A TOOL FOR THE DIRECT ASSESSMENT OF POKER DECISIONS

Darse Billings[1]
and Morgan Kan[2]

University of Alberta, Canada

ABSTRACT

The element of luck permeates every aspect of the game of poker. Random stochastic outcomes introduce a large amount of noise that can make it very difficult to distinguish a good player from a bad one, much less trying to quantify small differences in skill.

Good methods for directly assessing the quality of decisions exist for other stochastic games, including backgammon and blackjack. However, those are perfect information games, where the notion of an objectively best move is well-defined, which unfortunately is not the case for imperfect information games, in general.

The Ignorant Value Assessment Tool, DIVAT, uses perfect knowledge (hindsight) analysis to quantify the value of each player decision made in a game of two-player Limit Texas Hold'em. Comparisons are made against a realistic baseline betting sequence, which is based on quasi-equilibrium policies and game-theoretic invariant frequencies for raising and folding. Much of the relevant context involved in each decision is simply ignored, out of necessity; but enough is retained to provide a reasonably accurate yardstick, which is then applied equally to all players. The frequency and magnitude of expected value differences between player actions, relative to the baseline, provides a low-variance estimate of the long-term expected outcome between the players.

1. INTRODUCTION

The game of poker has many properties that present new challenges for Artificial Intelligence research, making it distinct from all previously studied games. Traditional games like chess and Go are two-player perfect information games with no element of random chance. Poker is at the opposite end of the spectrum: a multi-player imperfect information game with partial observability and stochastic outcomes. Deception, opponent modeling, and coping with uncertainty are indispensable elements of high-level strategic play.

The property of stochasticity compounds many of the complex problems in the domain. It also has a highly detrimental effect on performance assessment: on quantifying how well a player is playing, or simply determining who the better players are. The "signal to noise ratio" is low – a player can play extremely well and still lose over an extended period of time, just by bad luck. The normal variance in the game is so high that many thousands of games need to be played before obtaining even a modest degree of confidence in the result. Without reliable feedback on how effective a solution is, it is difficult to identify problem areas and make steady forward progress. Moreover, the problem will become more serious over time. As programs become stronger, and closer to each other in skill level, the the number of games, n, required to distinguish between them with a particular statistical confidence grows at a rate of $\Theta(n^2)$.

Some techniques can help ameliorate the problem of high variance. For example, *duplicate tournament systems* use independent tournaments with the same series of cards, shuffling the players to different positions for each replay, as described in Billings (1995). Since games between computer programs can be re-played with no memory, the total number of good and bad situations is equalized, at least to some degree. However, once the

[1] email: darse@cs.ualberta.ca
[2] email: mkan@cs.ualberta.ca

players deviate in their corresponding actions (particularly if one player folds while another continues with the hand), the subsequent actions are no longer directly comparable, and are subject to random outcomes. The problem of accurate assessment persists, even if other sources of noise are kept under control. Furthermore, duplicate systems cannot be used for assessing an individual human player, since they will recognize previously played hands and outcomes. Developing a tool to perform an objective and insightful analysis of the decisions made by each player would be of great value, because the effects of randomness could largely be factored out. Unfortunately, this turns out to be difficult to construct, both in theory and in practice, due to the implications of imperfect information in poker.

It is useful to compare poker to perfect information games that have a stochastic element. Two well-studied examples are backgammon and blackjack. In backgammon, both players have full access to the state of the game, but randomness is introduced by the roll of the dice. In blackjack, the dealer's face-down card has no effect on the player's decision making, and is drawn from the same probability distribution as the rest of the unknown cards. Since the dealer's decisions are completely determined by the rules of the game, blackjack can be viewed as a single-player stochastic game. In all perfect information games, whether stochastic or deterministic, the notion of a *best move*, or *set of maximal actions* is well-defined. In backgammon and blackjack, each choice has an objective game-theoretic *expected value* (EV), and usually there is exactly one choice that has a maximum EV. In deterministic perfect information games like chess and checkers, there is a set of moves that preserve the game-theoretic value of the game. Of these, there may be a "best practical move" that maximizes the probability of success in a contest between typical imperfect players.

