Math. Appl. **8** (2019), 131–149
DOI: 10.13164/ma.2019.09

PREDICTION AND EVALUATION IN COLLEGE HOCKEY USING THE BRADLEY–TERRY–ZERMELO MODEL

JOHN T. WHELAN AND ADAM WODON

Abstract. We describe the application of the Bradley–Terry model to NCAA Division I Men's Ice Hockey. A Bayesian construction gives a joint posterior probability distribution for the log-strength parameters, given a set of game results and a choice of prior distribution. For several suitable choices of prior, it is straightforward to find the maximum a posteriori point (MAP) and a Hessian matrix, allowing a Gaussian approximation to be constructed. Posterior predictive probabilities can be estimated by 1) setting the log-strengths to their MAP values, 2) using the Gaussian approximation for analytical or Monte Carlo integration, or 3) applying importance sampling to re-weight the results of a Monte Carlo simulation. We define a method to evaluate any models which generate predicted probabilities for future outcomes, using the Bayes factor given the actual outcomes, and apply it to NCAA tournament results. Finally, we describe an on-line tool which currently estimates probabilities of future results using MAP evaluation and describe how it can be refined using the Gaussian approximation or importance sampling.

1. Introduction

Sporting events, specifically games between pairs of teams, are a form of paired comparison experiment, where one team (the winner) is chosen over the other (the loser). The Bradley–Terry model associates the probability of the outcome of each paired comparison with the inherent strengths of each team. Estimates of these strengths can be used to construct a rating system which allows the ranking of teams based on game outcomes when imbalances in strength of schedule make simple winning percentage (fraction of games won) an unfair basis for ranking. Because the strength parameters are also associated with the probabilities of game results, they can be used to predict the outcome of future games. This paper considers the application of this technique to NCAA Division I Men's Ice Hockey, the highest level of college hockey competition in the United States.

The paper is organized as follows: In Section 1.1 we define the Bradley–Terry model and its use in constructing posterior estimates of team strengths. In Section 1.2 we describe the particulars of the college hockey season and postseason. In Section 2 we describe several methods for estimating the posterior predictive probabilities of the outcome of future games. In Section 3, we describe two applications of these techniques: the evaluation of the model constructed from the regular season by use of a Bayes factor associated with NCAA tournament results;

MSC (2010): primary 62F15; secondary 62F07.
Keywords: Bradley–Terry model, College Hockey.

and the *Pairwise Probability Matrix*, used during the season to construct predicted probabilities for a team's ranking associated with the NCAA tournament selection criteria.

1.1. The Bradley–Terry model

Given a set of t teams, the Bradley–Terry–Zermelo model [2, 19] associates with each team $i = 1, \ldots, t$ a log-strength parameter λ_i, and defines the probability of team i winning a given game with team j such that the odds ratio is the ratio of their strengths, i.e., the probability is

$$\theta_{ij} = \frac{e^{\lambda_i}}{e^{\lambda_i} + e^{\lambda_j}} = \operatorname{logistic}(\lambda_i - \lambda_j).$$

In this paper, we will work in terms of the log-strength $\lambda_i \in (-\infty, \infty)$.

Given a series of games among the teams in which a pair of teams i, j play $n_{ij} = n_{ji}$ times, the Bradley–Terry model defines a probability for a set of outcomes D which includes w_{ij} wins (and $n_{ij} - w_{ij}$ losses) for team i against team j:

$$P(D|\{\lambda_i\}) = \prod_{i=1}^{t}\prod_{j=1}^{t} \theta_{ij}^{w_{ij}} = \prod_{i=1}^{t}\prod_{j=i+1}^{t} \theta_{ij}^{w_{ij}} (1 - \theta_{ij})^{n_{ij} - w_{ij}}.$$

Note that if the order of the outcomes of games between pairs of teams is ignored, the sampling distribution for w_{ij} is

$$p(\{w_{ij}\}|\{\lambda_i\}) = \prod_{i=1}^{t}\prod_{j=i+1}^{t} \binom{n_{ij}}{w_{ij}} \theta_{ij}^{w_{ij}} (1 - \theta_{ij})^{n_{ij} - w_{ij}}$$

but the inferences about parameters $\{\lambda_i\}$ are unchanged.

The log-likelihood can be written in terms of the total number of wins $v_i = \sum_{j=1}^{t} w_{ij}$ as

$$\ln P(D|\{\lambda_i\}) = \sum_{i=1}^{t} v_i \lambda_i - \frac{1}{2}\sum_{i=1}^{t}\sum_{j=1}^{t} n_{ij} \ln\left(e^{\lambda_i} + e^{\lambda_j}\right)$$

so that the maximum likelihood equations are

$$v_i = \sum_{j=1}^{t} n_{ij} \frac{e^{\widehat{\lambda}_i}}{e^{\widehat{\lambda}_i} + e^{\widehat{\lambda}_j}} = \sum_{j=1}^{t} n_{ij} \widehat{\theta}_{ij}.$$

The maximum-likelihood log-strengths $\{\widehat{\lambda}_i\}$ are those for which the predicted number of wins $\sum_{j=1}^{t} n_{ij} \widehat{\theta}_{ij}$ for each team equals the actual number v_i. They can be found, e.g., by Ford's method, in which one iterates the equation [7]

$$\widehat{\lambda}_i = \ln\left(v_i \Bigg/ \sum_{j=1}^{t} \frac{n_{ij}}{e^{\widehat{\lambda}_i} + e^{\widehat{\lambda}_j}} \right).$$

Because the maximum-likelihood equations depend only on the differences $\widehat{\lambda}_i - \widehat{\lambda}_j$, the estimates $\{\widehat{\lambda}_i\}$ are defined only up to an overall additive constant.

Given a prior distribution $f(\{\lambda_i\}|I)$ for the log-strengths, the posterior distribution given the game results D will be

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

The maximum a posteriori (MAP) estimates $\{\widetilde{\lambda_i}\}$ of the log-strengths will be the solution to

$$v_i + \frac{\partial}{\partial \lambda_i} \ln f(\{\lambda_j\}|I)\bigg|_{\{\lambda_j=\widetilde{\lambda}_j\}} = \sum_{j=1}^{t} n_{ij}\widetilde{\theta}_{ij}.$$

For the sake of mathematical simplicity, we will often use the Haldane prior[1]

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

This is an improper prior but can be formally understood as the limiting form of a family of normalized priors.

Other convenient families of priors [17] are the generalized logistic prior

$$f(\{\lambda_i\}|I_\eta) = \prod_{i=1}^{t} \frac{\Gamma(2\eta)}{[\Gamma(\eta)]^2} \frac{1}{(1+e^{\lambda_i})^\eta (1+e^{-\lambda_i})^\eta}$$

and the Gaussian prior

$$f(\{\lambda_i\}|I_\sigma) = \prod_{i=1}^{t} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\lambda_i)^2}{2\sigma^2}\right).$$

The Haldane prior is the limit of the generalized logistic prior as $\eta \to 0$ and the Gaussian prior as $\sigma \to \infty$.

