

The Odds Calculators: Partial simulations vs. compact formulas

By Catalin Barboianu

As result of the expanded interest in gambling in past decades, specific math tools are being promulgated to support players' strategies or at least to provide relevant information about their chances of winning.

For example, so-called *odds calculators* have spread all over the Internet, whether as interactive software programs or just as *odds charts and tables*.

Whether the game is Poker, Blackjack, Roulette or Slots, these programs have dedicated sections on most gambling websites, on the pretense that they offer the final winning odds or the odds of reaching a certain game configuration at a certain moment of the game.

In reality, these sites do not deliver what they promise, namely odds and probabilities, as we shall see further. In fact, these programs have nothing to do with the applications of probability theory in gambling.

Most of these odds calculators are based on partial computer simulations. For convenience, we will abbreviate them as OCBPSs (odds calculators based on partial simulations).

How a partial simulation works

Each program reproduces on a computer the game under discussion (usually, a casino game) by using in its code all parameters that describe its rules, initial conditions, and progress.

Both external and internal users can simulate the game through one of these programs and some of the data from every simulation are recorded.

Specific gaming situations and certain events related to these situations are then tracked. For example, in a Blackjack simulation, the situation *one player holds (5J)* and the event *the player busts on the next hit* are tracked.

A certain gaming situation that requires the display of some statistical data is tracked and recorded each time it is encountered during the simulations. For the respective situation, the occurrence of the proposed event is also recorded in binary mode (1 if the event occurred, and 0 if not).

The OCBPS database is updated with every simulation, and what the program returns after each simulation is the ratio between the number of occurrences of the tracked event and the entire number of simulations that match the respective gaming situation until that moment.

Schematically, for a given gaming situation S and a related event A , if n is the number of simulations matching the situation S and m is the number of occurrences of event A during the n simulations, then the numerical return is m/n and is presented as "the odds of event A in situation S ".

For example, if we refer to the starting hands in Hold'em and the goal is the display of the final winning odds for the hand *(8 Q) suited* (this is the situation S), an OCBPS would perform complete simulations of the game of Hold'em, track the situation *(8Q) suited in one player's hand*, and record every match for it.

For every matched situation, the OCBPS would check and record whether that hand won the game (this is the event A).

If the simulations made 300 matches for situation *(8 Q) suited* as the starting hand and that hand won 60 times, the program would display the number $60/300 = 1/5 = 20\%$ as the winning chances of the hand *(8 Q) suited*.

Of course, after the 301-th matching simulation, the displayed percentage would be different even if that difference might not be so high.

What is a partial simulation from probability point of view

Now let us talk about where mathematics frames the partial simulation as an experiment generating probability-measurable events.

If we refer to probability as limit of a sequence of relative frequencies, as the Law of Large Numbers states, then an OCBPS builds the sequence of independent experiments having outcomes statistically recorded through its simulations for the respective game.

The experiments that match the given gaming situation (S) form a subsequence of that first sequence of experiments.

For this subsequence (which is built term by term over time with every simulation), the values of the ratio m/n are recorded. These recorded values are in fact the sequence of relative frequencies of the occurrence of event A .

Coming back to the Law of Large Numbers, which says that the sequence of relative frequencies converges toward probability, we can roughly state that the sequence $(m/n)_n$ built by an OCBPS tends to $P(A)$ (probability of occurrence of A). You will see later why we say *roughly*.

Therefore, what an OCBPS returns is not the probability of A (chances), but merely the terms of the sequence of relative frequencies of occurrences of A , in increasing rank.

The inputs and the returns of an OCBPS are what statisticians call a *practical statistic*, which are also a *partial statistic*, and not probabilities (odds) as they try to lead users to believe.

The difference is similar to the one between a sequence's term and sequence's limit. This difference can be at any value. Of course, as the rank increases, the difference diminishes.

This is true for every OCBPS, no matter what game is being simulated.

What, then, are the contradictions in an OCBPS as concerns its internal structure and its usage?