For these games, the quality of each player decision can be compared to the best available choice, using an *oracle* or *near-oracle* to determine the quantitative value of every possible move. A perfect oracle is based on exact enumeration over all reachable game states. In blackjack, simulation or enumeration using the exact composition of the remaining deck provides a sound basis for direct assessment of EV decisions, as shown by Wolfe (2002).

A near-oracle can be based on Monte Carlo simulations, using a strong program to provide accurate estimates of the objective value of each move. In backgammon, Monte Carlo *roll-outs* using a strong program like TD_GAMMON are considered to be the definitive assessment of game equity (Tesauro (1995), Tesauro (2002)). In the game of Scrabble, the program MAVEN has surpassed all human players in skill level, and is used in a similar fashion (Sheppard (2002b), Sheppard (2002a)).[3] Again, these are good baselines for excellent play because they are strongly correlated with the maximum EV play available in each situation. Poker programs are not yet strong enough to serve as a near-oracle; but the problem goes much deeper than that, because the concept of a single maximum EV play is not applicable.

In contrast to perfect information games, there is generally no single best choice in a given poker situation. The concept of a perfect oracle is not well-defined (or at least, is not very discriminating between the possible options). A sound strategy is dependent on the *relative balance* of deceptive plays (such as *bluffing* with a weak hand, or *trapping* with a strong hand). A balanced approach is essential, but the player has considerable flexibility in *how* to obtain that balance. In both theory and practice, radically different strategies can have the same objective EV, and can be equally viable. Analogously, the notion of a best move is not meaningful in the imperfect information game of Rock-Paper-Scissors – the most effective choice is entirely dependent on the preceding history between the players.

In practice, a good poker player's strategy is strongly dependent on their beliefs about the opponent's weaknesses and vulnerabilities. For example, if Alice bluffs too much, an appropriate counter-strategy is to call (or raise) more often with mediocre holdings. If Bob seldom bluffs, the appropriate response is to fold more often with mediocre holdings. We see the opposite response depending on the perception of the opponent's style; and a player must continue to adapt as the opponent changes their strategy over time. In other words, in chess it is possible and appropriate to simply "play the board" (ignoring the opponent), whereas in poker it is necessary to "play the player" to maximize the win rate. Since there cannot be a "perfect" way to do this in general, it introduces a speculative element to the game which is not consistent with defining an objective assessment system.

For these and other reasons (some of which are outlined in the next section), it becomes evident that developing a "perfect" assessment tool is essentially impossible. In fact, an expert might argue that the assessment problem

[3] Although Scrabble is technically a game of imperfect information, that property does not play a major role in the strategy of the game. As a result, Scrabble does not differ greatly from a perfect information stochastic game, and programs using traditional approaches such as Monte Carlo simulation are highly effective in practice.

is “POKER-complete”, meaning that it is at least as hard as any other problem in the domain, including playing perfectly (if such a notion is even well-defined).

However, it is possible to design an imperfect assessment system that is conceptually simple, objective, and highly effective at reducing the variance due to stochastic outcomes. The Ignorant Value Assessment Tool (DIVAT)[4] is one such system. The term “ignorant” refers to the fact that it implicitly *ignores* much of the context in each situation. The expected value estimates may not be especially well-informed, but they are *consistent*, and are applied in an egalitarian fashion to all players. In the two-player case, the same criteria are applied equally to both players, with the intention of obtaining an unbiased assessment, even if the specific criteria could be criticized as being somewhat arbitrary. After we developed the tool empirically, it was formally proven to be statistically unbiased by Zinkevich *et al.* (2006). This means that the long-term expected value from the DIVAT assessment is guaranteed to match the long-term expected value of money earnings.

