

BRADLEY–TERRY MODELING WITH MULTIPLE GAME OUTCOMES WITH APPLICATIONS TO COLLEGE HOCKEY

JOHN T. WHELAN AND JACOB E. KLEIN

Abstract. The Bradley–Terry model has previously been used in both Bayesian and frequentist interpretations to evaluate the strengths of sports teams based on win-loss game results. It has also been extended to handle additional possible results such as ties. We implement a generalization which includes multiple possible outcomes such as wins or losses in regulation, overtime, or shootouts. A natural application is to ice hockey competitions such as international matches, European professional leagues, and NCAA hockey, all of which use a zero-sum point system which values overtime and shootout wins as $2/3$ of a win, and overtime and shootout losses as $1/3$ of a win. We incorporate this into the probability model, and evaluate the posterior distributions for the associated strength parameters using techniques such as Gaussian expansion about maximum a posteriori estimates, and Hamiltonian Monte Carlo.

1. INTRODUCTION

The Bradley–Terry model [3, 19] has long been used for evaluating paired comparisons, such as games between pairs of teams in which one team or the other wins each game. The model assigns a strength parameter to each team, and the odds ratio associated with the probability of a team winning a game is equal to the ratio of the strengths. These strength parameters can be estimated based on the full set of game results and used to rank teams or make future predictions. The model has been extended by Davidson [7] to contests in which a tie or drawn contest is a possible outcome. For many years, such ties were a common occurrence in the sport of ice hockey, but recently tie-breaking methods such as an overtime period played under different rules and/or a shootout in which the teams alternate penalty shots are used to determine a winner. Results in overtime or shootouts can be evaluated differently from wins in regulation play. For instance, since 2006 [10] competitions organized by International Ice Hockey Federation (IIHF) have awarded three points in the standings to a team winning in regulation, two points for a win in overtime or a shootout, one point for a loss in overtime or a shootout, and no points for a loss in regulation, and many leagues have followed suit. Compared to the prior system which awarded two points for a win, one for a tie, and none for a loss, which effectively treated a tie as half a win and half a loss, the four-outcome system treats an overtime/shootout win as $2/3$

MSC (2020): primary 62F15; secondary 62F07.

Keywords: Bradley–Terry, College Hockey, International Hockey.

of a win and $1/3$ of a loss.¹ One possible approach to either of these situations (games with three or four outcomes) is to use standard Bradley–Terry and assign fractional wins as appropriate to the point system (see, e.g., [18]). However, this is unsatisfying, as it provides no way to assign a probability for a future game to end in a tie or overtime or shootout result. In this paper, we instead consider a generalization of the tie model of [7] which associates one strength parameter to each team, along with a single parameter describing the tendency for games to go into overtime.

The rest of this paper is organized as follows: In Section 2, we describe the three models (standard Bradley–Terry, Bradley–Terry–Davidson including ties, and a new model with four possible game outcomes including overtime/shootout wins and losses), and exhibit a generalization of the relevant formulas which describes all three cases. In Section 3, we describe methods for inferring the relevant parameters of these models given a set of game results: maximum likelihood estimation, and Bayesian inference using either a Gaussian approximation or Hamiltonian Monte Carlo. In Section 4, we demonstrate these methods using a recent set of game results: the 2020–2021 Eastern College Athletic Conference (ECAC) season. This season used the standard IIHF system with 3–2–1–0 points assigned for regulation wins, overtime/shootout wins, overtime/shootout losses, and regulation losses, respectively. For the purposes of illustration, we evaluate the ECAC results with the four-outcome model as well as with the other two models, treating in one case all wins the same, and in the other all overtime/shootout results as ties.

2. MODELS

In the standard Bradley–Terry model [3, 19] each team has a strength $\pi_i \in (0, \infty)$, and the modeled probability that team i will win a game with team j is

$$\theta_{ij}^W = \frac{\pi_i}{\pi_i + \pi_j}$$

so that the probability of a set of game outcomes D in which team i plays team j n_{ij} times and wins n_{ij}^W of those games is²

$$P(D|\{\pi_i\}) = \prod_{i=1}^t \prod_{j=i+1}^t (\theta_{ij}^W)^{n_{ij}^W} (\theta_{ij}^L)^{n_{ij}^L} = \left(\prod_{i=1}^t \prod_{j=1}^t (\theta_{ij}^W)^{n_{ij}^W} (\theta_{ij}^L)^{n_{ij}^L} \right)^{\frac{1}{2}},$$

where t is the number of teams, $n_{ij}^L = n_{ji}^W = n_{ij} - n_{ij}^W$ and $\theta_{ij}^L = \theta_{ji}^W = 1 - \theta_{ij}^W$.

¹We do not consider non-zero-sum point systems such as that used in association football (soccer) which awards three points for a win and one for a draw, so that drawn matches are only worth two points total rather than three. Likewise, the National Hockey League awards all wins two points and overtime/shootout losses one point; this 2–2–1–0 system awards three total points for games which go into overtime, but only two for games decided in regulation.

²The first form explicitly includes each pair of teams only once, while the second corrects for the double-counting, taking advantage of the fact that $n_{ii}^W = 0 = n_{ii}^L$. If the order of the games between pairs of teams is ignored, the sampling distribution for the $\{n_{ij}^W\}$ is instead

$$p(\{n_{ij}^W\}|\{\pi_i\}) = \left(\prod_{i=1}^t \prod_{j=1}^t \frac{(n_{ij})!}{(n_{ij}^W)!(n_{ij}^L)!} (\theta_{ij}^W)^{n_{ij}^W} (\theta_{ij}^L)^{n_{ij}^L} \right)^{\frac{1}{2}}.$$

Davidson [7] proposed an extension for competitions which include the probabilities of ties, in which the probabilities of the three possible outcomes of a game are

$$\begin{aligned}\theta_{ij}^W &= \frac{\pi_i}{\pi_i + \nu\sqrt{\pi_i\pi_j} + \pi_j}, \\ \theta_{ij}^T &= \frac{\nu\sqrt{\pi_i\pi_j}}{\pi_i + \nu\sqrt{\pi_i\pi_j} + \pi_j}, \\ \theta_{ij}^L &= \frac{\pi_j}{\pi_i + \nu\sqrt{\pi_i\pi_j} + \pi_j},\end{aligned}$$

where $\nu \in [0, \infty)$ is an additional parameter which describes how likely ties are to occur. (The probability of a tie in a game between evenly matched teams is $\frac{\nu}{2+\nu}$.) Evidently, $\theta_{ij}^L = \theta_{ji}^W$, $\theta_{ij}^T = \theta_{ji}^T$ and $\theta_{ij}^W + \theta_{ij}^T + \theta_{ij}^L = 1$. The probability of a given set of game outcomes in which the $n_{ij} = n_{ij}^W + n_{ij}^T + n_{ij}^L$ games between teams i and j result in n_{ij}^W wins, n_{ij}^T ties and n_{ij}^L losses for team i (where $n_{ij}^T = n_{ji}^T$ and $n_{ij}^L = n_{ji}^W$) is

$$P(D|\{\pi_i\}, \nu) = \left(\prod_{i=1}^t \prod_{j=1}^t (\theta_{ij}^W)^{n_{ij}^W} (\theta_{ij}^T)^{n_{ij}^T} (\theta_{ij}^L)^{n_{ij}^L} \right)^{\frac{1}{2}}.$$

We propose an extension appropriate for a system in which a win in overtime or a shootout is treated as 2/3 of a win and 1/3 of a loss. Writing the four possible game outcomes as RW for regulation win, OW for overtime/shootout win, OL for overtime/shootout loss, and RL for regulation loss, the modelled probability of each outcome would be³

$$\begin{aligned}\theta_{ij}^{\text{RW}} &= \frac{\pi_i}{\pi_i + \nu\pi_i^{2/3}\pi_j^{1/3} + \nu\pi_i^{1/3}\pi_j^{2/3} + \pi_j}, \\ \theta_{ij}^{\text{OW}} &= \frac{\nu\pi_i^{2/3}\pi_j^{1/3}}{\pi_i + \nu\pi_i^{2/3}\pi_j^{1/3} + \nu\pi_i^{1/3}\pi_j^{2/3} + \pi_j}, \\ \theta_{ij}^{\text{OL}} &= \frac{\nu\pi_i^{1/3}\pi_j^{2/3}}{\pi_i + \nu\pi_i^{2/3}\pi_j^{1/3} + \nu\pi_i^{1/3}\pi_j^{2/3} + \pi_j}, \\ \theta_{ij}^{\text{RL}} &= \frac{\pi_j}{\pi_i + \nu\pi_i^{2/3}\pi_j^{1/3} + \nu\pi_i^{1/3}\pi_j^{2/3} + \pi_j}.\end{aligned}$$

³The exponents are chosen to correspond to the share of the points (2/3 and 1/3, respectively) awarded for an overtime/shootout win or loss. This has the desirable feature that the maximum likelihood equation (3.3) becomes (after multiplying by 3)

$$\sum_{j=1}^t n_{ij} (3\hat{\theta}_{ij}^{\text{RW}} + 2\hat{\theta}_{ij}^{\text{OW}} + \hat{\theta}_{ij}^{\text{OL}}) = \sum_{j=1}^t (3n_{ij}^{\text{RW}} + 2n_{ij}^{\text{OW}} + n_{ij}^{\text{OL}}),$$

i.e., that the expected number of points for each team equals the actual number. See also the discussion in Section 5 about possible alternative models, including extended models in which the exponents are not fixed, but inferred from the data.

