A Bayesian Approach to Model Goal Scoring: Using the Poisson Distribution and Markov Chain Monte Carlo *

Kevin Lai †

December 29, 2018

Abstract

In this expository essay, I explore Markov Chain Monte Carlo (MCMC) in goal-scoring using the Poisson distribution, proposed by Everson and Goldsmith-Pinkham (2008). First, I explain the mathematical and intuitive aspects of the Gibbs Sampler, which uses a Gamma prior and Poisson likelihood along with a home-field advantage parameter, proposed in Everson and Goldsmith-Pinkham (2008), and discuss the results of their algorithm based on English Premier League soccer data. Then, I use a simplified version of the MCMC, which takes away the home-field advantage parameter, and apply it to 2017 Professional Ultimate Frisbee data (AUDL), a relatively newer sports scene which has a similar environment to soccer. I implement the MCMC for teams within a division, for 5 divisions. I find that this method is acceptable, but not all divisions are predicted properly. I suggest future steps of research, which include home-field advantage, star power, and other aspects of the sport.

***Acknowledgements**: I want to thank Noah Giansiracusa, Math Professor at Swarthmore College, for his feedback and input on this paper. I also want to thank Phil Everson, Statistics Professor at Swarthmore College, for teaching Markov Chain Monte Carlo in his STAT 111 class and for his input in this paper. Additional thanks to all my math professors and peers at Swarthmore College for their instruction and support. The views expressed in the paper are my own and do not reflect those of the authors of my references or Swarthmore College.

†Math and Statistics Department, Swarthmore College.

1 Introduction

Whether you are a sports fan or an analyst, how teams score in any sports is very important, especially when we talk about how good a particular team is. We can easily model how teams score in sports, especially in soccer and hockey because the basic scoring scheme is simple: you put the ball in the goal and it's a point. The literature points to modeling goal scoring in soccer and hockey (and other similar sports like lacrosse and waterpolo) using a Poisson model because of the sports' sparse scoring and timed nature. However, this model is often too simple, so it is pretty inaccurate outside of a perfect world. To supplement this simple model, we can use Markov Chain Monte Carlo (MCMC) algorithms, more specifically the Gibb's Sampler method, to fix the problems with using the simple model and more accurately predict goal scoring (a more Bayesian flavor[1]). Through this new model of the Poisson Distribution with a MCMC algorithm, we will not only be able to predict how well a team can score goals, but we can also relate how goal scoring can impact a team's offensive power and defensive power, thus giving a prediction for an overall "rating".

To check the accuracy of the model, we check our simulated data with that of the actual results by ranking the simulated results by strength (the combination of offensive and defensive power) and comparing them to the actual standings. The goal is to get the ranking of the strength from the simulated results to match closely to the actual standings. We rank on overall strength for the following reason: while it is good to score many goals, if the team allows many goals, then in terms of overall strength, the really strong offensive rating will be net out by a weak defensive rating. To extend the experimentation, although not done in this paper, we can simulate the season data many times to get at a more accurate final predicted standings.

The paper serves to explain all aspects of our model without, hopefully, confusing the reader too much. Section 2 will give the background terminology for the vehicles driving the model (i.e. Poisson Distribution, Markov Chain Monte Carlo, Bayesian inference). Section 3 will give an in depth look at the Everson - Goldsmith-Pinkham goal scoring model. Section 4 will apply the properties of the model to another less commonly known sport: Ultimate Frisbee. To do this, we use data from the 2017 American Ultimate Disc League (AUDL) season, a season of professional Ultimate Frisbee. While the majority of the discussion will be on the Everson - Goldsmith-Pinkham model using the 2006 English Premier League (EPL) data, we simplify the MCMC algorithm and derivations when we run it with the AUDL 2017 season data. Section 5 will conclude the paper and offer insights to the benefits of Bayesian inference in sports.

2 Background

In this section, I will explain the two major parts of our model: the Poisson Distribution and the Markov Chain Monte Carlo (MCMC). Then, I will introduce Bayesian Inference.

[1] Bayesian Statistics refer to a type of statistics that treats data as fixed and parameters as random. In other words, unknown parameters follow some distribution that is given by some expert. More background in **Section 2.3**.

2.1 Poisson Distribution

We will first go over the basics of the Poisson Distribution. Let X be a discrete random variable. The Poisson distribution $X \sim Poisson(\lambda)$ can be described with the following probability mass function (pmf):

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{1}$$

where $x \in (0, \infty)$ and λ is the rate parameter. The mean of a particular Poisson distribution can be expressed as

$$E(X) = \lambda \tag{2}$$

In order for the Poisson Distribution to hold, we must ensure that both of the following conditions are met

1. Events occur at random.
2. The average rate of occurance of events (the mean) must remain constant throughout the time intervals.

Based on the conditions, we can crudely estimate how often a particular team scores goals with the Poisson distribution because a goal can be scored at any time within the interval (thus the event of the goal is random) and the sport has a strict time interval (hockey has 3 quarters with fixed time interval, soccer is timed up to 90 minutes, etc.). However, based on the literature, there is a definite discrepency between scoring at home and away, thus breaking condition 2. To rectify this, we use the model that Everson and Goldsmith-Pinkham proposes building a composite model to partition offensive and defensive based on home-field advantage (more on this later).

Through the Poisson Model, we assume each soccer match to be the fixed time interval, where there is a probability to score at each time. Through this model, we can easily determine, though crudely, the mean of goals scored in all games. Mathematically, let $X_1, X_2, ...X_n$ be n independent soccer matches all distributed by a Poisson Distribution, each with some rate parameter λ. in other words, let $X_1, ..., X_n \overset{ind}{\sim} Poisson(\lambda_i)$ for $i = 1, 2, ..., n$. The means of each soccer game is

$$E(X_1) = \lambda_1, E(X_2) = \lambda_2, ..., E(X_n) = \lambda_n$$

While it is very crude to model goal scoring using the Poisson distribution, we can improve on it through partitioning on some condition (i.e. home vs away games) and adding other methods like using Bayesian processes such as MCMC.

