4-1-2008

Composite Poisson Models For Goal Scoring

Philip J. Everson
Swarthmore College, peverso1@swarthmore.edu

P. Goldsmith-Pinkham

Journal of Quantitative Analysis in Sports

Volume 4, *Issue* 2 2008 *Article* 13

Composite Poisson Models for Goal Scoring

Phil Everson, *Swarthmore College*
Paul S. Goldsmith-Pinkham, *Swarthmore College*

Recommended Citation:
Everson, Phil and Goldsmith-Pinkham, Paul S. (2008) "Composite Poisson Models for Goal Scoring," *Journal of Quantitative Analysis in Sports*: Vol. 4: Iss. 2, Article 13.

DOI: 10.2202/1559-0410.1107

Composite Poisson Models for Goal Scoring

Phil Everson and Paul S. Goldsmith-Pinkham

Abstract

Goal scoring in sports such as hockey and soccer is often modeled as a Poisson process. We work with a Poisson model where the mean goals scored by the home team is the sum of parameters for the home team's offense, the road team's defense, and a home advantage. The mean goals for the road team is the sum of parameters for the road team's offense and for the home team's defense. The best teams have a large offensive parameter value and a small defensive parameter value. A level-2 model connects the offensive and defensive parameters for the k teams. Parameter inference is made by imagining that goals can be classified as being strictly due to offense, to (lack of) defense, or to home-field advantage. Though not a realistic description, such a breakdown is consistent with our model assumptions and the literature, and we can work out the conditional distributions and generate random partitions to facilitate inference about the team parameters. We use the conditional Binomial distribution, given the Poisson totals and the current parameter values, to partition each observed goal total at each iteration in an MCMC algorithm.

KEYWORDS: gibbs sampler, jeffreys' prior, latent variables, poisson additivity, soccer, two-level model

1. Introduction and Summary

The sparse scoring in goal scoring sports such as hockey, lacrosse and soccer suggests a Poisson model might be a good description. The Poisson distribution arises as the limit of a Binomial distribution as the number of trials grows and the success probability shrinks in such a way that the mean number of "successes" approaches a constant θ. Each minute (e.g.) of a contest could be thought of as a "trial" with some probability of yielding a goal for either team. Of course, multiple goals could be scored in a single minute, and goals scored (or not scored) in different minutes are probably not independent outcomes with the same probability. The inter-arrival time for goals in soccer does not follow an exponential distribution, as is required for the Poisson process. However, a Poisson model is common in the literature for predicting soccer scores, such as in [2], and Figure 1 shows that the number of goals scored in the 2005-2006 English Premier League (EPL) season is well-approximated by a Poisson distribution with mean 1.23. While we will use the EPL data as an illustrative example, the primary goal of this paper is to lay out the framework to carry out inference for an additive Poisson model that could be applied to a variety of sports and data sets.

The literature currently contains models estimating team strength with Poisson distributions [2, 3] as well as models with differing home-field advantages across teams [1]. However, to our knowledge, our model is the first to use a hierarchical framework with Poisson additivity and differing home-field advantages.

Section 2 outlines the most general model, which allows for team offensive and defensive strengths, as well as different team offensive and defensive home advantages. Offensive home advantage is the propensity to score more goals when at home than when on the road (or on a neutral field) when playing comparable opponents. Defensive home advantage is the propensity to yield fewer goals when playing at home. If a common home advantage is assumed for all teams, as is commonly assumed for ranking purposes, then it is not possible to distinguish between offensive and defensive advantage, and a common offensive advantage is fit, with the defensive advantage set to zero. Section 3 describes a Gibbs sampler algorithm to estimate the parameters for the full model, and adjustments for fitting simpler models, such as the common home advantage or no home advantage models. Section 4 summarizes the inference for the EPL data and discusses future directions.

2. The Composite-Poisson Model

Let Y_{hi} represent the goals scored for the home team in game i, $i = 1, \ldots, N$, and let Y_{ri} represent the goals scored by the road team in game i (with an arbitrary home/road designation in events such as the World Cup, in which neither team has an obvious home field advantage). We assume that each of the k teams has an offensive parameter θ_{oj} and a defensive parameter θ_{dj}, and that these combine with

offensive and defensive home field advantage parameters δ_{oj} and δ_{dj} to determine the mean number of goals scored by a particular team. Let h_i and r_i represent the indices for the home and road teams in game i, so h_i and r_i take values $1, \ldots, k$. Our model for the goals scored in game i is then:

Level-1 Model

$$Y_{hi} \mid \theta_{oh_i}, \theta_{dr_i}, \delta_{oh_i}, \delta_{dr_i} \sim \text{Poisson}(\theta_{oh_i} + \theta_{dr_i} + \delta_{oh_i} + \delta_{dr_i}), \quad i = 1, \ldots, N,$$

independent of

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

This composite mean structure allows teams to vary in their ability both to produce goals and to prevent goals. Team j is one of the better teams if it has a large θ_{oj}, meaning they tend to score a lot, and a small θ_{dj}, meaning they tend to not allow many goals. Figure 2 displays the fitted θ_{oj} and θ_{dj} values for the $k = 20$ teams in the English Premier League. These values were estimated using the procedure described in Section 3, and predicts that Manchester United and Chelsea as the strongest teams that season, and Sheffield United as the weakest team. See Table 1 in Section 3 for a comparison to the actual results.