Since the generalized logistic prior has

$$\ln f(\{\lambda_j\}|I_\eta) = \sum_{j=1}^{t} \left[-\eta \ln(1+e^{-\lambda_j}) - \eta \ln(1+e^{\lambda_j})\right] + \text{constant}$$

and

$$\frac{\partial}{\partial \lambda_i} \ln f(\{\lambda_j\}|I_\eta) = \frac{\eta\, e^{-\lambda_i}}{1+e^{-\lambda_i}} - \frac{\eta\, e^{\lambda_i}}{1+e^{\lambda_i}} = \eta(1-2\theta_{i0})$$

where $\theta_{i0} = \text{logistic}(\lambda_i)$ is the probability that team i would win a game against a team with log-strength zero. This means that the MAP equations with the generalized logistic prior are

$$\eta + v_i = 2\eta\widetilde{\theta}_{i0} + \sum_{j=1}^{t} n_{ij}\widetilde{\theta}_{ij}.$$

This is just the same as we'd obtain from the maximum-likelihood equations after the addition of 2η "fictitious games" against a team with log-strength zero, half wins and half losses, for each team. As such, the MAP equations can be solved by a straightforward extension of Ford's method.

[1] So named because the marginal prior distribution for any θ_{ij} will follow the Haldane prior [8, 9], which is the limit of a Beta(α, β) distribution as $\alpha, \beta \to 0$.

With the Gaussian prior, the MAP equations become

$$v_i = \frac{\widetilde{\lambda}_i}{\sigma^2} + \sum_{j=1}^{t} n_{ij} \widetilde{\theta}_{ij}.$$

Note that these cannot in general be solved by iterating

$$\widetilde{\lambda}_i = \ln \left(\frac{v_i - \frac{\widetilde{\lambda}_i}{2\sigma^2}}{\sum_{j=1}^{t} \frac{n_{ij}}{e^{\widetilde{\lambda}_i} + e^{\widetilde{\lambda}_j}}} \right)$$

as suggested in [12] because, for small values of σ, the argument of the logarithm may become negative.

1.2. College Hockey

The NCAA (National Collegiate Athletic Association) Men's Division I Ice Hockey competition consists, at present, of 60 teams which play approximately 30 to 40 games each during the season. At the end of the season, 16 teams are selected (the champions of six conference tournaments, plus an additional 10 teams chosen according to a set of selection criteria related to the outcomes of their games) to play in a single-elimination tournament to determine the national champion. As the games during the season are played within six conferences (with one team currently competing as an independent), with additional non-conference games, in-season tournaments, and conference playoff tournaments, teams will typically face schedules of differing strengths. Rating systems have thus been devised to evaluate their game results more fairly than would be possible by simply comparing winning percentages (fraction of games won). The selection criteria for the NCAA tournament are of this sort, notably the ratings percentage index (RPI) which combines a team's winning percentage with average winning percentages of its opponents and opponents' opponents. The maximum-likelihood Bradley–Terry strengths are also used, under the name "Ken's Ratings for American College Hockey" (KRACH) [3].

In addition to a win or a loss, some college hockey games can end in a tie. In computing NCAA selection criteria, a team which wins (whether in regulation play or overtime) is awarded 2 points, a team which loses receives 0 points, and if a game ends in a tie (after overtime), each team receives 1 point. (Penalty shootouts, which may occur after a tie in some competitions, are not considered for NCAA selection purposes.) In principle, one could use an extension of the Bradley–Terry model with an additional parameter or parameters accounting for the probability of ties. [5, 10, 13] However, this is complicated by the fact that some college hockey games can end in ties, while others (mostly conference playoff and NCAA tournament games, but also some games in in-season invitational tournaments) continue in overtime until a winner is decided. Rather than keep track of the two sorts of games, in this work we perform all computations with ties contributing 0.5 to the win total and 0.5 to the loss total for each team. While this introduces a conceptual inconsistency (since the formulas were derived without consideration for the possibility of ties), it poses no impediment to the calculations, and ties are rare enough that no pathological conclusions have yet been encountered.

2. Posterior predictive probabilities

In this paper, we are interested in the calculation of posterior predictive probabilities. I.e., given a set of results D and a choice of prior I, so that the Bayesian Bradley–Terry model produces a posterior pdf $f(\{\lambda_i\}|D,I)$, what is the probability that some future games will have outcome O? This outcome may be a specified set of wins and losses, but it may also be a more coarse-grained alternative, such as that a particular team is chosen for the tournament field by winning its conference championship game or finishing the season well enough according to the selection criteria. If $P(O|\{\lambda_i\})$ is the probability of that outcome using the Bradley–Terry model with log-strengths $\{\lambda_i\}$ (which might itself be calculated according to some non-trivial summation technique) then the posterior predictive probability for O given the previous results D and prior information I is obtained by marginalizing over the log-strengths:

$$P(O|D,I) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(O|\{\lambda_i\})\, f(\{\lambda_i\}|D,I)\, d^t\lambda. \tag{2.1}$$

While the integrand of (2.1) may be straightforward to construct, the t-dimensional integral is in general impossible to evaluate analytically and impractical to compute via direct numerical integration. One approach is to draw samples from the posterior $f(\{\lambda_i\}|D,I)$ via Markov Chain Monte Carlo methods, such as the Hamiltonian Monte Carlo used in [12]. In this paper, we consider simpler techniques which apply approximation methods.

2.1. MAP evaluation

The simplest method is to use the Bradley–Terry model with the team strength parameters set to their maximum a posteriori values $\{\widetilde{\lambda}_i\}$, and evaluate $P(O|\{\widetilde{\lambda}_i\})$. This is equivalent to replacing the full posterior $f(\{\lambda_i\}|D,I)$ with a t-dimensional delta function at the MAP point. While this is convenient, it is clearly an oversimplification, since it fails to account for the posterior uncertainty in the team strengths.

2.2. Gaussian approximation

One way to quantify the uncertainty in the posterior, and obtain a better approximation, is to Taylor expand the log-posterior $\ln f(\{\lambda_i\}|D,I)$ about the MAP point and obtain a Gaussian approximation

$$g(\{\lambda_i\}|D,I) = \text{const} \times \exp\left(-\frac{1}{2}\sum_{i=1}^{t}\sum_{j=1}^{t}\left[\lambda_i - \widetilde{\lambda}_i\right] H_{ij} \left[\lambda_j - \widetilde{\lambda}_j\right]\right)$$

where the Hessian matrix $\{H_{ij}\}$ has elements

$$\begin{aligned} H_{ij} &= -\left.\frac{\partial^2}{\partial\lambda_i\,\partial\lambda_j} \ln f(\{\lambda_i\}|D,I)\right|_{\{\lambda_k=\widetilde{\lambda}_k\}} \\ &= -n_{ij}\widetilde{\theta}_{ij}\widetilde{\theta}_{ji} + \delta_{ij}\sum_{k=1}^{t}\widetilde{\theta}_{ik}\widetilde{\theta}_{ki} - \left.\frac{\partial^2}{\partial\lambda_i\,\partial\lambda_j}\ln f(\{\lambda_i\}|I)\right|_{\{\lambda_k=\widetilde{\lambda}_k\}}. \end{aligned}$$

