

ON THE QUASISUBJECTIVITY OF POOL STATS DATA COLLECTING: ADDRESSING QUASISUBJECTIVE DATA COLLECTION AND MAKING DATA OBJECTIVE

VINCENT RUSNELL

ABSTRACT. There have been a few outcries on how we handle our data collection. Mainly, there is a notion that our data entry is subjective and will not generalize to a larger degree of players. This paper seeks to address these notions about how we handle subjective input and the policies, guidelines, and algorithms we have in place to ensure generalized and objective data. In order to accommodate a larger reading audience, we keep this paper brief on the math since the concepts are relatively easy to understand and further rigorous explanation is unnecessary.

1. INTRODUCTION

In recent weeks, there has been a few remarks about how we collect and analyze our data. The main concern is that we make subjective input. This is not entirely true as will be shown in the subsequent sections. Educated guesses are the pinnacle of science and data collection, and it is with the interest of the billiards community that we follow on this scientific method for collecting and observing pool data.

2. DEFENSE AS AN ART

Pool is a different bread of sport for an all encompassing data collection approach. There is some situations where the data recorder has to make an educated guess as to the player's intent which is then confirmed with some deal of certainty after the shot has taken place. This player intent can be seen on deciding whether a given player was intending to play safe/defense or was intending to pocket the object ball. We are not the only institution that has to make an educated guess as to what the player's intent was in playing safe.

The APA [1] defines a defensive shot as follows:

Definition 2.1. A shot where the shooter deliberately misses, so as to pass his turn at the table to his opponent

Also as,

Definition 2.2. A shot where there is no intent by a player to pocket a ball of his category.

As you can see, we are not alone; billiard institutions have not found a 100% full-proof method of deciding a defensive shot. That's why the institution in question needs to have knowledgeable personal available to record the data as accurate as one can.

It is our belief that the player's intent can be resolved 95% of the time, with very few data points in the sample marked as questionable. This is statistics and an important

Key words and phrases. Pool Stats, Data Collection.

factor to statistics is having accurate data to represent the population. There will always be errors in data recording and input and it is our job, along with the job of other data scientists, to find this erroneous input and discard it from the sample if necessary.

3. SHOT DIFFICULTY AS A SCIENCE

Determining the difficulty of a shot, in a pool game, can be incorrectly viewed as a subjective stat recorded by our team or an individual - we will see why by our definitions and in the following section. We only use individuals with acute knowledge of the game (5+ years playing experience) to record our data. It helps if the individual recording the matches also has a background in analyzing pool match videos, either on youtube or through a personal collection. It may appear to be a a stat collected on a subjective basis; however, there are fundamental properties of physics which factor into the final decision of shot difficulty recorded in our databases.

The definition of shot difficulty is two-fold. Different properties are used to record a shot aimed at potting the object ball and to record a safe/defensive shot. The definitions are as follows:

Definition 3.1. Potting: Shot difficulty is measured by the *probability* i.e., a number in the open-set $(0, 1)$, that the object ball will be potted to a called pocket, or in the case of the 9-ball, any pocket. The following properties of physics is essential in determining and defining the *shot probability*:

- (1) The degree of the angle which requires the object ball to be potted. Some straight shots are a test of the player's mechanics as opposed to the same shot with a little bit of angle.
- (2) The number and degree of reflective angles to calculate needed to pot the object ball. i.e., kick shots.
- (3) The number and degree of the deflective angles to calculate needed to pot the object ball i.e., combination, carom and bank shots.
- (4) The distance of the object ball from the cue ball i.e., jump shots, long off-angle or straight shots.
- (5) Hampered cueing. i.e., bridging over a ball, or executing a shot from a snookered position.
- (6) The English needed to pot the object ball.
- (7) The extremeness of the English needed in positioning the cue ball after a successful pot.
- (8) The distance the cue ball travels after a successful pot.
- (9) Cue ball speed (to a lesser degree)

Definition 3.2. Safety: Shot difficulty is measured by the *probability* i.e., a number in the open-set $(0, 1)$, that the object ball will be contacted in order to prevent a foul from occurring. The following properties of physics is essential in determining and defining the *shot probability*:

- (1) The degree of the angle which requires the object ball to be struck.
- (2) The number and degree of reflective angles to calculate needed to strike the object ball. i.e., kick shots.
- (3) The number and degree of the deflective angles to calculate in order to strike the object ball.

- (4) The distance of the object ball from the cue ball i.e., jump shots, long off-angle or straight shots.
- (5) Hampered cueing. i.e., bridging over a ball, or executing a shot from a snookered position.
- (6) The English needed to strike the object ball.
- (7) The extremeness of the English needed in positioning the cue ball after a successful strike of the object ball.
- (8) The distance the cue ball travels after successfully striking the object ball.
- (9) Cue ball speed (to a lesser degree)
- (10) The player's intent on placement of the object ball and cue ball, if known, or can be determined with a high enough degree of certainty.