The home advantage parameters δ_{oj} and δ_{dj} indicate how much stronger a team is when playing at home compared to playing on the road or on a neutral field. The model requires that teams play at least as well at home (i.e., no negative δ's) so the offensive and defensive home advantages both appears as non-negative additions $(\delta_{oh_i} + \delta_{dr_i})$ to the home team's mean. The defensive advantages parameters are essentially a penalty for not playing at home. By forcing the defensive advantage to be positive, we maintain the additive property of the Poisson distribution and avoid possibly fitting a negative mean. The neutral-field means are therefore

Neutral Field Means

$$Y_{hi} \mid \theta_{oh_i}, \theta_{dr_i}, \delta_{oh_i}, \delta_{dr_i} \sim \text{Poisson}(\theta_{oh_i} + \theta_{dr_i} + \delta_{dr_i}), \quad i = 1, \ldots, N,$$

independent of

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

There are $4k$ parameters in this level-1 model: two θ's and two δ's for each of the k teams. For a given set of parameters, we can achieve the same mean structure if we add a constant c to every θ_o (e.g.) and subtract c from every θ_d. To facilitate inference (and to recognize that the parameters for the different teams should be more similar than different) we assume common *conjugate* Gamma distributions for the θ_{oj}'s and θ_{dj}'s, and common Gamma distribution for the δ_{oj}'s and the δ_{oj}'s.

Level-2 Model

$$\theta_{oj},\, \theta_{dj} \overset{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha_\theta,\, \alpha_\theta/\mu_\theta), \qquad j = 1, \ldots, k,$$

DOI: 10.2202/1559-0410.1107

independent of

$$\delta_{oj},\, \delta_{dj} \overset{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha_\delta,\, \alpha_\delta/\mu_\delta), \qquad j = 1, \ldots, k.$$

The team θ's and δ's are modeled to have mean μ and variance μ^2/α, so μ defines a typical value for the k teams, and increasing α makes the level-2 distribution more concentrated about μ.

3. Parameter Inference

In reality, there is no way to classify goals as being strictly offensive or defensive, or due to home field advantage. However, the Level-1 model in Section 2 is equivalent to a model for which "offensive", "defensive" and "advantage" goals occur according to independent Poisson processes and add to give the goal totals:

Hypothetical Partitioning

$$\begin{aligned}
\textbf{Home Goals:} \quad Y_{hi} &= Y_{hoi} + Y_{rdi} + Y_{hai} + Y_{rai}, \qquad i = 1, \ldots, N. \\
Y_{hoi} \mid \theta_{oh_i} &\overset{\text{indep}}{\sim} \text{Poisson}(\theta_{oh_i}) \\
Y_{rdi} \mid \theta_{dr_i} &\overset{\text{indep}}{\sim} \text{Poisson}(\theta_{dr_i}) \\
Y_{hai} \mid \delta_{oh_i} &\overset{\text{i.i.d.}}{\sim} \text{Poisson}(\delta_{oh_i}) \\
Y_{rai} \mid \delta_{oh_i} &\overset{\text{i.i.d.}}{\sim} \text{Poisson}(\delta_{dr_i}) \\
\textbf{Road Goals:} \quad Y_{ri} &= Y_{roi} + Y_{hdi}, \qquad i = 1, \ldots N. \\
Y_{roi} \mid \theta_{or_i} &\overset{\text{indep}}{\sim} \text{Poisson}(\theta_{or_i}) \\
Y_{hdi} \mid \theta_{dh_i} &\overset{\text{indep}}{\sim} \text{Poisson}(\theta_{dh_i})
\end{aligned}$$

The equivalence follows from the additivity of the Poisson distribution. If total goals are modeled as following a Poisson distribution, then it is reasonable to think of goals as the sum of two or four independent Poisson random variables, with rates that add to the desired means.

The simplicity of the distributions for the partitioned totals suggests that, were we to learn the values of these hypothetical, *latent* variables, we could easily estimate the unknown θ's and δ's using standard Poisson inference techniques. For example, we could estimate the offensive parameter for team j by $\hat{\theta}_{oj}$, the average of the offensive totals for team j in its n_j games. Conversely, were we to know all of the θ's and δ's, we could make inference about the unknown partitioned counts. Here we exploit another nice property of Poisson additivity: conditional on the total count, the distribution of the partial counts are Binomial.