For the Haldane prior, the last term vanishes; for the generalized logistic prior it is

$$-\left.\frac{\partial^2}{\partial\lambda_i\,\partial\lambda_j}\ln f(\{\lambda_i\}|I_\eta)\right|_{\{\lambda_k=\widetilde{\lambda}_k\}} = \delta_{ij}\,\widetilde{\theta}_{i0}(1-\widetilde{\theta}_{i0})$$

and for the Gaussian prior, it is

$$-\left.\frac{\partial^2}{\partial\lambda_i\,\partial\lambda_j}\ln f(\{\lambda_i\}|I_\sigma)\right|_{\{\lambda_k=\widetilde{\lambda}_k\}} = \frac{\delta_{ij}}{\sigma^2}.$$

The Gaussian approximation is a multivariate Gaussian distribution with mean $\{\widetilde{\lambda}_i\}$ and a variance-covariance matrix $\{\Sigma_{ij}\}$ which is the matrix inverse of the Hessian $\{H_{ij}\}$. Given a suitable prior distribution, $\{H_{ij}\}$ is invertible. Using the Haldane prior produces a Hessian matrix

$$H_{ij} = -n_{ij}\widehat{\theta}_{ij}\widehat{\theta}_{ji} + \delta_{ij}\sum_{k=1}^{t}\widehat{\theta}_{ik}\widehat{\theta}_{ki} = \sum_{k=1}^{t} h_k \ell_i^{(k)}\ell_j^{(k)}$$

where we have decomposed the Hessian matrix using its orthonormal eigenvectors $\sum_{i=1}^{t}\ell_i^{(k)}\ell_i^{(\ell)} = \delta_{k\ell}$, $\sum_{j=1}^{t}H_{ij}\ell_j^{(k)} = h_k\ell_i^{(k)}$, $h_i \le h_{i+1}$. There will be at least one zero eigenvalue $h_1 = 0$, corresponding to the eigenvector $\{\ell_i^{(1)}\} = \{\frac{1}{\sqrt{t}}, \frac{1}{\sqrt{t}}, \dots, \frac{1}{\sqrt{t}}\}$. If all of the maximum-likelihood estimates $\{\widehat{\lambda}_i - \widehat{\lambda}_j\}$ are finite and well-determined [1,4,14], which will nearly always be the case late in a season with as many games as college hockey, the other eigenvalues $\{h_i|i = 2,\dots,t\}$ will all be positive. Since the transformation $\lambda_i \to \lambda_i + a\ell_i^{(1)}$, for any $a \in \mathbb{R}$, doesn't change the probabilities $P(O|\{\lambda_i\})$, we can replace the Gaussian approximate distribution, which leaves $\sum_{i=1}^{t}\ell_i^{(1)}\lambda_i$ unspecified, and is therefore unnormalizable, with one which fixes $\sum_{i=1}^{t}\ell_i^{(1)}\lambda_i = 0$. This is a multivariate Gaussian distribution whose mean is $\{\widehat{\lambda}_i\}$ and whose variance-covariance matrix is the Moore–Penrose pseudo-inverse [11] of H_{ij}, i.e.,

$$\Sigma_{ij} = \sum_{k=2}^{t}\frac{\ell_i^{(k)}\ell_j^{(k)}}{h_k}\ ,$$

defined such that

$$\sum_{k=1}^{t}\Sigma_{ik}H_{kj} = \sum_{k=1}^{t}H_{ik}\Sigma_{kj} = \delta_{ij} - \ell_i^{(1)}\ell_j^{(1)}.$$

Depending on the specifics of the outcome O, it may be possible to evaluate the approximate integral

$$P(O|D,I) \approx \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} P(O|\{\lambda_i\})\,g(\{\lambda_i\}|D,I)\,d^t\lambda$$

analytically using the Gaussian approximation. More likely it will be necessary to use a Monte Carlo technique, drawing N samples $\{\lambda_i^{(s)}\}$ from the multivariate Gaussian distribution $N_t(\{\widetilde{\lambda}_i\},\{\Sigma_{ij}\})$ and estimating

$$P(O|D,I) \approx \frac{1}{N}\sum_{s=1}^{N} P(O|\{\lambda_i^{(s)}\}). \tag{2.2}$$

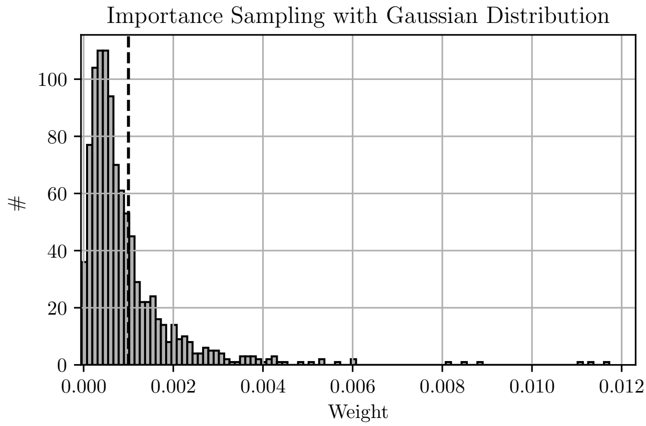

Figure 1. Histogram of weights in importance sampling using 1000 draws from a multivariate ($t = 60$, one degenerate degree of freedom) Gaussian sampling distribution with to approximate the Bradley–Terry posterior, starting with the Haldane prior and the results of the 2018-2019 NCAA Division I Men's Ice Hockey season prior to NCAA tournament selection. The dashed line indicates average weight of 0.001.

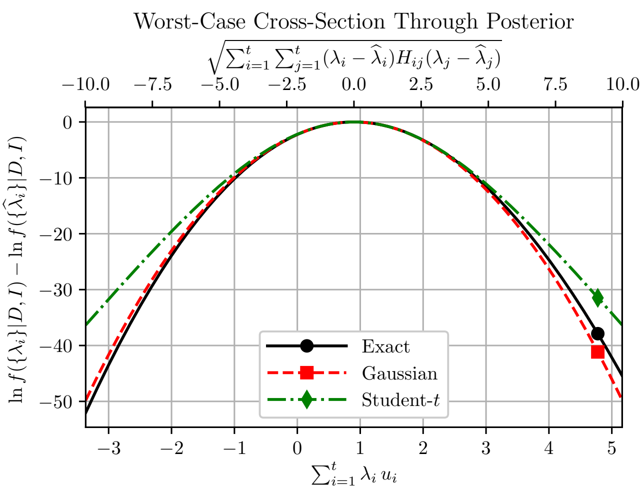

Figure 2. Cross-section (conditional distribution) of the log-posterior $\ln f(\{\lambda_i^{(s)}\}|D, I)$ through the point with the highest weight in the Gaussian importance sampling. While the multivariate Gaussian is a good approximation for some distance from the MAP point, the presence of $t-1 = 59$ meaningful parameters means that even seemingly large outliers can occur in the Gaussian sample.

