

Bradley–Terry model

Article

Talk



The **Bradley–Terry model** is a [probability model](#) for the outcome of pairwise comparisons between items, teams, or objects. Given a pair of items *i* and *j* drawn from some [population](#), it estimates the probability that the [pairwise comparison](#) *i* > *j* turns out true, as

$$\Pr(i > j) = \frac{p_i}{p_i + p_j} \tag{1}$$

where *p_i* is a positive [real-valued](#) score assigned to individual *i*. The comparison *i* > *j* can be read as "*i* is preferred to *j*", "*i* ranks higher than *j*", or "*i* beats *j*", depending on the application.

For example, *p_i* might represent the skill of a team in a sports tournament and **Pr(*i* > *j*)** the probability that *i* wins a game against *j*.^{[1][2]} Or *p_i* might represent the quality or desirability of a commercial product and **Pr(*i* > *j*)** the probability that a consumer will prefer product *i* over product *j*.

The Bradley–Terry model can be used in the forward direction to predict outcomes, as described, but is more commonly used in reverse to infer the scores *p_i* given an observed set of outcomes.^[2] In this type of application *p_i* represents some measure of the strength or quality of *i* and the model lets us estimate the strengths from a series of pairwise comparisons. In a survey of wine preferences, for instance, it might be difficult for respondents to give a complete ranking of a large set of wines, but relatively easy for them to compare sample pairs of wines and say which they feel is better. Based on a set of such pairwise comparisons, the Bradley–Terry model can then be used to derive a full ranking of the wines.

Once the values of the scores *p_i* have been calculated, the model can then also be used in the forward direction, for instance to predict the likely outcome of comparisons that have not yet actually occurred. In the wine survey example, for instance, one could calculate the probability that someone will prefer wine *i* over wine *j*, even if no one in the survey directly compared that particular pair.

^ History and applications



The model is named after [Ralph A. Bradley](#) and Milton E. Terry,^[3] who presented it in 1952,^[4] although it had already been studied by [Ernst Zermelo](#) in the 1920s.^{[1][5][6]} Applications of the model include the ranking of competitors in sports, [chess](#), and other competitions,^[7] the ranking of products in paired comparison surveys of [consumer choice](#), analysis of [dominance hierarchies](#) within animal and human communities,^[8] ranking of [journals](#), ranking of AI models,^[9] and estimation of the relevance of documents in [machine-learned search engines](#).^[10]

Definition

The Bradley–Terry model can be parametrized in various ways. Equation (1) is perhaps the most common, but there are a number of others. Bradley and Terry themselves defined exponential score functions $p_i = e^{\beta_i}$, so that^[2]

$$\Pr(i > j) = \frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}}.$$

Alternatively, one can use a [logit](#), such that^[1]

$$\text{logit } \Pr(i > j) = \log \frac{\Pr(i > j)}{1 - \Pr(i > j)} = \log \frac{\Pr(i > j)}{\Pr(j > i)} = \beta_i - \beta_j,$$

i.e. $\text{logit } p = \log \frac{p}{1-p}$ for $0 < p < 1$.

This formulation highlights the similarity between the Bradley–Terry model and [logistic regression](#). Both employ essentially the same model but in different ways. In logistic regression one typically knows the parameters β_i and attempts to infer the functional form of $\Pr(i > j)$; in ranking under the Bradley–Terry model one knows the functional form and attempts to infer the parameters.

With a scale factor of 400, this is equivalent to the [Elo rating system](#) for players with Elo ratings R_i and R_j .

$$\Pr(i > j) = \frac{e^{R_i/400}}{e^{R_i/400} + e^{R_j/400}} = \frac{1}{1 + e^{(R_j - R_i)/400}}.$$

Estimating the parameters

The most common application of the Bradley–Terry model is to infer the values of the parameters p_i given an observed set of outcomes $i > j$, such as wins and losses in a competition. The simplest way to estimate the parameters is by [maximum likelihood estimation](#), i.e., by maximizing the [likelihood](#) of the observed outcomes given the model and parameter values.

