# 4 Markov Chain Monte Carlo [20 pts] (Paul)

Nowadays, statistical modelling of sport data has become an important part of sports analytics and is often a critical reference for the managers in their decision-making process. In this part, we will work on a real world example in professional sports. Specifically, we are going to use the data from the 2013-2014 Premier League, the top-flight English professional league for men's football clubs, and build a predictive model on the number of goals scored in a single game by the two opponents. Bayesian hierarchical model is a good candidate for this kind of modeling task. We model each team's strength (both attacking and defending) as latent variables. Then in each game, the goals scored by the home team is a random variable conditioned on the attacking strength of the home team and the defending strength of the away team. Similarly, the goals scored by the away team is a random variable conditioned on the attack strength of the away team and the defense strength of the home team. Therefore, the distribution of the scoreline of a specific game is dependent on the relative strength between the home team A and the away team B, which also depends on the relative strength between those teams with their other opponents.

Table 1: 2013-2014 Premier League Teams

| | | | | | |
|---|---|---|---|---|---|
| Index | 0 | 1 | 2 | 3 | 4 |
| Team | Arsenal | Aston Villa | Cardiff City | Chelsea | Crystal Palace |
| Index | 5 | 6 | 7 | 8 | 9 |
| Team | Everton | Fulham | Hull City | Liverpool | Manchester City |
| Index | 10 | 11 | 12 | 13 | 14 |
| Team | Manchester United | Newcastle United | Norwich City | Southampton | Stoke City |
| Index | 15 | 16 | 17 | 18 | 19 |
| Team | Sunderland | Swansea City | Tottenham Hotspurs | West Bromwich Albion | West Ham United |

Here we consider using the same model as described by Baio and Blangiardo (2010) [4]. The Premier League has 20 teams, and we index them as in Table 1. Each team would play 38 matches every season (playing each of the other 19 teams home and away), which totals 380 games in the entire season. For the $g$-th game, assume that the index of home team is $h(g)$ and the index of the away team is $a(g)$. The observed number of goals $(y_{g0}, y_{g1})$ of home and away team is modeled as independent Poisson random variables:

$$y_{gj}|\theta_{gj} \sim \text{Poisson}(\theta_{gj}), \quad j = 0, 1 \tag{8}$$

where $\theta = (\theta_{g0}, \theta_{g1})$ represents the scoring intensity in the $g$-th game for the team playing at home ($j = 0$) and away ($j = 1$), respectively. We put a log-linear model for the $\theta$s:

$$\log \theta_{g0} = home + att_{h(g)} - def_{a(g)} \tag{9}$$

$$\log \theta_{g1} = att_{a(g)} - def_{h(g)} \tag{10}$$

Note that team strength is broken into attacking and defending strength. And home represents home-team advantage, and in this model is assumed to be constant across teams. The prior on the home is a normal distribution:

$$home \sim \mathcal{N}(0, \tau_0^{-1}) \tag{11}$$

where we set the precision $\tau_0 = 0.0001$.

The team-specific attacking and defending effects are modeled as:

$$att_t \sim \mathcal{N}(\mu_{att}, \tau_{att}^{-1}) \tag{12}$$

$$def_t \sim \mathcal{N}(\mu_{def}, \tau_{def}^{-1}) \tag{13}$$

We use conjugate priors as the hyper-priors on the attack and defense means and precisions:

$$\mu_{att} \sim \mathcal{N}(0, \tau_1^{-1}) \tag{14}$$

$$\mu_{def} \sim \mathcal{N}(0, \tau_1^{-1}) \tag{15}$$

$$\tau_{att} \sim \text{Gamma}(\alpha, \beta) \tag{16}$$

$$\tau_{def} \sim \text{Gamma}(\alpha, \beta) \tag{17}$$

where the precision $\tau_1 = 0.0001$, and we set parameters $\alpha = \beta = 0.1$.

This hierarchical Bayesian model can be represented using a directed acyclic graph as shown in Figure 2.