Conditional Partitions

$$\textbf{Home Goals:} \quad Y_{hi} = Y_{hoi} + Y_{rdi} + Y_{hai} + Y_{rai}, \qquad i = 1, \ldots, N.$$

$$Y_{hoi} \mid \theta, \delta, y_{hi} \sim \text{Binomial}\left(y_{hi},\ \phi_{hoi}\right), \qquad \phi_{hoi} = \frac{\theta_{oh_i}}{\theta_{oh_i} + \theta_{dr_i} + \delta_{oh_i} + \delta_{dr_i}}.$$

$$Y_{rdi} \mid \theta, \delta, y_{hi}, y_{hoi} \sim \text{Binomial}\left(y_{hi} - y_{hoi},\ \phi_{rdi}\right), \qquad \phi_{rdi} = \frac{\theta_{dr_i}}{\theta_{dr_i} + \delta_{oh_i} + \delta_{dr_i}}.$$

$$Y_{hai} \mid \theta, \delta, y_{hi}, y_{hoi}, y_{rdi} \sim \text{Binomial}(y_{hi} - y_{hoi} - y_{rdi},\ \phi_{hai}), \qquad \phi_{hai} = \frac{\delta_{oh_i}}{\delta_{oh_i} + \delta_{dr_i}}.$$

$$Y_{rai} \mid y_{hi}, y_{hoi}, y_{rdi}, y_{ha} = y_{hi} - y_{hoi} - y_{rdi} - y_{ha}$$

$$\textbf{Road Goals:} \quad Y_{ri} = Y_{roi} + Y_{hdi}, \qquad i = 1, \ldots N.$$

$$Y_{roi} \mid \theta, \delta, y_{ri} \sim \text{Binomial}\left(y_{ri},\ \phi_{roi}\right), \qquad \phi_{roi} = \frac{\theta_{or_i}}{\theta_{or_i} + \theta_{dh_i}}$$

As well as facilitating inference about the θ's, the simulated partitions of goal totals also enable us to make inference about μ_θ and α_θ, the level-2 parameters governing the distributions of the θ's, and about μ_δ and α_δ, the parameters governing home advantage. This leads to the following iterative simulation algorithm:

Outline of Gibbs Sampler Algorithm

(See appendix for details)

0. Determine starting values for the θ_{oj}'s, θ_{dj}'s, δ_{oj}'s and δ_{dj}'s, for $j = 1, \ldots, k$.

1. Use the θ's and δ's to generate random partitions of the Y_{hi}'s into Y_{hoi}'s, Y_{rdi}'s, Y_{hai}'s and Y_{rai}'s and of the Y_{ri}'s into Y_{roi}'s and Y_{hdi}'s.
2. Use the partitioned Y_i's to form estimates $\hat{\theta}_{oj}$, $\hat{\theta}_{dj}$, $\hat{\delta}_{oj}$ and $\hat{\delta}_{dj}$, for $j = 1, \ldots, k$.

DOI: 10.2202/1559-0410.1107

3. Generate $\{\alpha_\theta, \mu_\theta\}$ and $\{\alpha_\delta, \mu_\delta\}$ pairs from their joint posterior densities given the $\hat{\theta}_j$'s and $\hat{\delta}_j$'s using rejection sampling.

4. Generate θ_{oj}, θ_{dj}, δ_{oj}, δ_{dj}, $j = 1, \ldots, k$, from their conditional posterior distributions given the current $\hat{\theta}_j$'s, $\hat{\delta}_j$'s, $\alpha_\theta, \mu_\theta$, α_δ and μ_δ.

With appropriate prior specifications for $\alpha_\theta, \mu_\theta$, α_δ and μ_δ, repeating steps 1-4 constitutes a Gibbs sampler, and yields draws from the joint posterior distribution of the unknownθ's and δ's, given the goal totals. The details of the posterior calculations are given in the appendix.

4. Results of Estimation

In order to apply the algorithm described in the previous section, we took a dataset consisting of scores from the 2006 English Premier League season and split the dataset into two parts. Using the first 119 games, we generated estimates for each team's parameters: $\theta_o, \theta_d, \delta_o$, and δ_d. For ranking purposes, we used the simpler method with equal home-field advantage (HFA) parameters across teams, and then used the difference of the teams' θ_o and θ_d to rank them. The results can be seen in Table 1.

Using the full set of parameters generated with different team HFAs, we then predicted scores for the remaining 261 games in the season, and estimated the amount of points that each team would accumulate in the remaining games. Points in the EPL are given as follows: three points for a win, one point for a tie, and zero points for a loss. The predicted points for teams are given with 90% bands in Figure 2. The actual points fall within the 90% interval for 13 out of the 20 teams. The figure suggests that the fitted values were perhaps shrunk too strongly, with the most extreme observations falling outside of the prediction interval. This may imply an overestimate of the alpha's which could be a result of our prior specifications. It could also be an artifact of the conversion to the EPL point system, or a limitation of our model assumptions for this particular example.