The probability for a set of game outcomes will then be

$$P(D|\{\pi_i\}, \nu) = \left(\prod_{i=1}^t \prod_{j=1}^t (\theta_{ij}^{\text{RW}})^{n_{ij}^{\text{RW}}} (\theta_{ij}^{\text{OW}})^{n_{ij}^{\text{OW}}} (\theta_{ij}^{\text{OL}})^{n_{ij}^{\text{OL}}} (\theta_{ij}^{\text{RL}})^{n_{ij}^{\text{RL}}} \right)^{\frac{1}{2}}.$$

If we write $\lambda_i = \ln \pi_i \in (-\infty, \infty)$ and $\tau = \ln \nu \in (-\infty, \infty)$, we can describe all three models as special cases of a general model in which the probability of a game between teams i and j ending in outcome I is

$$\theta_{ij}^I = \frac{\pi_i^{p_I} \pi_j^{1-p_I} \nu^{o_I}}{\sum_J \pi_i^{p_J} \pi_j^{1-p_J} \nu^{o_J}} = \frac{(\pi_i/\pi_j)^{p_I} \nu^{o_I}}{\sum_J (\pi_i/\pi_j)^{p_J} \nu^{o_J}} = \sigma(\{p_J(\lambda_i - \lambda_j) + o_J \tau | J\})_I, \quad (2.1)$$

where

$$\sigma(\mathbf{x})_I = \frac{e^{x_I}}{\sum_J e^{x_J}}$$

is a vector equivalent of the logistic function known as the softmax function [4]. The probability for a set of game outcomes is

$$P(D|\{\pi_i\}, \nu) = \left(\prod_{i=1}^t \prod_{j=1}^t \prod_I (\theta_{ij}^I)^{n_{ij}^I} \right)^{\frac{1}{2}}.$$

Specifically,

- For the standard Bradley–Terry model, $p_W = 1$, $p_L = 0$, and $o_W = o_L = 0$.
- For the Bradley–Terry–Davidson model with ties, $p_W = 1$, $p_T = \frac{1}{2}$, $p_L = 0$, $o_W = o_L = 0$, and $o_T = 1$.
- For the model introduced in this paper, $p_{RW} = 1$, $p_{OW} = \frac{2}{3}$, $p_{OL} = \frac{1}{3}$, $p_{RL} = 0$, $o_{RW} = o_{RL} = 0$, and $o_{OW} = o_{OL} = 1$.

All of these models satisfy $\sum_I \theta_{ij}^I = 1$, and have “opposite” outcomes I and $-I$ such that $\theta_{ij}^{-I} = \theta_{ji}^I$, $p_{-I} = 1 - p_I$, and $o_{-I} = o_I$. They also satisfy $0 \leq p_I \leq 1$ and $o_I \in \{0, 1\}$. We confine ourselves below to cases where these properties hold.

3. INFERENCE OF PARAMETERS

3.1. Maximum likelihood

Maximum likelihood estimates (MLEs) of Bradley–Terry strength parameters [7, 8, 19] provide a straightforward way of associating a “rating” to each team based on their game results, and have been proposed as a replacement for less reliable ways of evaluating a team’s game results in light of the difficulty of their schedule [5].

We can consider the probability $P(D|\{\pi_i\}, \nu) = P(D|\{\lambda_i\}, \tau)$ as a likelihood function of the parameters $\{\lambda_i\}$ and τ , with log-likelihood

$$\ln P(D|\{\lambda_i\}, \tau) = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t \sum_I n_{ij}^I \ln \theta_{ij}^I.$$

We can use the identity

$$d \ln \sigma(\mathbf{x})_I = dx_I - \sum_J dx_J \sigma(\mathbf{x})_J$$

to show that

$$\frac{\partial \ln \theta_{ij}^I}{\partial \tau} = o_I - \sum_J o_J \theta_{ij}^J$$

and

$$\frac{\partial \ln \theta_{ij}^I}{\partial \lambda_k} = (\delta_{ik} - \delta_{jk}) \left(p_I - \sum_J p_J \theta_{ij}^J \right),$$

which means that

$$\frac{\partial \ln P(D|\{\lambda_i\}, \tau)}{\partial \tau} = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t \sum_I o_I n_{ij}^I - \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t n_{ij} \sum_I o_I \theta_{ij}^I \quad (3.1)$$

and

$$\frac{\partial \ln P(D|\{\lambda_i\}, \tau)}{\partial \lambda_k} = \sum_{i=1}^t \sum_I n_{ki}^I p_I - \sum_{i=1}^t n_{ki} \sum_I p_I \theta_{ki}^I. \quad (3.2)$$

Using these, we can write the maximum likelihood equations as

$$n^o = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t n_{ij} \sum_I o_I \hat{\theta}_{ij}^I = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t n_{ij} \sum_I o_I \frac{(\hat{\pi}_i / \hat{\pi}_j)^{p_I} \hat{\nu}^{o_I}}{\sum_J (\hat{\pi}_i / \hat{\pi}_j)^{p_J} \hat{\nu}^{o_J}}$$

and

$$p_k = \sum_{i=1}^t n_{ki} \sum_I p_I \hat{\theta}_{ki}^I = \sum_{i=1}^t n_{ki} \sum_I p_I \frac{(\hat{\pi}_k / \hat{\pi}_i)^{p_I} \hat{\nu}^{o_I}}{\sum_J (\hat{\pi}_k / \hat{\pi}_i)^{p_J} \hat{\nu}^{o_J}} \quad (3.3)$$

where

$$n^o = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t \sum_I o_I n_{ij}^I$$

can be interpreted in the models considered as the number of games which are tied or go to overtime, respectively, and

$$p_k = \sum_{i=1}^t \sum_I n_{ki}^I p_I$$

can be seen as the total number of “points” for team i . The maximum likelihood equation set each of these quantities equal to their expectation values.

We can solve the maximum likelihood equations by a generalization of the iterative method in [8]. writing them

$$\hat{\nu} = n^o \left/ \left(\frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t n_{ij} \frac{\sum_I o_I (\hat{\pi}_i / \hat{\pi}_j)^{p_I}}{\sum_J (\hat{\pi}_i / \hat{\pi}_j)^{p_J} \hat{\nu}^{o_J}} \right) \right.$$

(where we have used the fact that the only non-zero term in the numerator has $o_I = 1$) and

$$\hat{\pi}_k = p_k \left/ \left(\sum_{i=1}^t n_{ki} \frac{\sum_I p_I \hat{\pi}_k^{p_I - 1} \hat{\pi}_i^{-p_I} \hat{\nu}^{o_I}}{\sum_J (\hat{\pi}_k / \hat{\pi}_i)^{p_J} \hat{\nu}^{o_J}} \right) \right.$$

As in the standard Bradley-Terry model, the overall multiplicative scale of $\hat{\pi}_k$ is undefined (because θ_{ij}^I can be written so that the team strengths appear only in the combination π_j / π_i), so it is necessary to rescale the team strengths at each

iteration to preserve a property such as $\prod_{i=1}^t \hat{\pi}_i = 1$. Beyond that, there are conditions for the maximum likelihood estimates to be finite and well-defined, which are explored in, e.g., [2, 6, 15].

3.2. Bayesian approach

It is useful to move beyond maximum likelihood estimates, both to quantify uncertainty in the model parameters, and to make predictions about the outcome of future games. (For instance, [18] proposed simulating future games with probabilities drawn from a posterior distribution capturing the uncertainty in the strength parameters, rather than fixed probabilities generated from the MLEs of those parameters.)

A convenient framework for parameter estimates including uncertainties is Bayesian inference, which defines the posterior probability density for the parameters $\{\pi_i\}$ and ν , or equivalently $\{\lambda_i\}$ and τ , given a set of game results D and prior assumptions I , as

$$f(\{\lambda_i\}, \tau | D, I) = \frac{P(D | \{\lambda_i\}, \tau) f(\{\lambda_i\}, \tau | I)}{P(D | I)} \propto P(D | \{\lambda_i\}, \tau) f(\{\lambda_i\}, \tau | I).$$

A variety of choices can be made for the multivariate prior distribution on $\{\lambda_i\}$ [17] in the Bradley–Terry model, and likewise for the tie/overtime parameter τ . For simplicity, we work in this paper with the improper Haldane prior⁴

$$f(\{\lambda_i\}, \tau | I_0) = \text{constant},$$

which means that the posterior distribution is proportional to the likelihood:

$$f(\{\lambda_i\}, \tau | D, I_0) \propto P(D | \{\lambda_i\}, \tau).$$

With this choice of prior, the posterior probability density will be independent of the combination $\sum_{i=1}^t \lambda_i$, but otherwise will be normalizable under the same circumstances that lead to well-defined maximum likelihood estimates for the parameters.