– The main contradiction is a theoretical one: the program is not an odds calculator but rather a statistics displayer. If an OCBPS says that a displayed percent of $p\%$ represents the “winning odds”, in reality this means not even “this happens in $p\%$ from these situations”, but rather “this happened in $p\%$ of these situations until this moment”.

Moreover, the contradiction extends to the difference between the terms used and the function: the returned numerical value changes with each simulation, while the odds (probability) is a fixed value for a given event within a given probability space.

– Another practical contradiction is in the numerical difference between the real probability of the event and the partial statistic (the relative frequency).

– It is well known that, as the number of experiments grows, relative frequency approximates the probability (as the limit of the sequence) with higher accuracy, provided that the experiments are performed under identical conditions. In practice, this condition reverts to the fact that the simulations are totally random.

The producers of OCBPS are surely aware of this mathematical truth and thus try to base their software program on as many simulations as possible.

The statistics OCBPSs use have external entries (from end users), but may also use an initial own-generated base of simulations. This prebuilt database is part of the program at its launch to external users and its role is to give some minimal relevance to the first returns of the software.

We can state with certainty that the external entries are totally random. The internal entries have a high degree of randomness, but this randomness is not total.

However, neither hazard nor randomness has been rigorously mathematized yet.

Besides philosophy, mathematics did not succeed in providing at least a rigorous definition; moreover, it did not create a solid model for randomness and hazard.

Emile Borel stated that, unlike other objects from the surrounding reality for which the creation of models assumes an idealization that preserves their properties, this idealization is not possible in the case of hazard.

In particular, whatever the definition of a sequence formed by the symbols 0 and 1 is, this sequence will never have all the properties of a sequence created *at random*, except if it is experimentally obtained (for example, by tossing a coin in succession and putting down 0 if the coin shows heads and 1 if the coin shows tails).

Borel also proposes an inductive demonstration scheme for this affirmation regarding the *random sequence*.

Assume someone is building such an indefinite sequence 0010111001..., which has all properties of the sequences generated by experimental randomness.

Assume the first n terms of the sequence were built and follows to write the $n + 1$ term.

There are two options: the first n results are somehow taken into account or they are not.

The first option annuls the random character of the construction, because a precise rule for choosing that term exists.

The second option brings us to the same situation we would stand in at the beginning of the sequence's construction: How do we choose one of the symbols 0 or 1 without taking into account any difference between these?

This choice would be equivalent to a draw, which assumes an experimental intervention.

Using the reduction to the absurd method, we come to the conclusion that such a random sequence cannot be built.

Far from being totally rigorous with respect to a solid mathematical model, this proof underlines the theoretical difficulties of conception with regard to this subject. But even in this primary form, it tells us that a total randomness cannot be reproduced. (You can read more on this subject in the section titled *Relativity of probability* in the book *Understanding and Calculating the Odds*.)

The demonstration of Borel can be extended to more complex "random" sequences, like those that are software generated. These can be sequences of numbers, matrixes, card distributions, and the like.

The conclusion is that any random generating software cannot create a totally random sequence because of the experimental action.

Coming back to OCBPSs, in the light of these facts, we can state that any sequence of simulations pregenerated by an OCBPS does not submit to the conditions of the Law of Large Numbers: the experiments are independent but they are not performed under identical conditions.

As to the total number of simulations, to reach a good approximation of the real probability, this number must be huge; and most public odds calculators do not obey this condition, thus the number is otherwise unverifiable.

Even if a software program says that its returns are based on 1 million simulations, this figure is irrelevant because only a small portion of those simulations would match a given situation S .

Assume this fact for the example presented earlier (in a Hold'em game, the situation with (8Q) suited as the starting hand) and let us approximate the real number of experiments that generate the displayed relative frequency.

To do this, assume an average of five players in a game. The probability of one player being dealt (8Q) suited is $4/C(52,2) = 0.301\%$. By using the inclusion-exclusion principle, we find about 1.5% as the probability of at least one player from among the five being dealt (8Q) suited. This means we should expect at least one occurrence of (8Q) suited every 150 games. Therefore, from those supposed 1 million simulations, only about 6667 simulations match this situation. This is a very small number of experiments on which to base a good approximation of probability through relative frequency.