We wish to stress that, while we have emphasized goal scoring in sports as the motivation for this paper, we believe this approach of partitioning Poisson counts will have applications in many situations where Poisson models are used. For example, multiple detectors generate counts of cosmic rays in the presence of background radiation. The counts are therefore the sum of true counts and false counts (which often have a "known" rate). To compare the rates of true counts at the two locations, the corrupted counts could first be partitioned using the methods of this paper (and in this case the partition is not hypothetical). With larger numbers of means to be estimated, a hierarchical structure facilitates "borrowing strength from the ensemble" to compensate for small sample sizes for individual teams, detectors or groups. An obvious extension of this method will be to fit parameters for teams' abilities to generate and allow shots, shots on goal, or some combined classification

such as "potential goals". Then parallel hierarchical Binomial models could be fit for each team's propensities to convert shots into goals and to allow shots to be converted into goals. The goal totals would still be modeled as Poisson, but additional information about shot totals would also be incorporated in the model.

Appendix:

Computational Steps:

0: Set $\delta_{\mathrm{o}j} = \delta_{\mathrm{d}j} = \sum_{i=1}^{N}(Y_{\mathrm{h}i} - Y_{\mathrm{r}i})/(2N)$.

$$\text{Set } \theta_{\mathrm{o}j} = \left(\eta \sum_{h_i=j} Y_{\mathrm{ho}i} + \sum_{r_i=j} Y_{\mathrm{ro}i}\right)/n_j, \qquad \eta = \frac{\sum_i^N (Y_{\mathrm{h}i} - Y_{\mathrm{r}i})}{\sum_i^N Y_{\mathrm{h}i}}.$$

$$\text{Set } \theta_{\mathrm{d}j} = \left(\sum_{h_i=j} Y_{\mathrm{hd}i} + \sum_{r_i=j} Y_{\mathrm{rd}i}\right)/n_j$$

in the equations specified below, but the general idea is contained in the above "Basic Algorithm." Without the partitions, we can't estimate the parameters. And without the parameters, we can't estimate the partitions. But given a reasonable starting guess at the parameters, we can iterate between these two operations, and learn about both the hypothetical partitions and the unknown parameters.

Prior and Posterior densities

The algorithm for using Binomial draws to partition the goal counts, given the team θ's and δ, is outlined earlier in this section. Given these breakdowns, we can form an unbiased estimate for δ by averaging the N $Y_{\mathrm{a}i}$'s, and for each $\theta_{\mathrm{o}j}$ and $\theta_{\mathrm{d}j}$ by averaging the appropriate partitioned counts for the n_j games involving team j. This gives the following sampling distributions:

$$n_j\hat{\theta}_{\mathrm{o}j} \,|\, \theta_{\mathrm{o}j} \overset{\text{indep}}{\sim} \text{Poisson}(n_j\theta_{\mathrm{o}j}), \quad j = 1, \ldots, k$$

$$n_j\hat{\theta}_{\mathrm{d}j} \,|\, \theta_{\mathrm{d}j} \overset{\text{indep}}{\sim} \text{Poisson}(n_j\theta_{\mathrm{d}j}), \quad j = 1, \ldots, k$$

$$N\hat{\delta} \,|\, \delta \sim \text{Poisson}(N\delta)$$

Inference for δ

The likelihood function for δ given $\hat{\delta}$ is

$$L(\delta) = \frac{(N\delta)^{N\hat{\delta}}\, e^{-N\delta}}{(N\hat{\delta})!} \propto \delta^{N\hat{\delta}}\, e^{-N\delta}, \quad \delta > 0.$$

The posterior density for δ is proportional to this likelihood function multiplied by a *prior* density for δ. We would like to choose a prior density that is *non-informative* in that it adds little or no information. We could specify an improper

DOI: 10.2202/1559-0410.1107

Uniform distribution for δ, but this would imply a non-uniform density for a non-linear transformation of δ. One approach is to assign a Uniform prior density to a transformation of δ for which the expected curvature of the log-likelihood function does not change for different values of $\hat{\delta}$. The implied density for δ is known as "Jeffreys' prior", and is defined as the square root of the negative expected second derivative of the log-likelihood function. The expectation is taken with respect to the data, given the parameter value.

Jeffreys' prior

$$\begin{aligned} p(\delta) &= -E\left(\frac{\partial^2}{\partial\lambda^2}(N\hat{\delta}\log(\delta) - N\delta)\right)^{1/2} = \left(-E\left(\frac{\partial}{\partial\lambda}\quad\frac{N\hat{\delta}}{\delta} - N\right)\right)^{1/2} \\ &= E\quad\frac{\partial}{\partial\lambda}\quad\frac{N\hat{\delta}}{\delta^2}\Bigg)\Bigg)^{1/2} = \left(\frac{N\delta}{\delta^2}\right)^{1/2} \propto \delta^{-1/2}, \quad \delta > 0. \end{aligned}$$

This is equivalent to specifying a Uniform prior density for $\sqrt{\delta}$. With this specification, the posterior density is recognizable as a Gamma$(N\hat{\delta} + 1/2,\, N)$ distribution:

$$f(\delta \,|\, \hat{\delta}) \;\propto\; \delta^{N\hat{\delta}+1/2-1}\, e^{-N\delta}, \quad \delta > 0.$$

We notice that, despite using an improper prior density (its integral is infinite), it is guaranteed that the posterior density will be proper. Even if a particular random partition assigned no "advantage" goals, meaning that $\hat{\delta} = 0$, the posterior would still be a proper Gamma$(1/2, N)$ density (equivalent to that of a scaled Chi-square(1) random variable).