3.2.1. Gaussian Approximation. One convenient approach is to Taylor expand the log-posterior $\ln f(\{\lambda_i\}, \tau | D, I)$ about the maximum a posteriori solution (which, in this case, is the maximum likelihood solution $\{\hat{\lambda}_i\}, \hat{\tau}$).⁵ Truncating the expansion at second order gives a Gaussian approximation

$$f(\{\lambda_i\}, \tau | D, I_0) \approx f(\{\hat{\lambda}_i\}, \hat{\tau} | D, I_0) \exp \left(-\frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t H_{ij} (\lambda_i - \hat{\lambda}_i) (\lambda_j - \hat{\lambda}_j) - \sum_{i=1}^t H_{i\tau} (\lambda_i - \hat{\lambda}_i) (\tau - \hat{\tau}) - \frac{1}{2} H_{\tau\tau} (\tau - \hat{\tau})^2 \right),$$

⁴So named because the marginal prior distribution for probabilities such as θ_{ij}^I will follow the Haldane prior [9, 11], which is the limit of a Beta(α, β) distribution as $\alpha, \beta \rightarrow 0$.

⁵Note that this method does not assign special significance to the MAP estimates, but uses them as the starting point for a convenient approximation to the posterior probability distribution.

where \mathbf{H} is the $(t+1) \times (t+1)$ Hessian matrix

$$\mathbf{H} = \begin{pmatrix} \{H_{ij}\} & \{H_{i\tau}\} \\ \{H_{\tau j}\} & H_{\tau\tau} \end{pmatrix}$$

with elements⁶

$$\begin{aligned} H_{ij} &= - \left[\frac{\partial^2}{\partial \lambda_i \partial \lambda_j} \ln P(D|\{\lambda_k\}, \tau) \right]_{\{\lambda_k = \hat{\lambda}_k\}, \tau = \hat{\tau}}, \\ H_{i\tau} &= H_{\tau i} = - \left[\frac{\partial^2}{\partial \lambda_i \partial \tau} \ln P(D|\{\lambda_k\}, \tau) \right]_{\{\lambda_k = \hat{\lambda}_k\}, \tau = \hat{\tau}}, \\ H_{\tau\tau} &= - \left[\frac{\partial^2}{\partial \tau^2} \ln P(D|\{\lambda_k\}, \tau) \right]_{\{\lambda_k = \hat{\lambda}_k\}, \tau = \hat{\tau}}. \end{aligned}$$

To compute the elements of the Hessian matrix, we return to the first derivative (3.1) and differentiate them to get

$$- \frac{\partial^2 \ln P(D|\{\lambda_i\}, \tau)}{\partial \tau^2} = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t n_{ij} \sum_I o_I \frac{\partial \theta_{ij}^I}{\partial \tau} = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t n_{ij} \theta_{ij}^o (1 - \theta_{ij}^o),$$

where

$$\theta_{ij}^o = \sum_I o_I \theta_{ij}^I$$

is the probability of a tie or overtime game, depending on the model, and we have used the fact that $o_I^2 = o_I$ since $o_I \in \{0, 1\}$. Similarly, using the properties $\sum_I \theta_{ij}^I = 1$, $\theta_{ij}^{-I} = \theta_{ji}^I$, $p_{-I} = 1 - p_I$, and $o_{-I} = o_I$, we find

$$- \frac{\partial^2 \ln P(D|\{\lambda_i\}, \tau)}{\partial \tau \partial \lambda_k} = \sum_{i=1}^t n_{ki} \sum_I o_I \theta_{ki}^I \left(p_I - \sum_J p_J \theta_{ki}^J \right)$$

and, finally, differentiating (3.2) gives us

$$\begin{aligned} - \frac{\partial^2 \ln P(D|\{\lambda_i\}, \tau)}{\partial \lambda_k \partial \lambda_\ell} &= \delta_{k\ell} \sum_{i=1}^t n_{ki} \sum_I p_I \theta_{ki}^I \left(p_I - \sum_J p_J \theta_{ki}^J \right) \\ &\quad - n_{k\ell} \sum_I p_I \theta_{k\ell}^I \left(p_I - \sum_J p_J \theta_{k\ell}^J \right) \end{aligned}$$

so that the Hessian matrix has components

$$H_{\tau\tau} = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^t n_{ij} \sum_I o_I \hat{\theta}_{ij}^I (1 - \sum_J o_J \hat{\theta}_{ij}^J) \quad (3.4a)$$

⁶Note the similarity to the Fisher information matrix

$$I_{ij}(\{\lambda_k\}) = \sum_D P(D|\{\lambda_k\}, I) \frac{\partial^2}{\partial \lambda_i \partial \lambda_j} [-\ln P(D|\{\lambda_k\}, I)],$$

which differs from the Hessian in that H_{ij} depends on the observed data, while I_{ij} is a function defined on parameter space.

$$H_{\tau k} = H_{k\tau} = \sum_{i=1}^t n_{ki} \sum_I o_I \hat{\theta}_{ki}^I \left(p_I - \sum_J p_J \hat{\theta}_{ki}^J \right) \quad (3.4b)$$

$$H_{k\ell} = \delta_{k\ell} \sum_{i=1}^t n_{ki} \sum_I p_I \hat{\theta}_{ki}^I \left(p_I - \sum_J p_J \hat{\theta}_{ki}^J \right) - n_{k\ell} \sum_I p_I \hat{\theta}_{k\ell}^I \left(p_I - \sum_J p_J \hat{\theta}_{k\ell}^J \right). \quad (3.4c)$$

Note that in the case of the Bradley–Terry model, where the only outcomes are win and loss, the condition $o^I = 0$ simplifies the Hessian to $H_{\tau\tau} = H_{\tau k} = 0$ (since the τ parameter is not actually part of the likelihood), and

$$H_{k\ell} = \delta_{k\ell} \sum_{i=1}^t n_{ki} \hat{\theta}_{ki}^W (1 - \hat{\theta}_{ki}^W) - n_{k\ell} \hat{\theta}_{k\ell}^W (1 - \hat{\theta}_{k\ell}^W),$$

which is the form seen in, e.g., [18].

The Hessian matrix in (3.4) is singular since $\sum_{\ell=1}^t H_{\tau\ell} = 0$ and $\sum_{\ell=1}^t H_{k\ell} = 0$, which ultimately arise from the fact that the probabilities $\{\theta_{ij}^I\}$, and thus the likelihood, are unchanged by adding the same constant to all the $\{\lambda_i\}$. This can be handled computationally by constructing a variance-covariance matrix $\Sigma = \mathbf{H}^+$ which is the Moore–Penrose pseudo-inverse [13]⁷ of the Hessian matrix, and approximating the posterior as a multivariate Gaussian with a mean of $\{\hat{\lambda}_i\}, \hat{\tau}$ and a variance-covariance matrix Σ . This has the effect of enforcing the constraint

$$\sum_{i=1}^t \lambda_i = \sum_{i=1}^t \hat{\lambda}_i = 0$$

on the combination of the parameters which has no influence on the model.

This Gaussian approximation can be used to produce analytic estimates of quantities of interest, or used for Monte Carlo sampling, as illustrated in Section 4. It can also be used as a starting point for importance sampling of the sort discussed in [18]. For the present work, we consider a different Monte Carlo method for sampling from the exact posterior.

3.2.2. Hamiltonian Monte Carlo. Markov-chain Monte Carlo methods provide a convenient way to draw samples from a posterior distribution. We demonstrate in this paper how to draw posterior samples for the Bradley–Terry extensions considered, using Hamiltonian Monte Carlo as implemented in the Stan library [16]. There are a few technical considerations. Because the posterior on $\{\lambda_i\}$ and τ is improper, trying to draw from it directly will lead to chains which never converge. Any probabilities constructed from the samples will be well-behaved since only the meaningless degree of freedom $\sum_{i=1}^t \lambda_i$ is unconstrained, but these apparent errors make it more difficult to detect other potential problems. It is thus useful instead to consider only variables $\gamma_{ij} = \lambda_i - \lambda_j$ (and τ) which contribute to the probability model via (see (2.1))

$$\theta_{ij}^I = \sigma(\{p_J \gamma_{ij} + o_J \tau | J\})_I.$$

⁷For a real symmetric matrix with a complete eigenvalue decomposition, this operation replaces each non-zero eigenvalue with its reciprocal while leaving zero eigenvalues unchanged.