In addition, if most of those 1 million simulations were automatically pregenerated, the approximations might be altered by the non-random character of the generating program.

Still, there are OCBPSs that return very good approximations of the real probability for certain gaming situations.

These programs work best for games with very large audiences, like Texas Hold'em and Blackjack. Owing to the high number of external users, such calculators have already cached a huge number of simulations, which helps ensure a satisfactory approximation of the real odds.

Still, in these programs this does not happen for every gaming situation or every event within a respective game. More often, they are confined to those frequently inputted by the users, usually for configurations from the first part of the game.

In fact, what an OCBPS substitutes through these partial simulations is a complete probability calculus. For some games, this calculus is hard and complex, but for others it is relatively easy.

If we refer to concrete gaming situations, the probability calculus for the attached events can be performed by anyone with a high-school mathematics background, even if some applications require a lot of time. For example, in a one-deck Blackjack game, calculate the odds of being hit with a card with a value less than 5, if you hold (A6) and the other seen cards on the table are 3, 8, 10 and J.

However, if we refer to categories of situations and larger events, it is necessary to create a rigorous mathematical model for each respective application, to identify the variables that completely describe the gaming situation and the event to measure, and to work out the compact formulas that give the probability of that event as a function of that variable. This level of complexity is a mathematician's job. For example, in Draw Poker, calculate the probability of at least one of your opponents holding two pair, as a function of your own seen cards – initial hand plus replacements – and the number of your opponents.

What does a compact formula mean

A compact formula is a real function given by a unique algebraic expression that holds one or more variables. Thus, a compact formula is not given in a recurring or iterative mode, it is merely a block expression given in an explicit mode.

Any probability calculus involving categories of situations aims to find a compact formula that returns the probability of any event after inputting the parameters that describe the concrete given situation.

Thus, to use a compact probability formula means compressing a very large number of concrete situations by extracting from them only the parameters that represent the formula's variables.

For example, if we want to calculate the probability of one opponent holding a three of a kind formation in a Draw Poker game after the first card distribution, we must quantify the five seen cards from your own hand. Three of a kind is a formation made of values, so the variables of the probability formula are the distributions of values. (Read more about the quantification of the information to use in Draw Poker in the section titled *The Supporting Mathematics* in the book *Draw Poker Odds*.)

There are only six possible distributions of values for the initial five cards:

4-1-0-0-0-0-0-0-0-0-0-0-0-0-0, 3-2-0-0-0-0-0-0-0-0-0-0-0-0-0, 3-1-1-0-0-0-0-0-0-0-0-0-0-0-0,
2-2-1-0-0-0-0-0-0-0-0-0-0-0-0, 2-1-1-1-0-0-0-0-0-0-0-0-0-0-0 and 1-1-1-1-1-0-0-0-0-0-0-0-0-0-0.

The compact formula that returns the searched probability looks like this:

$$P(A) = \frac{1}{12C_{47}^5} \sum_{i=1}^{13} (2 - c_i)(3 - c_i)(4 - c_i)[(42 + c_i)(43 + c_i) - 2S + (3 - c_i)(4 - c_i)],$$

where $S = \frac{1}{2} \sum_{j=1}^{13} (3 - c_j)(4 - c_j)$.

The variables are the 13-size vectors $(c_1, c_2, \dots, c_{13})$, which are the distributions of the values within the seen five cards.

While any card distribution (or situation) can be represented by one of these six vectors, by using this formula the complete number of situations is reduced to 6.

Now imagine how an OCBPS would work this kind of problem:

Assume that we search for the situation in which you hold two pair as your initial hand and want to find the odds for one opponent holding three of a kind at that moment. From the total number of possible initial five-card hands of $C_{52}^5 = 2598960$, only $C_{13}^2 C_4^2 C_4^2 (52 - 8) = 123552$ matches two pair. So, in practice about 1 out of 21 simulations matches the desired situation. For each, the event of one opponent holding three of a kind is recorded as occurred (or not occurred) and then the relative frequency is calculated.