Suppose we know the outcomes of a set of pairwise competitions between a certain group of individuals, and let w_{ij} be the number of times individual i beats individual j . Then the likelihood of this set of outcomes within the Bradley–Terry model is $\prod_{ij} [\Pr(i > j)]^{w_{ij}}$ and the [log-likelihood](#) of the parameter

vector $\mathbf{p} = [p_1, \dots, p_n]$ is^[1]

$$\begin{aligned} l(\mathbf{p}) &= \ln \prod_{ij} [\Pr(i > j)]^{w_{ij}} = \sum_{i=1}^n \sum_{j=1}^n \ln \left[\left(\frac{p_i}{p_i + p_j} \right)^{w_{ij}} \right] \\ &= \sum_{ij} w_{ij} \ln \left(\frac{p_i}{p_i + p_j} \right) = \sum_{ij} [w_{ij} \ln(p_i) - w_{ij} \ln(p_i + p_j)]. \end{aligned}$$

Zermelo^[5] showed that this expression has only a single maximum, which can be found by differentiating with respect to p_i and setting the result to zero, which leads to

$$p_i = \frac{\sum_j w_{ij}}{\sum_j (w_{ij} + w_{ji}) / (p_i + p_j)}.$$

(2)

This equation has no known closed-form solution, but Zermelo suggested solving it by simple iteration. Starting from any convenient set of (positive) initial values for the p_i , one iteratively performs the update

$$p'_i = \frac{\sum_j w_{ij}}{\sum_j (w_{ij} + w_{ji}) / (p_i + p_j)}$$

(3)

for all i in turn. The resulting parameters are arbitrary up to an overall multiplicative constant, so after computing all of the new values they should be normalized by dividing by their geometric mean thus:

$$p_i \leftarrow \frac{p'_i}{\left(\prod_{j=1}^n p'_j\right)^{1/n}}.$$

(4)

This estimation procedure improves the log-likelihood on every iteration, and is guaranteed to eventually reach the unique maximum.^{[5][11]} It is, however, slow to converge.^{[1][12]} More recently it has been pointed out^[13] that equation (2) can also be rearranged as

$$p_i = \frac{\sum_j w_{ij} p_j / (p_i + p_j)}{\sum_j w_{ji} / (p_i + p_j)},$$

which can be solved by iterating

$$p'_i = \frac{\sum_j w_{ij} p_j / (p_i + p_j)}{\sum_j w_{ji} / (p_i + p_j)},$$

(5)

again normalizing after every round of updates using equation (4). This iteration gives identical results to the one in (3) but converges much faster and hence is normally preferred over (3).^[13]

Worked example of solution procedure



Consider a sporting competition between four teams, who play a total of 22 games among themselves. Each team's wins are given in the rows of the table below and the opponents are given as the columns:

Results

	A	B	C	D
A	–	2	0	1
B	3	–	5	0
C	0	3	–	1
D	4	0	3	–

For example, Team A has beat Team B twice and lost to team B three times; not played team C at all; won once and lost four times against team D.

We would like to estimate the relative strengths of the teams, which we do by calculating the parameters \mathbf{p}_i , with higher parameters indicating greater prowess. To do this, we initialize the four entries in the parameter vector \mathbf{p} arbitrarily, for example assigning the value 1 to each team: $[1, 1, 1, 1]$. Then we apply equation (5) to update \mathbf{p}_1 , which gives

$$p_1 = \frac{\sum_{j(\neq 1)} w_{1j} p_j / (p_1 + p_j)}{\sum_{j(\neq 1)} w_{j1} / (p_1 + p_j)} = \frac{2 \frac{1}{1+1} + 0 \frac{1}{1+1} + 1 \frac{1}{1+1}}{3 \frac{1}{1+1} + 0 \frac{1}{1+1} + 4 \frac{1}{1+1}} = 0.429.$$

Now, we apply (5) again to update \mathbf{p}_2 , making sure to use the new value of \mathbf{p}_1 that we just calculated:

$$p_2 = \frac{\sum_{j(\neq 2)} w_{2j} p_j / (p_2 + p_j)}{\sum_{j(\neq 2)} w_{j2} / (p_2 + p_j)} = \frac{3 \frac{0.429}{1+0.429} + 5 \frac{1}{1+1} + 0 \frac{1}{1+1}}{2 \frac{1}{1+0.429} + 3 \frac{1}{1+1} + 0 \frac{1}{1+1}} = 1.172$$

Similarly for \mathbf{p}_3 and \mathbf{p}_4 we get

$$p_3 = \frac{\sum_{j(\neq 3)} w_{3j} p_j / (p_3 + p_j)}{\sum_{j(\neq 3)} w_{j3} / (p_3 + p_j)} = \frac{0 \frac{0.429}{1+0.429} + 3 \frac{1.172}{1+1.172} + 1 \frac{1}{1+1}}{0 \frac{1}{1+0.429} + 5 \frac{1}{1+1.172} + 3 \frac{1}{1+1}} = 0.557$$

$$p_4 = \frac{\sum_{j(\neq 4)} w_{4j} p_j / (p_4 + p_j)}{\sum_{j(\neq 4)} w_{j4} / (p_4 + p_j)} = \frac{4 \frac{0.429}{1+0.429} + 0 \frac{1.172}{1+1.172} + 3 \frac{0.557}{1+0.557}}{1 \frac{1}{1+0.429} + 0 \frac{1}{1+1.172} + 1 \frac{1}{1+0.557}} = 1.694$$

