STATS 50: Mathematics of Sport **Spring 2019**

Week 6 – Game to Game Consistency

Lecturer: Maxime Cauchois

Warning: *these notes may contain factual errors*

1 The limits of mean reporting

In a wide variety of sports, performance is generally acknowledged and reported in two different ways. On the one hand, crowds tend to favour raw numbers as a measure of greatness, ranking a player with more "scoring events" (goals, baskets, touchdowns, winners, hits...) over players whose work for the team may be more obscure and hidden. In some way, there is nothing really wrong with this approach insofar as those players are also the ones allowing a team to enlarge its fanbase and attract more people to the stadium, eventually yielding more money for the franchise. At the same time, in terms of pure sport performance (i.e if one is only interested in winning), we have previously discussed to what extent these numbers are not sufficiently representative of a player's impact on the floor, and how other more accurate measures can be derived.

In the lectures concerning Regression to the mean or Bayesian estimation, we extensively focused on the problem of estimating the population mean for each player based on roughly three main pieces of information: the number of attempts, the number of successes and some prior information about the population of players. However, no matter how accurate our estimation may be, we are still summarizing a player's game with a single number. We saw when discussing variance estimation how the mean can sometimes yield little to none information about the player's real abilities, in particular how it fails to account for any kind of consistency (see figure 1). Without going as far as trying the predict whether or not the next attempt will be a success, which seems unrealistic, we'd like to offer a more precise description of a player's performance in a specific domain (shooting, passing, ...), especially from one game to another.

In most of our previous studies, we had this very strong assumption that all the attempts in every were not only independent, but had the same probability of success equal to the player's real success percentage. However, in practice, players arguably don't show the same level of performance and consistency at every game. For instance, in basketball, analysts agree on the existence of "good" and "bad" shooting nights. While these fluctuations might also be explained with sheer luck, a quick statistical analysis would reject the hypothesis of "uniform" level during each game. In other terms, if one wanted for any player to test whether or not the observed performance is compatible with our (too) strong prior assumption, this hypothesis would much likely end up being rejected. While this forces us to come up with a new model to account for this new variability, it also gives us more freedom and potentially more accuracy in describing a player's game.

2 Evaluating game to game variability

For simplicity, we will assume here that we consider a player whose number of attempts N is constant across each game. Typically, one would have N around 10 in basketball, or 3 in baseball, and so on... This assumption can be relaxed to the case of unequally weighted games, but it will make our analysis more readable for now.

The most simple model assumed that during each game $i \in \{1, ..., n\}$, the player had a number of successes $X_i \sim \text{Binomial}(N, p)$ where p is the real success percentage for the player. In other

