A Holistic Statistical Test for Fairness in Video Poker

Evan Kleiner

May 17, 2014

Abstract

Electronic gambling has been around for decades and has gained popularity in recent years on the internet. In particular, video poker is a popular game where computers are used to simulate traditional poker. This paper addresses the possibility of rigging electronic games by artificially manipulating the probabilities associated with the game. Such modifications are unethical and illegal. This paper develops a statistical test capable of detecting this cheating by casinos. We begin with an introduction to video poker and the statistical methods used. The final part of the paper describes the test and details simulation results.

1 Introduction

1.1 Video Poker

Video Poker is a casino game played around the world and online. In contrast with regular poker, video poker consists of only one player gambling against the house on a machine. The player is dealt five cards, and then chooses to keep none, some, or all of the cards. If they choose to keep some or none of the cards, the player may discard the unwanted cards, and get the same number of new ones. This new hand is then assigned a score by the machine. Unlike traditional poker, each hand in video poker has an inherent point value, and the machine will pay out a pre-specified amount of money whenever the player gets that hand.

Figure 1: This figure shows the payout table for one version of video poker. [1]

Figure 1 shows the payouts on a $1 bet for the 'Jacks or Better' version of video poker. Though the concepts discussed in this paper could be applied to any version of video poker, all ensuing analysis has been completed using this version. In 'Jacks or Better', a two pair is a pair of pairs (e.g. KK337). A straight is five cards that are in order (e.g. 45678), and suits are irrelevant. A flush is 5 cards of the same suit, and numbers are irrelevant. A full house is a hand with a three of a kind and a pair (e.g. AAA33). A straight flush is a hand that is both a straight and a flush. Finally, a royal flush is a hand consisting of the top five cards of a suit (e.g. AKQJ10♠).

1.2 Motivation and Goals

Casinos can make a profit with games like Video Poker in two different ways. The first method involves selecting the payouts for certain hands such that the game pays less for good hands. For example, Figure 1 says that the payout for a Full House is $9. If the casino wanted to make more money off of this game, they could always decrease this

1

payout to \$8. If no other changes are made, the casino would make more money in the long term for each dollar put into the machine.

Alternatively, the casino could modify the chances of getting particular hands, by changing the way machine draws cards. For instance, the house could decrease the probability that a player gets a 4 of a kind. Say a player draws 3 Jacks. The casino could simply decrease the probability that the player draws a forth Jack, thus ensuring that the player gets a 4 of a kind less often. This would mean that the game is no longer fair; that is, the probability of drawing any one card is no longer equal to the probability of drawing any other card, as is true with a fair deck. We will refer broadly to this method of increasing profit as **Rigging the Cards**. It is illegal for a casino to modify the game in this way [3]. A player who walks into a casino should be able to assume that the probabilities associated with video poker are the same as those associated with regular poker. If these changes are subtle enough, the player would never notice the unfairness in the game; and would continue to play the game not knowing that the casino was hiding the true probabilities behind the game in order to make additional profit. The aim of this paper is to develop a statistical test that would discover this type of modification.

1.3   General Procedure

The procedure to build such a test is theoretically straight forward. We will first determine what is called an **Optimal Strategy** for Jacks or Better. This is a strategy for the game such that the player's winnings are maximized. Once this is completed, we will know the theoretical probabilities of getting each hand as an outcome. If we then play enough games on a questionably fair machine, we should be able to tell from the results whether or not the outcome differs significantly from the theoretically predicted outcome. If it does, we will be able to conclude that the machine is unfair. The remainder of this paper gives a detailed procedure for the development of such a test. It will begin with an introduction to much of the terminology and concepts that will be used when describing the test.

2   Probability Theory Background

2.1   Probability Basics

The following is an introduction to the probability theory that will be used throughout the paper, mostly drawn from [2]. We first define an event:

Definition 2.1. An Event *is an outcome or a set of outcomes of an experiment for which a probability can be assigned.*

Definition 2.2. *The* **Sample Space** *of an experiment is the set of all possible outcomes of the experiment.*

For example, suppose that one die is rolled in an experiment. The sample space S is the set of all possible outcomes of the roll, that is $S = \{1, 2, 3, 4, 5, 6\}$. An event is any subset of the sample space, such as a 3 being rolled or an odd number being rolled. $P(A)$ refers to the probability that event A occurs. Likewise, $P(AB)$ refers to the probability that both events A and B occur.