In Section 2, we examine the strengths and limitations of perfect information hindsight analysis, provide some background on previous attempts to use this approach, and present a concrete example of a game in order to ground the discussion. In Section 3 we define the terminology and explain the metrics used in each of the components, and re-visit the example to perform a detailed quantitative analysis using the DIVAT system. In Section 4 we illustrate some of the uses of the tool, and demonstrate its power in variance reduction, while providing an unbiased estimate of the difference in skill between two players playing a series of games. We conclude with a discussion of future work and generalization of the technique to other domains.

2. MOTIVATION

In this section we will focus our attention on a specific example of a game of poker, in order to identify some of the considerations and obstacles an analysis tool will need to cope with. We examine the reasons that an imperfect information domain cannot simply be broken down into a collection of perfect information instances. Then we look at our previous attempts to use perfect knowledge hindsight analysis and identify their major failings, which leads us to the design of the DIVAT system.

2.1 Example Game

We now look at an example of one complete game of Limit Texas Hold’em. We will call Player 1 (P1) Alfred, and Player 2 (P2) Betty. Although we will try to motivate some of their decisions, the objective analysis will, of course, be completely free of any such subjective interpretation.

A succinct summary of the game is:

Alfred: A♣-K♣ Betty: 7♡-6♡ Board: K♠-5♡-3♢ T♣ 4♡

Betting: SlRrC/kBrC/bC/bRc

All monetary units will be expressed in terms of *small bets* (sb), which is the size of all bets and raises in the first two rounds of play (the bet size doubles for the last two rounds of play). With the *reverse-blinds* format of two-player Limit Texas Hold’em, P2 posts a *small blind* bet of 0.5 sb, P1 posts a *big blind* bet of 1 sb, and P2 must make the first betting decision – to either fold, call 0.5 sb, or raise another 1 sb. The betting notation is: ‘s’ for small blind, ‘l’ for large blind, ‘k’ for check, ‘b’ for bet, ‘f’ for fold, ‘c’ for call, ‘r’ for raise, and ‘/’ is used as a delimiter for the four betting rounds.[5] Upper-case letters are used to more easily distinguish the actions of the second player (Betty).

Betty is dealt the **7♡-6♡**, which is not a particularly strong hand, but is certainly worth calling to see the *flop*. In this game, Betty elects to raise with the hand, for purposes of deception (called a *semi-bluff*). Alfred has a very strong hand, the **A♣-K♣**, and re-raises, to Betty’s detriment. Betty calls.

The flop is the **K♠-5♡-3♢**, giving Alfred a strong hand with a pair of Kings. Alfred decides to try for a *check-raise trap*, perhaps based on the belief that Betty bets too aggressively after a check (and is not as aggressive

[4] The first letter of the acronym corresponds to the first author’s initial.

[5] A *check* is functionally equivalent to a *call*, but “check” is used when the current amount to call is zero. Similarly, a *bet* is functionally equivalent to a *raise*, but “bet” is used when the current amount to call is zero.

when it comes to raising a bet). Betty bets (another semi-bluff) and Alfred completes his plan by raising. Betty has a weak hand, but the *pot* is now large (nine small bets) so she calls in the hope of improving (with a **4**, **6**, or **7**; or any heart, **8**, or **9**).

The *turn* card is the **T♣**, and Alfred simply bets his strong hand, since he believes that trying for another check-raise trap is unlikely to succeed. Betty still has nothing, and cannot be certain whether or not making a pair with a **6** or **7** will give her the best hand. If those outcomes will win (10/44 in total), then she has a correct call (2 sb to win 12 sb); if not, then she is better off folding. She elects to call, perhaps because Alfred has frequently followed this betting pattern in previous games without having a strong hand.