Finally, another nice property of the Poisson distribution is called *Poisson additivity*. Under the assumption of independence, we can simply add together the distributions, thus adding together the parameters. Mathematically, let $X \sim \text{Poisson}(\lambda_1)$ be independent from $Y \sim \text{Poisson}(\lambda_2)$. Thus, we can say that

$$X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2) \tag{3}$$

We will definitely exploit this property in the explanation of the goal-scoring model.

2.2 Markov Chain Monte Carlo (MCMC)

Next, we will go over Markov Chain Monte Carlo (MCMC). First, we will define a Markov Chain, then go into detail on specifically Markov Chain Monte Carlo, a type of Markov Chain.

2.2.1 Markov Chain

Markov Chains is a process in which the outcome of a given experiment can affect the outcome of the next experiment. In past analyses, we assume independence among all random variables, which is not very realistic. In a Markov Chains, random variables exhibit a one-step dependence where the random variables X_{i+1} depend on X_i. Because of the one-step dependence, Markov Chains can add a layer of realism since they are a good compromise between complete indepedence and complete dependence. We use this technique to simulate complex distributions, especially using Markov Chain Monte Carlo (discussed in the next section).

2.2.2 Markov Chain Monte Carlo

Markov Chain Monte Carlo is a set of powerful algorithms that enable us to simulate from complicated distributions using Markov Chains. In other words, we are able to build our own Markov chain so that the distribution of interest is the stationary distribution of the chain (think of stationary distribution as the equilibrium distribution or end goal). As an added bonus, we can use the stationary distribution that is easy to describe without having to know the normalizing constant of the distribution.

There are two main MCMC algorithms: Metropolis-Hastings algorithm and Gibbs Sampling. We will, as well as in the Everson and Goldsmith-Pinkham model, focus on Gibbs Sampling.

2.2.3 Gibbs Sampling

Gibbs sampling is a type of MCMC algorithm that obtains approximate draws from a joint distribution, based on sampling from conditional distributions, derived using the full joint distribution, one at a time. Since goal-scoring is discrete, for the sake of consistency, we will show an arbitrary example of the process of Gibbs sampling with two discrete random variables.

Consider two discrete random variables X and Y with the joint distribution

$$p_{X,Y}(x, y) = P(X = x, Y = y) \tag{4}$$

where x and y are known values.

Under Gibbs sampling, we update X conditional on Y, then use that value to update Y conditional on X, then repeat the process until we get to the stationary distribution. Mathematically, we carry out the following steps:

0. Start with a current state: $(X_i, Y_i) = (x_i, y_i)$

1. Draw a value x_{i+1} from $P(X|Y = y_i)$ and set $X_{i+1} = x_{i+1}$.

2. Draw a value y_{i+1} from $P(Y|X = x_{i+1})$ and set $Y_{i+1} = x_{i+1}$.

3. Repeat steps 1-2 until convergence to $p_{X,Y}$.

We can extend this beyond two random variables, which will lengthen the algorithm because of the need to sample from more conditional distributions. Adding more steps will improve the quality of our sample.

So, when do we use Gibbs sampling and why do we use it? Gibbs sampling is used when the conditional distributions worked with are pleasant to work with. In most cases, this usually means whether the conditional distributions are parametric (i.e. Poisson, Gamma, etc.). In other words, if the distributions we use to model the problem are parametric, Gibbs sampling is the easiest way to converge at the desired distribution.

2.3 Bayesian Inference

Footnote 1 describes the primary goal of Bayesian Statistics. In this section, we will walk through how Bayesian Statistics work in practice. As the name suggests, Bayesian Statistics employs the use of Bayes' Rule[2] given by:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{5}$$

Now, consider we have an unknown parameter θ and some gathered data $\mathbf{X}$. If we want to find the unknown θ using our data $\mathbf{X}$, we can use Bayes Rule to find what our unknown parameter is

$$P(\theta|\mathbf{X}) = \frac{P(\mathbf{X}|\theta)P(\theta)}{P(\mathbf{X})} \tag{6}$$

Now, we assume that θ and $\mathbf{X}|\theta$ follow some distribution and $\mathbf{X}$ is known. We can simplify **Equation (5)** further since $P(\mathbf{X})$ is some constant, we have

$$P(\theta|\mathbf{X}) \propto P(\mathbf{X}|\theta)P(\theta) \tag{7}$$

We have three components of the Bayesian model

1. *Prior*: A parameter that has a probability distribution where an expert can add basic knowledge to our model. This is $P(\theta)$.

2. *Likelihood*: The data conditioned on the probability distribution of the prior. This is $P(\mathbf{X}|\theta)$.

3. *Posterior*: An updated prior where we incorporate data with the expert's knowledge. This is $P(\theta|\mathbf{X})$.

Equation (6) makes it easier to discern parameteric posteriors from parametric priors and likelihood functions. Furthermore, using **Equation (6)**, we have the following nice property:

Conjugates: Under Bayesian paradigm, there are families of distributions such that prior and posterior distributions are related. This is super nice since we can properly identify prior and posterior pairings to get the final distribution we want.

[2] Also known as Bayes' Theorem or Bayes' Law. A common application of this theorem is identifying the probability of cancer based on how accurate a test is. Mathematically, if A is the event of cancer and B is how accurate the test is, we can determine $P(A|B)$ from $P(B|A)$

3 Model

In this section, we will use the framework from **Section 2** to lay out Everson's and Goldsmith-Pinkham's proposed framework for the Composite Poisson Model. First, we will cover the parameters used in the model. Next, we will cover the Composite Poisson Model (i.e. how the multiple Poisson distributions and other related distributions work together). Finally, we will build the distributions for the Gibb's Sampling algorithm and discuss the results of the algorithm.