Inference for the θ_j's

For estimating the θ_{oj}'s and θ_{dj}'s, we have two equivalent problems of the form:

$$X_j \,|\, \theta_j \quad \overset{\text{indep}}{\sim} \quad \text{Poisson}(n_j\theta_j), \quad j = 1, \ldots, k$$

$$\theta_j \,|\, \lambda \quad \overset{\text{i.i.d.}}{\sim} \quad \text{Exp}(\lambda), \quad j = 1, \ldots, k.$$

For example, for the θ_{oj}'s, each X_j represents $n_j\hat{\theta}_j$, the total number of "offensive" goals for team j in its n_j games.

Given λ, the posterior distribution for each θ_j is Gamma$(x_j + 1, n_j + \lambda)$:

$$f(\theta_j \,|\, X_j = x_j, \lambda) \;\propto\; \left(\frac{(n_j\theta_j)^{x_j}}{x_j!}\, e^{-n_j\theta_j}\right)\left(\lambda\, e^{-\lambda\theta_j}\right) \;\propto\; \theta_j^{x_j+1-1}\, e^{-(n_j+\lambda)\theta_j}, \quad \theta_j > 0.$$

The marginal likelihood function for the parameter λ can be found by integrating

the θ_j's out of the joint likelihood function:.

$$\begin{aligned} L(\lambda) \;=\; P(X_1 = x_1, \ldots, X_k = x_k \,|\, \lambda) \;&=\; \prod_{j=1}^{k} \left(\frac{n_j^{x_j}\lambda}{x_j!} \int_0^\infty \theta_j^{x_j+1-1} e^{(n_j+\lambda)\theta_j}\, d\theta_j \right) \\ &=\; \prod_{j=1}^{k} \left(\frac{\lambda}{n_j+\lambda} \right) \left(\frac{n_j}{n_j+\lambda} \right)^{x_j}, \quad \lambda > 0. \end{aligned}$$

This likelihood function simplifies considerably if all of the k teams have all played the same number of games, meaning $n_j = n$, $j = 1, \ldots, k$. This is the case for the entire EPL season, with $n = 38$ games for each of the $k = 20$ teams. There are also several times during the season at which point all of the teams have played the same number of games. We focus on this special case in order to simplify the computations.

With equal n_j's, it is convenient to work with the transformation $\phi = \lambda/(n+\lambda)$. Then the marginal probability distribution of each X_j is given by

$$\begin{aligned} P(X_j = x_j \,|\, \lambda) \;&=\; \frac{n^{x_j}\lambda}{x_j!} \int_0^\infty \theta_j^{x_j+1-1} e^{(n+\lambda)\theta_j}\, d\theta_j \\ &=\; \frac{n^{x_j}\lambda}{(n+\lambda)^{x_j+1}} \;=\; \phi(1-\phi)^{x_j}, \quad x_j = 0, 1, \ldots, \end{aligned}$$

This is the distribution of the number of failures (e.g.) before the first success for independent trials with success probability ϕ on each trial. The inverse transformation is $\lambda = n\phi/(1-\phi)$, so the expectation of each θ_j given ϕ is $1/\lambda = (1-\phi)/(n\phi)$. The expected value of each X_j given θ_j is $n\theta_j$, so the marginal mean of each X_j is

$$E(X_j \,|\, \phi) \;=\; E(E(X_j \,|\, \theta_j) \,|\, \phi) \;=\; E(n\theta_j \,|\, \phi) \;=\; \frac{1-\phi}{\phi}.$$

The marginal likelihood function for ϕ is

$$L(\phi) \;\propto\; \phi^k (1-\phi)^s, \quad s \;=\; \sum x_j, \quad 0 < \phi < 1.$$

This leads to the following log-likelihood function and derivatives. These define Jeffreys' prior distribution for the parameter $\phi = \lambda/(n+\lambda)$:

$$\log(L(\phi)) \;=\; c + k\log(\phi) + s\log(1-\phi), \quad 0 < \phi < 1.$$

$$\begin{aligned} \frac{\partial}{\partial\phi}\log(L(\phi)) \;&=\; \frac{k}{\phi} - \frac{s}{1-\phi}; \quad \frac{\partial^2}{\partial\phi^2}\log(L(\phi)) \;=\; -\frac{k}{\phi^2} - \frac{s}{(1-\phi)^2} \\ p(\phi) \;&\propto\; \left(\frac{k}{\phi^2} + \frac{k}{\phi(1-\phi)} \right) \;\propto\; \phi^{-1}(1-\phi)^{-1/2}, \quad 0 < \phi < 1. \end{aligned}$$