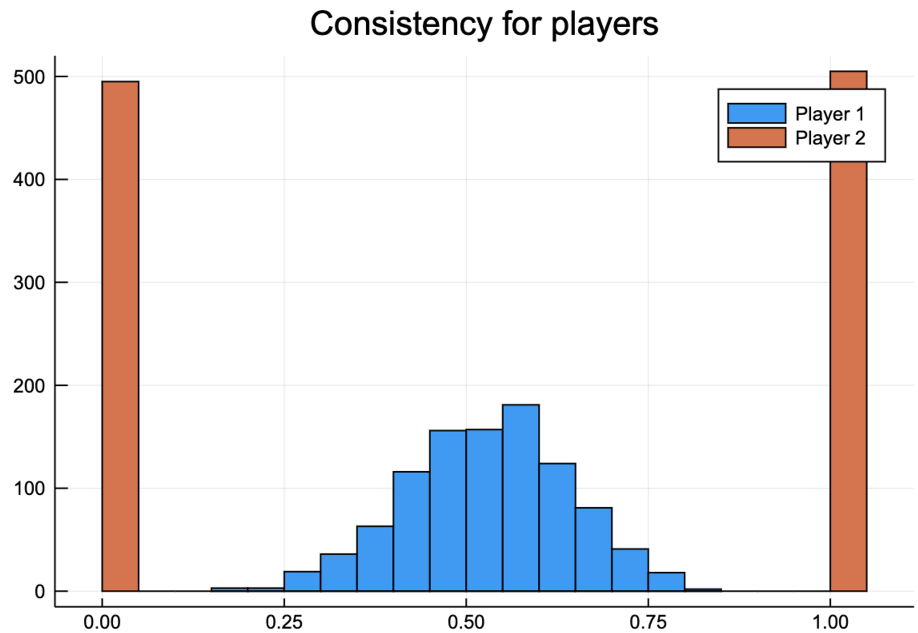


Figure 1. Histogram of players' shooting percentage during each game, assuming that both of them attempted 20 shots at each game

terms, at each game, the number of successes follows the same distribution, where every attempt is independent from the others and made with probability p.

Now, we are considering instead that at each game i, the player has a probability $p_i \in (0,1)$ of making each attempt, where the probabilities, p_i are now allowed to vary between each game. To be precise, within each game, all the attempts are still realized independently, but the probabilities of success are allowed to vary from one game to another.

In summary, we are allowed to observed a set of independent variables $X_1, ..., X_n$ such that: representing the number of successes at each game, such that for each $i \in \{1, ..., n\}$:

$$X_i \sim \text{Binomial}(N, p_i).$$

Our model now becomes very flexible, because one could potentially the player's ability each game with a different number p_i. However, remember that we are only allowed to observe N different attempts, where N is potentially as small as 3, so it is completely hopeless to estimate well each p_i individually. Indeed, remember that if $p_i = 0.5$ and $N = 10$, then one still has a non-zero probability (around 18%) of observing only 3 successes or less, so with large probability, one would estimate $\hat{p}_i \leq 0.3$. In short, the number of attempts during each game is way too small to mitigate the variance due to pure luck.

The good news is we do not actually want to estimate each p_i individually: we are much more interested in computing their distribution altogether, or if you prefer how they are scattered in the $[0,1]$ interval. Indeed, in figure 1, we have two extreme cases with one player whose all p_i are equal to 0.5, and on whose half of them are equal to 0 and half of them to 1. In a more realistic setting, you would expect the p_i to be grouped around the overall mean (the $p \in [0,1]$ we previously wanted to estimate) with more or less variance depending on the player consistency. For instance, if one assumes that at each game, p_i is drawn itself from a $Beta(6,9)$ (to get 40% accuracy), figure 2 gives a plausible histogram for the distribution of p_i.