Table 1. Results of the 2020–2021 Eastern College Athletic Conference (ECAC) season, showing the number of regulation wins n_{ij}^{RW} and overtime/shootout wins n_{ij}^{OW} for each team against each opponent. From these, we can derive the total number of results of each type (RW, OW, OL and RL) for each team, which are used, for example, to generate the standings in the 3–2–1–0 point system.

Team i	$n_{ij}^{\text{RW}} \ (n_{ij}^{\text{OW}})$				$n_i^I = \sum_{j=1}^t n_{ij}^I$			
	Cg	Ck	Qn	SL	RW	OW	OL	RL
Colgate (Cg)	—	1(1)	1(0)	2(1)	4	2	3	9
Clarkson (Ck)	3(1)	—	1(2)	1(0)	5	3	4	2
Quinnipiac (Qn)	4(1)	1(2)	—	4(1)	9	4	2	3
St. Lawrence (SL)	2(1)	0(1)	1(0)	—	3	2	2	7

Of course, the full set of $\frac{t(t-1)}{2}$ values γ_{ij} are not independent. Instead, they are determined by the $t-1$ parameters $\omega_i = \lambda_i - \lambda_{i+1}$ for $i = 1, \dots, t-1$. Given the $\{\omega_i\}$, we can construct $\gamma_{ij} = \sum_{k=i}^{j-1} \omega_k$.

In Appendix A, we show the code of the Stan model used to perform Hamiltonian Monte Carlo simulations of all three models.

4. DEMONSTRATION USING GAME RESULTS

We now illustrate the application of the models described in this paper using game results from a competition which used the 3–2–1–0 point system: the 2020–2021 Eastern College Athletic Conference (ECAC) season. While the league ordinarily plays a balanced round-robin schedule in which each team plays each other team the same number of times, the season in question ended up being unbalanced due to cancellations of games arising from the COVID-19 pandemic. In Table 1, we show the results for the ECAC season, in the form of n_{ij}^{RW} and n_{ij}^{OW} for each team against each opponent, along with the total number of results of each type for each team, $n_i^I = \sum_{j=1}^t n_{ij}^I$.

4.1. ECAC: Standard Bradley–Terry model

As a first demonstration, we consider the standard Bradley–Terry model applied to the ECAC results with regulation and overtime/shootout wins being counted as simply “wins” and regulation and overtime/shootout losses being counted as “losses”. I.e., we define $n_{ij}^{\text{W}} = n_{ij}^{\text{RW}} + n_{ij}^{\text{OW}}$ and $n_{ij}^{\text{L}} = n_{ij}^{\text{RL}} + n_{ij}^{\text{OL}}$. The resulting maximum-likelihood solutions $\{\hat{\lambda}_i\}$ and associated probabilities $\{\hat{\theta}_{ij}^{\text{W}}\}$ are shown in Table 2, along with the uncertainties and correlations encoded in the variance-covariance matrix $\{\Sigma_{ij}\}$ of the Gaussian approximation to the posterior distribution.

Since the log-strengths $\{\lambda_i\}$ have an arbitrary additive scale, a more meaningful understanding of the posterior distributions is obtained by considering the marginal distribution of the difference of a pair of team strengths $\gamma_{ij} = \lambda_i - \lambda_j$. In Figure 1, we illustrate the maximum likelihood estimate and posterior distribution of this quantity for two of the six pairs of teams: Quinnipiac-Colgate and Quinnipiac-Clarkson. We show the posterior in the Gaussian approximation (for

Table 2. The maximum likelihood estimates and parameters of the the Gaussian approximation to the posterior distribution for the Bradley–Terry model applied to the 2020–2021 ECAC results, with regulation and overtime/shootout results counted the same. The maximum likelihood estimate $\hat{\lambda}_i$ for each team’s log-strength has an associated one-sigma uncertainty $\sqrt{\Sigma_{ii}}$. The variance-covariance matrix $\{\Sigma_{ij}\}$ can be converted to a correlation matrix $\rho_{ij} = \Sigma_{ij} / \sqrt{\Sigma_{ii}\Sigma_{jj}}$. Note that the information included in Σ_{ij} is also influenced by the constraint $\sum_{i=1}^t \lambda_i = 0$, so for example the anti-correlation of the different log-strengths is somewhat artificial. We also show the maximum-likelihood estimates $\{\hat{\theta}_{ij}^W\}$ for the head-to-head win probabilities between pairs of teams.

Team	$\hat{\lambda}_i$	$\sqrt{\Sigma_{ii}}$	$\hat{\theta}_{ij}^W$				$\rho_{ij} = \Sigma_{ij} / \sqrt{\Sigma_{ii}\Sigma_{jj}}$			
			Cg	Ck	Qn	SL	Cg	Ck	Qn	SL
Cg	−0.55	0.39	—	0.29	0.22	0.49	1.00	−0.31	−0.39	−0.21
Ck	0.32	0.43	0.71	—	0.40	0.70	−0.31	1.00	−0.22	−0.50
Qn	0.74	0.40	0.78	0.60	—	0.78	−0.39	−0.22	1.00	−0.35
SL	−0.51	0.45	0.51	0.30	0.22	—	−0.21	−0.50	−0.35	1.00

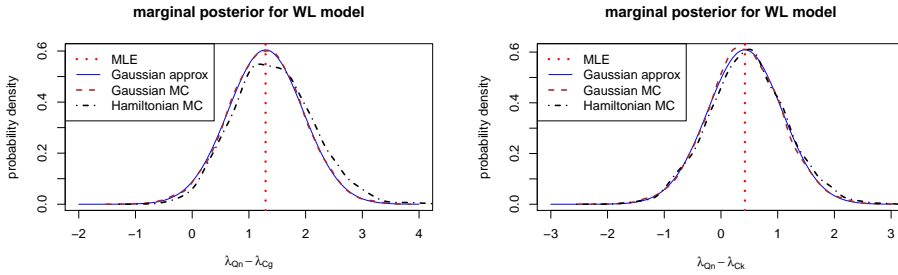


Figure 1. Posterior probability density for difference in log-strengths $\gamma_{ij} = \lambda_i - \lambda_j$ between selected pairs of teams (left: Quinnipiac and Colgate; right: Quinnipiac and Clarkson), based on 2020–2021 ECAC game results in the standard Bradley–Terry model with regulation and overtime/shootout wins treated the same. The dotted red vertical line shows the maximum likelihood estimate $\hat{\gamma}_{ij}$. Since the Haldane prior used is uniform in the $\{\lambda_i\}$, this is also the maximum a posteriori (MAP) value. The curves show the approximate Gaussian posterior from expanding about the MAP value (solid blue line), along with density estimates from a set of Monte Carlo samples drawn from that distribution (dashed brown line), and a set of samples drawn from the exact distribution using Hamiltonian Monte Carlo (dot-dash black line). Differences between the Gaussian approximation and the samples from the exact posterior are small, but can be noticeable, especially if the maximum likelihood estimate $\hat{\gamma}_{ij}$ is far from zero. For reference, note that the “Gaussian approx” and “Gaussian MC” curves should only differ due to Monte Carlo errors in the construction of the latter.

which the marginal posterior on γ_{ij} is also a Gaussian), in a Monte Carlo drawn from the approximate multivariate Gaussian distribution, and in posterior samples from the exact posterior generated using Hamiltonian Monte Carlo with the Stan library [16]. We can transform the posterior on a difference γ_{ij} in log-strength into a posterior on the corresponding probability $\theta_{ij}^W = \text{logistic}(\gamma_{ij})$; this is shown in Figure 2 for the two sets of posterior samples. In all cases, the exact marginal

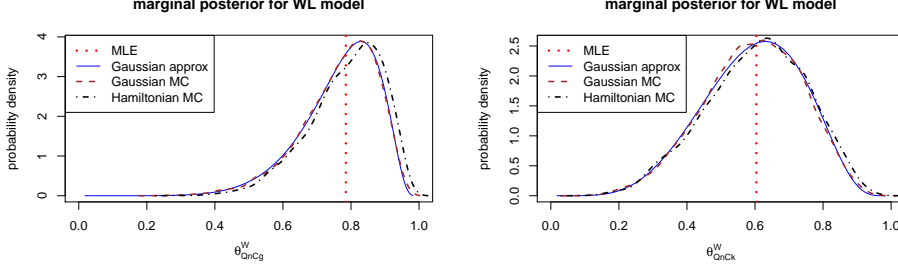


Figure 2. Posterior probability density for the win probability $\theta_{ij}^W = \text{logistic}(\gamma_{ij})$ predicted by the Bradley–Terry model for selected pairs of teams, as in Figure 1. Note that, due to the transformation of the probability density, the maximum of the posterior probability density in this parameter is not the maximum likelihood value as it was in Figure 1.