One thing to notice is that in definition 3.1, specifically for the game of 9-ball, there is an element of luck that can occur in potting a ball. We are working on ways to handle this variable in future adaptations of Pool Stats Pro and our online Metabase Suite.

We note that, presently, the various factors in determining shot probability cannot be calculated exactly at the time of data input. These factors are approximated by intelligent individuals of the game when recording data for us, or for themselves using the Pool Stats Pro app. This is what makes shot probability a *quasisubjective* statistic.

In 3.2.10 we again see the term *intent*. As with the APA definitions on a defensive shot, the term intent is used here when determining the *shot probability*. It cannot be fully known, but with the aid of knowledgeable announcers, in the instance of professional pool matches, analyzing the shot after it has occurred and prior to submitting the data, it can be well approximated with a high enough degree of certainty.

We are aware that the *shot probability* for a given player may be different than that of another player. So, the question remains, how can one ensure that there is a generalization of *shot probability* among all players? Using the properties of physics in deciding difficulty as outlined in the above definitions is one very good approach, but more is needed to generalize the difficulty for players who excel at different shots. We outline the methods of *generalized shot probability* in the next section.

4. MAKING QUASISUBJECTIVE DATA OBJECTIVE

When it comes to deciding if a player was playing safe/defense or not, there is not much in our community one can do when recording this data. An approximation based on [1] together with a player's body language and psychology is the best we can do at this time. Combing through hundreds of thousands of data points to discard a maximum of 5% of the defensive shots is unrealistic and a waste of time since we already have a high enough degree of certainty in the sample.

What we can make objective, generalized, and more certain for the group of player's as a whole, is *shot probability*. Outlined below is the constructs of an algorithm we have in place to ensure an objective and generalized approach to *shot probability* as defined in the last section. We call this algorithm *Generalized* or *Adjusted Shot Probability - AdjSP* for short.

Algorithm 4.1. *Generalized Shot Probability*

- (1) Create a column in the **gamedata** table and label it **AdjSP**.
- (2) Count the number of each case of shot probability i.e, # of .1, # of .25 ... # of .9
- (3) Count the number of made, missed, and defensive shots for each shot probability case.
- (4) Compute the shooting percentage [4] of each probability case; this gives **AdjSP** for each probability case.
- (5) Identify and populate the new column **AdjSP** of the **gamedata** table with the respective shooting percent of each probability case.

This gives an account of the percentage of shots made for a given *shot probability* as defined in the previous section. With this adjusted shot probability, we effectively make the data objective for all player's on a whole. The reasoning behind it is because we have calculated the rate at which a recorded shot probability is occurring, thereby giving a true measure of the probability of any given shot. Users of our online web portal can then use this adjusted shot probability to compute stats more accurately for every player on the whole sample of data, as we do.

5. FURTHER REMARKS

Currently, we have not populated the **AdjPM** column of our database. This is due to insignificant amateur player data. We hope that people and tournaments of amateur pool players will begin adopting our app and recording data for us to analyze and provide results. This will give us the data necessary to accurately compute **AdjPM** for the range of amateur, league, and professional players alike. Please join us in reaching out to your league in hopes of recording vital stats that players should have. The algorithm is in place and ready for deployment as soon as the sample size increases to a level we are comfortable with.

Furthermore, we also have an algorithm that will allow us to compute the correct *shot probability* for any give player based on a large enough sample size. This will highlight which players excel at shots of various difficulty. The algorithm is the same just grouped on a player's data instead of the entire sample. This same algorithm can be used to calculate **AdjSP** on player level, league level, FargoRate level, and professional pool player level. Many things are possible when generalizing stats and we hope to bring a lot to the table for the pool community with our methods.

REFERENCES

- [1] American Poolplayers Association: *Defensive Shots and APA Scorekeeping*
<https://www.youtube.com/watch?v=VL8YxKFQd-o>
- [2] James, Bill: *The Bill James Handbook*
2014, Baseball Info Solutions; ACTA Sports
- [3] Pomeroy, Ken: *Pomeroy College Basketball Ratings*
<https://kenpom.com>
- [4] Rusnell, Vincent: *Shooting Percentage: A simple statistic of the rate of made shots*
Not yet published; preprint.

POOL STATS L.L.C., DEPARTMENT OF MATHEMATICS, STATISTICS, AND COMPUTER SCIENCE,
DETROIT, MI

Email address: vrusnell@umich.edu, vrusnel@clemsn.edu