The *river* is a perfect **4**♡ for Betty, giving her the best possible hand (she could tie, but she cannot lose). Of course, Alfred has no way of knowing that this card was a disaster for him, and bets with what he believes to be the best hand. Betty raises, and Alfred elects to only call, perhaps based on his prior experience where Betty's raises on the final betting round are usually meaningful. The final pot size is 22 sb, giving Betty a net of +11 sb, and Alfred a net of -11 sb.

As an outside observer, we would like to know if one of these players played better than the other. We would like to filter out the stochastic luck element, and have an objective means of quantifying the perceived difference in skill. For a rational and accurate assessment, we would like to know how each player played *in general*, not just in hindsight, with all of the cards known. In this example game, most poker players would agree that Alfred played his hand well. However, most of his decisions were relatively easy,[6] so it is difficult to say how competent he is based on this one game. From the perspective of an average player, Betty might appear to be a wild gambler; but to an expert observer, she might appear to be either a world-class player *or* a weak player. Much of the assessment is subjective opinion, and is neither right nor wrong. Moreover, the opinions cannot be well-informed without a full historical context – Betty's decisions will depend very strongly on Alfred's style of play, and vice-versa. Real poker is highly non-Markovian in nature (*i.e.*, is not memoryless). A single game taken in isolation does not provide sufficient context for a well-informed assessment. However, for practical purposes, an analysis tool will likely have to ignore this fact (or temper it with some form of approximation).

2.2 EVAT and LFAT

Since 2001, we have used an assessment procedure called the Expected Value Assessment Tool (EVAT) to try to gain a more accurate picture of the results of short matches. The EVAT is based on hindsight analysis, where each player's decisions are compared to what the maximal action would have been if all players had known all of the cards. Thus we are comparing imperfect information decisions to the actions that would have been taken in the perfect information variant of the game. This provides a perspective that is quite distinct from the money line, and can occasionally provide useful insights. Unfortunately, it also suffers from some serious drawbacks, and tends to be rather unreliable as a predictor of future outcomes. Because of these limitations, the tool has not found its way into regular use.

The EVAT view is analogous to what poker author David Sklansky Sklansky (1992) calls "The Fundamental Theorem of Poker" (FToP), which states:

> Every time you play a hand differently from the way you would have played it if you could see all your opponents' cards, they gain; and every time you play your hand the same way you would have played it if you could see all their cards, they lose. Conversely, every time opponents play their hands differently from the way they would have if they could see all your cards, you gain; and every time they play their hands the same way they would have played if they could see all your cards, you lose.

Sklansky is suggesting that the long-term expected value in poker is equivalent to the differences in perfect information decisions. Similar assertions have been made in the past for many domains involving imperfect or incomplete information. However, strictly speaking this is not true, and the FToP is not a theorem. An imperfect information game cannot be expressed merely as a collection of perfect information instances. Some of the key differences between these classes of problems were exemplified for the game of bridge by Frank and Basin

[6] In general, it is easier to make good decisions with a strong hand than it is with a hand near the borderline between fold and not fold.

(2001), and by Ginsberg (2001).[7] The FToP is also qualitative in nature, not quantitative.[8] It is, however, a useful heuristic that can guide human players to formulating better decisions.

The EVAT is essentially a quantitative comparison between actual decisions and the ideal perfect information counterparts. Each time the actions are different, the player is assessed a penalty, equal in magnitude to the difference in expected value. The sum of all such "misplays" for each player are compared, yielding a quantitative estimate of the difference in skill.

Unfortunately, the EVAT is based on a highly unrealistic baseline for comparison, because the players are implicitly expected to have omniscient knowledge, without regard to the actual conditions that exist in each instance. As an extreme example, a player with the second-best possible hand on the river would be expected to *fold* whenever the opponent happens to hold the best possible hand, and raise otherwise. Conversely, a player with the second-worst possible hand would be expected to fold *except* when the opponent happens to hold the worst possible hand.