2.3. Importance sampling

Since $g(\{\lambda_i\}|D, I)$ is only an approximation to $f(\{\lambda_i\}|D, I)$, a natural correction to (2.2) is to use importance sampling, weighting the probability $P(O|\{\lambda_i^{(s)}\})$ coming from each Monte Carlo draw by a factor

$$w_s \propto \frac{f(\{\lambda_i^{(s)}\}|D,I)}{g(\{\lambda_i^{(s)}\}|D,I)}$$

so that

$$P(O|D,I) \approx \frac{\sum_{s=1}^{N} w_s P(O|\{\lambda_i^{(s)}\})}{\sum_{s=1}^{N} w_s} \ .$$

If there are values of $\{\lambda_i\}$ for which

$$\frac{g(\{\lambda_i^{(s)}\}|D,I)}{f(\{\lambda_i^{(s)}\}|D,I)} \gg \frac{1}{N}\sum_{s'=1}^{N}\frac{g(\{\lambda_i^{(s')}\}|D,I)}{f(\{\lambda_i^{(s')}\}|D,I)},$$

the importance sampling procedure may produce erratic results. We see this effect when we use the Gaussian approximation for importance sampling. We see in Figure 1 that a few outliers produce large weight factors to try to adjust for the fact that the tail of the posterior is longer than that of the approximate Gaussian distribution.

This is illustrated in Figure 2, which shows a slice through the log-posterior including the maximum-likelihood point $\{\widehat{\lambda}_i\}$ and the point $\{\lambda_i^{(\hat{s})}\}$ with the largest importance sampling weight. The lower x-axis shows the projection of the vector $\{\lambda_i\}$ onto the unit vector $\{u_i\}$ defined by

$$u_i = \frac{\lambda_i^{(\hat{s})} - \widehat{\lambda}_i}{\sqrt{\sum_{j=1}^{t}(\lambda_j^{(\hat{s})} - \widehat{\lambda}_j)^2}}.$$

The upper x-axis shows the normalized distance

$$\sqrt{\sum_{i=1}^{t}\sum_{j=1}^{t}\left[\lambda_i - \widehat{\lambda}_i\right] H_{ij} \left[\lambda_j - \widehat{\lambda}_j\right]}$$

from the maximum-likelihood point. The largest importance sampling rate occurs at a normalized distance of 9.072-sigma from the ML point. This is not as large an outlier as it might seem. With $t-1=59$ meaningful parameters, the normalized distance in a Gaussian Monte Carlo will be a chi-distributed random variable with $t-1=59$ degrees of freedom, whose probability density function is shown in Figure 3. It shows that the samples are overwhelmingly likely to be found between 5 and 10 sigma from the ML point. This means that using a “heavy-tailed” distribution such as the multivariate Student-t distribution will not improve the situation. While the Student-t distribution has more support at large distances relative to the maximum-likelihood point, essentially none of the random samples will be near the ML point. Instead, the differently-shaped tails of the t-distribution would cause points less far from the ML point to be undersampled relative to points farther out, producing larger outliers in the normalized weights. This is illustrated in Figure 2 for a multivariate Student-t distribution with $\nu = 59$ degrees of freedom. The covariance matrix has been scaled up so that a one-dimensional “slice” through the maximum, i.e., the conditional distribution [6] is a Student-t

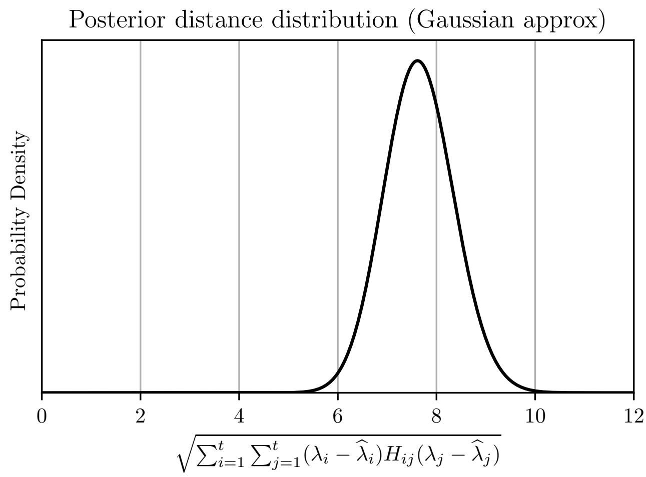

Figure 3. Marginal probability density function for the normalized distance from the maximum-likelihood point under multivariate Gaussian random sampling, which is a $\chi(59)$ random variable. Essentially all the probability weight is between 5 and 10 sigma from the ML point.

distribution with $\nu + (t-1) - 1$ degrees of freedom and (pseudo-)inverse scale matrix $\{H_{ij}\}$.

Given that the departure of the posterior from normality (at least in this example) is a matter of slight skewness than heavy tails, an avenue for future exploration is importance sampling with a skew distribution, as proposed in [15, 16].

3. Applications

3.1. Evaluation via Bayes factor

In Section 2, we described methods to calculate or approximate the probability of future outcomes O using the Bradley–Terry model. We now describe a simple method for evaluating any set of predictions. Suppose $P(O|M_1, D, I)$ and $P(O|M_2, D, I)$ are the probabilities assigned to a future outcome O by two different methods M_1 and M_2, given past results D and any additional information I. (These should be defined so that $\sum_O P(O|M, D, I) = 1$ for any exhaustive set of mutually exclusive outcomes $\{O\}$.) A general method for comparing M_1 and M_2 is the Bayes factor

$$B_{12} = \frac{P(O|M_1, D, I)}{P(O|M_2, D, I)}$$

which is the factor by which the posterior odds ratio for M_1 over M_2 increases relative to the prior odds ratio:

$$\frac{P(M_1|O, D, I)}{P(M_2|O, D, I)} = \frac{P(O|M_1, D, I)}{P(O|M_2, D, I)} \frac{P(M_1|I)}{P(M_2|I)} = B_{12} \frac{P(M_1|I)}{P(M_2|I)}.$$

We can apply this technique to any method of generating probabilities for future outcomes of hockey games (not just Bradley–Terry). We can think of the results

D as "training data" and the outcome O as describing the "evaluation data" of the rest of the games. We consider a straightforward example, where the training data are the games of each season prior to tournament selection and the evaluation data are the NCAA tournament games, with O being the actual sequence of results which occurred. Note that for this evaluation calculation, we don't actually need to know $P(O|M,D,I)$ for each possible outcome, only for the exact sequence of results which occurred. For convenience, we compare each model to a "tossup model" M_0 in which each team is assigned a 50% chance to win each game, for which $P(O|M_0,D,I) = 2^{-n_O}$ where n_O is the number of games in the evaluation data set. Evidently $B_{12} = B_{10}/B_{20}$.