3.1 Parameters

First, we want to describe some modifications to simply using the Poisson distribution. As mentioned in **Section 2**, there is a discrepency between scoring at home vs. scoring away. This discrepency will clearly affect a team's scoring power, thus impacting its offensive and defensive strengths. Because of this, we use the following parameters in our model:

1. Offense "strength" parameter (call this θ_{oj})
2. Offensive home field advantage (call this δ_{oj})
3. Defensive "strength" parameter (call this θ_{dj})
4. Defensive home field advantage (call this δ_{dj})

θ_{oj} and θ_{dj} refer to a team's offensive and defensive strength for team j. A large θ_{oj} means the team tends to score a lot, while a large θ_{dj} means the team tends to not allow many goals. δ_{dj} and δ_{oj} refer to how much stronger a team is when playing at home compared to playing on the road. We assume all parameters to be positive (i.e. $\theta_{oj}, \theta_{dj}, \delta_{oj}, \delta_{dj} > 0$).

By construction, we want to assume these parameters take on some distribution. By doing so, we get some nice Bayesian properties (i.e. having friendly *priors*, *posteriors*, and *conjugates*) when we condition our random variable (call it Y) on our parameters. We can even further partition the index j of all θ terms into home and road teams for game i (i.e. h_i and r_i respectively). By doing so, we cover all possible cases: the offensive strength of a home team vs the defensive strength of the road team and the offensive strength of the road team vs the deffensive strength of the home team. This gives us a total of 6 parameters to watch. We call these parameters:

1. $\theta_{oh_i}, \theta_{dr_i}, \delta_{oh_i}, \delta_{dr_i}$ are the parameters that determine the goals of a home team scored in game i
2. $\theta_{or_i}, \theta_{dh_i}, \delta_{oh_i}, \delta_{dr_i}$ are the parameters that determine the goals of a road team scored in game i

Intuitively, the grouping of these parameters make sense. Statement 1 takes the offensive strength of the home team and the defensive strength of the road team, while also factoring the home field advantage parameters δ, to determine the number of goals the home team scores. Statement 2 is the exact opposite, where we take the offensive strength of the road team and the defensive strength of the home team, while also factoring home field advantage parameters δ, to determine the number of goals the road team scores.

Section 3.2 will cover how these parameters work together to give the mean number of goals scored by a particular team.

3.2 Composite Poisson Model

This subsection will cover how to construct the Poisson distributions using the parameters from **Section 3.1**. Here, we will choose to use a composite structure. The composite structure is great for the goal-scoring model since it allows teams to vary in their abilities to score and prevent goals. This is shown by having θ_{oj} and θ_{dj} parameters separate offensive and defensive power.

There are two levels to the model. The first level uses the Poisson distribution to represent the mean number of goals scored for the home and road teams. The second level of the model will help facilitate the inference of the first level by giving the θ and δ parameters (i.e. $\theta_{oj}, \theta_{dj}, \delta_{oj}, \delta_{dj}$) some distribution.

Level-1 Model

Let Y_{hi} be the number of goals scored for the home team for game i, where $i = 1, 2, ..., N$. Let Y_{ri} be the number of goals scored for the road team for game i. Let there be a total of k teams. Using the parameters outlined in **Section 3.2**, we can model the Y's with the following parametric distributions:

$$Y_{hi}|\theta_{oh_i}, \theta_{dr_i}, \delta_{oh_i}, \delta_{dr_i} \sim \text{Poisson}(\theta_{oh_i} + \theta_{dr_i} + \delta_{oh_i} + \delta_{dr_i}), \qquad i = 1, 2, ..., N \tag{8}$$

$$Y_{ri}|\theta_{or_i}, \theta_{dh_i}, \delta_{oh_i}, \delta_{dr_i} \sim \text{Poisson}(\theta_{or_i} + \theta_{dh_i}), \qquad i = 1, 2, ..., N \tag{9}$$

Intuitively, these distributions make sense. As mentioned before, the Poisson Model is a good crude estimate for modeling goal-scoring when there is some fixed time parameter. Equation (8) shows that goal scoring for the home team is conditioned on the offensive power of the home team, the defensive power of the road team, and the home-field advantage parameters. Thus, when we want to find the mean goals scored of the home team, we add together the offensive power of the home team, the defensive power of the road team, and the home-field advantage parameters. Equation (9) shows that goal scoring for the road team is conditioned on the offensive power of the road team, the defensive power of the home team, and the home field advantage parameters. When we want to calculate the mean goals scored of the away team, it is simply the sum of the offensive power of the road team and the defensive power of the home team. Note that for the road team, we do not add the home-field advantage parameters since the road team does not have such an advantage.

From equations (8) and (9), we can clearly see that the road team is disadvantaged in terms of scoring. Now, the Poisson model (or in this case the Composite Poisson model) coincides with the findings from literature[3].

We can think of these distributions as the *likelihood* (or data) of our model. However, under the Bayesian paradigm, we need starting points for the θ and δ parameters, so we include another level to the model to help us set the θ's and δ's. We can consider these as *prior* distributions.

Level-2 Model

Once again, let us assume that there are k teams. In the second level of the model, we model

[3] Recall in **Section 2.1** that the literature points to a definite scoring discrepency between scoring at home vs. scoring on the road.

the θ's and δ's using a Gamma distribution. We assume the Gamma distribution because it has nice conjugate properties in conjunction with the Poisson distribution. Thus we arrive at the following distributions for the j-th teams:

$$\theta_{oj}, \theta_{dj} \overset{iid}{\sim} \text{Gamma}\left(\alpha_\theta, \frac{\alpha_\theta}{\mu_\theta}\right), \qquad j = 1, 2, ..., k \tag{10}$$

$$\delta_{oj}, \delta_{dj} \overset{iid}{\sim} \text{Gamma}\left(\alpha_\delta, \frac{\alpha_\delta}{\mu_\delta}\right), \qquad j = 1, 2, ..., k \tag{11}$$

where μ is the mean of θ and δ respectively[4], and α is the degree of concentration about μ. By construction, the θ's and δ's have mean μ and variance μ^2/α.

