A Strong Information-Theoretic Bound for Providing Secure Signs in Baseball

Professor Anton Dahbura, Jack Pausic ‘22

Johns Hopkins University | Whiting School of Engineering | Baltimore, MD

Design Day 2021

JOHNS HOPKINS
WHITING SCHOOL *of* ENGINEERING

Introduction

One of the most durable practices in the game of baseball is the utilization of hand signals to call plays such as pitches (by the catcher to the pitcher) and bunts, steals, and hit-and-runs (by the third base coach to the batter and baserunners) while concealing one team’s strategy from its opponent. Many teams attempt to intercept useful information about opponents’ strategy by decoding these signals. In this project, we formalize the processes of conveying and observing signals and show that there exist optimal strategies for both parties that can gain a competitive advantage. This project builds on the information-theoretic bound established in previous semesters of **n - m + 1** sequences before meaningful information can be obtained, where n is the length of the alphabet and m is the number of signs in one sequence.

Objectives

To explore the differences observed in our bound when an allowance is introduced for human errors to occur during the transmitting and receiving of signals. A practical application of this project is to assess the robustness of sign systems employed by different teams.

Materials and Methods

Analysis was conducted primarily through construction of theoretical scenarios in baseball games. These scenarios formalize the processes of transmitting, receiving, and attempting to decode signs, each role being carried out by a different actor: the sender, receiver, and observer, respectively. The signals transmitted prior to each pitch are represented by a sequence of integers that typically correspond to body touches a coach performs in order to make the signals visible to their team. Each sign is chosen from a set, called the alphabet, which contains all the possible signs the coach can give. One of the signs contained in the alphabet, the key, corresponds to an action performed by the receiver upon reception of the sequence. Knowledge of the identity of this key is the most complete information the observer can have about their opponents’ sign system. Our scenario is modeled as a game in which the sender attempts to transmit signs for as many sequences as possible before the observer can obtain useful information about the system, or code, the sender is employing. This semester, we generalized the scenario into one that allows for a certain number of human errors to occur during transmitting and receiving.

Results

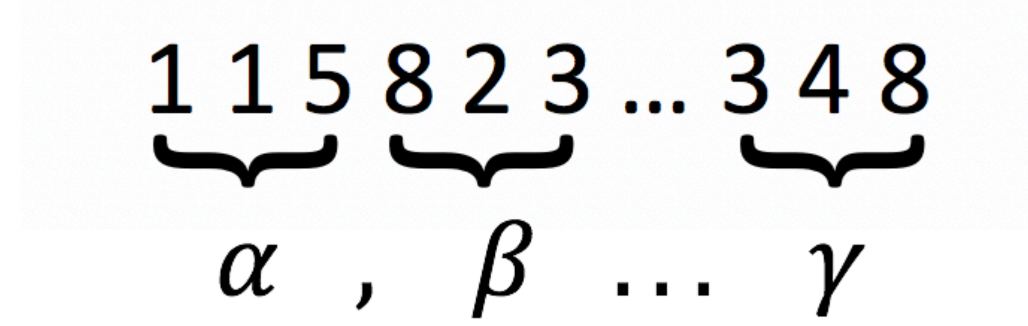


Figure 1— Generalization of More Complex Sign Systems

Of course, the system being used to encode the information being conveyed in a sequence of signs varies by level of competition. Signs considered can be each individual signal conveyed via body touch, or they can be composed by strings of such signals in more complex systems (if, for example, string of 3 touches was the key). If the observer knows which of these systems is being used, or checks all of them by running their system concurrently, then the information from the number of touches being considered a sign can be represented with just one sign, simplifying our problem back to the original case where each sign is considered individually.

Background—Error Tolerance

Over the course of a game, errors may occur on the part of the team sending and receiving signs. These errors may be attributable to misinterpreting a sequence or the receiver, usually a player, exercising their own discretion over the play. Regardless, the deviation of the result of a sequence from the sender’s intention is misinformation that would break our algorithm and force it to start over from the beginning. To address this possibility, a version of the algorithm was introduced that allows for a certain number of two types of errors: those of commission (t_c) and omission (t_o). An error of commission can be thought of as an action occurring that was unintended by the sender, while one of omission is the failure of the receiver to execute the action intended by the sender. However, we also hypothesized that, if the sender knows that such an allowance exists, that they could manipulate their transmitting strategy to take advantage of the allowance and gain a few additional sequences of obscurity.