We say that events A and B are **Independent** if and only if the outcome of A has no effect on the outcome of B, and if the converse is also true.

Theorem 2.1. *Two events A and B are independent if and only if $P(AB) = P(A)P(B)$.*

For example, let S denote the set of all possible outcomes when two dice are rolled. Let A be the event that an even number is rolled on the first die, and let B be the event that a 5 is rolled on the *second* die. Note that the roll of the first die has no impact on the roll of the second die. We conclude that events A and B are independent. Thus,

$$P(AB) = P(A)P(B) = \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) = \frac{1}{12}.$$

For contrast, let A be the event that an even number is rolled on the first die, and let B be the event that a 5 is also rolled on the *first* die. Since 5 is not an even number, these two events cannot occur at the same time. The outcome of either event has an impact on the other, thus the two events are said to be **dependent**, that is they are not independent. In this instance, the events are **mutually exclusive**, but this is not always the case when two events are dependent.

2.2 Results commonly used in this paper

It is far simpler to calculate the probability of independent events than dependent events. The following theorem gives a formula to calculate the probabilities of a **Compound Event**, which is just an event that only occurs if several other smaller events also occur. Note that this is a generalization of Theorem 2.1.

Theorem 2.2. The Multiplication Rule for Independent Events *Suppose that event E is the event that events $\{A_1, A_2, \ldots, A_n\}$ all occur. Furthermore, assume that the events $\{A_1, A_2, \ldots, A_n\}$ are independent. Then*

$$P(E) = P(A_1)P(A_2)\cdots P(A_n).$$

This theorem should make intuitive sense. If we draw two cards from a deck **with replacement**, then the probability of getting an ace will not change between draws. Thus, the probability of drawing 3 aces in a row will just be the probability of drawing an ace once cubed.

There are other conditions that make probabilities very easy to calculate. The following theorem is used frequently when discussing probabilities in gambling.

Theorem 2.3. *Suppose that for some probabilistic experiment, all outcomes are equally likely. Then for some event, E,*

$$P(E) = \frac{\#\ of\ experiment\ outcomes\ where\ E\ occurs}{\#\ of\ experiment\ outcomes\ in\ the\ sample\ space\ S}$$

It is best to illustrate this theorem with an example. Suppose that 5 cards are drawn **without replacement** out of a standard deck of cards. Note that the probability of drawing any particular hand of 5 cards is the same as the probability of drawing any other hand. Recall that a **Royal Flush** is a poker hand that consists of the highest 5 cards in any one suit. Thus a royal flush in diamonds consists of the ace, king, queen, jack, and 10 of diamonds (abbreviated $AKQJ10\diamondsuit$). We want to calculate $P(\text{Royal Flush is drawn})$. Note that there are 4 such combinations of cards that make up a royal flush. Since each poker hand is equally likely to be drawn, we can apply Theorem 2.3 to derive the probability of drawing a royal flush:

$$P(\text{Royal Flush is drawn}) = \frac{4}{\text{Total number of possible 5-card hands}}.$$

The denominator of this fraction is easily calculated using **binomial coefficients** [4]. This concept lets us determine exactly how many ways there are to choose 5 cards out of the 52. Thus,

$$P(\text{Royal Flush is drawn}) = \frac{4}{\binom{52}{5}} \simeq .00000154$$

Recall that this entire calculation was made possible by the fact that every 5-card poker hand is equally likely to be drawn. This theorem will be essential as we examine probabilities involving a deck of cards.

2.3 Random Variables and Expected Value

We will now introduce some terminology and concepts that will come in handy when building an optimal strategy for jacks or better.