To obtain 10000 simulations of the tracked situation, more than 210000 random simulations are needed. And even this many may not be enough.

Now imagine a simple software calculator using this formula. Instead of dealing with 2100000 results, it would only deal with 6 (one for each possible distribution of values). And, what is most important, it would return the real probability and not a relative frequency.

Now let us take a simple example from Blackjack:

Assume you are the only player at the table in a one-deck game. You hold (J7) and the dealer's face-up card is 8. You want to evaluate the probability of not busting after two additional hits.

The calculus is very simple and can be done rapidly by anyone:

The two-card combinations worth less than 5 points in total are of type: (1A), (11), (AA), (21), (2A), (13), (3A) and (22) numbering $16 + 6 + 6 + 16 + 16 + 16 + 16 + 6 = 98$ card combinations, from $C(49,2) = 1176$. The probability is then $98/1176 = 8.33\%$.

For more complex situations (involving more players, more decks, more played cards), there is a simple compact formula with two variables that returns the probabilities of hitting a favorable card at any moment in a Blackjack game. (This formula is detailed in the *Blackjack* section of the book *Probability Guide to Gambling*.)

Because of its simplicity, this formula can replace any Blackjack odds calculator, whether implemented in a software calculator or used directly, so OCBPSs for Blackjack are completely unjustified. Still, they exist.

The situation from our example (J and 7 held as the first two cards, 8 as the dealer's card) appears about once in every 1035 initial simulations, so an OCBPS would need more than 10 million random simulations to acquire 10000 simulations matching that situation!

Is this collecting effort worth it, when the result is not even a probability but an approximation, and the real probability can be found through a simple arithmetic calculation? My answer is definitely no.

For Texas Hold'em poker, OCBPSs have collected millions of simulations, especially those dedicated to ranking the starting hands by their "probability" of winning a game. Owing to the explosion in popularity of this poker variation in recent years and the relatively small number of possible combinations for the two pocket cards (1326 as value and symbol and only 91 as value), the OCBPS database is complete enough to return good approximations of the real winning odds. This is why you will see about the same percentages for the same starting hand with different major OCBPSs over the Internet, including at those integrated into online poker room software.

But there are gaming situations in Hold'em where mathematics can more rapidly solve the odds problem than an OCBPS and where the OCBPS database is not large enough to return values close to the real odds. These situations are about hitting the various card formations by river, for your own hand and for your opponents' hands.

The probability of these events can be calculated through compact formulas. Moreover, the results can be combined to obtain the final probability of winning for a given starting hand. But the mathematical work for this goal is huge. One person could spend from a few months to a year solving this problem. A team of mathematicians could do it in a reasonable time, but a computer can help in this mathematical effort in other ways than by using partial simulations. But, as we said, this work is not of greater worth except for the sake of rigorousness, because the major OCBPSs have already reached good approximations of the winning odds for starting hands in Texas Hold'em.

Any gambling probability application can be solved analytically by using the results of probability theory in a finite sample space. Some of them are complicated and require a lot of time to solve completely, especially for card games. But, as we said, a team of mathematicians and a computer packed with mathematical software can solve any probability aspect of a game in a reasonable time.

Why developers choose the partial simulations

Obviously, it is far easier to write a code for a software program that performs statistical records of simulations than to solve all the probability formulas of a respective game and then implement them.

This latter option assumes the existence of a team of mathematicians to do all the math in a reasonable time, then of a qualified programmer to implement the formulas correctly in the program.

Therefore, an OCBPS can be developed by programmers, while a software program based on compact formulas requires one or more mathematicians.

Why OCBPSs are used on such a large scale

As we mentioned, there exist OCBPSs having returns that approximate real probabilities with high accuracy. These can be found among Hold'em and Blackjack odds calculators.

For the rest of OCBPSs, the conditions in which the simulations are performed and the entire number of statistical entries are questionable.

