

A Conceptualization and Quantification of Skill

Ryne A. Sherman
University of California, Riverside

Author's Note: This article was originally published under the title of "Towards a Skill Ratio" in the 2+2 Internet Magazine Vol. 3 No. 5. Correspondence regarding this article can be addressed to the author at rsher002@ucr.edu.

A Conceptualization and Quantification of Skill

A typical debate surrounding poker is the question of skill. Specifically, how much skill is involved? Many arguments have been made in an effort to “prove” how much skill is involved. However, few have considered that the first step in understanding how much skill is involved in poker, or anything else for that matter, involves defining the term “skill.”

So what is skill? There can be many different definitions of skill. Here are a few provide by Dictionary.com: (1) “the ability, coming from one's knowledge, practice, aptitude, etc., to do something well” (2) “competent excellence in performance; expertness; dexterity” (3) “an ability that has been acquired by training” and (4) “ability to produce solutions in some problem domain.” Though these definitions are useful, they don’t seem to be concrete; meaning that these definitions don’t describe what it means to be skillful. For example, shooting three-point shots in basketball can be considered skillful. However, we certainly wouldn’t consider everyone who could make 1 three-point basket in 5,000 attempts to be skilled. Rather, skill requires the *consistent* ability to do something well. Taken one step further, the skillfulness of any game can be conceptualized as follows: ***Games in which persons make conscious behavioral efforts are considered more skillful if there are consistent individual differences in performance.*** From this, it also follows that games in which persons make conscious behavioral efforts are considered less skillful if performances are inconsistent and/or if individual differences are small or non-existent.

Before going any further, let us breakdown this conceptualization of skill. First, games with *greater consistent individual differences* are considered more skillful. Consider for example the lottery and chess. In the lottery, individual differences are minute even over a large number of trials. Although the central limit theorem predicts that individual differences will exist in lottery performance over a large number of trials, as the number of trials gets larger, individual differences diminish. In chess however, individual differences are much larger even over a large number of trials. These consistent individual differences are due to skill.

Additionally, it is important to note that for something to be considered skillful, a conscious behavioral effort must be made. For example, one might play the hypothetical game of “who is taller.” For simplicity consider it a two player game. Assuming neither person grows during the course of this game, nor are the two players the same height, one person will always

win. Despite demonstrating consistent individual differences, no conscious behavioral effort has been made. Therefore, being taller than another person is not skillful.

For the rest of this article, I consider a quantification of skill. From the preceding paragraphs, it is apparent that consistent individual differences constitute skill. By individual differences, I mean performance from one person to another differs, such that increased differences among people in a particular game is evidence for that game being more skilled. Keeping with the above examples, individual differences in long term lottery performance are clearly smaller than individual differences in long term chess performance.

By consistent I mean that skill is indexed by the ability of the same individual differences emerge over multiple trials. Keeping with the examples above one would expect greater consistency (using the outcome win or don't win) while playing chess than while playing the lottery from trial to trial.

So how can consistency among individual differences be quantified? The answer is actually quite simple and has been in existence for quite some time. The Pearson Product-Moment Correlation coefficient (r) can be used to estimate the consistency of individual differences for any game. The equation for r is shown below:

$$r = (\sum Z_x Z_y) / N,$$

where Z_x are standardized scores for one variable, Z_y are the standardized scores for the other variable and N is the number of pairs of scores. For those unfamiliar with the correlation, a simple web search will yield many results explaining its function. As a simple definition, consider the correlation a measure of correspondence between two variables which ranges from -1.00 (perfect opposite correspondence) to 1.00 (perfect correspondence) with 0.00 meaning no correspondence whatsoever.

Consider again the two examples of the lottery and chess. In the simplest case, let us assume 100 persons have played both the lottery and chess 2 times each (not against each other). Consider the first time playing the lottery trial x and the second time trial y with the outcome being money won. The correlation between these two trials would not likely be very high. Now consider the first time playing chess trial x and the second time trial y with the outcome being pieces captured from one's opponent. While the correlation between these two trials might not be

very high, it is certainly higher than that of the lottery. This is because individual differences in the ability to play chess are more consistent than individual differences in the ability to play the lottery.

