Official ASA 2024 ELO Documentation

Tabulation

The ASA rubric consists of **4 major sections**: Musical Composition, Vocal Execution, Visual Execution, and South Asian Representation. At each competition, judges will assign scores to teams in each category, which will subsequently be converted to a 100-point scale for convenience.

Normalization

Each team's raw score per judge *is then normalized by simply computing its Z-score*. To do this, the average score for each judge as well as the standard deviation, or spread from the average, is calculated. For each team's raw score per judge, subtract the average score and then divide by the standard deviation of the respective judge. Next, compute the median of the normalized scores for each team. These median values will be used as an additional unbiased data point to account for outliers in judge scoring and act as a synthetic judge. At the end of each competition, each team will now have a set of normalized scores for each judge + median of aforementioned scores at that competition.

ELO Ratings

Team ratings will be determined by a generalized approach to the ELO rating system. ELO is a renowned system that is a way to keep track of how good a player is at a game. Essentially, a team's ELO rating is represented by a number that either increases or decreases depending on how many other teams it outperforms. Teams with lower ELO ratings will win more points from higher-rated teams they beat and conversely, teams with high ELO ratings will win fewer points from lower-rated teams they beat. Traditionally, ELO is used widely for 1-on-1 matchups as in chess tournaments, for example, but we apply a generalized ELO system that works well for multi-team competitions.

The primary reason for using this system is that we want a framework that is not only mathematically sound but also accurate at gauging the skill levels of each team. Employing ELO allows us to standardize the rating calculations between competitions and enables us to provide the means in which teams can now much more easily visualize their growth through the circuit. This therefore allows them to better understand what they may need to do in each competition to succeed. Note that ELO will only be used for selection to A3 while scoring/ranking at individual bid competitions will remain the same.

Following sections document the necessary steps for generating ELO ratings and discuss the system's advantages.

Rating Initialization

At the beginning of the ASA competition circuit, every team begins with the same arbitrary rating of 1500, which is reflected by the variable R_i.

Expected Scores

Expected scores, E_A, are used to predict the outcome of a competition, or essentially how likely a team is to win. Standard ELO can only compare two players at a time, so let's instead think of a multi-team a cappella competition as a series of one-on-one matches. For example, if there are three teams, we have three separate matches (A vs. B, A vs. C, and B vs. C). In general, if there are N teams, there are $N(N-1)/2$ individual matches.

To figure this out, we'll use the regular ELO method to predict the scores for each pair of teams using a logistic function as depicted below. Then, we'll add up the expected scores for each team from all their matches to find their overall expected score. Finally, we'll adjust the scores so that they add up to 1 for all teams. This helps us see the scores as probabilities, like we do in the standard ELO system.

$$E_A = \frac{\sum_{1 \leq i \leq N, i \neq A} \frac{1}{1+10^{(R_i - R_A)/D}}}{N(N-1)/2}$$

Actual Scores

For each competition, the results of the matchups are observed. In standard ELO, the results follow a binary representation where one player is assigned a score of 1 if they win and 0 if they lose. For multi-team a cappella competitions, the results are encoded with a score in a similar manner. The scores themselves abide by certain conditions: 1) They must be monotonically decreasing, 2) Last place must have a score of 0, and 3) The scores must sum to 1 across all teams.

For a given judge at a competition, find their minimum normalized score. Subtract the judge's minimum score from each team's normalized score to get a delta score, depicting how much better each team performed vs. the lowest-scoring team. Next, calculate the sum of all teams' delta scores. Then, divide each delta score by the overall sum. The final resulting scores are the actual scores assigned to each team, retaining the weighting of how much higher each team scored vs. the lowest-scoring team. Repeat the above steps for each judge at a competition.

Rating Update

$$R'_A = R_A + K(N - 1)(S_A - E_A)$$

To get the new rating for a team, R'_A, simply add the current team's rating, R_A, with the product of a constant *K* factor and the difference in expected and actual scores. The *K* factor in the ELO rating system determines how much a team's rating changes after a competition. A higher *K* factor means bigger rating swings, making it useful for new or rapidly improving teams. Established teams might use a lower *K* factor, resulting in more gradual changes.

At a competition, the above calculations for new team ratings are computed for each team. For newer teams, or those who have been to either 0 or 1 competition, the *K* factor is set at an arbitrary value of 20. For teams that have already competed at two or more competitions, the *K* factor is set to 10. The varying *K* factor assignment is made to simulate convergence, or the rating system's way of making sure a team's rating changes become steadier and more reflective of their actual skill as they compete at more competitions.

Once all of the new ratings have been calculated, every team should expect to have a list of updated ratings whose size is equal to the total number of actual judges plus the one synthetic judge. Next, we compute a weighted sum where an arbitrary 10% weighting is placed on each of the actual judges (ie. judge weight) and remaining percentage associated with the synthetic judge (ie. synthetic weight). To produce the final rating update for a team, R'_A, the updated ratings associated with the actual judges are averaged out and multiplied by the judge weight and then summed up with the product of the updated rating associated with the synthetic judge and the synthetic weight. Repeat the above steps for each team at the competition.

Advantages

The use of the ELO rating system provides several key advantages. One of the foremost advantages is that the strength of schedule for each team is taken into account. Consecutively winning against lower-rated teams will not gain a team as many points as winning against many higher-rated teams will. Also, competing fewer times than other teams does not put a team at a major disadvantage and teams that do compete a lot aren't rewarded by just competing. Finally, teams that consistently improve over the season are rewarded.

Qualification

Once a competition has ended, a list of updated ratings for each competing team is produced and then sorted by rating. After all bid competitions have ended, all competitive teams and their final ratings are compiled together and ranked accordingly. The top *n* highest-rated teams are then selected as the teams that have successfully qualified for A3.

Rank	Team	Initial Rating	Final Rating
1	Team A	1500	1530
2	Team B	1500	1521
3	Team C	1500	1519
4	Team D	1500	1516
5	Team E	1500	1505
6	Team F	1500	1502
7	Team G	1500	1489