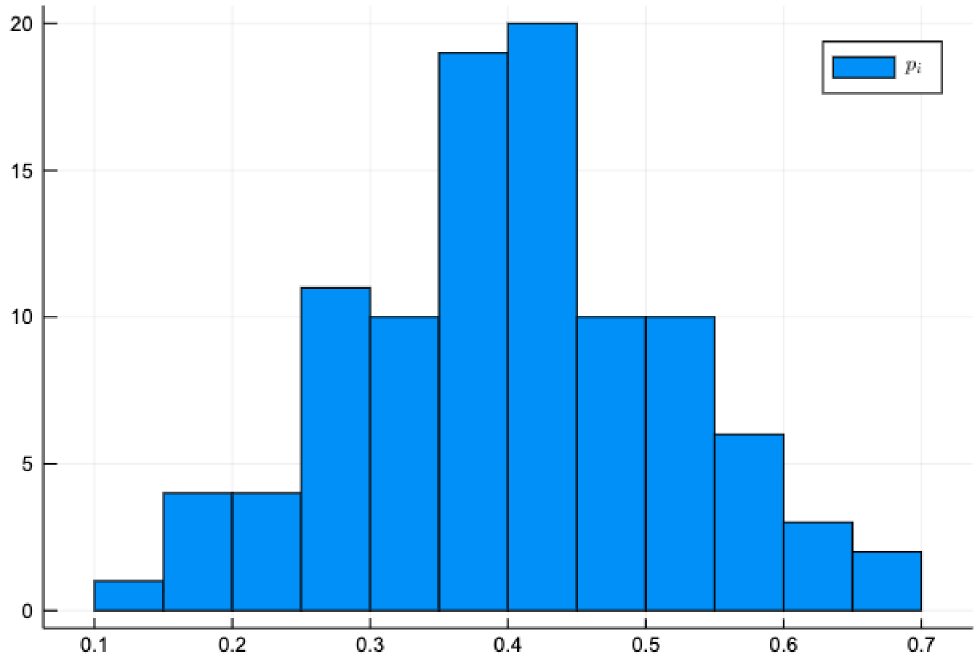


Figure 2: Simulated histogram of real p_i drawn from a $Beta(6, 9)$ distribution

The main question is now to estimate this histogram as accurately as possible. Indeed, at first view, one would think that the histogram of X_i/N should be a good estimator for the p_i but some examples should convince you that it is actually a very poor representative. First, in the case where all the p_i are equal to 0.5, we know that the histogram of the X_i/N yet won't be peaked on 0.5 and will have a much wider standard deviation (see figure 1). However, one could still think that this case is somehow too specific and that if the p_i are more scattered, then it should be the case that the empirical distribution of the X_i approaches the one of the p_i. As the figure 3 shows, this is not the case for several reasons. First, X_i only takes on discrete values $(1, ..., N)$, where N is much smaller than n, which explains why both histograms look so different. Then, the X_i/N are often much more scattered on the whole interval $[0, 1]$ that the p_i, mostly because we actually add noise when we simulate each X_i based on p_i, increasing the variability within our data. For these reasons, it is unlikely that our histogram of X_i yields a good estimate of the distribution of p_i.

3 Hints about the method of moments

We want to estimate the distribution of the p_i. There exists a nice mathematical result which states that the vector $(p_1, ..., p_n)$ is uniquely determined up to its order when all its "moments" α_k are fixed, where:

$$\alpha_k = \frac{1}{n} \sum_{i=1}^{n} p_i^k$$

In other terms, for any sequence $\alpha_1, ..., \alpha_n$, there is only a single vector $(p_1, ..., p_n)$ able to generate those moments. Now, because the sum of p_i^k is a sort of averaging, it turns out these quantities can be estimated pretty well!