Table 3. The maximum likelihood estimates for the Bradley–Terry–Davidson model applied to the 2020–2021 ECAC results, with all overtime games counted as ties. The maximum likelihood estimates $\{\hat{\lambda}_i\}$ and $\hat{\tau}$ of the log-strengths and log tie parameter are used to compute the estimated probability $\hat{\theta}_{ij}^W$ for a win and $\hat{\theta}_{ij}^T$ for a tie between each pair of teams. Note that the estimated probability of a game between evenly-matched teams to end in a tie is $\hat{e}^{\hat{\tau}}/(2 + \hat{e}^{\hat{\tau}}) = 0.39$, and it is lower the more different the two teams’ strengths are.

Team i	$\hat{\lambda}_i$	$\hat{\theta}_{ij}^W \ (\hat{\theta}_{ij}^T)$			
		Cg	Ck	Qn	SL
Cg	−0.73	—	0.13 (0.33)	0.11 (0.32)	0.33 (0.38)
Ck	0.70	0.54 (0.33)	—	0.28 (0.38)	0.56 (0.32)
Qn	0.89	0.57 (0.32)	0.34 (0.38)	—	0.59 (0.31)
SL	−0.85	0.29 (0.38)	0.12 (0.32)	0.10 (0.31)	—
$\hat{\tau} = 0.23$					

posterior, as estimated by the Hamiltonian Monte Carlo, is only slightly different from the Gaussian approximation. This is similar to the results found using importance sampling in [18].

4.2. ECAC: Bradley–Terry–Davidson model with ties

Moving on to the Bradley–Terry–Davidson model with ties, we now consider inference of the log-strength parameters $\{\lambda_i\}$ along with the log-tie parameter τ . We illustrate the methods by reanalyzing the 2020–2021 ECAC results, with all overtime games treated as ties, so that now $n_{ij}^W = n_{ij}^{RW}$, $n_{ij}^T = n_{ij}^{OW} + n_{ij}^{OL}$, and $n_{ij}^L = n_{ij}^{RL}$. The maximum likelihood solutions $\{\hat{\lambda}_i\}$ and $\hat{\tau}$ are shown in Table 3, along with the associated probabilities $\{\hat{\theta}_{ij}^W\}$ for a win and $\{\hat{\theta}_{ij}^T\}$ for a tie in contests between pairs of teams. In Table 4, we show the maximum-likelihood estimates along with the uncertainties in and correlations among the log-strengths $\{\lambda_i\}$ and the log-tie parameter τ , which are encoded in the variance-covariance matrix $\{\Sigma_{ij}, \Sigma_{i\tau}, \Sigma_{\tau\tau}\}$ of the Gaussian approximation to the posterior distribution.

Table 4. The parameters of the the Gaussian approximation to the posterior distribution for the Bradley–Terry–Davidson model applied to the 2020–2021 ECAC results, with all overtime games counted as ties. In addition to the log-strength parameters considered for the Bradley–Terry model in Table 2, there are uncertainties and correlations associated with the log-tie parameter τ .

Team i	$\hat{\lambda}_i$	$\sqrt{\Sigma_{ii}}$	$\rho_{ij} = \Sigma_{ij} / \sqrt{\Sigma_{ii}\Sigma_{jj}}$				
			Cg	Ck	Qn	SL	τ
Cg	−0.73	0.50	1.00	−0.35	−0.40	−0.16	−0.22
Ck	0.70	0.57	−0.35	1.00	−0.16	−0.53	0.19
Qn	0.89	0.51	−0.40	−0.16	1.00	−0.38	0.26
SL	−0.85	0.58	−0.16	−0.53	−0.38	1.00	−0.22
τ	0.23	0.40	−0.22	0.19	0.26	−0.22	1.00

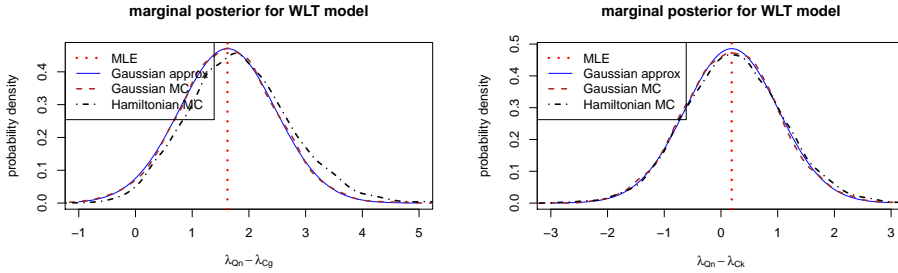


Figure 3. Posterior probability density for difference in log-strengths $\lambda_i - \lambda_j$ between selected pairs of teams (left: Quinnipiac and Colgate; right: Quinnipiac and Clarkson), based on the Bradley–Terry–Davidson model applied to the 2020–2021 ECAC results, with all overtime games counted as ties. Curves are as defined in Figure 1.

As with the standard Bradley–Terry model, we can show the marginal posterior distributions on the differences $\{\gamma_{ij} = \lambda_i - \lambda_j\}$ between pairs of log-strength parameters, and we do this in Figure 3 for the same pairs of teams as before. Once again, samples drawn from the multivariate Gaussian approximation capture the shape of that distribution well, and samples drawn from the exact posterior using Hamiltonian Monte Carlo are slightly different but similar.

We cannot convert γ_{ij} directly into a probability, however, since probabilities depend on the log-tie parameter τ as well. In Figure 4, we plot the marginal posterior on τ . The parameter τ can be transformed into a probability $\frac{\nu}{2+\nu}$, where $(\nu = e^\tau)$ for a game between evenly-matched teams to be tied, and we plot the posterior for this as well. Finally, in Figure 5, we illustrate the joint marginal posterior in γ_{ij} and τ for our selected pairs of teams.

To illustrate the posterior on the probabilities $\{\theta_{ij}^I | I = W, T, L\}$ for a pair of teams, we note that the constraint $\sum_I \theta_{ij}^I = 1$ means that the space is actually two dimensional. The natural visualization for the behavior of three quantities which sum to one is a ternary plot, and we contour plot density estimates of the

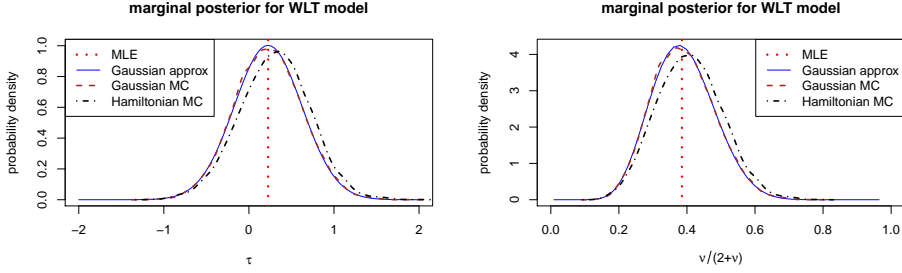


Figure 4. Posterior probability density for the log-tie parameter τ (left) and the associated probability $e^\tau/(2 + e^\tau)$ (right) of a tie game between evenly matched teams, in the Bradley–Terry–Davidson model applied to the 2020–2021 ECAC results, with all overtime games counted as ties. As in Figure 1 and Figure 3, the dashed vertical line is the maximum-likelihood estimate, the solid blue line is a Gaussian approximation to the posterior by expanding about the MAP point $\hat{\tau}$, and the dashed brown and dot-dash black lines are density estimates, respectively, constructed from a Monte Carlo sample from the approximate Gaussian distribution and from the exact distribution using Hamiltonian Monte Carlo. Note that while the MLE is the maximum of the full posterior, it is not quite the maximum of the marginal posterior on τ due to the effects of the marginalization integral. (Note the asymmetrical shape of the exact posteriors in Figure 5.) Additionally, the transformation of τ into $e^\tau/(2 + e^\tau)$ means that the maximum of the posterior probability density is at a different parameter value.