3.3 Gibbs Sampling Distributions

This section will cover the derivations of the Gibbs sampling distributions, particularly the derivation of the joint posterior distribution which the Gibbs sampling algorithm will draw from. First, we will further partition the distributions of Y. Then, we will derive the joint posterior distribution.

3.3.1 Further Partitions of Y

While in theory, we would like to just directly use the distributions given in the Composite Poisson model, it is more practical to further partition Y to have much nicer distributions to deal with. In addition, these partitions will also make the algorithm easier to follow. We assume the independence of the Poisson distributions, so we can simply add together the distributions and parameters.

For *home goals*, we let $Y_{hi} = Y_{hoi} + Y_{rdi} + Y_{hai} + Y_{rai}$, for games $i = 1, ..., N$. Thus, the

[4] Let $X \sim \text{Gamma}(\alpha, \beta)$. The mean of X is

$$E(X) = \frac{\alpha}{\beta}$$

and the variance is

$$Var(X) = \frac{\alpha}{\beta^2}$$

Thus, in this case, the mean and variance for θ and δ are given respectively:

$$E(\theta \text{ or } \delta) = \frac{\alpha}{\alpha/\mu} = \alpha \cdot \frac{\mu}{\alpha} = \mu$$

$$Var(\theta \text{ or } \delta) = \frac{\alpha}{(\alpha/\mu)^2} = \alpha \cdot \frac{\mu^2}{\alpha^2} = \frac{\mu^2}{\alpha}$$

distributions for the partitions are as follow:

$$Y_{hoi}|\theta_{oh_i} \sim \text{Poisson}(\theta_{oh_i}) \quad \text{(Partition 1)}$$
$$Y_{rdi}|\theta_{dr_i} \sim \text{Poisson}(\theta_{dr_i}) \quad \text{(Partition 2)}$$
$$Y_{hai}|\delta_{oh_i} \sim \text{Poisson}(\delta_{oh_i}) \quad \text{(Partition 3)}$$
$$Y_{rai}|\delta_{oh_i} \sim \text{Poisson}(\delta_{dr_i}) \quad \text{(Partition 4)}$$

Intuitively, these partitions similarly describe the Level-1 model described in **Section 3.2**. Partitions 1 and 2 reflect the scoring and defensive parameters of some home team. Partitions 3 and 4 reflect the home-field advantage for the home team. Thus, the number of goals scored by the home team, Y_{hi}, is the sum of Partitions 1-4.

We can do the same for *road goals*. Here, we let $Y_{ri} = Y_{roi} + Y_{hdi}$ for games $i = 1, ..., N$. Thus, the distributions for these partitions are as follow:

$$Y_{roi}|\theta_{or_i} \sim \text{Poisson}(\theta_{or_i}) \quad \text{(Partition 5)}$$
$$Y_{hdi}|\theta_{dh_i} \sim \text{Poisson}(\theta_{dh_i}) \quad \text{(Partition 6)}$$

Once again, intuitively, these partitions describe the Level-1 model described in **Section 3.2**. Partitions 5 and 6 reflect the scoring and defensive parameters of some road team. The number goals scored by the road team, Y_{ri}, is the sum of partitions 5-6.

Comparing Y_{hi} and Y_{ri}, we again see the property that the home team scoring more goals than the road team scoring, which is reflected in the literature.

Everson and Goldsmith-Pinkham take the model further by exploiting the property of the Poisson model: by using Poisson distributions to approximate Binomial distributions and continuing to use Poisson additivity. By doing so, we create a new conjugate prior-posterior pair: the "Beta-Binomial". While the analyses come out to be the same, it is important to note that this step is unnecessary. We can easily get at what we want by using the partitions of the Poisson distributions.[5]

3.3.2 Deriving the Joint Posterior Distribution

Now, we want to derive the joint posterior for the model. This is the distribution that the Gibbs sampling will converge to. We break this down into inference of δ and θ. The paper goes through this, but I still want to highlight some key aspects of the derivation.[6]

First, it is impotant we choose the correct prior. Consider the inference of δ. Our goal is to find some posterior density for δ, so after deriving the likelihood for δ, we want to consider picking a prior for δ. The prior density we want to choose is *non-informative*, that is, it adds little to no information, and *proper* such that the posterior comes out "nice". Consider the inference for θ. Once again, we want to pick a prior such that it is non-informative and

[5]I actually talked to Phil Everson about this, and even he said he kind of got "carried away" when he decided to use the Binomial distributions to approximate Poisson distributions.

[6]The derivation of the Joint Posterior distribution is found in the paper: Everson, Phil and Goldsmith-Pinkham, Paul S. (2008). "Composite Poisson Models For Goal Scoring". *Journal Of Quanititative Analysis In Sports.* Volume 4, Issue 2. `http://works.swarthmore.edu/fac-math-stat/124`, pp.6-9

"nice". The methodology is the same, and we can come up with the likelihood in terms of λ, thus finding the right the prior for λ.[7]

Second, it is important to not let the math bog down the core mechanics of Bayesian inference. At the end of the day, we are looking for the following results:

$$f(\delta|\hat{\delta}) = p(\delta)L(\delta) \qquad \text{(Inference for } \delta\text{)}$$

$$f(\theta|\lambda) = p(\lambda)L(\lambda) \qquad \text{(Inference for } \theta\text{)}$$

where $p(\cdot)$ are prior functions and $L(\cdot)$ are likelihood functions. The parameters λ and $\hat{\delta}$ come from the estimation changes defined in **Section 3.3.1**.