In the example game, Alfred's entirely reasonable bet on the river is a technical misplay, incurring an EVAT penalty, since his action is different from what he would have done if he had known all the cards. Moreover, he is expected to play the hand differently in otherwise identical situations, based only on the hidden information. This goes against fundamental principles of imperfect information games, in that the same policy must be employed within the same *information set* of the *game-tree*. The perfect hindsight view is largely irrelevant to the imperfect information reality of the situation, and illustrates one of the biggest flaws with the EVAT policy: when a player has a strong second-best hand, then with reasonable play they *should* lose money, in expectation.

Another way of looking at the problem of variance is in terms of *opportunities*. An immediate consequence of a high-variance stochastic game is that it can take a very long time before each player has had their "fair share" of good and bad situations. One of the reasons that the money line is a weak estimator of the long-term EV of the players is that the opportunities for each player do not balance out quickly. To a large extent, the EVAT analysis suffers from the same limitation.

The Luck Filtering Analysis Tool (LFAT) is a complementary system that was developed to address some of the short-comings of the EVAT. Using the same methods to compute the expected value of each situation, the LFAT compares each player's *pot equity*[9] before and after each chance event (*i.e.*, cards being dealt). Thus the LFAT analysis occurs between the betting rounds, while EVAT occurs between the chance events, so the analysis is split into the natural alternating phases of chance outcomes and player actions. The LFAT quantifies the effects of the stochastic outcomes alone, without regard to implications of betting decisions. Conversely, the EVAT method considers only the betting decisions themselves, and simply disregards the fluctuations due to luck, which is a highly desirable feature for reducing variance.

However, the separation is not perfect. Although the stochastic luck has been eliminated from consideration, the situations that arise truly are a consequence of all previous chance events that have occurred. Thus the two are not independent, and treating them as such introduces some degree of error.

3. DEVELOPMENT

In this section we define terminology and explain the components of the the Ignorant Value Assessment Tool. We present specific details of how the DIVAT system is constructed, and then work through the concrete example, illustrating its use.

3.1 Definitions and Metrics

In order to analyze the play of poker games, we will need metrics to assess the value of a hand. This is a complex problem, and good methods are often multi-dimensional. For example, *adjusted hand strength* is a combination of

[7] A simple example is the case of a two-way finesse: the declarer can win a key trick and secure the contract in every perfect information lay of the cards (a 100% chance of success), but is forced to *guess* in the real imperfect information game (for a 50% chance of success).

[8] As stated, the FToP is also untrue for a number of mundane reasons, and does not hold for multi-player situations, but that is of little concern to this discussion.

[9] The concept of equity is explained in section 3.1.2.

hand strength, which is a weighted probability of currently holding the best hand, and *hand potential*, which is a measurement of future outcomes (Billings *et al.* (2002)). For our purposes, we want to develop a *one-dimensional metric of hand goodness*, on a scale from zero to one.

For a given situation in the middle of a game, it will be necessary to estimate the *equity* for each player – the net amount that will be won or lost by the end of the game, on average. Most simplistic metrics do not consider the impact of future betting rounds (known as *implied odds* in the terminology of poker theory). While those methods tend to be easy to compute, and are sufficient for some purposes, we will also need to develop better-informed metrics which take future betting rounds into account.

3.1.1 IHR, 7cHR, and EHR

The Immediate Hand Rank (IHR) is the relative ranking of a hand, on a scale from zero to one, compared to all other possible holdings at the current point in a game. IHR considers only the number of opponent hands that are currently ahead, behind, or tied with the player's hand, and ignores all future possibilities. For two-player games, the formula is: IHR = (ahead + tied/2) / (ahead + tied + behind). Before the flop, there are 1225 possible two-card combinations for opponent hands. After the flop cards are known, there are 1081 possible opponent hands. There are 1035 combinations on the turn, and 990 combinations on the river. No inferences or assumptions are made about the opponent's cards, so all hands implicitly have uniform probability in hand rank calculations.