If M_{mle} is the MAP evaluation method of Section 2.1, in which all probabilities are independently assigned using the maximum-likelihood Bradley–Terry estimates (the KRACH ratings),

$$B_{\text{mle}0} = \prod_{g=1}^{n_O} 2\widehat{\theta}_{w_g l_g}$$

where w_g is the winner and l_g the loser of game g. So we see that for each game predicted "correctly" (winner assigned a greater than 50% probability), the Bayes factor increases by a factor of up to 2. However, for each game predicted "incorrectly" (winner assigned a less than 50% probability), the Bayes factor decreases. If a result occurs which the model considered impossible, the Bayes factor is zero. We can illustrate this with the results of the 2019 NCAA tournament, in Figure 4 we see that the Bayes factor using all the results of the tournament is actually slightly lower than 1. This is because the upset of American International College defeating St. Cloud State was such a surprise according to the model.

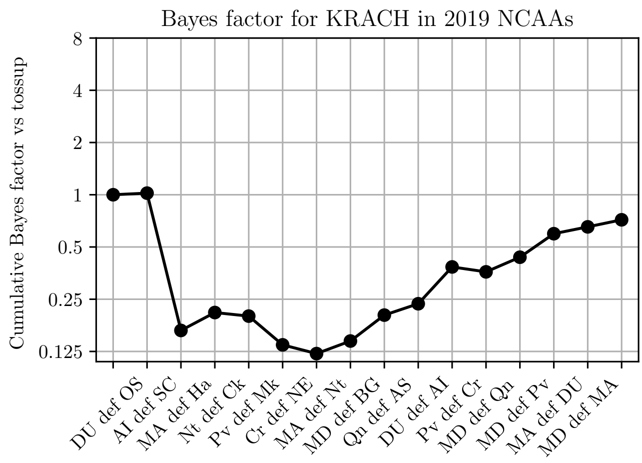

Figure 4. Evolution of the Bayes factor for the predictions of the maximum likelihood Bradley–Terry model (KRACH) over the 2019 NCAA tournament.

If we compute the Bayes factor using the predictions and outcomes of multiple NCAA tournaments (using the game results from each season to produce probabilities for that season's tournament), we begin to see distinctions between models. In Figure 5 we plot the evolution of this cumulative Bayes factor over the NCAA tournaments from 2003 (the first year of the current 16-team format) to 2019. In addition to the maximum likelihood/KRACH model, we plot the Bayes factor for a model with a generalized logistic prior with $\eta = 1$ (estimated using the Gaussian approximation and 20,000 Monte Carlo draws)[2], along with a simple model based on the win ratios $v_i/(n_i - v_i)$ for each team, where the probability that team i will beat team j is assumed to be $\theta_{ij}^{\mathrm{wr}}$, where

$$\frac{\theta_{ij}^{\mathrm{wr}}}{\theta_{ji}^{\mathrm{wr}}} = \frac{\theta_{ij}^{\mathrm{wr}}}{1 - \theta_{ij}^{\mathrm{wr}}} = \sqrt{\frac{v_i}{n_i - v_i} \cdot \frac{n_j - v_j}{v_j}}$$

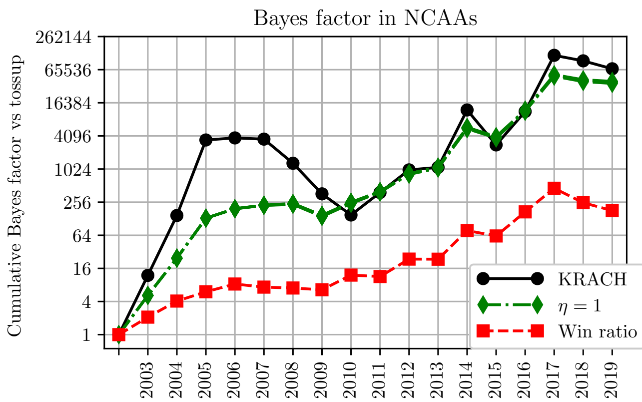

Figure 5. Evolution of the Bayes factor for the predictions of the three models over the NCAA tournament since the 16-team format was introduced in 2003: the maximum likelihood Bradley–Terry model (KRACH), a model with a generalized logistic prior with $\eta = 1$ and a naïve model where the probability for a team to win a game is proportional to the square root of its win ratio (wins divided by losses) without regard to strength of schedule.

We can see that 17 tournaments of 15 games each are enough to show that the Bradley–Terry model is clearly preferred to the model using win ratios without including strength of schedule, which is in turn better than declaring each game a tossup. It is not enough, however, to establish a preference between the Haldane and generalized logistic priors, although their predictions have not always been identical.

[2]Note that Figure 5 contains the results of four different Monte Carlo simulations (each with 20,000 draws for each season) plotted on top of one another, to illustrate that the integrals of the Gaussian-approximated posterior have been estimated accurately. If a similar exercise is performed with Gaussian importance sampling, the four simulations give vastly different posterior predictive probabilities, indicating the algorithm is not stable enough to estimate the small probability associated with one particular sequence of results.

3.2. The Pairwise Probability Matrix

The Pairwise Probability Matrix [18] is a tool to predict the probability that each team will make the NCAA tournament. It typically runs with a few weeks remaining before the end of the conference tournaments and the selection of the tournament field. In its current configuration (2018-2019 season), it takes a set of Bradley–Terry log-strengths $\{\lambda_i\}$ and estimates the probability $P(O|\{\lambda_i\})$ for an outcome O (typically a team being selected for the NCAA tournament) as follows:

(1) A set of $N = 20{,}000$ Monte Carlo trials are run. In each trial:
   (a) The remaining games of the season are simulated; in each game, a winner is randomly chosen according to the probability predicted by the Bradley–Terry model. For instance, if team i plays team j, the probability that i will win is modelled as $\theta_{ij} = \text{logistic}(\lambda_i - \lambda_j)$, and team i is assigned as the winner if a Uniform$(0, 1)$ random draw is less than θ_{ij}.
   (b) The games to be played are not pre-determined, but may depend on the results of other games earlier in the simulation (e.g., the loser of a game may be eliminated from a conference tournament).
   (c) When all the games have been simulated, teams are evaluated according to the NCAA selection criteria, including an ordering based on pairwise comparisons, and automatic qualification for the winners of conference tournaments.

(2) The probability of an outcome O is approximated as the fraction of Monte Carlo simulations in which it occurs.

At present, the ratings used are the maximum likelihood estimates $\{\widehat{\lambda}_i\}$, expressed as KRACH ratings $\{100\, e^{\widehat{\lambda}_i}\}$, so that the probability of a future outcome given past game results D is approximated as in Section 2.1:

$$P(O|D, I) \approx P(O|\{\widehat{\lambda}_i\}) \approx \frac{1}{N} \sum_{s=1}^{N} I^{(s)}(O),$$

where $I^{(s)}(O) = 1$ if O occurs in Monte Carlo trial s and 0 if not. An excerpt of a typical display, taken before the final day of games of the 2018–2019 season, is shown in Figure 6.