Then we normalize all the parameters by dividing by their geometric mean

to get the estimated parameters

$\mathbf{p} = [0.516, 1.413, 0.672, 2.041]$.

To improve the estimates further, we repeat the process, using the new \mathbf{p} values. For example,

Repeating this process for the remaining parameters and normalizing, we get

$\mathbf{p} = [0.677, 1.034, 0.624, 2.287]$. Repeating a further 10 times gives rapid convergence toward a final solution of $\mathbf{p} = [0.640, 1.043, 0.660, 2.270]$. This indicates that Team D is the strongest and Team B the second strongest, while Teams A and C are nearly equal in strength but below Teams B and D. In this way the Bradley–Terry model lets us infer the relationship between all four teams, even though not all teams have played each other.

^ See also



- [Ordinal regression](#)
- [Rasch model](#)
- [Scale \(social sciences\)](#)
- [Elo rating system](#)
- [Thurstonian model](#)

^ References



1. ^ [a](#) [b](#) [c](#) [d](#) [e](#) Hunter, David R. (2004). "MM algorithms for generalized Bradley–Terry models" [↗](#). *The Annals of Statistics*. **32** (1): 384–406. [CiteSeerX 10.1.1.110.7878](#) [↗](#). doi:10.1214/aos/1079120141 [↗](#). JSTOR 3448514 [↗](#).
2. ^ [a](#) [b](#) [c](#) Agresti, Alan (2014). *Categorical Data Analysis*. John Wiley & Sons. pp. 436–439.
3. ^ E.E.M. van Berkum. "Bradley-Terry model" [↗](#). *Encyclopedia of Mathematics*. Retrieved 18 November 2014.
4. ^ Bradley, Ralph Allan; Terry, Milton E. (1952). "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons". *Biometrika*. **39** (3/4): 324–345. doi:10.2307/2334029 [↗](#). JSTOR 2334029 [↗](#).
5. ^ [a](#) [b](#) [c](#) Zermelo, Ernst (1929). "Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung". *Mathematische Zeitschrift*. **29** (1): 436–460. doi:10.1007/BF01180541 [↗](#). S2CID 122877703 [↗](#).
6. ^ Heinz-Dieter Ebbinghaus (2007), *Ernst Zermelo: An Approach to His Life and Work*, Springer, pp. 268–269, ISBN 9783540495536
7. ^ Shev, A.; Fujii, K.; Hsieh, F.; McCowan, B. (2014). "Systemic testing on Bradley-Terry model against nonlinear ranking hierarchy" [↗](#). *PLOS One*. **9** (12): e115367. doi:10.1371/journal.pone.0115367 [↗](#). PMC 4274013 [↗](#). PMID 25531899 [↗](#).
8. ^ Boyd, Robert; Silk, Joan B. (1983). "A method for assigning cardinal dominance ranks". *Animal Behaviour*. **31** (1): 45–58. doi:10.1016/S0003-3472(83)80172-9 [↗](#). S2CID 53178779 [↗](#).
9. ^ "Chatbot Arena: New models & Elo system update | LMSYS Org" [↗](#). *lmsys.org*. Retrieved 2024-01-30.
10. ^ Szummer, Martin; Yilmaz, Emine (2011). *Semi-supervised learning to rank with preference regularization* [↗](#) (PDF). CIKM.
11. ^ Ford, Jr., L. R. (1957). "Solution of a ranking problem from binary comparisons". *American Mathematical Monthly*. **64** (8): 28–33. doi:10.1080/00029890.1957.11989117 [↗](#).
12. ^ Dykstra, Jr., Otto (1956). "A note on the rank analysis of incomplete block designs". *Biometrics*. **12**: 301–306. doi:10.2307/2334029 [↗](#). JSTOR 2334029 [↗](#).
13. ^ [a](#) [b](#) Newman, M. E. J. (2023). "Efficient computation of rankings from pairwise comparisons" [↗](#). *Journal of Machine Learning Research*. **24** (238): 1–25.