DOI: 10.2202/1559-0410.1107

Combining this prior density with the likelihood function for ϕ yields a posterior density that is Beta(k, $s + 1/2$):

$$\begin{aligned} f(\phi \,|\, X_1, \ldots, X_k) &= \frac{P(X_1 = x_1, \ldots, X_k = x_k \,|\, \phi) p(\phi)}{\int_0^1 P(X_1 = x_1, \ldots, X_k = x_k \,|\, \phi) p(\phi)\, d\phi} \;\propto\; L(\phi) p(\phi) \\ &\propto \phi^{k-1} (1-\phi)^{s+1/2-1}, \qquad 0 < \phi < 1. \end{aligned}$$

Once again we are guaranteed a proper posterior density. Even if a particular random partition assigned no "offensive" goals (for example) in any of the games, meaning all of the x_j's are 0 and therefore $s = 0$, the posterior for ϕ would still be a proper Beta(k, $1/2$) density.

We can now fill in the gaps in the Basic Algorithm:

Detailed Algorithm:

0. **Starting values:**
   Set $\delta^{(0)}$ to be the difference in the average home team score and the average road team score in the N games. After correcting every home goal total $Y_{\mathrm{h}i}$ by subtracting off $\delta^{(0)}$, set $\theta_{\mathrm{o}j}^{(0)}$ equal to be half the average goals scored by team j and $\theta_{\mathrm{d}j}^{(0)}$ to be half the average goals allowed by team j in its n games, for each $j = 1, \ldots, k$.

1. **Given $\boldsymbol{\theta}_{\mathrm{o}\boldsymbol{j}}^{(\boldsymbol{t-1})}$ and $\boldsymbol{\theta}_{\mathrm{d}\boldsymbol{j}}^{(\boldsymbol{t-1})}$, $\boldsymbol{j = 1, \ldots k}$, generate $\boldsymbol{\lambda}_{\mathrm{o}}^{(\boldsymbol{t})}$, $\boldsymbol{\lambda}_{\mathrm{d}}^{(\boldsymbol{t})}$, $\boldsymbol{\theta}_{\mathrm{o}\boldsymbol{j}}^{(\boldsymbol{t})}$ and $\boldsymbol{\theta}_{\mathrm{d}\boldsymbol{j}}^{(\boldsymbol{t})}$.**

2. **Random partitions:**
   Partition each home team goal total $Y_{\mathrm{h}i}$ into $Y_{\mathrm{ho}i}^{(t)}$, $Y_{\mathrm{rd}i}^{(t)}$ and $Y_{\mathrm{a}i}^{(t)}$ using $\theta_{\mathrm{o}h_i}^{(t-1)}$, $\theta_{\mathrm{d}r_i}^{(t-1)}$ and $\delta^{(t-1)}$ (the values at iteration $t-1$) to compute the probabilities for the breakdown.

3. **Conditional estimates:**
   Set $\hat{\delta}^{(t)} = \sum Y_{\mathrm{a}i}^{(t)} / N$.
   Set $\hat{\theta}_{\mathrm{o}j}^{(t)} = (\sum Y_{\mathrm{ho}h_i=j}^{(t)} + \sum Y_{\mathrm{ro}r_i=j}^{(t)})/n$ and $\hat{\theta}_{\mathrm{d}j}^{(t)} = (\sum Y_{\mathrm{hd}h_i=j}^{(t)} + \sum Y_{\mathrm{rd}r_i=j}^{(t)})/n$ for $j = 1, \ldots, k$, summing over the games i for which each team j is a home team ($\mathrm{h}_i = j$) or a road team ($\mathrm{r}_i = j$), as appropriate.

4. **Update $\boldsymbol{\lambda}$'s:**
   Compute $s_{\mathrm{o}}^{(t)} = \sum \hat{\theta}_{\mathrm{o}j}^{(t)}$ and $s_{\mathrm{d}}^{(t)} = \sum \hat{\theta}_{\mathrm{d}j}^{(t)}$.
   Generate $\phi_{\mathrm{o}}^{(t)} \sim \mathrm{Beta}(k,\, s_{\mathrm{o}}^{(t)} + 1/2)$ and $\phi_{\mathrm{d}}^{(t)} \sim \mathrm{Beta}(k,\, s_{\mathrm{d}}^{(t)} + 1/2)$.
   Set $\lambda_{\mathrm{o}}^{(t)} = n\phi_{\mathrm{o}}^{(t)}/(1-\phi_{\mathrm{o}}^{(t)})$ and $\lambda_{\mathrm{d}}^{(t)} = n\phi_{\mathrm{d}}^{(t)}/(1-\phi_{\mathrm{d}}^{(t)})$.