3.4 Gibbs Sampling Algorithm

Here, we use Partions 1-6 in **Section 3.3.1** and our priors, given by equations (10) and (11), to generate the steps for the Gibbs sampling algorithm. The algorithm is given by the following steps:

0. Determine starting values for the θ_{oj}, θ_{dj}, δ_{oj}, and δ_{dj} for the k teams
1. Given θ's and δ's, generate the random partitions of Y_{hi} into $Y_{hoi}, Y_{doi}, Y_{hai},$ and Y_{rai} and Y_{ri} into Y_{roi} and Y_{hdi}
2. Use the partitioned Y_i values to generate estimates of $\hat{\theta}_{oj}, \hat{\theta}_{dj}, \hat{\delta}_{oj}$ and $\hat{\delta}_{dj}$.
3. Use $\hat{\theta}_{oj}, \hat{\theta}_{dj}, \hat{\delta}_{oj}$ and $\hat{\delta}_{dj}$ to generate $(\alpha_\theta, \mu_\theta)$ and $(\alpha_\delta, \mu_\delta)$ pairs from the joint posterior distribution
4. Generate new values of θ_{oj}, θ_{dj}, δ_{oj}, and δ_{dj} from the conditional posterior distributions given $\hat{\theta}$'s, $\hat{\delta}$'s, and $(\alpha_\theta, \mu_\theta)$ and $(\alpha_\delta, \mu_\delta)$ pairs.
5. Repeat steps 1-4 until convergence.

We assume that the parameters converge after 1000 iterations.

To clarify the notation of the algorithm, the values with hats are estimates from the iteration and the values without hats are unknown probability distributions. Like we explained in **Section 2**, the MCMC should converge to the joint probability distribution derived in the previous section.

Now, we want to determine how to determine the starting values. We set δ to be the difference between mean home team score and mean road team score in all the games. Mathematically:

$$\delta^{(0)} = \frac{\sum Y_{hi} - \sum Y_{ri}}{N} \tag{12}$$

We then correct Y_{hi} by subtracting off δ. Now, we set θ_{oj} to be half the goals scored by team j and θ_{dj} to be half the goals allowed by team j in the team's n games for all k teams. So for each team:

$$\theta_{oj}^{(0)} = \frac{(\sum Y_{hi} - \delta^{(0)}) + \sum Y_{ri}}{2n} \tag{13}$$

[7]The priors chosen are called *Jeffrey's Prior*. These are special priors that are defined as the square root of the negative expected second derivative of the log-likelihood function. Sounds like a load of jargon, but the key thing to grasp is we use this prior because it is non-informative.

$$\theta_{dj}^{(0)} = \frac{\sum \text{Points Against}_{ij}}{2n} \quad (14)$$

Intuitively, since δ reflects home-field advantage, so it makes sense that it starts as being the difference between the mean home goals and mean road goals. θ values are a function of scoring and points allowed means, and so it also makes sense that they start as simply being the mean of points scored and mean points against.

3.5 Model Results

In this section, we cover the results of the Gibbs sampling algorithm in **Section 3.4** (shown in **Table 1**). The data Everson and Goldsmith-Pinkham use is the 2006 English Premier League soccer data. This dataset is ideal in terms of experiemental design since in the EPL, teams play each other twice, once at home and once on the road. The teams are ranked by strength from estimating θ_o and θ_d, and the actual standings are also shown for comparison.

id	Team Name	θ_o	θ_d	Strength	Final Standing
6	Man United	1.18	0.16	1.03	1
5	Chelsea	0.84	0.22	0.62	2
8	Arsenal	0.73	0.23	0.50	4
3	Portsmouth	0.58	0.24	0.34	3
7	Aston Villa	0.47	0.26	0.21	11
10	Wigan	0.62	0.44	0.18	17
16	Everton	0.53	0.38	0.16	6
18	Bolton	0.40	0.24	0.16	7
12	Fulham	0.51	0.41	0.10	16
2	Man City	0.21	0.28	-0.07	14
1	Liverpool	0.32	0.4	-0.08	3
9	Charlton	0.31	0.42	-0.11	19
19	Tottenham	0.24	0.36	-0.12	5
17	Reading	0.41	0.59	-0.18	8
13	Watford	0.22	0.44	-0.21	20
11	Middlesbrough	0.30	0.51	-0.22	12
15	Newcastle	0.19	0.43	-0.25	13
20	Blackburn	0.26	0.52	-0.26	10
4	West Ham	0.22	0.55	-0.33	15
14	Sheffield United	0.18	0.55	-0.38	18

Table 1: 2006 English Premier League (EPL) predicted standings vs actual standings. Figure taken from Everson and Goldsmith-Pinkham (2007).

Based on the results, the MCMC model properly predicted the first and second place teams perfectly. The teams in the top four place and bottom teams were all fairly close, but the model doesn't perform as well once we get down to the middle-tier teams. The shortcoming of the model can be summed up with the following (and is true in pretty much any sport): *just because you score the most points doesn't mean you finish near the top.*

More specifically, we can come up with possible errors within the model itself:

1. The possibility of "lucky" draws when drawing samples from the conditional distributions.

2. The lack of variables indicating great play. This may include strings of passes, the effectiveness of having some "star player", and physical defensive play.

Error 1 comes from the possibility of drawing a value in the tail of the distribution (where the value is high and the probability of drawing the value is low but still possible). As we can imagine, using a very big value can impact the next distribution we draw from, so having this happen a couple times can lead to higher (or lower) than usual θ's. Error two reflects parameters we can't see in the model. So far, the model only uses goals scored and goals allowed to predict offensive and defensive power, but there are other factors that also affect offensive and defensive play. Good offensive flow, such as stringing a series of passes or having a star striker dribble through a defense, should positively influence θ_o and thus contribute to a team's final standings. On the flip side, good defense, such as number of saves a goalkeeper makes or not allowing the other team's offense to take shots at the goal, should positively influence θ_d and thus also contribute to a team's final standings. Clearly, these factors are not reflected in the model, and so we end up with potentially wrong final standings.