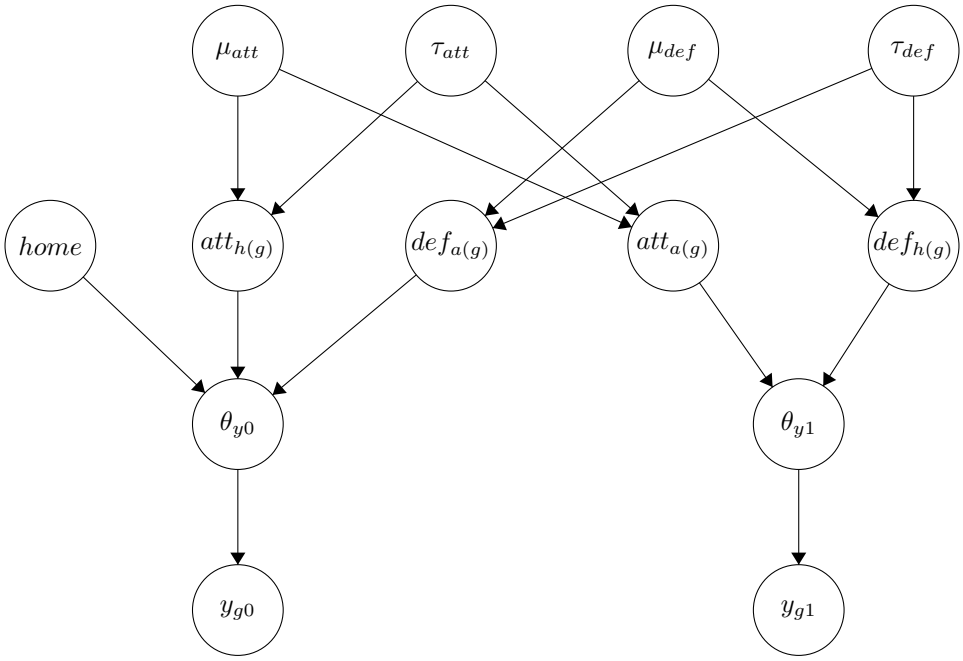



Figure 2: The DAG representation of the hierarchical Bayesian model. Figure adapted from [4].

The goals of each game are $\mathbf{y} = \{y_{gj} | g = 0, 1, ..., 379, j = 0, 1\}$ are the observed variables, and parameters $\boldsymbol{\theta} = \{home, att_0, def_0, ..., att_{19}, def_{19}\}$ and hyper-parameters $\boldsymbol{\eta} = (\mu_{att}, \mu_{def}, \tau_{att}, \tau_{def})$ are unobserved variables that we need to make inference on. To ensure identifiability, we enforce a corner constraint on the parameters (pinning one team's parameters to 0,0). Here we use the first team as reference and assign its attacking and defending strength to be 0:

$$att_0 = def_0 = 0 \tag{18}$$

In this question, we want to estimate the posterior mean of the attacking and defending strength for each team, i.e. $\mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{\eta}|\mathbf{y})}[att_i]$, $\mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{\eta}|\mathbf{y})}[def_i]$, and $\mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{\eta}|\mathbf{y})}[home]$.

1. **[4 points]** Find the joint likelihood $p(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta})$.

Solution

**Answer:** $p(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta}) = p(\mu_{att})p(\tau_{att})p(\mu_{def})p(\tau_{def})p(home)p(att_{h_g}|\mu_{att}, \tau_{att})$
$p(def_{a_g}|\mu_{def}, \tau_{def})p(att_{a(g)}|\mu_{att}, \tau_{att})p(def_{h(g)}|\mu_{def}, \tau_{def})p(\theta_{y0}|home, att_{h(g)}, def_{a(g)})$

$p(\theta_{y1}|att_{a(g)}, def_{h(g)})p(y_{g0}|\theta_{y0})p(y_{g1}|\theta_{y1})$

2. **[4 points]** Write down the Metropolis-Hastings algorithm for sampling from posterior $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\mathbf{y})$, and derive the acceptance function for a proposal distribution of your choice (e.g. isotropic Gaussian).