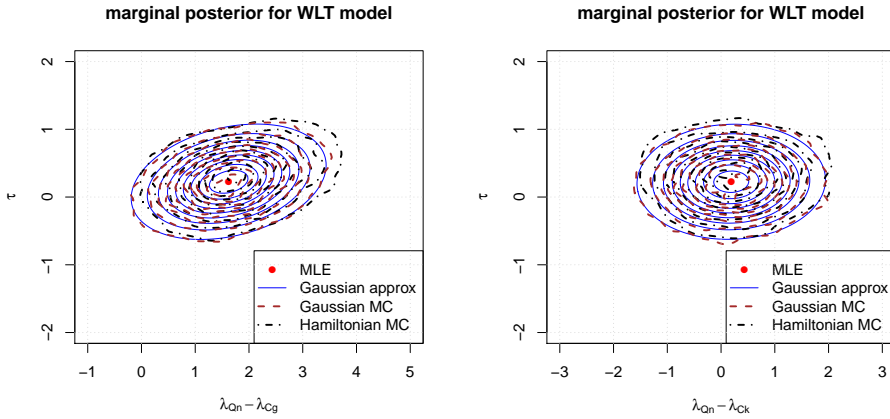


Figure 5. Contours of the joint posterior probability density of the log-strength differences $\gamma_{ij} = \lambda_i - \lambda_j$ shown in Figure 3 and the log-tie parameter τ shown in the left panel of Figure 4. The red circle is the MLE $\hat{\gamma}_{ij}, \hat{\tau}$. The solid blue curves are contours of the Gaussian approximation, and the dashed brown curves are density contours of a Monte Carlo sample drawn from that approximate distribution. The dot-dashed black curves are density contours of a sample from the exact distribution drawn using Hamiltonian Monte Carlo. As with the Bradley–Terry model, the exact and approximate posteriors are comparable, but differences are detectable beyond the level of the Monte Carlo uncertainties illustrated by the difference between the “Gaussian approx” and “Gaussian MC” contours.

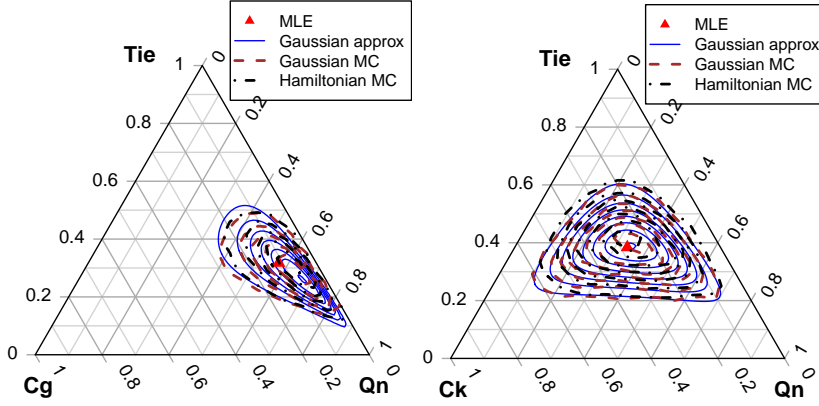


Figure 6. Ternary plots illustrating the joint posterior on θ_{ij}^W , θ_{ij}^T , and θ_{ij}^L , based on the Bradley–Terry–Davidson model applied to the 2020–2021 ECAC results, with all overtime games counted as ties. The horizontal gridlines correspond to lines of constant θ_{ij}^T , with $\theta_{ij}^T = 1$ labelled as “Tie”; the diagonal gridlines correspond to lines of constant θ_{ij}^W or θ_{ij}^L , with $\theta_{ij}^W = 1$ labelled with the abbreviation for team i (“Qn” for Quinnipiac in both cases) and $\theta_{ij}^L = 1$ labelled with the abbreviation for team j (“Cg” for Colgate and “Ck” for Clarkson). The red triangle is the maximum likelihood point $\hat{\theta}_{ij}^T$. Note that for a given set of game results, the maximum likelihood point for all pairs of teams will lie along a one-dimensional curve in the Ternary plot since, for a fixed $\hat{\tau}$, the maximum-likelihood probabilities are functions of the single value $\hat{\gamma}_{ij}$. The three sets of contours are as defined in Figure 5. Note that the MLE is no longer the maximum of the posterior probability density after transforming parameters from γ_{ij}, τ to $\theta_{ij}^W, \theta_{ij}^T$, and $\theta_{ij}^L = 1 - \theta_{ij}^W - \theta_{ij}^T$.

posterior and its Gaussian approximation in Figure 6, along with the maximum likelihood estimates $\{\hat{\theta}_{ij}^T\}$.

4.3. ECAC: Bradley–Terry-like model with overtime/shootout results

Having developed the mechanisms to characterize the posterior distribution for the Bradley–Terry–Davidson model with three outcomes (win, tie, and loss), we apply similar analogues for the model with four outcomes: regulation win (RW), overtime/shootout win (OW), overtime/shootout loss (OL), and regulation loss (RL), now applied to the full 2020–2021 ECAC results shown in Table 1. As before, there is a log-strength parameter λ_i for each team, and τ is now the log of a parameter associated with overtime results. We show the maximum likelihood estimates in Table 5 along with the probabilities $\{\hat{\theta}_{ij}^{RW}\}$ for a regulation win and $\{\hat{\theta}_{ij}^{OW}\}$ for an overtime/shootout win in contests between pairs of teams. In Table 6, we show the parameters of the Gaussian approximation to the posterior.

As in the Bradley–Terry–Davidson model, we can plot the marginal parameters for the differences $\{\gamma_{ij} = \lambda_i - \lambda_j\}$ between pairs of log-strength parameters (Figure 7), the log-overtime parameter τ or equivalently the probability $\frac{\nu}{1+\nu}$, where $(\nu = e^\tau)$ for a game between evenly-matched teams to go to overtime (Figure 8), and the joint marginal posterior in γ_{ij} and τ for our selected pairs of teams (Figure 9).

Table 5. The maximum likelihood estimates for a Bradley–Terry-like model with four game outcomes applied to the 2020–2021 ECAC results. The maximum likelihood estimates $\{\hat{\lambda}_i\}$ and $\hat{\tau}$ of the log-strengths and log overtime parameter are used to compute the estimated probability $\hat{\theta}_{ij}^{\text{RW}}$ for a regulation win and $\hat{\theta}_{ij}^{\text{OW}}$ for an overtime/shootout win between each pair of teams. Note that the estimated probability of a game between evenly-matched teams to go to overtime is $\frac{e^{\hat{\tau}}}{1+e^{\hat{\tau}}} = 0.38$, and it is lower the more different the two teams’ strengths are.

Team i	$\hat{\lambda}_i$	$\hat{\theta}_{ij}^{\text{RW}} (\hat{\theta}_{ij}^{\text{OW}})$			
		Cg	Ck	Qn	SL
Cg	−0.74	—	0.14 (0.13)	0.11 (0.12)	0.32 (0.19)
Ck	0.60	0.53 (0.21)	—	0.26 (0.18)	0.53 (0.20)
Qn	0.93	0.57 (0.20)	0.36 (0.20)	—	0.58 (0.20)
SL	−0.79	0.30 (0.19)	0.13 (0.13)	0.10 (0.11)	—
$\hat{\tau} = -0.49$					

Table 6. The parameters of the the Gaussian approximation to the posterior distribution for the Bradley–Terry-like model with four game outcomes applied to the 2020–2021 ECAC results. In addition to the log-strength parameters $\{\lambda_i\}$, there are uncertainties and correlations associated with the log-overtime parameter τ .

Team i	$\hat{\lambda}_i$	$\sqrt{\Sigma_{ii}}$	$\rho_{ij} = \Sigma_{ij} / \sqrt{\Sigma_{ii}\Sigma_{jj}}$				τ
			Cg	Ck	Qn	SL	
Cg	−0.74	0.48	1.00	−0.34	−0.41	−0.17	−0.19
Ck	0.60	0.54	−0.34	1.00	−0.17	−0.52	0.14
Qn	0.93	0.50	−0.41	−0.17	1.00	−0.38	0.23
SL	−0.79	0.56	−0.17	−0.52	−0.38	1.00	−0.18
τ	−0.49	0.39	−0.19	0.14	0.23	−0.18	1.00

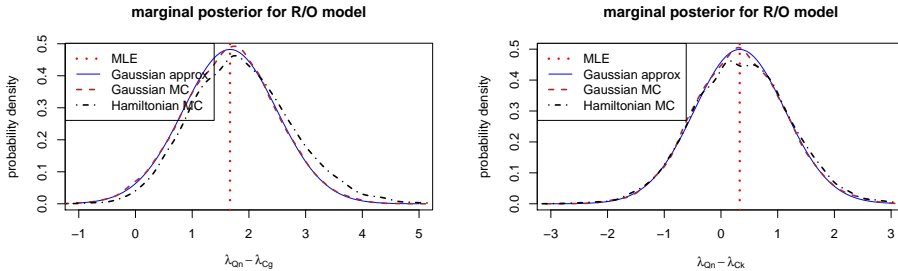


Figure 7. Posterior probability density for difference in log-strengths $\lambda_i - \lambda_j$ between selected pairs of teams (left: Quinnipiac and Colgate; right: Quinnipiac and Clarkson), based on the Bradley–Terry-like model with four game outcomes applied to the 2020–2021 ECAC results.

Curves are as defined in Figure 1.

