A commentary on *Z-score* methodology to detect bridge cheating

Richard Johnson
bridge@richardjohnson.io

November 15, 2020

Introduction

This is a commentary on the *Z-score* methodology developed by Paul Barden (PB) which has been applied as evidence to help detect cheating in the game of bridge.
The commentary is divided into three parts:

1. A non-technical overview clarifying what is measured and why it is of interest.
2. Details of the PB methodology. Technical statistical clarification, and free, standard, open-source versions of programs used by PB.
3. Observations on strengths and weaknesses of the method.

Author

Background: I have made my living from bridge for 20 years, and hold two masters degrees in Mathematics, including one in Statistics.

I have collected a set of free, open-source Python tools to assist technically-minded readers who wish to perform the calculations in the PB methods independently here:
https://github.com/xorich/karapet

1 Overview

1.1 Motivating example

Consider South's contract of 6♡ here (*example* 1):

♠AK9
♡QJT98
♢QJ10
♣32

N
S

♠QJ10
♡AK765
♢AK9
♣K4

1. If West holds the ♣A, then the fate of the slam will be determined in the play. The plausible lead of the ♣A allows the slam to make; only if West refrains from a club lead will the contract fail.

2. If East holds the ♣A, and no void, then West's lead doesn't matter - however expert the defence, they are unable to beat the slam - *even if they could see each other's hand and defend perfectly* (double-dummy). Only if declarer chooses an unsuccessful line of play will the contract fail.

When a layout is such that not even a perfect defence could beat the slam this is an *a priori* property of the deal, unaffected by the auction or the relative skill of the players.
If declarer bids slam in scenario 2, when the defence's cards lie favourably, irrespective of the fate of the contract, this is a **success** in PB's terminology.
If declarer bids slam in scenario 1, when the layout is dangerous, irrespective of the fate of the contract, this does **not** count as a **success**.
Over many deals, one would expect inexpert, club defenders to allow an expert to make slams that a perfect defender might have beaten. Similarly, one would expect an expert to bid good slams that should fail when the opponents' cards lie unfavourably - an unlucky trump break, for example. Neither of these is possible when declarer's slam is a **success**.

To play against a declarer who only bids slams that are **successes** as the cards lie is to play poker against a man named Doc who only bets when you hold a bad hand.

1.2 PB methodology

It is remarkable to hold a ten-card suit, but they are occasionally dealt to innocent players. Similarly, it is remarkable if the proportion of declarer's slams that are **successes** is much higher than expected, but this expectation requires careful analysis.
Which deals of a set of boards are **successes** is routinely printed on hand-records and can be displayed on dealing machines. Unfortunately access to this information before play can obviously therefore lead to suspicion.
PB's work attempts to quantify the probability p of a given proportion of **successes** arising

by chance. He uses the terminology *Z-score* for the reciprocal of this probability. As a numerical example, a probability of $p = 0.1$, is equivalent to a *Z-score* of $\frac{1}{0.1} = 10$, a 1 in 10 chance. Note that the PB terminology should not be confused with other usages of the term *Z-score* in Statistics.

2 Details of PB Methodology

2.1 Formal approach

Formally, PB adopts a hypothesis testing approach in his analysis. He uses the number of **successes** s in n trials as a test statistic.
He then models the distribution of the test statistic, using Monte Carlo simulation to approximate necessary parameters (see below). The technical clarification offered by this author is to identify this distribution as exactly the Poisson Binomial probability distribution.
The p-value associated with this statistic is the reciprocal of the *Z-score* quoted in PB analysis, which is presented as evidence against a null hypothesis that the observed number of **successes** occured purely by chance.

2.2 Simulation of p_i

Consider example 1 above. Considering the North South hands alone, how likely was this contract **X** to be a **success**? PB estimates this probability using simulation, using a dealing program to generate 1000 possible East-West layouts, and estimating the probability of the contract being a **success** in the natural way:

$$P(\mathbf{X} \text{ is a success}) = \frac{1}{1000} \sum_{i=1}^{1000} \delta_0 \left(\mathbf{X}_j \right)$$

where δ_0 is an indicator function equal to 1 if simulation $\mathbf{X}_j$ is a success and 0 otherwise.

The parameter p_i is the simulated probability that deal i would be a success. For each $i, \quad i = 1, \dots n$ for a set of n deals, the corresponding p_i is simulated as above. Given these parameters p_i for each deal, the distribution of the test statistic, s (the number of successes in n deals), can be calculated.
The footer gives a link to a free, open-source Python implementation of a well-known dealing program which can assist the reader to perform these simulations, and also an integrated Python implementation of a well-known Double Dummy solver which is necessary for the analysis of whether a contract is a success.
Applying these methods to example 1 gave a simulated probability **X=0.484** (48.4%). This example code is available as `example.py` in the footer link.

2.3 Distribution of the test statistic

PB has written his own code to evaluate probability of observing s successes in n independent Benouilli trials, with varying probability $p_i \quad i = 1, \dots n$. He provides an algebraic justification of his approach in the note *9540 Calculation of Probabilities*.

Fortunately, it is not difficult to independently verify this aspect of the calculations, as this is a well-known problem in engineering and actuarial science, recognisable as the Poisson Binomial probability distribution. Multiple powerful, numerically stable methods to calculate the characteristics of this distribution are available, for example, by Hong Hong [2013]. For convenience, the author has included a Python implementation of this as a submodule in the github link in the footer.

3 Observations

3.1 Strengths

PB's method is an ingenious, powerful tool that is useful metric in investigations where improper knowledge of the opponents' cards is alleged.
As the test statistic is a property of the deal rather than the players, it suggests an objective, direct link to indicate prior knowledge of the deal valid irrespective of the style or strength of the players at the table.
The analysis treats both members of a partnership as a unit. This makes the analysis conservative when the partner of an accused is innocent, as the innocent player may bid slams that a guilty player would have avoided, thereby reducing the proportion of **successes**.

3.2 Weaknesses

In hypothesis testing, interpretation of p-values should be done carefully in the broader context of the analysis, for example necessary corrections if conducting multiple testing.
The determination of a result as statistically significant is not predefined and hence requires care, as does the interpretation of the strength of evidence that such a result provides. If a player bids slams infrequently, a longer time period will be necessary to build the sample.
As originally presented, it is not straightforward to reproduce the analysis of PB. However, high quality resources are freely available that enable the reader to conduct a private analysis, as indicated through this text.
The simulation of the p_i does not include the auction, which might be a legitimate source of information about the opponents' hands for declarer. For example, in example 1 a club preempt by East/West might locate the ♣A. PB comments on deals manually which addresses this issue; another approach would be a sensitivity analysis where conservative estimates of p_i are made and a calculation is presented of the p-value for the test statistic under those assumptions.
A comparison of the statistics of other players, or of the same player in different settings, might provide useful intuition about the test statistic. Clarity and transparency is important for the accused, as it may not be obvious to a player, for example, that neither the *table result*, nor the *relative expertise of the players* are relevant to this analysis.

References

Yili Hong. On computing the distribution function for the poisson binomial distribution. *Computational Statistics and Data Analysis*, 59:41 – 51, 2013. ISSN 0167-9473.