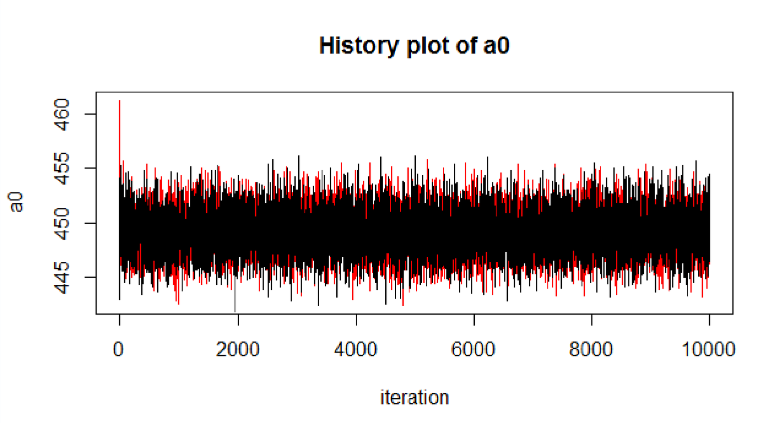
Finally, we sample from the conditional posterior of σ^2:

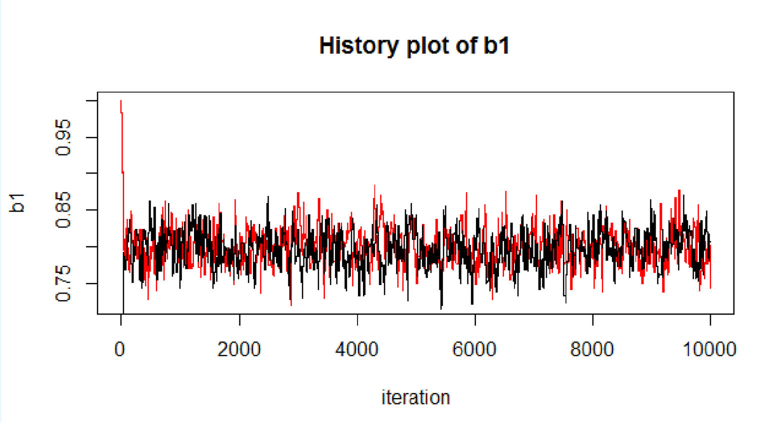
$$p(\sigma^2|X, Y, a_{0,t}, b_{1,t}, b_{2,t}) = \Gamma^{-1}(\alpha_1, \beta_1)^*$$

After all, we remove the burn-in period.

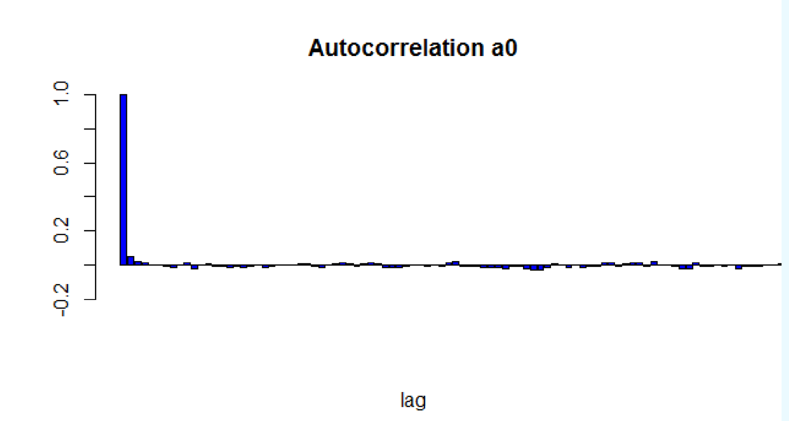
Convergence

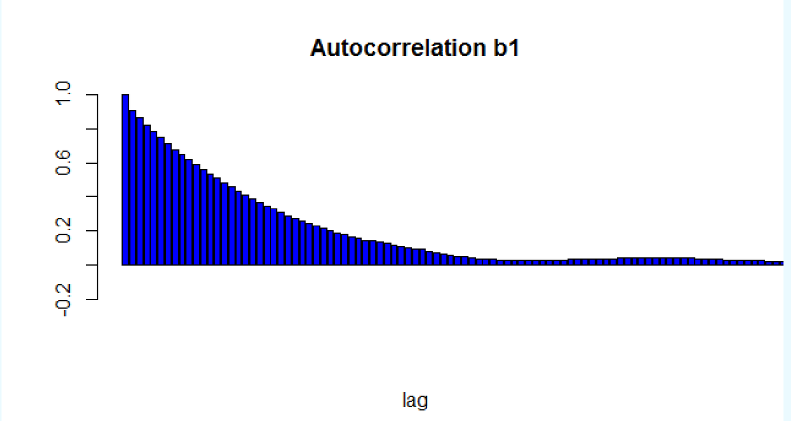
History plots:





Autocorrelations:





Conclusion: According to the history plots (oscillating around the parameter estimate) and the autocorrelations (decreasing to zero), the Gibbs sampler with the M-H step **convergence seems to be reached**.

DIC

To choose the best model in terms of fit and complexity we compute the DIC, obtaining the following results:

$$DIC = Dhat + 2p_D$$

	Dhat	Dbar	p_D	DIC
Model 1	4880.3	4881.4	1.081	4882.5
Model 2	5270.6	5272.2	1.580	5273.7

Then, **in terms of fit and complexity, the first model including two predictors is better according to the DIC criteria.**

Estimates and credible intervals

The estimates are shown in the following table:

Bayesian	Mean	2.50%	97.50%
a_0	449.21	446.54	451.90
b_1	0.798	0.763	0.834
b_2	1.044	1.033	1.055
σ	41.76	39.62	44.07

Conclusions:
a_0:There is a 95% probability that a player with average FTA and FGA scores between 446.5 and 451.9 points with a mean of 449.2.
b_1: The mean of the posterior of b_1 is 0.798. According to my belief, there is a 95% probability that the increment in points by shooting 1 more FT is between 0.763 and 0.834. Then, this is a meaningful predictor.
b_2: The mean of the posterior of b_2 is 1.044. I am 95% certain that an increment in 1 FG leads to an increment between 1.033 and 1.055 in points. Consequently, b_2 is also a meaningful predictor.

Comparison with frequestist estimates and confidence intervals:

Frequentist	Estimate	2.50%	97.50%
a_0	431.8334	428.41	435.26
b_1	0.799	0.736	0.861
b_2	1.044	1.021	1.067

For b_1 and b_2 we obtained similar results with slightly broader confidence intervals than credible intervals. For a_0, the mean of the posterior distribution is shifted towards the mean of the prior. Furthermore, the credible interval is broader than the confidence interval.
Differences are mainly caused by the effect of the priors previously defined.

Bayes Factor

To know how much the data supports the hypothesis that the coefficient of FGA, b_2, is greater than the FTS one, b_1, we compute its Bayes factor.

H_0: $b_2>b_1$, H_1: b_1,b_2

We define new priors with same mean and variance (aprox. average of previous ones) to compute the Bayes factor. The resulting joint distribution of b_1 and b_2 is a bivariate normal with means equal to 1, variances equal to 0.0064 and no covariances, a good approximation of the previous joint distribution.
Sampling from this prior joint distribution and computing the proportion of iterations in agreement with H_0 we obtain the **complexity**.

c = 0.503

Sampling from the posterior using the Gibbs sampler and computing the proportion of iterations in agreement with H_0 we obtain the **fit**.

f = 1

The resulting **Bayes factor is 1.988**. This is, there is a small evidence in favor of the null hypothesis $b_2>b_1$.
Given the previous hypotheses, the **posterior model probability** of H_0 is 0.665 and the posterior model probability of H_1 is 0.335.

Posterior Predictive P-value

Model assumption tested: **Absence of outliers**

Test is based on:
-Residuals (e_i) - Normal distribution of residuals - Chebyshev's inequality
-Existence of one outlier implies the assumption is not met.
Formal definition:

$$\boldsymbol{test(data) := \max_i(|e_i|) - 3\sigma}$$

The higher is the test value, the more isolated is the most extreme observation of the dataset.

Results and interpretation:

Bayesian	Mean	2.50%	97.50%
p-value	0.00	0.00	0.00

The posterior predictive p-value mean is the share of iterations in which the test applied to the replicated datasets is greater than the test applied to the original dataset. This is, the proportion of iterations in which the replicated dataset contains a more extreme case than the original dataset.
Given that **mean(p-value)=0**, there are no values in the replicated dataset more extreme than in the original one. Consequently, we consider that **the assumption of absence of outliers is not met.**

*For further information and R code: j.claramuntgonzalez@students.uu.nl