The posterior distribution on the probabilities

$$\{\theta_{ij}^I | I = \text{RW}, \text{OW}, \text{OL}, \text{RW}\}$$

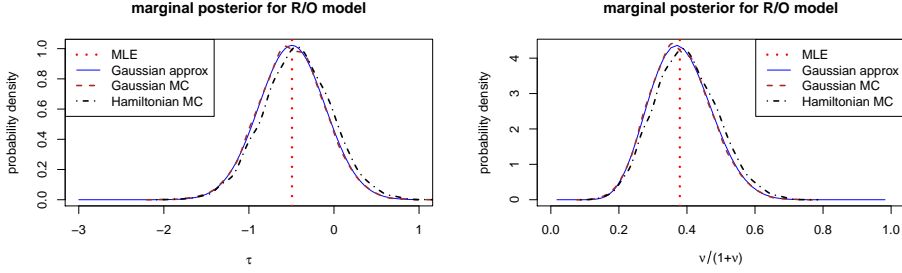


Figure 8. Posterior probability density for the log-overtime parameter τ (left) and the associated probability $\frac{e^\tau}{1+e^\tau}$ (right) of an overtime game between evenly matched teams, in the Bradley–Terry-like model with four game outcomes applied to the 2020–2021 ECAC results. Curves are as defined in Figure 4. Note that the posterior on the overtime probability is very similar to that for the tie probability in the right panel of Figure 4. This is not surprising, since the two calculations are based on different interpretations of the same set of game results, and the “ties” used to generate Figure 4 are just the overtime games in the current computation. The estimates on τ appear different in the two models, but that is mostly because $\nu = e^\tau$ is a measure of the probability of each type of overtime result compared to each type of regulation result, and there are two overtime results in this model and only one in the Bradley–Terry–Davidson model with ties.

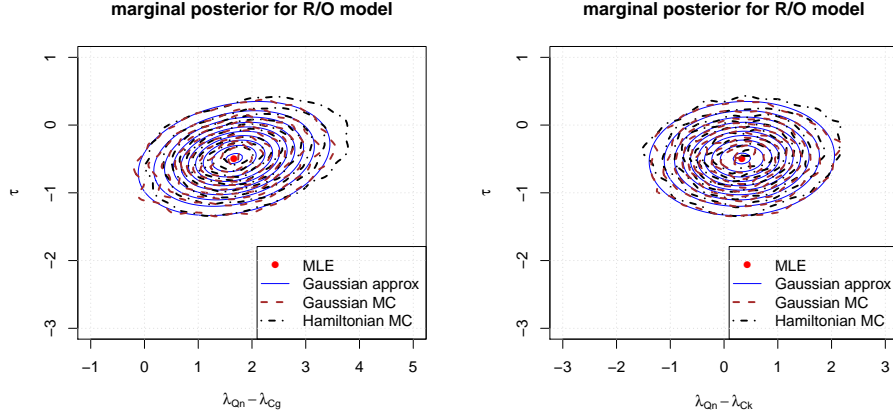


Figure 9. Contours of the joint posterior probability density of the log-strength differences $\gamma_{ij} = \lambda_i - \lambda_j$ shown in Figure 7 and the log-overtime parameter shown in the left panel of Figure 8. Contours are as defined in Figure 5.

is more difficult to visualize, because we have four probabilities which sum to 1, so the posterior can be thought of as defined on the interior of a tetrahedron, which is an example of an Aitchison simplex [1]. However, since all four probabilities are determined by two parameters γ_{ij} and τ , they must lie on a (curved) two-dimensional subsurface of the simplex, defined by the constraint $\frac{\theta_{ij}^{\text{OW}}}{\theta_{ij}^{\text{OL}}} = \left(\frac{\theta_{ij}^{\text{RW}}}{\theta_{ij}^{\text{RL}}} \right)^{1/3}$

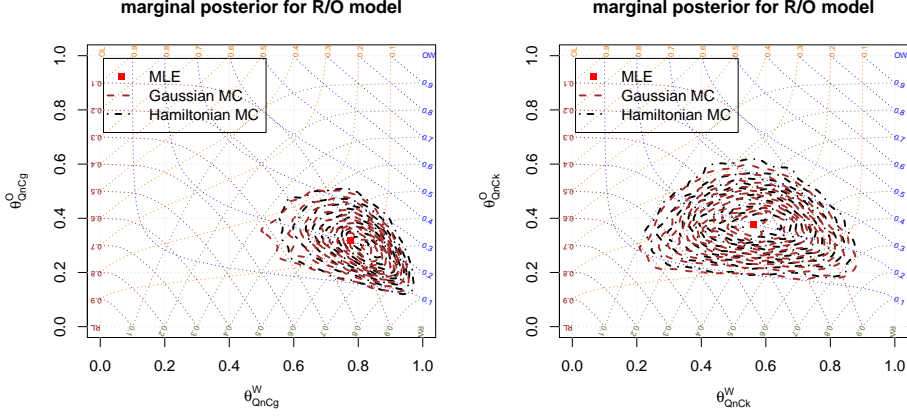


Figure 10. Density contours from the joint posterior probability distribution of $\theta_{ij}^W = \theta_{ij}^{RW} + \theta_{ij}^{OW}$, $\theta_{ij}^O = \theta_{ij}^{OW} + \theta_{ij}^{OL}$, transformed from the joint probability on γ_{ij} and τ shown in Figure 9. Each point on this plot can be converted into a set of probabilities $\{\theta_{ij}^I | I\}$ using the relations $\frac{\theta_{ij}^{OW}}{\theta_{ij}^{OL}} = \left(\frac{\theta_{ij}^{RW}}{\theta_{ij}^{RL}} \right)^{1/3}$ and $\theta_{ij}^{RW} + \theta_{ij}^{OW} + \theta_{ij}^{OL} + \theta_{ij}^{RL} = 1$. The red square is the maximum-likelihood estimate $\hat{\theta}_{ij}^W, \hat{\theta}_{ij}^O$. The dashed brown curves are density contours of a Monte Carlo sample drawn from the Gaussian approximation to the posterior distribution on $\{\lambda_i\}$ and τ . The dot-dashed black curves are density contours of a sample from the exact distribution drawn using Hamiltonian Monte Carlo. As usual, while the MLE is the maximum a posteriori point in the parameters γ_{ij} and τ , it is not so in the parameters shown here due to the transformation of the posterior probability density.

as well as $\theta_{ij}^{RW} + \theta_{ij}^{OW} + \theta_{ij}^{OL} + \theta_{ij}^{RL} = 1$. In Figure 10, we illustrate one possibility for a two-dimensional plot of the marginal posterior on θ_{ij}^I by plotting posterior density contours in $\theta_{ij}^W = \theta_{ij}^{RW} + \theta_{ij}^{OW}$ (the probability of any sort of a win) and $\theta_{ij}^O = \theta_{ij}^{OW} + \theta_{ij}^{OL}$ (the probability of an overtime result). This has the conceptual advantage that each side of the square corresponds to an edge of the tetrahedron, and each vertex of the square corresponds to a vertex of the tetrahedron, at which $\theta_{ij}^I = 1$ for some result I . However, the conversion of a point $\theta_{ij}^W, \theta_{ij}^O$ into θ_{ij}^I is nontrivial and cannot be written in closed form, so further investigation of methods of presenting the posterior is called for.

5. DISCUSSION AND CONCLUSIONS

We have defined a generalization of Davidson’s extension to the Bradley–Terry outcome that handles the set of game outcomes currently distinguished in ice hockey: regulation wins, overtime/shootout wins, overtime/shootout losses, and regulation losses. We have explicitly computed maximum likelihood estimates, constructed a Gaussian approximation to the likelihood, and drawn posterior samples directly from the Gaussian approximation or from the exact posterior using the Hamiltonian Monte Carlo method implemented in Stan. For the data

sets examined, the Gaussian approximation produced similar (but slightly different) results to the exact posterior. The differences in log-team strengths were qualitatively similar among the original Bradley–Terry model (Section 4.1), the Bradley–Terry–Davidson model with ties (Section 4.2), and the new model including regulation and overtime/shootout results (Section 4.3), when applied to the same set of results (albeit with overtime/shootout results interpreted differently). However, these computations are not meant to determine a “best” model, but to illustrate the capabilities of the algorithm. (By definition, we consider the appropriate model to be the one that corresponds to how the league actually assigns values to the results of games in the standings.)

We now wish to discuss some limitations of the work to date, and possible approaches to addressing them: the use of an improper non-informative Haldane prior, the choice of the parameters $\{p_I\}$ in the probability model, and the application of the model to predict the outcomes of playoff games, which may not be played under the same conditions with overtime and shootouts.

First, for simplicity, we worked with a non-informative Haldane prior which was uniform in the log-parameters $\{\lambda_i\}$ and τ , so that the posterior probability distribution in those variables was proportional to the likelihood. There are a number of options for normalizable priors on the distribution of log-strengths $\{\lambda_i\}$ in the Bradley–Terry model (see [17] for a discussion), of which two promising options are a Gaussian prior or a generalized logistic prior [12, 14, 18], each of which has a hyperprior which can be fixed to previous seasons’ data or estimated in a hierarchical model as in [14]. Similar options suggest themselves for the prior on the log-overtime parameter τ , although the situation is somewhat different in that τ has a meaningful origin, so one has to consider a possible location parameter. In particular, it is not clear whether the most natural “origin” for $\nu = e^\tau$ is 1, 2, or something else.