Definition 2.3. *A **Random Variable** is a variable that measures some quantifiable aspect of an experiment. Random Variables are typically denoted by a capital letter.*

For instance, suppose that our probabilistic experiment is drawing a card from a standard deck. Then we could define a random variable X to be the face value (1 - 13) of the card that is drawn. Some experiments can have multiple random variables associated with them. We could define a second random variable Y to be the suit of the card that is pulled. There are standard ways to refer to the probabilities associated with random variables. For instance, we would say that

$$P(Y = \text{Heart}) = \frac{1}{4} \qquad\qquad P(X = \text{King}) = \frac{1}{13}$$

3

since a fourth of the cards in the deck are hearts, and a thirteenth of the cards in a deck are kings. When we discuss random variables, it is important to note the probability that the random variable will equal each one of its possible values. This prompts the following definition, which prescribes a way that we can formalize the probabilities associated with random variables.

Definition 2.4. *Let x be any possible value for the random variable X. Then the* **Probability Mass Function** *(PMF) of X is defined as some function $f(x)$ such that*

$$f(x) = P(X = x)$$

for all possible values x, of X.

For example, let X be a random variable representing the number of the card drawn from a standard deck of cards. Then

$$f(x) = \begin{cases} \frac{1}{13}, & \text{x = (Ace, 2, 3, \ldots, 10, Jack, Queen, King)} \\ 0, & \text{otherwise} \end{cases}$$

Since X has to take on one of the values in the domain of the its PMF, we note that

$$\sum_{x: f(x) > 0} f(x) = 1$$

This is a simple consequence of the fact that for any given instance of an experiment, a random variable can equal one and only one of its potential values.

We will now discuss the concept of **Expected Value** associated with random variables. This is informally defined as a measure of the average value we can expect a random variable to take on. Thus, expected values are essential to predicting the value of random variables, and thus the outcome of experiments.

Definition 2.5. *The* **Expected Value** *of a random variable X, denoted by $E(X)$ is defined as follows:*

$$E(X) = \sum_{x: f(x) > 0} x f(x)$$

Suppose (again) that X is a random variable equal to the number on a card drawn from a standard deck of cards. If we assign the values 11, 12, and 13 to jacks, queens, and kings, respectively, then

$$E(X) = \sum_{x=1}^{13} x f(x) = \sum_{x=1}^{13} x \frac{1}{13} = \frac{1}{13} \sum_{x=1}^{13} x = 7$$

On average, then, we expect to draw a 7. The expected value for a random variable X does not have to be in the domain of X. Rather, the expected value is a reflection of what we think will happen in the long run of an experiment.

2.4   House Advantage and Optimal Strategy

The concepts in the above section allow us to calculate the **House Advantage** and **Optimal Strategy** for video poker. This will be essential for the development of a statistical test to detect cheating.

Definition 2.6. ***House Advantage*** *for video poker is defined as the expected long-run house profit expressed as a percentage of a player's original bet. For example, if the house advantage in video poker is known to be 5%, then the house expects to make $.05 of every $1 that players spend on blackjack. [5]*

Of course, the house advantage depends on the choices made by the player. Recall that the player is allowed to choose which cards to keep before the remaining cards are discarded and re-drawn from the deck. Thus, the house can expect to make significantly more money from players who make poor choices. To maximize the quality of our test, we will assume that a player is playing according to the optimal strategy of video poker. This assumption will allow us to detect more minute degrees of cheating on the part of the house.

Definition 2.7. *For each video poker hand $h \in H$, there is some set of choices $C = \{c_1, c_2, \ldots, c_{32}\}$ that a player can make to keep cards. Let X_{c_i} be a random variable equal to the possible payouts received for hand h, given that a player made choice c_i. A player is playing* **optimal strategy** *if for each hand h, they choose c_i such that $E[X_{c_i}] \geq E[X_{c_j}]$ for all $c_j \in C$.*