A second example may make this clearer. Consider again two games, say batting average in baseball and the ability to roll a “7” using two six-sided regular dice. Imagine that the same 100 Major League Baseball players are involved in both games. In the batting average game, randomly select $\frac{1}{2}$ of a player’s at bats and compute the average. Then take the remaining $\frac{1}{2}$ of that player’s at bats and compute the average. Do this for each player such that each player has two batting averages. Then have each player roll the two dice, say 500 times scoring a yes if the player rolled a 7 and a no if the player did not roll a 7. Randomly select $\frac{1}{2}$ of those dice rolls and take the average number of 7s rolled. Do the same for the other $\frac{1}{2}$. Do this for all 100 players. Now, compute the correlation (amount of correspondence) between the two batting average scores and the two dice scores across all 100 players. If batting requires more skill than rolling 7s, we would expect that there would be a stronger correlation between the two batting averages than between the two dice rolls. In fact, if one were to do this project, this is exactly what one would find. This stronger correlation is evidence that batting in baseball is more skilled than rolling 7s.

Application to Poker

Thus far this article has demonstrated that the skillfulness of games is defined by the ability for people to show consistent individual differences in performance. Additionally, we have seen that the simple Pearson Product-Moment correlation coefficient can be used to estimate the consistency of individual differences, and therefore estimate the amount of skill involved in any game.

So how can this be applied to poker? The answer is quite simple. First one must define a poker trial. This can be anything from a single hand of poker to, a single session, a single tournament, or a single year’s winnings. Next one needs to identify a number of persons completing this trial. Then get a second trial from these same people. The correlation between trial 1 and trial 2 is quite simply the amount of skill involved. Without much thought, one will quickly realize that there isn’t much skill involved from one poker hand to the next, probably near zero. However, there is a great deal of correspondence, or skill, involved in one’s yearly winnings to the next.

Thus far I have shown how the correlation coefficient can be used to estimate the skill in a game by using two trials of the said game. However, very often we have many trials of the said games. Especially in poker where people have thousands to hundreds of thousands of hands played. Using a special formula, one can estimate the skill over any number of hands. This formula is Spearman-Brown Reliability formula and it is defined as follows:

$$r_{xy} = (M_r * N) / 1 + (N - 1)M_r,$$

where M_r = the average correlation among all the trials (or measures), N = the number of trials (or measures), and r_{xy} is the estimated correspondence or skill. Using this formula, one could do two things.

First, one could compute the amount of skill involved in a given number of poker hands, say 100. First gather how much was won/lost for 100 poker hands from a large sample of people and compute the correlations between each of those 100 hands. This will yield 4,950 correlations ($100 * 99 / 2$). One would then compute the average of these correlations. This number could then be plugged into the above formula as M_r . The number of trials (hands) is plugged in as N , in this case 100. The resulting number would be an estimate of the amount of skill involved in 100 poker hands.

Secondly, one could also estimate the number of poker hands needed to achieve a particular skill level. For instance, say one already knew the skill level of some other game (say batting in baseball). One would simply insert that number into the formula as r_{xy} and solve for N , the number of poker hands needed to approximate the skill of batting in baseball. The algebra of solving for N is already done in the formula below:

$$N = r_{xy} (1 - M_r) / (M_r (1 - M_r)),$$

where again M_r = the average correlation among all the trials (or measures), N = the number of trials (or measures), and r_{xy} is the estimated correspondence or skill.

Conclusions

Skill requires the consistent ability to do something well. By definition then, games which have greater consistent individual differences are more skillful. This article has

demonstrated how a common measure of individual difference consistency, the Pearson Product-Moment Correlation, can be used to estimate the skill between any two given trials. Additionally, this article has demonstrated how the Spearman-Brown Reliability Formula can be used to estimate the skill of a game over any given number of trials. In next month's magazine, I plan to further demonstrate the utility of these measures using real sports and gaming data.