Dropping Lowest and Highest Score in Debate Adjudication Processes

Ann Author, Bee Author, and Cee Author*
Göttingen Institute for Contemporary Debate Research
(Dated: February 15, 2018)

Debating is a competetive sport where the winner is decided based upon points given from adjudicators. A common practice in sports with similar setup is to drop the highest and lowest scoring for each contestant. In this work, we take a statistical approach to analyze the effect of this measure on the example of the debating system *Offene Parlamentarische Debatte*.

I. INTRODUCTION

Debating is a highly competitive ... *Offene Parlamentarische Debatte* (OPD) adjudicators ..., (cite Regelwerk)

Similar types of competitive sports exist where the sport does not include any direct quantitive measure (like e.g. goals in football), but rather the performance in a sport is decided through adjudicators and is based on criteria, where a full objective measure is not possible. A common procedure in a number of these sports is that for each player or participant, both the highest and the lowest rating are dropped. Examples are gymnastics [1], ski jumping [2] and diving competitions [3].

In the preliminary rounds often a high difference in experience is present, which is accounted for by the chief adjudicators and chairs ... We therefore focus on the judging of finals (and potentially quarter and semi finals, if existant), where the so called *break* assures a certain standarization ... In addition, for simplicity, we solely focus on a single integer, the total speaker points, as well as the points of a team, that is the sum of three speaker points. We, however, believe the results will without loss of generality be usable for team points or single speaker categories.

We make the following assumptions:

1. Every adjudicator awards all points to the best of his / her knowledge and abilities. This means in no case does any adjudicator award points despite the knowledge that they do not match the speaker's or the team's performance.

2. No communication between adjudicators. If the results, and especially the adjudicators who have given the highest and lowest points, change during the adjudication discussion, this may have implications. For now, this study will focus on speakers whose result are not discussed in the panel.

3. The "true" strength of a speaker is defined by the average number points that a speaker would get from a high or infinite number of adjudicators with the experience to break into a final.

4. The points awarded by an adjudicator spread in a Gaussian fashion for a single speaker. This means, that if the same adjudicator would listen to the identical speech again (without remembering) he / she would probably not give identical points, but rather a number drawn from a probablility distribution spread around a mean that would be reached if the adjudicator was (conceptually) allowed to listen to the same speech an infinite number of times.

5. This aformentioned mean again is spreaded for different adjudicators. It is openly recognized that different adjudicators award systematically higher or lower points than others.

We assume the distribution of point 3. and 4. as Gaussian. This is justified through the Central Limit Theorem, that states that for a high number of parameters ..., the probability distribution always approaches a Gaussian. In OPD, five categories go into the total speaker points, while a large number of criteria, that is differently evaluated and weighted by each adjudicator, contributes to each category.

II. ANALYTICAL CONSIDERATIONS

A. Single Adjudicator

A single speaker has a true strength of μ_{true}. The adjudicator n has a point level of μ_n. It is drawn from a Gaussian distribution with mean μ_{true}, and standard deviation σ_s, i.e. the probability density function (PDF) is

$$p(\mu_n|\mu_{true}) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(-\frac{(\mu_n - \mu_{true})^2}{2\sigma_s^2}\right) \quad (1)$$

Equation **??** therefore gives the distribution that would be obtained for an infinite number of adjudicators is therefore centred around the true value. The speaker points awarded for the speaker by adjudicator n denoted by x_n is again drawn from a Gaussian distribution with mean μ_n and σ_n:

$$p(x_n|\mu_n) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(x_n - \mu_n)^2}{2\sigma_n^2}\right) \quad (2)$$

Equation 2 reflects the fact that an adjudicator has a some uncertainty, i.e. would possibly award the same

_____
* corresponding.author@institution.edu

speech different points each time he / she saw and forgot it agian. Now, the probability density function for such a speaker is the convolution of equation 1 and equation 2 and integrating out the probability to obtain a certain adjudicator.

$$p(x_n|\mu_{true}) = \int_0^{100} d\mu_n \, p(x_n|\mu_n) \, p(\mu_n|\mu_{true}) \quad (3)$$
$$(4)$$