Solution

**Algorithm 1** Metropolis-Hastings Algorithm

**Input :** $X = (\theta, \eta), Y = \mathbf{y}$, Q(x) as the proposal function, $P(X|Y)$ as the sampled function

**Initialization:** initialize starting state $x^{(0)}.t = 0$.

**for** *t = 1,2,...T* **do**

  $x = x^t, t = t + 1$,

  sample $x^* \in Q(x^*|x)$ draw sample from proposal.

  sample $u \in U(0, 1)$, which is to draw acceptance threshold.

  $A(x^*|x) = min(1, \frac{P(x^*,y)Q(x|x^*)}{P(x,y)Q(x*|x)})$

  **if** $u < A(x^*|x)$ **then**

    $x^t = x^*$ transition to next sample

  **else**

    $x^t = x$ stay in current state

  **end**

**end**

3.

**Output :** Samples from P(X—Y) for t = 1:T

Acceptance function can be written as:

$$A(x^*|x) = min(1, \frac{P(\theta^*, \eta^*, y)Q(\theta, \eta|\theta^*, \eta^*)}{P(\theta, \eta, y)Q(\theta^*, \eta^*|\theta, \eta)})$$

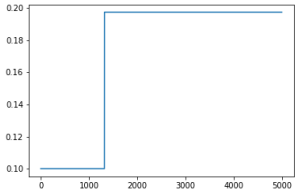
4. **[12 points]** Implement the Metropolis-Hastings algorithm to inference the posterior distribution. The data can be found from `premier_league_2013_2014.dat`, which contains a $380 \times 4$ matrix. The first column is the number of goals $y_{g0}$ scored by the home team, the second column is the number of goals $y_{g1}$ scored by the away team, the third column is the index for the home team $h(g)$, and the fourth column is the index for the away team $a(g)$.

- Use an isotropic Gaussian proposal distribution $\mathcal{N}(0, \sigma^2 I)$ and use 0 as the starting point.
- Run the MCMC chain for 5000 steps to burn in and then collect 5000 samples with $t$ steps in between (i.e., run M-H for $5000t$ steps and collect only each $t$-th sample). This is called thinning, which reduces the autocorrelation of the MCMC samples introduced by the Markovian process. The parameter sets are $\sigma = 0.005, 0.05, 0.5$, and $t = 1, 5, 20, 50$.
- Plot the trace plot of the burn in phase and the MCMC samples for the latent variable *home* using proposal distributions with different $\sigma$ and $t$.
- Estimate the rejection ratio for each parameter setting, report your results in a table.
- Comment on the results. Which parameter setting worked the best for the algorithm?

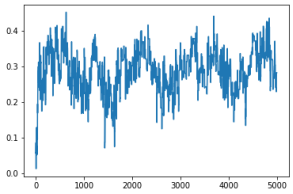
- Use the results from the optimal parameter setting:
  (a) plot the posterior histogram of variable *home* from the MCMC samples.
  (b) plot the estimated attacking strength $\mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{\eta}|\mathbf{y})}[att_i]$ against the estimated defending strength $\mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{\eta}|\mathbf{y})}[def_i]$ for each the team in one scatter plot. Please make sure to identify the team index of each point on your scatter plot using the index to team mapping in Table 1.

Solution

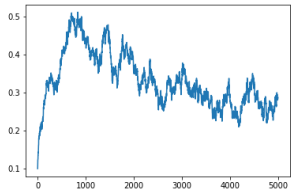
**Answer:** Trace plot for t = [1,5,20,50] and $\sigma$ = [0.005, 0.05, 0.5]