3.2.1. Shortcomings of the MLE probabilities. As demonstrated in Section 3.1, the KRACH/MLE Bradley–Terry model produces reasonably accurate predictions when applied late in the college hockey season, based on the use of the model to assign probabilities to NCAA tournament outcomes. However, it can lead to some potentially inaccurate probabilities, in particular in underestimating the probabilities of unlikely events or sequences of events. As an illustration, we consider the situation on 2018 March 9, when Cornell and Quinnipiac began a best-of-three playoff series. Their respective KRACH ratings were 415.3 and 93.30, so the estimated probability of Cornell winning the game was 81.7%. However, there was still some uncertainty in the difference of their Bradley–Terry log-strengths, as illustrated in Figure 7. Applying the Gaussian approximation of Section 2.2,

Pairwise Probability Matrix

These are the results of **20,000** **Monte Carlo simulations** of the remaining games prior to Selection Day. The winner of each game in the simulation was determined randomly, weighted by **KRACH**. When that simulation was completed -- playing out the six conference tournaments -- a **Pairwise** was calculated based upon those results.

The numbers in the chart represent the percentage of times (among the total simulations run) each team placed in that spot in the Pairwise. Note that just placing in the top 16 does not indicate the team made it, due to automatic bids (AQ).

Please see below the chart for more information.

Last updated: March 23, 2:58 am

RK	TEAM	In	AL	AQ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
1	St. Cloud State	100%	32%	68%	100%																	
2	Minnesota State	100%	34%	66%		45%	21%	34%														
3	Massachusetts	100%	100%			23%	56%	21%														
4	Minnesota-Duluth	100%	68%	32%		32%	23%	45%														
5	Quinnipiac	100%	100%						13%	50%	38%											
6	Northeastern	100%	26%	74%					36%	38%		4%	13%	9%								
7	Clarkson	100%	49%	51%					51%			4%	21%	24%								
8	Ohio State	100%	100%							6%	30%	44%	20%									
9	Denver	100%	100%							7%	27%	25%	12%	15%	12%	1%						
10	Cornell	100%	51%	49%							5%	24%	20%		34%	18%						
11	Arizona State	100%	100%										13%	51%	36%							
12	Harvard	100%	100%													5%	61%	30%	4%			
13	Bowling Green	82%	48%	34%											19%	15%			66%			
14	Penn State	49%		49%												32%	17%			51%		
15	Notre Dame	51%		51%												29%	20%	2%		49%		
16	Providence	91%	91%														2%	68%	30%			
17	Union																				53%	47
18	Western Michigan																				47%	53
19	Minnesota																					
20	North Dakota																					

Figure 6. An excerpt of the Pairwise Probability Matrix display, displaying estimated probabilities entering the final day of games (2019 March 23) before NCAA tournament selection. Excerpt from `https://www.collegehockeynews.com/ratings/probabilityMatrix.php`.

we get posterior predictive probability for Cornell to defeat Quinnipiac of

$$\int_0^1 \theta_{\mathrm{CrQn}}\, f(\theta_{\mathrm{CrQn}}|D,I)\, d\theta_{\mathrm{CrQn}} \approx 80.0\%.$$

An additional effect of the uncertainty is that the game results are not independent. If Cornell lost one game with Quinnipiac, it would mean the difference of their Bradley–Terry log-strengths was more likely to be below $\widehat{\lambda}_{\mathrm{Cr}} - \widehat{\lambda}_{\mathrm{Qn}}$ than above it, and the adjusted posterior probability they would lose another game would be higher.[3] This is reflected in Figure 8, which shows that while the probabilities from the KRACH ratings would give Cornell a 91.1% chance to win two out of three games with Quinnipiac, the actual posterior predictive probability (using the Gaussian approximation) is

$$\int_0^1 \left[\theta_{\mathrm{CrQn}}^2 + 2(1-\theta_{\mathrm{CrQn}})\theta_{\mathrm{CrQn}}^2\right] f(\theta_{\mathrm{CrQn}}|D,I)\, d\theta_{\mathrm{CrQn}} \approx 88.2\%.$$

[3]In practice, one calculates the probability for the whole sequence of results, but it can be conceptually understood according to the Bayesian updating of posteriors, where $P(O_1,O_2|D,I) = P(O_2|O_1,D,I)P(O_1|D,I)$.

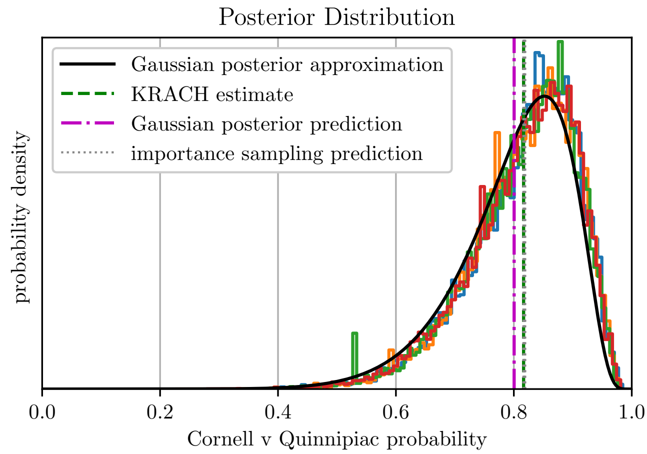

Figure 7. Marginal posterior distribution on θ_{CrQn}, the probability for Cornell to defeat Quinnipiac using the Haldane prior. Note that the maximum likelihood estimate $\widehat{\theta}_{\mathrm{CrQn}}$ associated with the KRACH ratings is not the maximum of the marginal posterior, because of the transformation of the probability density function. The theoretical curve uses the Gaussian approximation, and the histograms are four replications of the marginal posterior estimated using importance sampling with the Gaussian approximation and 20,000 samples each, as described in Section 3.2.3.

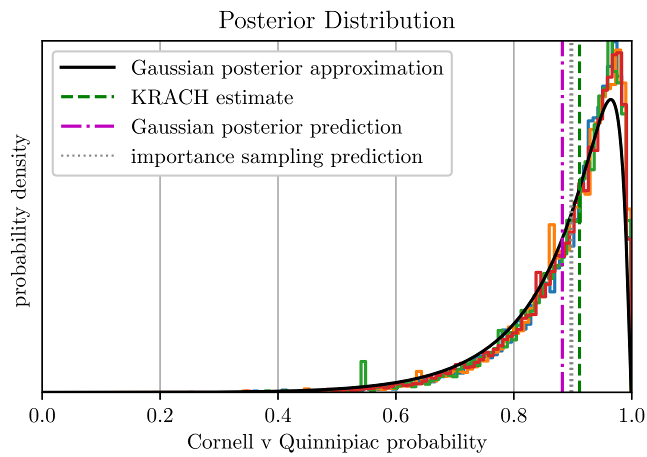

Figure 8. Marginal posterior distribution on the probability $\theta_{\mathrm{CrQn}}^2 + 2(1-\theta_{\mathrm{CrQn}})\theta_{\mathrm{CrQn}}^2$ for Cornell to defeat Quinnipiac in a best-of-three series, using the Haldane prior. The theoretical curve uses the Gaussian approximation, and the histograms are four replications of the marginal posterior estimated using importance sampling with the Gaussian approximation and 20,000 samples each, as described in Section 3.2.3.