4 AUDL and the Goal Scoring Model

In this section, we want to see if the framework of the model that Everson and Goldsmith-Pinkham lay out can be applied to a lesser known sport: professional Ultimate Frisbee (otherwise referred to as the American Ultimate Disc League or AUDL for short). The AUDL follows the properties of soccer: the score is measured by single points and the games are timed. To do this, we simplify the model and iterate the MCMC algorithm to see if we can get similar results.

4.1 Background

Before we get into the model, we first want to give a brief overview of Ultimate Frisbee. Briefly, the rules are as follows:

1. There are 7 players on the field per team, so there are a total of 14 players on the team.
2. The offense moves the disc down the field through throwing it from player to player. A player with the disc cannot run with the disc, and can only throw the disc with a pivot foot.
3. The offense loses possession of the disc when a pass is not completed (dropped, defensive block, thrown out of bounds, etc.).
4. A team scores 1 point when a player catches the frisbee in the endzone.

In collegiate and club levels, the games are not timed, but rather go until a team reaches 13 points or hard cap goes off (time cap that occurs after the game starts). Because of the scoring, this would break the model.[8] Luckily, Professional Ultimate Frisbee, most notably AUDL which is founded in 2010, is played with 4, 12-minute quarters, which lends itself

[8]In conversation with Paul Goldsmith-Pinkham, this makes modeling really hard since we can't use the Poisson Distribution due to not having a consistent fixed time period.

to using the Poisson distribution. Teams in the AUDL play each other within a division (Midwest, South, West, East) 3 times over the course of the regular season. At the end of the regular season, there is a playoff bracket to crown the year's champion.

4.2 Building the Model

The data we use in this exercise is the AUDL 2017 Season Data. The data is scraped from the AUDL website, and includes win-loss, division, points scored, points allowed, and point differential.

Before we begin to build our model, we create a set of assumptions. Because games are played regionally only and traveling is quite limited, we will assume that home and road teams have equal opportunity to score, so we drop the home-field advantage parameter. Next, we assume that an expert from the AUDL gives us a some prior information, so our prior estimates are consistent throughout the model (one α and μ value). With these two assumptions we can build our model.

Without homefield advantage, our likelihood function becomes:

$$Y_{ij}|\theta_{oj}, \theta_{dj} \sim \text{Poisson}(\theta_{oj} + \theta_{dj}) \tag{15}$$

where θ_{oj} and θ_{dj} are independent. The priors are given by

$$\theta_{oj}, \theta_{dj} \overset{iid}{\sim} \text{Gamma}\left(\alpha, \frac{\alpha}{\mu}\right) \tag{16}$$

We can find our posterior by multiplying the likelihood with the priors and get $\theta_{oj}, \theta_{dj}|Y_{ij}$. Since θ_{oj} and θ_{dj} are independent, we can partition the likelihood into $Y_{oij}|\theta_{oj}$ and $Y_{dij}|\theta_{dj}$. Multiplying the paritioned Y_{ij} distributions with the matching prior will get us the following:

$$\theta_{oj}|Y_{oij} \propto \theta_{oj}(Y_{oij}|\theta_{oj}) \tag{Offense}$$
$$\theta_{dj}|Y_{dij} \propto \theta_{dj}(Y_{dij}|\theta_{dj}) \tag{Defense}$$

Because these distributions are conjugates, the posteriors are

$$\theta_{oj}|Y_{oij} \overset{iid}{\sim} \text{Gamma}\left(\sum Y_{oij} + \alpha, n_j + \frac{\mu}{\alpha}\right) \tag{17}$$

$$\theta_{dj}|Y_{dij} \overset{iid}{\sim} \text{Gamma}\left(\sum Y_{dij} + \alpha, n_j + \frac{\mu}{\alpha}\right) \tag{18}$$

where n_j is the number of games played by each team.[9]

Before we fill out the algorithm, we need to identify our starting conditions. We set the

[9]Here, n_j might be different for each team since the data scraped contains playoff games, so teams that make the playoffs playoff games. We are still assuming independence from game to game.

following starting values:

$$\theta_{oj}^{(0)} = \frac{\text{Total Points Scored}}{2n_j}$$
$$\theta_{dj}^{(0)} = \frac{\text{Total Points Allowed}}{2n_j}$$
$$\alpha = 0.01$$
$$\mu = \text{mean(Total Points)}$$

θ_{oj}, θ_{dj} are set similar to that of the Everson - Goldsmith-Pinkham model. α is set to be a non-informative prior. μ is set to be the mean points, as described in the Everson- Goldsmith-Pinkham paper. Both α and μ are our prior specifications, and can be subjected to change depending on the information given by an expert.

Now that we have our conditional distributions, we can set up the iterative Gibbs sampler steps.

0. Start with $\theta_{oj}^{(0)}$ and $\theta_{dj}^{(0)}$. Let $Y_{oij}^{(0)}$ start as the total number of goals scored and $Y_{dij}^{(0)}$ start as the total number of goals allowed for team j.
1. Draw $\theta_{oj}^{(1)}$ and $\theta_{dj}^{(1)}$ given $Y_{oij}^{(0)}$ and $Y_{dij}^{(0)}$ from Gamma $\left(\sum Y_{oij}^{(0)} + \alpha, n_j + \frac{\mu}{\alpha}\right)$ and Gamma $\left(\sum Y_{dij}^{(0)} + \alpha, n_j + \frac{\mu}{\alpha}\right)$ respectively.
2. Draw $Y_{oij}^{(1)}$ and $Y_{dij}^{(1)}$ given $\theta_{oj}^{(1)}$ and $\theta_{dj}^{(1)}$ from Poisson$(\theta_{oj}^{(1)})$ and Poisson$(\theta_{dj}^{(1)})$ respectively.
3. Repeats step 1 and 2 until convergence.

Once again, we assume that convergence occurs after 1000 iterations.