$\sigma = 0.5$

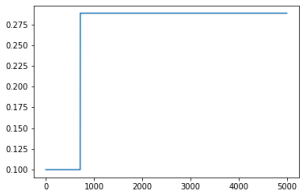


$\sigma = 0.05$

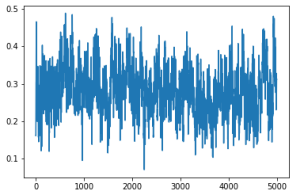


$\sigma = 0.005$

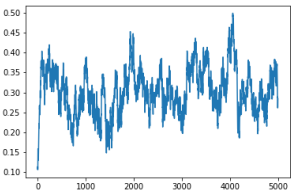
Figure 3: t = 1



$\sigma = 0.5$

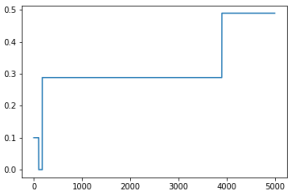


$\sigma = 0.05$

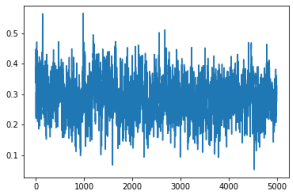


$\sigma = 0.005$

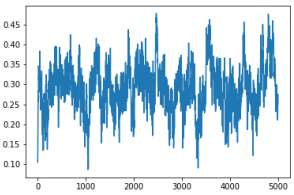
Figure 4: t = 5



$\sigma = 0.5$

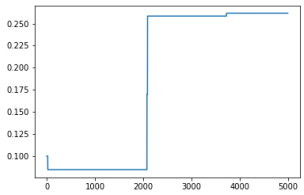


$\sigma = 0.05$

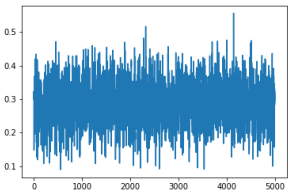


$\sigma = 0.005$

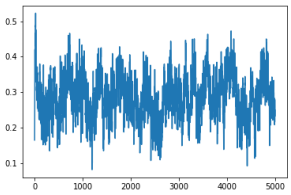
Figure 5: t = 20




$\sigma = 0.5$




$\sigma = 0.05$




$\sigma = 0.005$

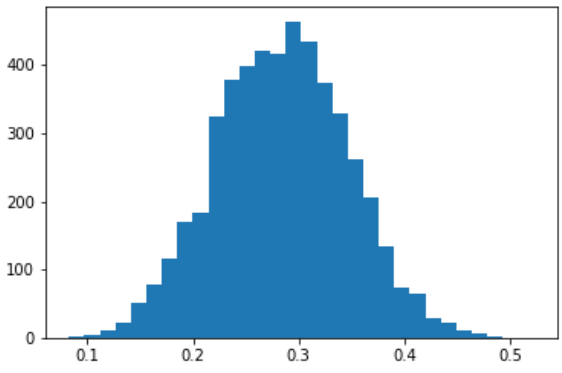
Figure 6: t = 50

Rejection ratio:

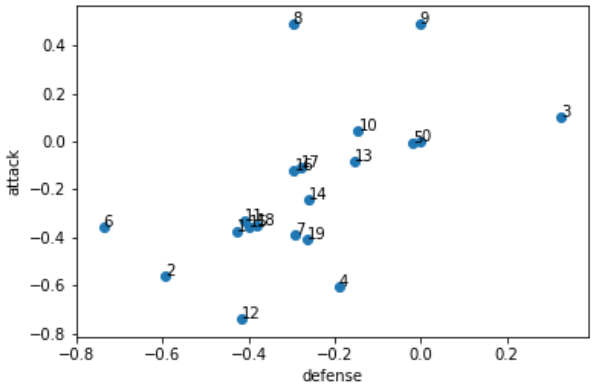
| Rejection ratio | $\sigma$=0.005 | $\sigma$=0.05 | $\sigma$=0.5 |
|---|---|---|---|
| t = 1 | 0.1423 | 0.7978 | 0.9998 |
| t = 5 | 0.1066 | 0.8141 | 0.9999 |
| t = 20 | 0.1046 | 0.8360 | 0.9999 |
| t = 50 | 0.1052 | 0.8407 | 0.9999 |

Comments on the result:
We can see clearly from the table and graph that when $\sigma = 0.5$, the reject ratio is too high to generate useful samples. $\sigma = 0.005$ works best for this problem. Besides, for differnt t, as t is bigger, data correlation is decreasing. t = 20 and 50 have better results.

Optimal parameter setting: $\sigma = 0.005$, t = 50
histogram of variable home:




Defense Attack




5. [**0 points**] Despite what the data says, conclude that Manchester United is the best team in the Premier League!

**Note: You are free to use Python or MATLAB for your implementation. You are NOT allowed to use any existing implementations of Metropolis-Hastings in this problem. Please include all the required results (figures + tables) in your writeup PDF submission, as well as submit your code to Gradescope separately.**