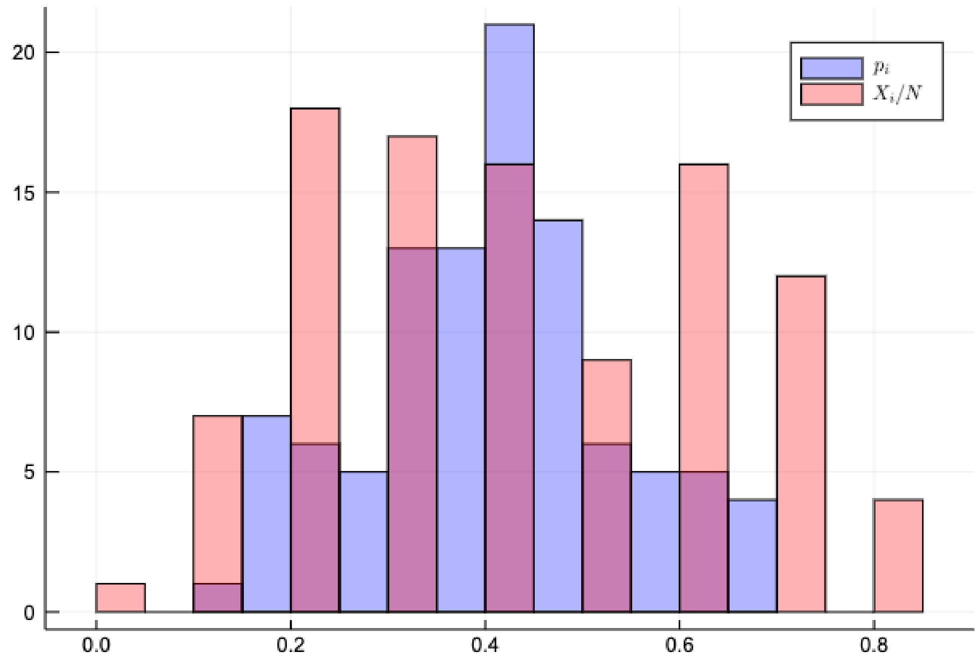


Figure 3. Simulated histogram of real p_i drawn from a $Beta(6, 9)$ distribution, along with the realized X_i/N

More specifically for any $k \in \{1, ..., N\}$:

$$\mathbb{E}(\beta_k) = \alpha_k \tag{1}$$

where:

$$\beta_k = \frac{1}{n} \sum_{i=1}^{n} \frac{\binom{X_i}{k}}{\binom{N}{k}} \tag{2}$$

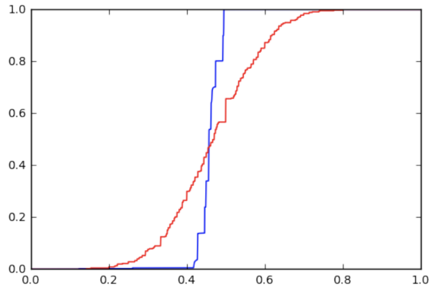
and with high probability β_k and α_k are close to each other. Because all β_k are entirely dependent on observed quantities, this implies that we are able to accurately recover all the moments of order k up to $k = N$.

All that we are "left" to do to estimate the p_i is to find a distribution $\hat{p}_i$ that matches these computed moments as well as possible, which can be done through a Convex Problem described more extensively in [1].

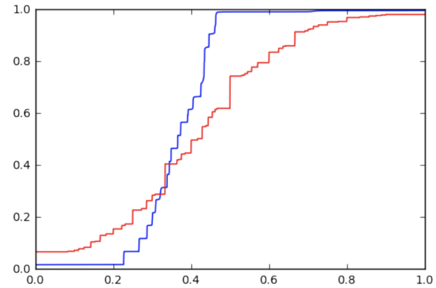
4 Basketball estimation

In [1], they also apply their algorithm to the case of two different NBA players, Stephen Curry and Danny Green, and compare their findings. As one would expect, one player shows much more consistency across all of his games than the other. The findings are shown in figure 4: one is very consistent almost every night, while the other alternates more between on and off shooting nights.

Eventually, how can this form of results be applied? There is a very nice Bayesian interpretation to this problem. Suppose that we found an accurate distribution $\hat{p}_i$ representative of a player's talent, or even that somehow we have an oracle who gave us the distribution from which the p_i are



(a) Estimated CDF of Curry's game-to-game shooting percentage (blue), empirical CDF (red), n=457 games.



(b) Estimated CDF of Green's game-to-game shooting percentage (blue), empirical CDF (red), n=524 games.

Figure 4: Consistency comparison between two NBA players

drawn. Now, on any given night, a specific p_i is drawn from this distribution, but I don't get to observe it. As a coach, I can however observe how the game is going and obtain new information based on the first attempts. More precisely, suppose that the player I am interested in missed the first 4 attempts. If the player in question is Steph Curry, it appears that almost every night, his probability is close to 43% (because he is really consistent), so I can just attribute these misses to bad luck and I would argue that he should continue shooting even then. If the player is Danny Green, the situation is different because I know his performances are very volatile, and the fact he missed his first four attempts gives me strong indication that he currently is in a bad shooting night, where he is likely to shoot between 20 and 30%. In this case, there might be other players on the floor to shoot the ball better...

References

[1] Kevin Tian, Weihao Kong, and Gregory Valiant, Learning Populations of Parameters. NIPS, 2017.