Error-Free Case

Sequence	Result	
Sequence 1	Null ⟵	Since no action occurred, no signs contained in the sequence can be the key. Eliminate all of them from consideration.
Sequence 2	Action ⟵	Since action occurred, key must be contained in the sequence. Possible keys = (signs in sequence) ∩ (signs not eliminated).

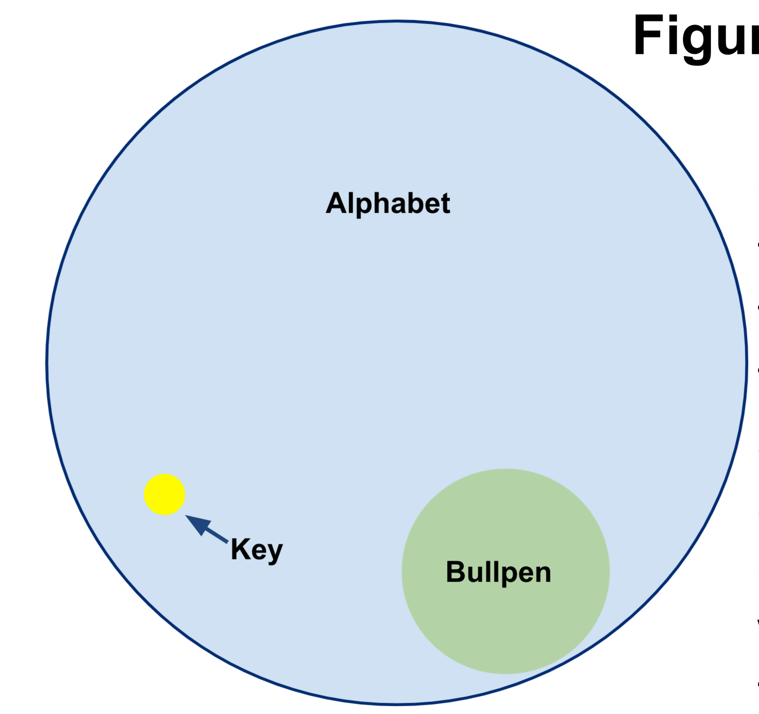
Considering Errors of Commission (t_c = j)

Sequence	Result	
Sequence 1	Null ⟵	The only type of error being considered results in an action, so as in error-free case, eliminate all signs from consideration.
Sequence 2	Action ⟵	We are allowing for up to j errors of commission, so we must consider the possibility that the runner carried out an action without receiving the correct sign up to j times.

Considering Errors of Omission (t_o = j)

Sequence	Result	
Sequence 1	Null ⟵	We are allowing for up to j errors of omission, so we must consider the possibility that the runner did not carry out an action but should have.
Sequence 2	Action ⟵	The only type of error being considered does not result in an action, so as in error-free case, possible keys = (signs in sequence) ∩ (signs not eliminated.

Figure 3—Diagram Showing Bullpen Set-Up



Our hypothesis surrounding the exploitation of an error tolerance further posits that the sender will set aside some additional number of signs, called a “bullpen” that can be used in a particular way as decoys to obscure the true identity of the key.

The allowance takes the form of assigning strikes to potential keys rather than eliminating them with a ”one-and-done” approach. The sender must attempt to maximize the number of strikes his bullpen, or decoy, signs have while also minimizing the number of strikes the next sequence they transmit will cost. Having a larger bullpen gives more starting strikes available, but if there are too many, they can’t be shuffled effectively, and the number of strikes lost after each sequence will be high. We then sought to determine the optimal number of signs to be stored in the bullpen, assuming our strategy is to be executed effectively by the sender.

Optimal Bullpen Results

In the case of errors of commission, we determined that it is optimal to hold out just one additional sign to include as a decoy. Doing so added **m** additional sequences of utility to the sender’s sign system.

Conclusion

Given that our findings concluded that the additional sequences a sender can add when the error tolerance is considered is significant, an observer should certainly consider whether they believe the actual probability of an error occurring is significant before choosing to employ such an allowance. Future work for this project includes analysis of the optimal strategy for exploiting errors of omission as well as continuing work on an academic paper to be submitted to sports analytics conferences. A possible extension of this project would be to employ image processing to recognize the body touches and encode them automatically before being used in our algorithm.