5. **Update $\boldsymbol{\theta}$'s:**
   Generate $\theta_{\mathrm{o}j}^{(t)} \sim \mathrm{Gamma}(\hat{\theta}_{\mathrm{o}j}^{(t)} + 1, n + \lambda_{\mathrm{o}}^{(t)})$ and $\theta_{\mathrm{d}j}^{(t)} \sim \mathrm{Gamma}(\hat{\theta}_{\mathrm{d}j}^{(t)} + 1, n + \lambda_{\mathrm{d}}^{(t)})$, for $j = 1, \ldots, k$.

The algorithm works by iterating between steps 1-4 until the simulated values appear to have converged to stationary distributions. There is extensive theory on how to determine this, but we simply ran 1000 iterations and did not see a reason to doubt that the Markov Chain had converged. After convergence, we collected the parameter values generated on every 100th iteration until we had collected 1000 sets of values. Taking only every 100th draw reduces the auto-correlation in the Markov Chain, and gives us a nearly independent sample from the joint posterior distribution of the unknown parameters.

DOI: 10.2202/1559-0410.1107

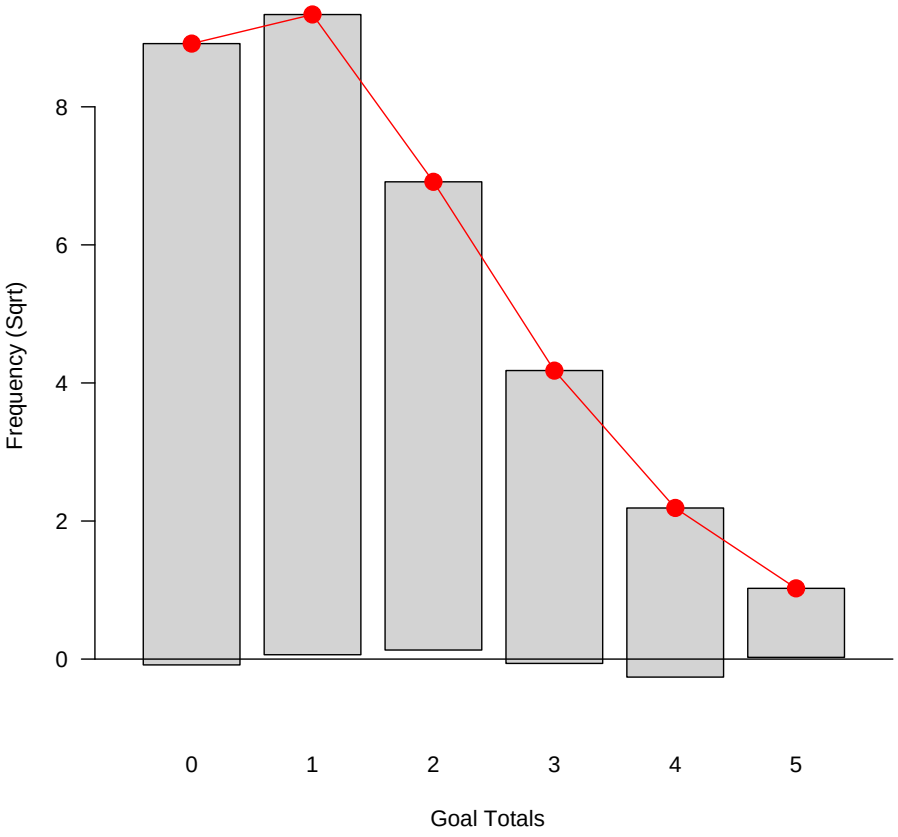

Figure 1: **Hanging Rootogram of 2006 EPL Goal Scoring.**
The square roots of the counts of goal totals for the two teams in each of the 380 matches played during the 2006 EPL season are compared to the expected root-counts for a fitted Poisson distribution. With a Poisson model, the counts of individual goal totals are distributed Binom$(n = 2 * 380, p(x))$, where $p(x)$ is the Poisson probability of a particular goal total x. The variance of each count is approximately $np(x)$, and a square-root transformation stabilizes the variances for the different values of x and $p(x)$. The dots mark the square roots of the expected Poisson counts for 760 draws from a Poisson distribution with mean 1.23 (the overall average of the 760 goal totals). The height of each bar corresponds to the square-root of the actual count and the bars "hang" from the dots. So the deviations from 0 indicate prediction errors, on a scale with roughly constant variance for all goal totals. The data are very consistent with a Poisson distribution.