To avoid any confusion, a player is playing optimal strategy if they are making choices such that they maximize the amount of money they make on the game, which will typically be a negative number. To find an optimal strategy for video poker, we have to derive the best choice that could be made with any given hand. For some hands, this is simple and relatively intuitive. However, the game is complicated enough that few if any people could play optimal strategy without the use of extremely detailed strategy tables. For example, suppose that a person is dealt J77xx (a Jack, two 7s, and two 'irrelevant' cards). A perfect strategy for video poker should tell us whether to keep the Jack, or the two 7s (recall that a pair only gives a payout if it is Jacks or higher). Let X be the possible payouts when we keep the pair of 7s, and let Y be the possible payouts when we keep the Jack. Then the greater of $E[X]$ and $E[Y]$ should dictate the choice that the player makes when selecting cards to keep. To calculate these expected values, we need to consider every possible hand that could result from keeping the Jack or the pair of 7s. For both choices, this includes a 3 of a kind, a 4 of a kind, and a Full House. If we keep the Jack, a Royal Flush, pair of Jacks or higher, and a Straight Flush are added to the list. Thus, this calculation quickly becomes daunting.

Practically, these calculations had to be performed using a computer program. The program first generated every possible initial hand for jacks or better. It then tried keeping every possible combination of cards in the hand that could be kept. For each choice of kept cards, it then checked all of the possible draws that could result, and then recorded the outcome. In this fashion, it was able to calculate the average winnings resulting from each set of kept cards. The program then reported back that we should keep whichever combination of cards resulted in the greatest average winnings.

3 The Test

3.1 Introduction to the Statistical Test

The theory behind the development of the test is fairly straightforward. Suppose that the fair probability of getting a 4 of a kind in video poker is $1/100$. Now, suppose that the casino has rigged the cards such that the probability of getting a 4 of a kind is $1/150$. Such a change would be difficult to detect by a player enjoying the game for fun. A player would need to be well grounded in the probabilities associated with the game (which is surprisingly rare...), and they would also need to play many games before they could have any chance of noticing the difference.

A computer, on the other hand, can play games very quickly and keep precise track of how often various hands appear as outcomes. Note that the calculations below depend on the computer playing optimal strategy. Suppose that a computer plays 3000 games of video poker on the rigged machine. Let N_r be a random variable equal to the number of 4 of a kinds the computer gets in 3000 games on the rigged machine. Since the rigged machine gives a 4 of a kind with probability $1/150$, we find that

$$E[N_r] = 3000 \left(0 \left(\frac{149}{150} \right) + 1 \left(\frac{1}{150} \right) \right) = 20$$

Now we will contrast this expected value with that of a fair machine. Let N_f be a random variable equal to the number of 4 of a kinds the computer gets in 3000 games on a fair machine. Since the probability of getting a 4 of a kind on a fair machine is $1/100$, we calculate the expected value as follows:

$$E[N_f] = 3000 \left(0 \left(\frac{99}{100} \right) + 1 \left(\frac{1}{100} \right) \right) = 30$$

Intuitively, the difference between these two numbers seems significant. If a player played 3000 games in a casino with perfect strategy, and only got 20 4 of a kinds, they would have reason to be suspicious of the machine. However, it is *possible* to only get 20 four of a kinds on a fair machine. Because the dealing of the cards is a random process, it is technically *possible* to get no 4 of a kinds in 3000 hands, albeit highly unlikely. The job of a good statistical test is to quantify how unlikely each of these outcomes is on a fair machine. If we discover that the probability of a given outcome is highly unlikely under the assumption that the machine is playing fairly, then we would have significant evidence that the machine is rigged.

5

It is fairly easy to find the probability discussed in the above paragraph, however it will require the introduction of some statistical terminology. Suppose that we expect to get 30 4 of a kinds on a fair machine in 3000 trials, but only get 20 on the questionable machine. In this example, the probability of getting a 4 of a kind on a fair machine is 1/100. This is a textbook application of a binomial random variable:

Definition 3.1. *Suppose there is some probability experiment where event E occurs with probability p. Thus, event E^c occurs with probability $1 - p$. Additionally, suppose that the experiment is conducted n times, where each trial is independent of the others. Then the* **Binomial Random Variable** X *represents the number of times that event E occurs, and follows the distribution below.*

$$P(X = x) = \binom{n}{x} (p^x)(1 - p)^{(n-x)}$$

In the case of our example, $n = 3000$ and $p = 1/100$. The **Null Hypothesis**, denoted H_0 refers to what should happen on a fair machine. The **Alternative Hypothesis**, denoted H_a states the hypothesis of the test that we are conducting. Our alternative hypothesis suggests that the probability of getting a 4 of a kind on the test machine is less than the theoretical probability. Stated symbolically,

$$H_0 : p = \frac{1}{100} \qquad\qquad H_a : p < \frac{1}{100}$$

Our statistical test will attempt to reject the null hypothesis in favor of the alternative hypothesis. Our P-**Value**, P, will give the probability of getting 20 or fewer 4 of a kinds on a fair machine. We will say that if $P < .05$, then we can reject the null hypothesis and conclude that the probability of getting a 4 of a kind on the machine is indeed less than 1/100.
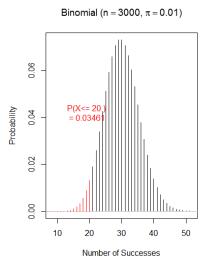

Figure 2: The plot of a Binomial Distribution testing the probability of getting 20 or fewer 4 of a kinds.

The above figure is a graph of the binomial distribution for our experiment. The lines shaded in red represent the section of the graph where the player got 20 or fewer 4 of a kinds. The p-value is indicated in red, and it is indeed less than .05. Thus, for this example, we can conclude that the probability of getting a 4 of a kind on the machine in question is indeed less than 1/100. This leads us to suspect that the house has indeed 'rigged the cards' on the machine.

3.2 Formalized Statistical Test

Unlike in the above example, our goal is to develop a test that will analyze all of the hand probabilities simultaneously. That is, it will compare the theoretical probability of a 3 of a kind with the observed probability, while doing the

same for each other potential outcome listed in Figure 1. We will use a statistical metric called the **Chi-Squared Test Statistic** to accomplish this more holistic test. Suppose that E_i is the number of times that we expect to see a particular poker hand out of n trials, each played with optimal strategy. For instance, if $n = 10000$, then E_1 might be the number of royal flushes we expect to see in 10000 trials. Let O_i be the number of times that we actually observe a particular hand in n trials. Using the same example from above, O_1 might be the number of royal flushes that we actually observe in 10000 trials. Thus, we can now define a random variable X^2:

$$X^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i}$$

The 10 in the upper bound of summation corresponds to the 10 different outcomes that are possible in video poker. This random variable is a holistic measure of how far observed quantities are from expected quantities. For example, if we expect to see 20 royal flushes and get 19, the 'royal flush component' of the sum will be equal to $\frac{1}{20}$. However, if we expect 20 royal flushes and get 10, then the component will be equal to 5. Thus, a large X^2 value for an experiment means that the number of observed results differed from the number of expected results significantly. The next logical question is to ask how large X^2 has to be in an experiment for us to conclude that a machine is not generating the cards in a truly random manner.

To answer this question, we will first build a pool of X^2 values generated from trials on a machine that we know is random. That is, if we use a sample size of 1000, we would run 1000 trials, say 1000 times, and generate the X^2 value for each set of 1000 trials. This set of experiments could be described as a 'control group'; it is a set of trials that we can compare our experiments to. After performing trials on an experimental machine, we will compare the resulting X^2 value to this database of fair X^2 values, and use that comparison to determine relatively how large the experimental X^2 value is. This database will be referred to as a **Null Distribution**. If an experimental X^2 value is larger than 95% of the values in the null distribution, we will conclude that the machine in question is unfair. To use the statistical terminology introduced above, the null hypothesis will be that the probability of getting each hand on the experimental machine is equal to the theoretical probability of getting each hand. The alternative hypothesis will be that this is not the case. For sufficiently large X^2 values, we will reject the null hypothesis.

A graph of such a null distribution is included in Figure 3. As mentioned above, we will conclude that a machine is not fair if the test on the machine gives a chi-squared test statistic greater than 95% of the values in this distribution. For the graph in Figure 3, this corresponds to chi-square values greater than 28.60. In this context, the p-value for an experiment reports what percentile of this distribution a chi-square statistic is in. For example, if an experiment gives a p-value of .75, then the corresponding chi-square statistic is greater than 75% of the numbers in the distribution.

