Assessing Ranking Fidelity of Competition Structures via Weighted Mutual Information

Zach Culp, Josie Peterburs, Dr. Gregory Matthews, Dr. Ryan McShane

2025-10-09

Introduction

- There exists numerous competition formats across all professional sports
- Which competition structure is best?
- The “best” competition structure depends on how well it reflects the competitors’ true strengths
- We will propose a metric that quantifies how effectively different tournament structures convey information about competitors’ true strengths

Game Simulation

- Given the assigned "known strength", θ, we simulate a game using the Bradley-Terry Model

$$\Pr(i > j) = \frac{e^{\theta_i}}{e^{\theta_i} + e^{\theta_j}} = \frac{1}{1 + e^{\theta_i - \theta_j}}$$

Tournament Simulations

- Each competition structure is simulated given the number of teams, n, and θ
- Below is an example of the output from one simulation of a single elimination tournament structure

```
##   r  theta r_hat game_wins game_losses
## 1 1  1.221     6         0           1
## 2 2  0.765     1         3           0
## 3 3  0.431     4         1           1
## 4 4  0.140     2         2           1
## 5 5 -0.140     7         0           1
## 6 6 -0.431     5         0           1
## 7 7 -0.765     8         0           1
## 8 8 -1.221     3         1           1
```

Mutual Information

- If one views the true ranking of the teams in a tournament as a message to be sent to a receiver and the outcome of the tournament as a message that is received, we can measure the "goodness" of a tournament in terms of its ability to accurately transmit the true ranking.
- The notation for mutual information is:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

- In our case, it can be simplified to:

$$I(r,\hat{r}) = n! \sum_{\hat{r}} p(r,\hat{r}) \log \frac{p(r,\hat{r})}{(\frac{1}{n!})^2} = n! \log\left((n!)^2\right)$$

Weighted Mutual Information

- A limitation of mutual information is that it does not distinguish between the direction of information
- In other words, a ranking of $(1, 2, 3, 4)$ is equivalent to $(4, 3, 2, 1)$
- Because of this, we proposed a weighting function $w(r, \hat{r})$
- So, the new formula is:

$$wI_m(r, \hat{r}) = \sum_{r} \sum_{\hat{r}} w(r, \hat{r}) * p(r, \hat{r}) \log \frac{p(r, \hat{r})}{p(r)p(\hat{r})}$$

Weighting Function

- In our original paper submission, we used squared error loss as the weighting function:

$$w(r, \hat{r}) = \begin{cases} \frac{1}{\hat{r}' r}, & r \neq \hat{r} \\ 1, & r = \hat{r} \end{cases}$$

- However, after comments and reviews from people at CSAS and CMSAC, we have decided to use a weighted Kendall's Tau

Weighted Kendall

$$\tau_w = \frac{\sum_{i \neq j} W_{ij} \times \text{sgn}(i-j) \times \text{sgn}(R_i - R_j)}{\sum_{i,j} W_{ij} - \sum_i W_{ii}}$$

- W is a $n \times n$ lower-triangular matrix, where

$$W_{ij} = w_{\min(i,j)}, \text{ for } i,j = m, 2m, \dots, n \times m$$

- m: vector of length n representing the "salient weights", or the importance of correctly getting each ranking
- If $\tau_w = 1$, then the ranking is identical to the true ranking; if $\tau_w = 0$, then the order of teams is perfectly inversely proportional to the true rankings

Weighted Kendall Example

- If $\hat{r} = (2, 1, 3, 4)$ and m $= (3, 1, 1, 0)$, then:

$$W = \begin{pmatrix} 3 & 3 & 3 & 3 \\ 3 & 1 & 1 & 1 \\ 3 & 1 & 1 & 1 \\ 3 & 1 & 1 & 0 \end{pmatrix}$$

$$\sum_{i,j} W_{ij} = 29; \sum_{i} W_{ii} = 5; \sum_{i \neq j} W_{ij} \times \mathrm{sgn}(i - j) \times \mathrm{sgn}(R_i - R_j) = 18$$