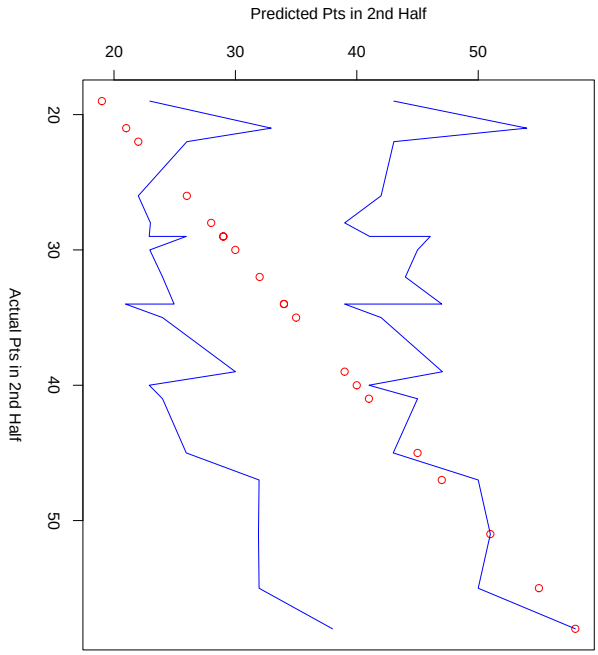

Figure 2: Red points represent the actual points accumulated in EPL. Blue lines represent the 90% interval predicted by algorithm.

DOI: 10.2202/1559-0410.1107

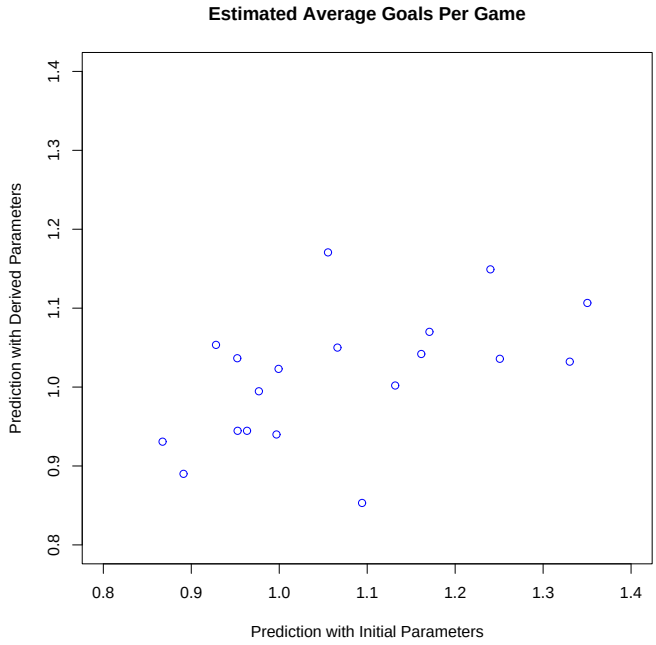

Figure 3: A comparison of predicted game scores using the initial parameters vs. the derived parameters.

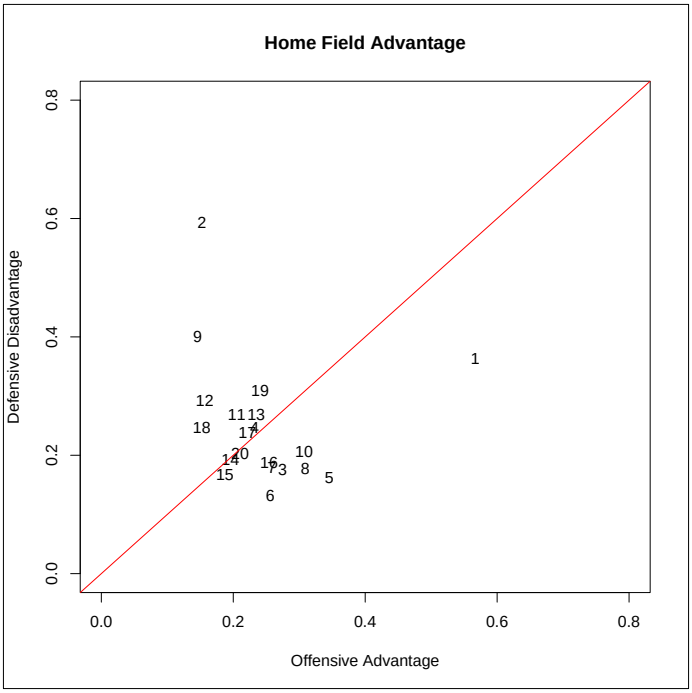

Figure 4: A comparison of team home field advantages. Team IDs are given in Table 1. Offensive advantage is an additional propensity for a team to score goals when playing on its home field. Defensive advantage is a penalty for a team not playing on its home field, and enters the model as an additional propensity to allow goals when not playing at home. The difference between a teams Offensive and Defensive advantages is an indication of the overall home advantage it enjoys.

DOI: 10.2202/1559-0410.1107

Table 1: Team Ranking using Equal Home Advantage

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

References

[1] Stephen R. Clarke and John M. Norman. Home ground advantage of individual clubs in english soccer. *The Statistician*, 1995.

[2] Mark J. Dixon and Stuart G. Coles. Modeling association football scores and inefficiencies in the football betting market. *Applied Statistics*, 1997.

[3] John Goddard. Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 2005.