3.2.2. Proposed modification to the Pairwise Probability Matrix. Recalculating the KRACH ratings to account for each simulated game result would address the correlations between game results, but not the uncertainties arising from the asymmetries of the marginal distributions for probabilities like θ_{CrQn}. It would also be rather computationally intensive. Instead, we propose to modify the Monte Carlo algorithm to improve the estimate of probabilities using the Gaussian approximation of Section 2.2 or the importance sampling method of Section 2.3. The modified Monte Carlo workflow would be

(1) The multivariate Gaussian approximation is constructed to the posterior distribution from the Bradley–Terry log-strengths using the Haldane prior with the constraint $\sum_{i=1}^{t}\lambda_i = 0$; the peak is at the maximum-likelihood point $\{\widehat{\lambda}_i\}$ and the variance-covariance matrix is the pseudo-inverse $\{\Sigma_{ij}\}$ of the Hessian matrix $H_{ij} = -n_{ij}\widehat{\theta}_{ij}\widehat{\theta}_{ji} + \delta_{ij}\sum_{k=1}^{t}\widehat{\theta}_{ik}\widehat{\theta}_{ki}$.

(2) A set of $N = 20{,}000$ Monte Carlo trials are run. In each trial:

   (a) A random draw $\{\lambda_i^{(s)}|i = 1,\dots,t\}$ is made from the multivariate normal $N_t(\{\widetilde{\lambda}_i\},\{\Sigma_{ij}\})$.

   (b) If importance sampling is to be used, the ratio $w_s \propto \frac{f(\{\lambda_i^{(s)}\}|D,I)}{g(\{\lambda_i^{(s)}\}|D,I)}$ of the exact posterior to the sampling distribution, at the point $\{\lambda_i^{(s)}\}$, is recorded.

   (c) The games are simulated using a win probability matrix

$$\theta_{ij}^{(s)} = \operatorname{logistic}(\lambda_i^{(s)} - \lambda_j^{(s)})$$

   as in the current algorithm.

   (d) The sequence of games and NCAA selection criteria are created from the series of game results as they are now.

(3) If importance sampling is not used, the probability of an outcome O is approximated as the fraction of Monte Carlo simulations in which it occurs

$$P(O|D,I) \approx \frac{1}{N}\sum_{s=1}^{N} I^{(s)}(O).$$

If importance sampling is used, the outcomes are weighted by the ratio w_s, normalized such that $\sum_{s=1}^{N} w_s = 1$:

$$P(O|D,I) \approx \frac{\sum_{s=1}^{N} w_s\, I^{(s)}(O)}{\sum_{s=1}^{N} w_s} = \sum_{s=1}^{N} w_s\, I^{(s)}(O).$$

3.2.3. Demonstration of modifications. The modifications proposed in Section 3.2.2 have not yet been integrated into the generation of the Pairwise Probability Matrix. However, we can demonstrate their impact by recomputing the probabilities shown in Section 3.2.1. Using the game results of the 2018–2019 season prior to 2018 March 9, we construct the Gaussian approximation $g(\{\lambda_i\}|D,I)$ to the posterior $f(\{\lambda_i\}|D,I)$. We then draw $N = 20{,}000$ samples $\{\lambda_i^{(s)}\}$ from this distribution, and calculate the weights $w_s \propto \frac{f(\{\lambda_i^{(s)}\}|D,I)}{g(\{\lambda_i^{(s)}\}|D,I)}$. For each sample, we have a probability $\theta_{\text{CrQn}}^{(s)} = \operatorname{logistic}(\lambda_{\text{Cr}}^{(s)} - \lambda_{\text{Wn}}^{(s)})$ that Cornell will defeat Quinnipiac in a game, and a probability

$$\pi_{\text{CrQn}}^{(s)} = (\theta_{\text{CrQn}}^{(s)})^2 + 2(1-\theta_{\text{CrQn}}^{(s)})(\theta_{\text{CrQn}}^{(s)})^2$$

Table 1. Results of simulations using the Gaussian approximation with and without importance sampling to estimate the probability, expressed as a percentage, of Cornell winning a game, or a best-of-three series, with Quinnipiac, as of 2018 March 9.

Game (KRACH probability = 81.7%)

	integration	MC integration	MC simulation
Gaussian approx	80.0	80.0, 80.1, 80.0, 79.9	80.2, 80.0, 79.7, 79.8
Importance sampling		81.8, 81.7, 81.4, 81.5	81.8, 81.6, 81.8, 81.9

Series (KRACH probability = 91.1%)

	integration	MC integration	MC simulation
Gaussian approx	88.2	88.2, 88.3, 88.1, 88.0	88.3, 88.5, 87.7, 87.9
Importance sampling		89.8, 89.8, 89.5, 89.6	89.8, 89.9, 89.6, 89.7

that Cornell will win a three-game series.

We simulate a subset of the Pairwise Probability Matrix Monte Carlo as follows. For each sample s we make three draws from a Bernoulli distribution with probability $\theta^{(s)}_{\mathrm{CrQn}}$. If the first draw, $W^{(s)}_{\mathrm{CrQn}}$, is one, we assign that sample as a win for Cornell; if it is zero, we assign that as a loss for Cornell. If two or more of the three draws for a sample are one, we set $S^{(s)}_{\mathrm{CrQn}} = 1$ (series win for Cornell); otherwise we set $S^{(s)}_{\mathrm{CrQn}} = 0$ (series loss for Cornell).

To estimate the probability of Cornell winning a single game against Quinnipiac, assuming the Gaussian approximation (which was analytically computed in Section 3.2.1 to be 80.0% by numerical integration of the marginal Gaussian posterior on $\lambda_{\mathrm{Cr}} - \lambda_{\mathrm{Qn}}$), we can perform two different calculations: A Monte Carlo average $\frac{1}{N}\sum_{s=1}^{N} \theta^{(s)}_{\mathrm{CrQn}}$ of the single-game probability, or the fraction $\frac{1}{N}\sum_{s=1}^{N} W^{(s)}_{\mathrm{CrQn}}$ of the of simulations in which Cornell wins the first game. The latter is the analogue of what would be computed in the Pairwise Probability Matrix. To adjust the single-game computation using importance sampling, we again have two options: A weighted Monte Carlo average $\sum_{s=1}^{N} w_s\, \theta^{(s)}_{\mathrm{CrQn}}$ of the single-game probability, or the sum of the weights $\sum_{s=1}^{N} w_s\, W^{(s)}_{\mathrm{CrQn}}$ of the of simulations in which Cornell wins the first game. The latter is the analogue of what would be computed in the Pairwise Probability Matrix. To test these, and estimate Monte Carlo errors, we performed four replications of the whole process (with 20,000 Monte Carlo samples each). The histograms of the simulated probabilities $\theta^{(s)}_{\mathrm{CrQn}}$, weighted by w_s, are plotted in Figure 7. We see that there are some heavily-weighted outliers (the largest weights in the four replications are 0.00399, 0.00686, 0.00488, and 0.00509, compared to an average weight of 0.00005). However, when we estimate the single-game probabilities using importance sampling, summarized in Table 1, they all come out consistently, slightly below 82%, and we can distinguish a small difference between the probability predicted with and without importance sampling. This is also reflected in the histograms in Figure 7, where we can see that, despite the outliers, the overall shape of the estimated posterior appears to be skewed a bit further right than the one derived from the Gaussian approximation on $\{\lambda_i\}$. Note that the effects of including the posterior uncertainty, and going from the Gaussian approximation to the posterior estimated by importance sampling, cancel out, and the estimated probability is quite close to that calculated from the maximum-likelihood/KRACH approximation. We will see when we consider three-game series that this cancellation is a coincidence.