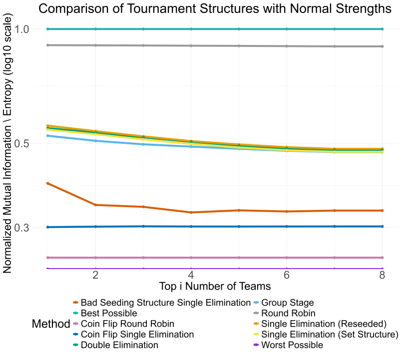
- So, τ_w for this example would be $\frac{18}{29-5} = 0.75$

Normalizing Mutual Information

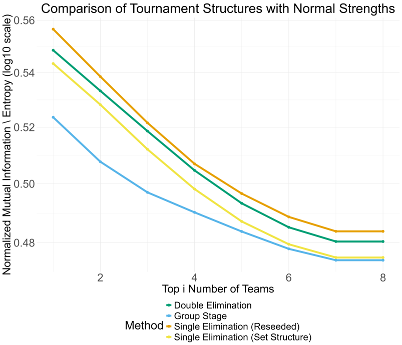
- Mutual Information can be normalized by dividing by the joint entropy
- Entropy: $-p(x) \log p(x)$
- Assuming r and $\hat{r}$ are independent, joint entropy can be written as $H(r, \hat{r}) = H(r) + H(\hat{r})$
- $H(r)$ and $H(\hat{r})$ are theoretically equivalent, leaving $H(r, \hat{r}) = -2 \sum_r p(r) \log(p(r)) = -2(n!) \times (1/n!) \times \log_2(1/n!) = \log_2((n!)^2)$
- So, the final formula for the metric is:

$$\frac{wI_m(r, \hat{r})}{H(r) + H(\hat{r})} = \frac{\sum_r \sum_{\hat{r}} w(r, \hat{r}) * p(r, \hat{r}) \log \frac{p(r, \hat{r})}{p(r)p(\hat{r})}}{\log_2((n!)^2)}$$

Results



Comparison of Tournament Structures with Normal Strengths
Normalized Mutual Information \ Entropy (log10 scale)
1.0
0.5
0.3
2
4
6
8
Top i Number of Teams
Method
Bad Seeding Structure Single Elimination
Best Possible
Coin Flip Round Robin
Coin Flip Single Elimination
Double Elimination
Group Stage
Round Robin
Single Elimination (Reseeded)
Single Elimination (Set Structure)
Worst Possible



Comparison of Tournament Structures with Normal Strengths
Normalized Mutual Information \ Entropy (log10 scale)
0.56
0.54
0.52
0.50
0.48
2
4
6
8
Top i Number of Teams
Method
Double Elimination
Group Stage
Single Elimination (Reseeded)
Single Elimination (Set Structure)

Limitations

- The "known" strengths are assumptions that can be estimated but never truly known
- Ties are not addressed in any of the games
- Ties in the $\hat{r}$ are randomly assigned

Future Work

- Evaluate structures using various distributions and "known" weights
- Estimate the θs in a real-world example and compare recent changes to the old format

References

- Appleton, David R. 1995. "May the Best Man Win?" Statistician 44 (4): 529.
- Csato, L. 2021. Tournament Design: How Operations Research Can Improve Sports Rules. 1st Ed. Cham, Switzerland: Palgrave Macmillan.
- Devriesere, Csató, K., and D. Goossens. 2025. "Tournament Design: A Review from an Operational Research Perspective." European Journal of Operational Research 324 (1): 1–21.
- Glenn, W A. 1960. "A Comparison of the Effectiveness of Tournaments." Biometrika 47 (3-4): 253–62.

References

- Guiasu, Silviu. 1977. Information Theory with Applications. New York, NY: McGraw-Hill.
- Johnson, Sidney, and Rodney Fort. 2022. "Match Outcome Uncertainty and Sports Fan Demand: An Agnostic Review and the Standard Economic Theory of Ports Leagues." Int. J. Emp. Econ. 01 (02).
- Lasek, Jan, and Marek Gagolewski. 2018. "The Efficacy of League Formats in Ranking Teams." Stat. Modelling 18 (5-6): 411–35.
- Sziklai, Balázs R, Péter Biró, and László Csató. 2022. "The Efficacy of Tournament Designs." Comput. Oper. Res. 144 (105821): 105821.