Second, we made something of an arbitrary choice by setting $p_{\text{OW}} = \frac{2}{3}$ and $p_{\text{OL}} = \frac{1}{3}$. In the Bradley–Terry–Davidson model with ties, the requirement that $p_{\text{T}} = p_{-\text{T}} = 1 - p_{\text{T}}$ means $p_{\text{T}} = \frac{1}{2}$ is the only option, as there is only one zero-point system in the three-outcome model. With four outcomes, however, $p_{\text{OW}} = \frac{2}{3}$ is a choice. This choice was of course informed by the point system used for the standings, so that the maximum likelihood equations would enforce that the expected number of points for each team equals its actual number. Other point systems are possible, however. In an earlier experiment with shootouts, the Central Collegiate Hockey Association awarded 5 points for a win in regulation or overtime, 3 for a shootout win, 2 for a shootout loss, and 0 points for a loss in regulation or overtime, so analysis of that season might have used $p_{\text{SW}} = \frac{3}{5}$ and $p_{\text{SL}} = \frac{2}{5}$. Similarly, the NCAA, for tournament selection purposes, considers a win in 3-on-3 overtime worth 0.55 of a win, and treats games decided in a shootout as a tie. Capturing this in a model would require two parameters in addition to the team-strengths: one for overtime games and one for ties, and would have parameters like $p_{\text{RW}} = 1$, $p_{\text{OTW}} = 0.55$, $p_{\text{SO}} = 0.50$, $p_{\text{OTL}} = 0.45$, and $p_{\text{RL}} = 0$. One avenue for future investigation would be to define an extended model in which the unconstrained values of $\{p_I\}$ are treated as additional parameters to be

estimated from the data. For instance, in the four-outcome model, p_{OW} could be treated as a parameter with prior support on the interval $\frac{1}{2} < p_{OW} < 1$.

Finally, the model has assumed all games are played under the same conditions, with 3-on-3 overtimes and shootouts. However, in a number of hockey leagues, playoffs and other postseason games are played to conclusion with overtimes played under the same set of rules with a full squad on the ice, and shootouts are not possible. To produce probabilities for such a game, one would have to decide what probability to assign to a win or a loss. The natural model is probably to use $\theta_{ij}^{POW} = \pi_i/(\pi_i + \pi_j)$, i.e., the conditional probability of winning a game given that it is not decided in (3-on-3) overtime or a shootout. Likewise, if any playoff games are included in the results used for inference, their contribution to the likelihood would need to be adjusted.

APPENDIX A. STAN MODEL

Here, we show the Stan model implementing the family of Bradley-Terry-like models described in this paper. The generalization allows a single Stan dynamic shared object (DSO) [16] to be used for all three models. This is computationally convenient, because compiling the DSO is often the most time-consuming part of a Stan simulation.

```
data {
  int<lower=1> nteams;
  int<lower=1> nres;
  int<lower=0> n_ttR[nteam,nteam,nres];
  int<lower=0,upper=1> o_R[nres];
  real p_R[nres];
}
parameters {
  real omega_t[nteam-1];
  real tau;
}
model {
  int n_tt[nteam,nteam];
  real gamma_tt[nteam,nteam];
  real denom_tt[nteam,nteam];
  int sumo;
  vector[nres] gamma_ttR[nteam,nteam];
  vector[nres] theta_ttR[nteam,nteam];
  sumo = 0;
  for (R in 1:nres) {
    sumo += o_R[R];
  }
  // Hack to keep tau from going crazy if the model does not have ties
  if (sumo == 0) {
    target += std_normal_lpdf(tau);
  }
  for (i in 1:(nteam-1)) {
    for (j in (i+1):nteam) {
      n_tt[i,j] = 0;
    }
  }
}
```

```

for (R in 1:nres) {
  n_tt[i,j] += n_ttR[i,j,R];
}
if ( n_tt[i,j]>0 ) {
  gamma_tt[i,j] = 0;
  for (k in i:(j-1)) {
    gamma_tt[i,j] += omega_t[k];
  }
  for (R in 1:nres) {
    gamma_ttR[i,j,R] = o_R[R] * tau + p_R[R] * gamma_tt[i,j];
  }
// Unfortunately only supported in Stan 2.24 and above
//   n_ttR[i,j] ~ multinomial_logit(gamma_ttR[i,j]);
  denom_tt[i,j] = 0;
  for (R in 1:nres) {
    denom_tt[i,j] += exp(gamma_ttR[i,j,R]);
  }
  for (R in 1:nres) {
    theta_ttR[i,j,R] = exp(gamma_ttR[i,j,R]) / denom_tt[i,j];
  }
  n_ttR[i,j] ~ multinomial(theta_ttR[i,j]);
}
}
}
}

```

Acknowledgment. JTW wishes to thank Adam Wodon, Kenneth Butler, Gabriel Phelan, the attendees of the UP-STAT conferences, and the members of the Schwerpunkt Stochastik at Goethe University, Frankfurt am Main, for useful discussions. Game results for the computations in this paper were collected from <https://www.collegehockeynews.com/>.

The R code used to perform the simulations in this paper is available at <https://gitlab.com/jtwsma/bradley-terry>.

REFERENCES

- [1] J. Aitchison, *The statistical analysis of compositional data*, J. R. Stat. Soc., Ser. B **44** (1982), 139–177.
- [2] A. Albert and J. A. Anderson, *On the existence of maximum likelihood estimates in logistic regression models*, Biometrika **71** (1984), 1–10.
- [3] R. A. Bradley and M. E. Terry, *Rank analysis of incomplete block designs: The method of paired comparisons*, Biometrika **39** (1952), 324–345.
- [4] J. S. Bridle, *Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition*, in: F. F. Soulié and J. Hérault (eds.), *Neurocomputing*, NATO ASI Series, Series F: Computer and Systems Sciences **68**, Springer, Berlin, Heidelberg, 1990, pp. 227–236.
- [5] K. Butler, Ken’s Ratings for American College Hockey, HOCKEY-L mailing list post, 1993.
- [6] K. Butler and J. T. Whelan, *The existence of maximum-likelihood estimates in the Bradley–Terry model and its extensions*, arXiv:math/0412232, 2004.
- [7] R. R. Davidson, *On extending the Bradley–Terry model to accommodate ties in paired comparison experiments*, J. Am. Stat. Assoc. **65** (1970), 317–328.

- [8] L. R. Ford, Jr., *Solution of a ranking problem from binary comparisons*, Am. Math. Mon. **64** (1957), 28–33.
- [9] J. B. S. Haldane, *A note on inverse probability*, Math. Proc. Camb. Philos. Soc. **28** (1932), 55–61.
- [10] International Ice Hockey Federation, *Tie games are history; a win earns three points for teams*, IIHF 100 Top Stories of the Century, 2006.
- [11] H. Jeffreys, *Theory of Probability*, 3rd ed., Clarendon Press, Oxford, 1961.
- [12] T. Leonard, *An alternative Bayesian approach to the Bradley–Terry model for paired comparisons*, Biometrics **33** (1977), 121–132.
- [13] R. Penrose, *A generalized inverse for matrices*, Math. Proc. Camb. Philos. Soc. **51** (1955), 406–413.
- [14] G. C. Phelan and J. T. Whelan, *Hierarchical Bayesian Bradley–Terry for applications in major league baseball*, Math. Appl. **7** (2017), 71–84.
- [15] T. J. Santner and D. E. Duffy, *A Note on A. Albert and J. A. Anderson’s conditions for the existence of maximum likelihood estimates in logistic regression models*, Biometrika **73** (1986), 755–758.
- [16] Stan Development Team, *RStan: the R interface to Stan*, R package version 2.28, 2019.
- [17] J. T. Whelan, *Prior distributions for the Bradley–Terry model of paired comparisons*, arXiv:1712.05311, 2017.
- [18] J. T. Whelan and A. Wodon, *Prediction and Evaluation in College Hockey using the Bradley–Terry–Zermelo Model*, Math. Appl. **8** (2019), 131–149.
- [19] E. Zermelo, *Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung*, Math. Z. **29** (1929), 436–460.

John T. Whelan, School of Mathematical Sciences and Center for Computational Relativity and Gravitation, Rochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, New York 14623, USA, and Institute for Theoretical Physics, Goethe University Frankfurt, Max-von-Laue Str. 1, D-60438 Frankfurt am Main, Germany
e-mail: `jtwsm@rit.edu`

Jacob E. Klein, School of Mathematical Sciences, Rochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, New York 14623, USA
e-mail: `jek8543@rit.edu`