To estimate the probability of Cornell winning a best-of-three series against Quinnipiac, assuming the Gaussian approximation (which was computed in Section 3.2.1 using numerical integration as 88.2%), we can again perform two calculations: A Monte Carlo average $\frac{1}{N}\sum_{s=1}^{N}\pi_{\mathrm{CrQn}}^{(s)}$ of the single-game probability, or the fraction $\frac{1}{N}\sum_{s=1}^{N}S_{\mathrm{CrQn}}^{(s)}$ of the of simulations in which Cornell wins the first game. The latter is the analogue of what would be computed in the Pairwise Probability Matrix. Similarly, when we use importance sampling, we can compute either a weighted Monte Carlo average $\sum_{s=1}^{N} w_s\,\pi_{\mathrm{CrQn}}^{(s)}$ of the best-of-three probability, or the sum of the weights $\sum_{s=1}^{N} w_s\,S_{\mathrm{CrQn}}^{(s)}$ of the of simulations in which Cornell wins the three-game series. The latter is the analogue of what would be computed in the Pairwise Probability Matrix. The histograms of the simulated probabilities $\pi_{\mathrm{CrQn}}^{(s)}$, weighted by w_s, are plotted in Figure 8, and the probabilities are summarized in Table 1. We see that the Monte Carlo simulations of the games are somewhat more robust than the Monte Carlo averages, but the results are consistent, just below 90%, and noticeably different from both the Gaussian approximation (about 88%) and the maximum likelihood/KRACH probability of 91.1%.

4. Discussion

We have illustrated some applications of the Bradley–Terry model to college hockey. The model, in its maximum-likelihood form, is already used to rank teams as the basis of the KRACH ratings. Because the Bradley–Terry strength parameters naturally produce probabilities of game outcomes, the model can also be used for the prediction of future outcomes based on past results. In Section 2 we showed how to go beyond the maximum-likelihood values of these probabilities to account for posterior uncertainties in the parameters and estimate posterior predictive probabilities. One can avoid the use of full Markov Chain Monte Carlo methods by approximating the relevant marginalization integrals using a multivariate Gaussian approximation to the posterior and/or importance sampling.

In Section 3 we exhibited two applications of these posterior predictive probabilities. The first used these probabilities to evaluate the models (Bradley–Terry or otherwise) generating them. Constructing a Bayes factor for the NCAA tournament results using the probabilities predicted using the pre-tournament results shows, over time, the superiority of Bradley–Terry models to a naïve model based only on each team's win/loss ratio. It would be illuminating to compare the Bradley–Terry model to more sophisticated alternatives such as the Ratings Percentage Index (RPI), but that is non-trivial because the RPI doesn't naturally produce predictive probabilities.

Finally, the Pairwise Probability Matrix is a natural application of the Bradley–Terry model to assign probabilities to the outcome of the last few weeks of a college hockey season, in terms of which teams qualify for the NCAA tournament. The current application uses Monte Carlo simulation to estimate these probabilities from the maximum-likelihood Bradley–Terry parameters (KRACH ratings). We have proposed a modification to this program where, at each Monte Carlo iteration the ratings are also randomly drawn from an approximation to their posterior distribution. This should more accurately account for posterior uncertainties in the parameters and induced correlations of future game outcomes.

Acknowledgment. J. T. Whelan wishes to thank 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. Parts of Section 3.2.1 were inspired by a discussion on the eLynah forum. Game results for the computations in this paper were collected from `https://www.collegehockeynews.com/schedules/composite.php`

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

References

[1] A. Albert and J. A. Anderson, *On the existence of maximum likelihood estimates in logistic regression models*, Biometrika **71** (1984), 1–10.

[2] R. A. Bradley and M. E. Terry, *Rank analysis of incomplete block designs: The method of paired comparisons*, Biometrika **39** (1952), 324–345.

[3] K. Butler, *Ken's Ratings for American College Hockey*, HOCKEY-L mailing list post, 1993.

[4] K. Butler and J. Whelan, *The existence of maximum-likelihood estimates in the Bradley–Terry model and its extensions*, arXiv:math/0412232.

[5] R. R. Davidson, *On extending the Bradley–Terry model to accommodate ties in paired comparison experiments*, Journal of the American Statistical Association **65** (1970), 317–328.

[6] P. Ding, *On the conditional distribution of the multivariate t distribution*, The American Statistician **70** (2016), 293–295.

[7] L. R. Ford, Jr., *Solution of a ranking problem from binary comparisons*, The American Mathematical Monthly **64** (1957), 28–33.

[8] J. B. S. Haldane, *A note on inverse probability*, Mathematical Proceedings of the Cambridge Philosophical Society **28** (1932), 55–61.

[9] H. Jeffreys, *Theory of Probability*, 3rd ed., Clarendon Press, Oxford, 1961.

[10] H. Joe, *Extended use of paired comparison models, with application to chess rankings*, Journal of the Royal Statistical Society, Series C (Applied Statistics) **39** (1990), 85–93.

[11] R. Penrose, *A generalized inverse for matrices*, Mathematical Proceedings of the Cambridge Philosophical Society **51** (1955), 406–413.

[12] G. C. Phelan and J. T. Whelan, *Hierarchical Bayesian Bradley-Terry for applications in Major League Baseball*, Mathematics for Applications **7** (2017), 71–84.

[13] P. V. Rao and L. L. Kupper, *Ties in paired-comparison experiments: A generalization of the Bradley–Terry model*, Journal of the American Statistical Association **62** (1967), 194–204.

[14] T. J. Santner D. E. and 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.

[15] T. B. Swartz, *Importance sampling with skew-normal distributions*, in: Proceedings of the 37th Winter Simulation Conference, Orlando, FL, USA, December 4–7, 2005, IEEE Computer Society, p. 32.

[16] S. Wang and T. B. Swartz, *Adaptive importance sampling from restricted skew-Student distributions*, unpublished manuscript.

[17] J. T. Whelan, *Prior distributions for the Bradley–Terry model of paired comparisons*, arXiv:1712.05311.

[18] A. Wodon, *Re-introducing the Pairwise Probability Matrix*, 2017 https://www.collegehockeynews.com/news/2017/02/21_re-introducing_the_pairwise.php.
[19] E. Zermelo, *Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung*, Mathematische Zeitschrift **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: jtwsma@rit.edu

Adam Wodon, College Hockey News
e-mail: adamw@collegehockeynews.com, https://www.collegehockeynews.com/