The combination of the terms again leads to the form of a Gaussian integral. The integration is conducted for bounds of $\pm\infty$. We assume the probability density around the real bounds of 0 and 100 sufficiently small, such that the contribution of the integral beyond these points is negligible. The solution is calculated as:

$$p(x_n|\mu_{true}) = \frac{1}{2\pi\sigma_s\sigma_n} \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a} + c\right) \quad (5)$$
$$(6)$$

with

$$a(\sigma_n, \sigma_s) = \frac{1}{2}\left(\frac{1}{\sigma_s^2} + \frac{1}{\sigma_n^2}\right) \quad (7)$$

$$b(x_n, \sigma_n, \sigma_s, \mu_{true}) = \frac{x_n}{\sigma_s^2} + \frac{\mu_{true}}{\sigma_n^2} \quad (8)$$

$$c(x_n, \sigma_n, \sigma_s, \mu_{true}) = -\frac{1}{2}\left(\frac{x_n^2}{\sigma_s^2} + \frac{\mu_{true}^2}{\sigma_n^2}\right) \quad (9)$$

The probability function over x_n can be evaluated, given the parameters σ_s, σ_n and μ_{true}. We show a plot for an example range of parameters in figure xy.

B. Multiple Adjudicators

1. Without Point Dropping

For a number of adjudicators N, the resulting score $\bar{x}$ calculated through a simple average, i.e.

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n \quad (10)$$

Therefore, assuming the points are uncorrelated (no communication between adjudicators) the PDF for the total score can be calculated as

$$p(\bar{x}|\mu_{true}) = \frac{1}{Z}\prod_{n=1}^{N} p(x_n|\sigma_n, \sigma_s, \mu_{true}) \quad (11)$$

where Z is the normalization constant, i.e.

$$Z = \int_\Gamma \prod_{n=1}^{N} p(x_n|\sigma_n, \sigma_s, \mu_{true}) \quad (12)$$

where the integration is conducted over whole Γ, which denotes the adjudication space of dimension N. Again, this can easily be evaluated for a given set of parameters and examples are shown in figure xyz.

Now, we calculate the accuracy of the method through

$$< (\bar{x} - \mu_{true}) > \quad (13)$$

2. With Point Dropping

Now, we introduce a function $f(x_n)$ with

$$f(x_n) = 0 \text{ if } x_n \text{ smallest or highest element} \quad (14)$$
$$= 1 \text{ else} \quad (15)$$

Now, analogous to before:

$$\bar{x} = \frac{1}{N-2}\sum_{n=1}^{N} f(x_n) \quad (16)$$

For the adjudication space, this means a projection of $\Gamma_N \to \Gamma_{N-2}$, i.e. from the N dimensional space to the $N-2$ dimensional space.

III. NUMERICAL VALIDATION

We here simulate a high number of finals (excluding independent speakers).

A. Single Speaker

It is the goal to obtain a probability distribution of the points a speaker would obtain given his true strength, and subsequently calculate the accuracy. The true value is assigned to 50. Procedure:

- Draw five values μ_i discrete values from a Gaussian distribution with mean 50 and standard deviation 2. These are the means of the adjudicators.

- For each adjudicator, draw a value x_i from a Gaussian distribution with mean μ_i and standard deviation σ_i.

B. Team Win

Procedure:

- Draw discrete values denoted by μ_i from a Gaussian distribution with mean 50 and standard deviation 2. These are the means of the adjudicators.

- For each adjudicator, draw a value x_i from a Gaussian distribution with mean μ_i and standard deviation σ_i.

- In one case, take the mean of x_i as result, in case b), cross out the highest and the lowest ranking and calculate the result through the remaining three.

- Repeat a high number of times

- Calculate $< (correct - measured)^2 >$.

IV.   CONCLUSION

ACKNOWLEDGMENTS

Appendix A: Appendixes

[1] Https://usagym.org/pages/events/pages/fig_scoring.html.
[2] Https://www.pyeongchang2018.com/en/sports/ski-
jumping.
[3] Http://www.nbcolympics.com/news/diving-101-scoring.