4.3 Results and Analysis

First, we check the results of the posterior θ distributions to make sure they follow a gamma distribution. Figures 1 and 2 are depict the posterior distributions of θ_o and θ_d respectively, which both follow the shape of a gamma distribution.[10]

Table 2 gives the results of the algorithm on the AUDL 2017 season data. Because the standings divides the 24 teams into 4 divisions, we compare the predicted standings to the actual standings by division. Immediately, we notice a couple things. First, the θ's are much bigger. This is because in AUDL Ultimate Frisbee games are higher scoring. Second, the South and West division prediiction was the best while the Midwest division prediction was really bad. In the Eastern division, the MCMC algorithm correctly predicted the first and last place teams. However, we can still see that the results were similar to those of **Table 1**. Once again, we see that the model best predict top and bottom teams, but fail to closely provide accurate estimates of the middle teams.[11]

[10]We could have burned some iterations, which was not done here. A burn is taking out an iteration at some interval (i.e. every 10th, 50th, or 100th iteration) in order to improve independence.

[11]I realize that the Midwest Division results were completely wrong and the West Division results properly guessed the top 3 and bottom 3 teams in the wrong order.

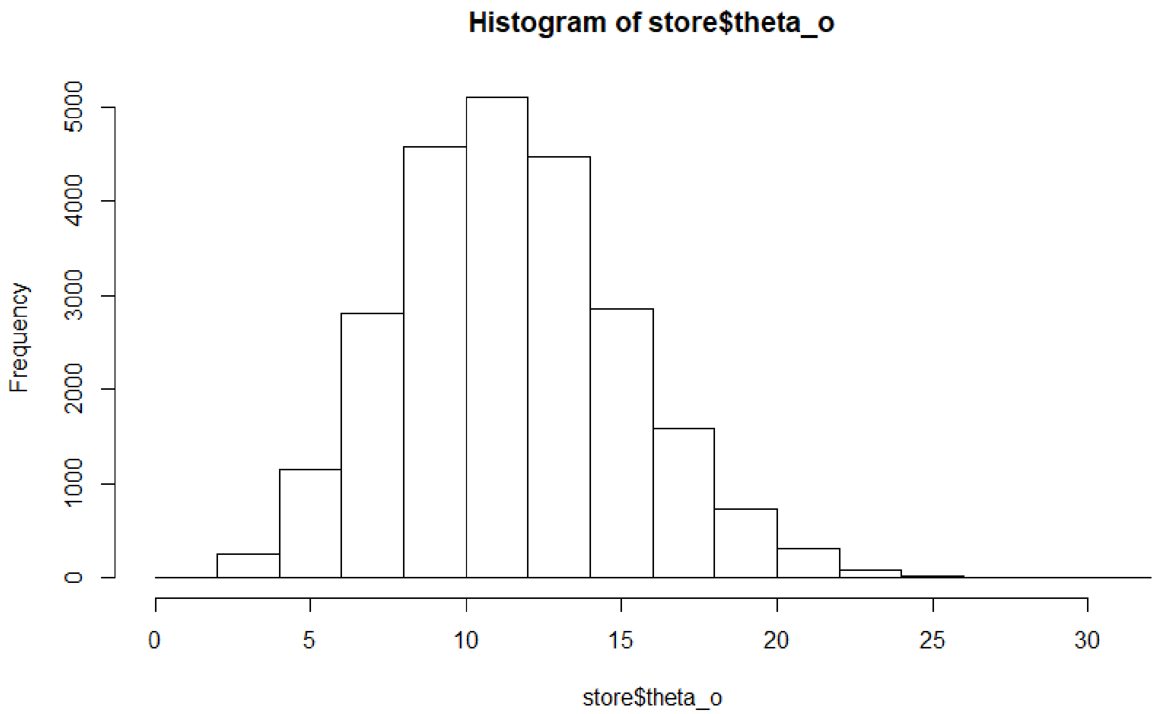

Figure 1: Distribution of posterior θ_o draws.

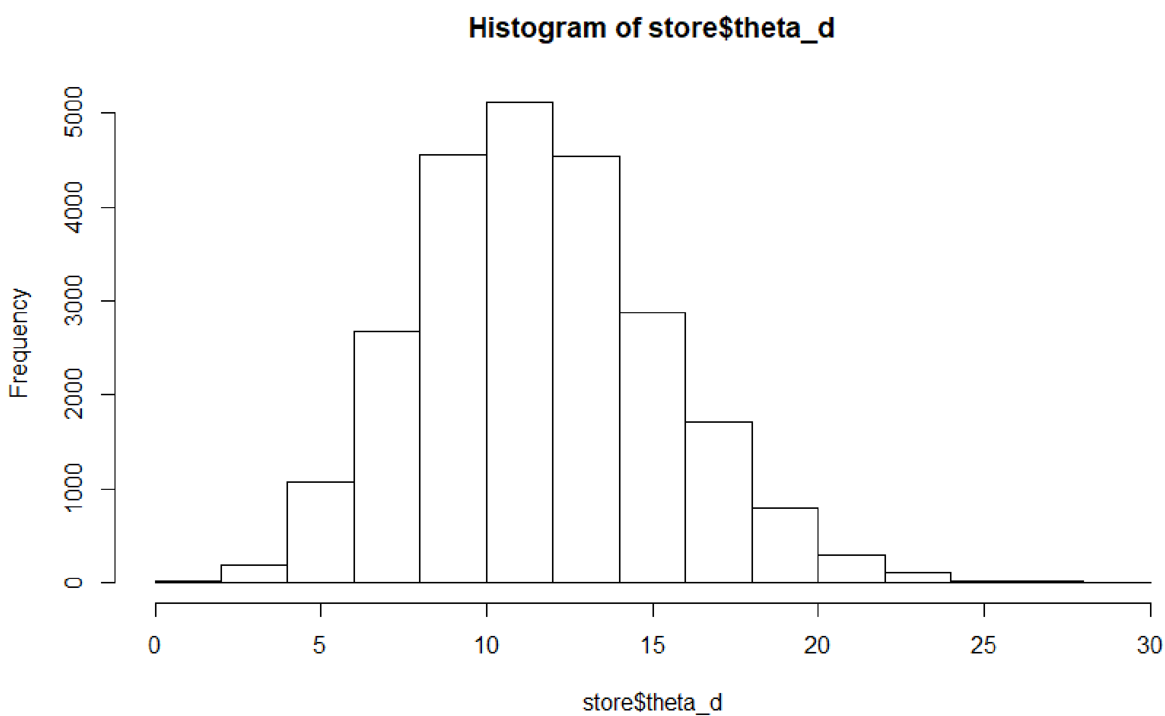

Figure 2: Distribution of posterior θ_d draws.