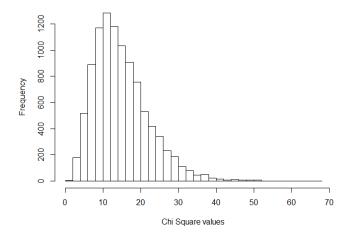**Null Distribution of Chi Square Values for 10000 Sample Size**

Figure 3: A histogram of 10,000 Chi Square Test Statistics generated by a fair machine. The sample size for each statistic was also 10,000.

3.3 Testing the Test

Unfortunately, casinos are not overly eager to have this test tried out on their video poker machines. Even if they were willing to cooperate, it would be practically difficult to carry this test out manually. This method would more likely be used by a regulatory agency with the proper authority and equipment. However, this does not prevent us from evaluating the quality of the test; we can still provide a theoretical proof of concept. Essentially, we can modify the random number generator in a program to simulate a rigged machine. We can then run the test on the 'rigged machine', and see if the test picks up any changes that were made to the random number generator. Using such a method, it will be easy to determine how rigged a machine has to be in order for the test to tell that the machine is unfair.

To carry out such an evaluation, we must first determine how to simulate a rigged machine. More specifically, we need a set of parameters to change such that the house advantage in the game goes up. These changes have been implemented in an otherwise fair video poker game to determine the sensitivity of our test. For the simulations used below, we built a parameter which we called the **Adjusted Pair Probability** parameter. This is best illustrated with an example. We will denote our adjusted pair probability with A. Suppose that $A = .95$. Whenever a pair is drawn in the initial hand, the probability of drawing additional cards with the same number of the pair is multiplied by a coefficient of .95. Suppose we draw a pair of 2's. Then the probability of drawing another 2 will be 95% of what it would normally be. It is easy to see that this change will increase the house advantage of the game, as the player will get fewer three of a kinds, full houses, and four of a kinds.

It is important to note that for all of the simulations described below, results are considered <u>significant</u> if they lie in the 95th percentile of the null distribution. Consequently, the test will return a 'false positive' 95% of the time on a fair machine. This is typical amoung many statistical tests. For $A = .95$, we found that the change was detected 66% of the time with a sample size of $1,000,000$. The change was detected 100% of the time with a sample size of $10,000,000$. It is interesting to note that this degree of cheating would increase house profits by about $9,650 over the course of $1,000,000$ games. For $A = .99$, we found that the change was detected 32% of the time with a sample size of $100,000,000$. This degree of cheating would make the house approximately $2,320 over the course of $1,000,000$ games. Analyzing the profit made by the cheat is interesting, as it introduces some 'risk and reward' analysis into the problem. It would be interesting to investigate how much the house could reliably make off of a cheat with very little probability of the cheat being detected.

4 Conclusion

This test, or a similar one, would allow regulators to automate a large portion of the video poker testing process. Rather than manually examining every machine (or a sample of machines), the agency could apply a test such as the one presented above to every machine, and quickly gather some preliminary results. After these results had been examined, they could then look at the code on any machines that they suspect could be cheating based off of the test results. Such a process could cut down on costs for the agency as well as ensuring that more machines are examined.

There are a few possible areas for future exploration on this subject. First and foremost, more simulations should be run on different types of rigged machines, so that the test's effectiveness can be further evaluated. It would also be helpful to try the test out on a real video poker machine. Secondly, more statistical metrics for measuring the fairness of machines could be developed. For instance, it would be interesting to run a test based only on the payoff received in some number of trials. I suspect that this may be very effective, and also much easier to understand and implement.

References

[1] http://www.videopokerballer.com/tools/videopoker-com/mid-play-jacks-or-better.jpg

[2] Ross, Sheldon M. A First Course in Probability. Upper Saddle River, NJ: Prentice Hall, 2010. Print.

[3] http://www.gambling-law-us.com/State-Laws/Nevada/

[4] http://mathworld.wolfram.com/BinomialCoefficient.html

[5] http://definitions.uslegal.com/h/house-advantage-gaming-law/