Still, most gamblers do not notice this difference and use any OCBPS with an attractive presentation.

It is a fact that any software program is used as a result of the marketing presentation made by its developer and what its interface shows. So no matter the program, the user never knows what is behind it because he or she has no access to its code. OCBPSs also submit to this rule.

In addition, some OCBPSs do not say that they use only partial simulations, so their users do not know what kind of "odds" they display and how are they obtained.

Many OCBPSs say they use partial simulations but present their returns as odds and probabilities, which often misleads average users who have no probability background.

Besides these aspects related to the poor quality of information, there is another factor at play here, and that is the general psychological behavior of humans to think of practical statistics as degree of belief instead of actual probability when making decisions.

What is, in fact, the motivation behind such general behavior of appealing to statistics? Is there a rational motivation or it is only a matter of human biological structures?

The answer is somewhere at the middle and can be explained in large part by psychology.

All along, man felt safe as an individual only when he was grounded on something sure and perceptible. The human mind also submits to this principle.

Practical statistics is a collection of *sure* results; namely, frequencies of events that *already happened*, and these happenings are a certainty.

Unlike statistics, the prediction, by estimating a degree of belief, refers to events that have *not happened* yet and their occurrence is an uncertainty, so the human mind classifies them in the category of *unsafe* and tends to increase somehow their sureness by transferring sure things (the statistical results) upon them.

Although humans act in the real world in uncertain conditions, in an unsure environment, the human mind perceives this thing as an anomaly and tries to ameliorate it by elaborating degrees of belief that come from sure environments, such as practical statistics. (Read more on this subject in the section titled *Psychology of probability* in the book *Understanding and Calculating the Odds*).

Applying this to a gambler's behavior, the consequence is that players use OCBPSs even when they are aware of how they work, just to base on past statistics – which they consider to be “sure” information – instead of predictions for the future.

A last motivation for using OCBPSs is a lack of any substitute. Players use such programs because there are still not enough odds calculators based on compact formulas on the market.

Beside software programs, several books on the mathematics of gambling have been published that cover the basic probability aspects of gaming strategy, but their approach to this subject is still preponderantly statistical. In addition, a software program is easier to use from a practical point of view.

Regarding the books in this field, it is sad that so many consecrated authors chose to use the returns of OCBPSs in their probability sections instead of collaborating with a mathematician.

When we talk about the role of probability as prediction within a strategy, a certain mathematical rigorousness is expected. And ignoring the mathematical probability for a partial statistic does not mean rigorousness.

Infarom Publishing developed in 2006 and launched in 2007 the first odds calculator entirely based on compact formulas. Called the *Draw Poker Odds Calculator*, it was built as result of solving all the probability formulas for this variation of poker.

We hope this example is followed by other developers because there is still a gap in the market for professional (mathematical) odds calculators. These real odds calculators would enable more players to refer to real odds and not to past statistics in their strategies.

Technically, I have nothing against OCBPSs. There is nothing wrong in their way of functioning and their returns are correct. What I contest is the way in which returned information is presented, which misleads users, and some of the descriptions made by their developers.

As I mentioned earlier, some OCBPSs are useful, at least because they replace some of the unsolved formulas thus far.

The purpose of this article is to make a comparison from both a mathematical and a practical point of view between the returns of partial simulations and the returns of the compact probability formulas. Many times, the inadvertence is only conceptual and hints of rigorousness. The target practical aspect is a good approximation of the real probability and is accomplished in many cases.

However, if we want to use mathematics in gaming strategies or just to provide pure information, we must preserve its character of rigorousness. In fact, games of chance stand as a perfect base for the application of probability theory, which thus far is the only consistent theory that models the hazard with the goal of providing a mathematical measure of the uncertainty. While this measure is the only objective one we have, we must not distort it in any way.

Catalin Barboianu is an applied mathematician and author of the books *Understanding and Calculating the Odds*, *Probability Guide to Gambling*, *Texas Hold'em Odds* and *Draw Poker Odds*. He maintains a probability website with gambling applications at <http://probability.infarom.ro> .