Team	Division	theta_o	theta_d	strength	actual rank
Toronto Rush	East	13.0545	8.421165	4.633332	1
New York Empire	East	9.627919	7.371351	2.256568	4
Philadelphia Phoenix	East	9.26103	10.42659	-1.16556	5
DC Breeze	East	8.143843	11.49049	-3.34665	2
Montreal Royal	East	9.071118	13.86056	-4.78944	3
Ottawa Outlaws	East	7.177045	22.83352	-15.6565	6
Chicago Wildfire	Midwest	10.47871	9.458743	1.019965	5
Minnesota Wind Chill	Midwest	8.093713	8.790123	-0.69641	2
Detroit Mechanix	Midwest	8.211483	9.663305	-1.45182	6
Madison Radicals	Midwest	9.728869	14.18819	-4.45932	1
Indianapolis AlleyCats	Midwest	6.888513	13.37453	-6.48601	4
Pittsburgh Thunderbirds	Midwest	3.46643	12.92368	-9.45725	3
Dallas Roughnecks	South	16.21295	6.856465	9.356489	2
Raleigh Flyers	South	13.59556	9.842208	3.753354	1
Austin Sol	South	11.26263	9.718254	1.54438	5
Jacksonville Cannons	South	9.893341	9.983914	-0.09057	3
Atlanta Hustle	South	9.37349	12.75551	-3.38202	4
Nashville Nightwatch	South	8.267487	16.26286	-7.99537	6
LA Aviators	West	21.33397	11.62847	9.705502	2
San Jose Spiders	West	16.06393	13.70414	2.359793	3
San Francisco FlameThrowers	West	11.75547	12.05126	-0.29579	1
Seattle Cascades	West	8.27846	12.73397	-4.45551	4
Vancouver Riptide	West	5.40495	9.876772	-4.47182	6
San Diego Growlers	West	6.720606	15.30406	-8.58346	5

Table 2: 2017 American Ultimate Disc League (AUDL) predicted standings vs actual standings by division

In the future, to improve the predictions, we could better use the prior α parameter. In the algorithm above, we set α to be uninformative, so we are allowing values to vary randomly. In other words, there is no concentration around μ. By allowing α to be non-zero, we will get different results since we allow for the draws closer around the μ, which means the concentration around μ increases .

Furthermore, we could include a weighted home field advantage parameter, δ, since the games are played at a 2-1 home-road split. We can hypothesize that the δ parameter will help fix some of the errors in the simplified algorithm. Also, we hope to have better access to game-by-game data, which could help cut down some of the assumptions being made earlier. This data could also help us set better prior parameters for better posterior results.

5 Conclusion

In conclusion, in both using the Everson - Goldsmith-Pinkham and simplified algorithms, we see that the model does well in predicting the top and bottom teams, but doesn't do well in predicting the middle teams. We attribute the error to three possible things: lucky draws from the intermediate steps of the Gibbs sampler, not accounting for all parameters that impact scoring, and, in the case of the simplified algorithm, lack of great data. These errors can be potentially resolved upon making changes to the original model.

Scoring is an important aspect of any sport, so how would this model help a sports analyst (or I guess also a well-informed, die-hard sports fan)? It is true that the Gibbs sampling algorithm used *ex post* season scoring, that is the algorithm is applied to "predict" the season standings after the season has happened. The first motivation of this is to merely see how well the model performs compared to the true result. In reality, we can use the previous season's θ's and set some arbitrary number of points a team might score per game, then iterate through the algorithm.[12] The second motivation highlights the beauty of Bayesian Statistics: we can simulate a sports season multiple times. For example, while the results shown in **Figure 2** only highlights one possible ranking of the 2017 AUDL season, we can imagine running the algorithm another 1000 times, each with different $\hat{\theta}_{oj}$ and $\hat{\theta}_{dj}$ estimates. In fact, running the algorithm 1000 times (each with different draws in the intermediate steps of the Gibbs sampler) may even help fix the absolutely crazy standings predictions in some divisions! Under the a standard statistics approach, as one might learn in an introductory statistics course, we would need the same season to occur 1000 times, then derive estimates by sampling, which is clearly impossible in reality. While under the Bayesian paradigm it is possible, and usually a really good idea, to simulate the season 1000 times, we can usually get at the right "answer" through one run of the algorithm.

[12] Here, we assume all else constant (same players, same level of play, etc.). It will be an interesting exercise/ extension to see what changes when certain factors change, like if a player gets traded or signs with a new team.

References

1. Blitzstein, Joseph and Hwang, Jessica. (2015) *Introduction to Probability.* CRC Press/Taylor & Francis Group.

2. Everson, Phil and Goldsmith-Pinkham, Paul S. (2008). "Composite Poisson Models For Goal Scoring". *Journal Of Quanititative Analysis In Sports.* Volume 4, Issue 2. `http://works.swarthmore.edu/fac-math-stat/124`

3. Pollard, Richard. (1985). "69.9 Goal-Scoring and the Negative Binomial Distribution." *The Mathematical Gazette* vol. 69, no. 447 pp. 45-47. JSTOR, JSTOR, `www.jstor.org/stable/3616453`.

4. Yildrim, Ilker. (2012). "Bayesian Inference: Gibbs Sampling". *Department of Brain and Cognitive Sciences, University of Rochester*