The 7-card Hand Rank (7cHR) performs a complete enumeration of all possible future *board cards*, computing the hand rank on the river for each instance, and reporting the overall average outcome. It amounts to an unweighted enumeration of unweighted hand ranks, and can be computed efficiently through the use of clever caching techniques (Billings *et al.* (2002)). Much greater efficiency was obtained by pre-computing look-up tables for all *pre-flop*, flop, and turn possibilities, mapping to and from canonical representatives to take advantage of suit isomorphisms to reduce storage requirements (Billings and Bard (2005)).

Since 7cHR is an average over all possible future chance outcomes, it contains a mixture of *positive potential* (the chance of improving to the best hand when behind), and *negative potential* (the chance of falling behind when ahead). However, it is not a particularly good balance of the two, because it implicitly assumes that all hands will proceed to a *showdown*, which is patently false. In practice, 7cHR tends to overestimate the value of weak no-pair hands. For example, **P1: 3♣-2♣ Board: K♠-T♡-7♢**, has 7cHR = 0.1507, based on two chances of making a small pair. Since the hand could be forced to fold on the turn, it is incorrect to assume a two-card future. In general, we should only be looking forward to the next decision point, which is a one-card future.[10] One possibility is to use 6cHR as an estimate of hand value on the flop. In practice, a better measure for Limit Hold'em is to take the *average* of 5cHR and 7cHR, because that captures the extra value from good two-card combinations. For example, **P1: 3♡-2♡ Board: K♠-T♡-7♢**, has 6cHR = 0.0729, but 7cHR = 0.1893 due to the possibility of two running hearts. Since hitting the first heart will provide a flush draw that almost certainly has enough value to proceed to the river, these indirect combinations have definite value. With respect to the DIVAT folding policy on the flop, we define the Effective Hand Rank (EHR) to be the average of IHR and 7cHR: EHR = (IHR + 7cHR) / 2.

With respect to the DIVAT betting policies (when the hand is sufficiently strong), we define EHR to be the *maximum* of IHR and 7cHR: EHR = max (IHR, 7cHR). The reason the maximum is a better measure of hand value for this purpose is that a hand with high negative potential generally has a greater *urgency* to bet. In the terminology of poker theory, a hand that degrades in strength as the game progresses (*i.e.*, 7cHR $<$ IHR) is said to have a high *free-card danger*. This means that there is a relatively high risk in *not betting* (or raising), because allowing a free draw could be potentially disastrous in terms of EV (such as losing the pot when a bet would have forced the opponent to fold).

Since 7cHR does not distinguish cases where the opponent does or does not bet, the assessment is somewhat optimistic in practice, even for one-card futures from the turn forward. For example, **P1: 3♣-2♣ Board: K♠-T♡-7♢ 4♣**, has 7cHR = 0.0620, but could easily be *drawing dead* against a one-pair hand or better. Making a small pair on the river will not be strong enough to bet, but will be strong enough to warrant a call. Thus the hand stands to lose an additional two small bets whenever the opponent has a strong hand, but gain nothing when it is the best. In the terminology of poker theory, 7cHR does not adequately account for the *reverse implied odds* in most situations.

[10] This is especially imperative in No-Limit poker, where the opponent can apply much greater leverage with a large subsequent bet.

Despite these drawbacks, 7cHR is a convenient metric that is simple to compute, and the short-comings can be dealt with easily enough. The use of these metrics is illustrated in the detailed example of section 3.4.

3.1.2 AIE and ROE

The *all-in equity* (AIE) is measured with the fraction of all future outcomes each player will win, given perfect knowledge of the cards held by all players.[11] This fraction can be multiplied by the current pot size as a simple measure of (gross) *pot equity* – the "fair share" of the current pot that each player is entitled to, based on the complete enumeration of future outcomes. For two-player Texas Hold'em, there are 44 cases to consider for turn situations, 990 cases for the flop, and 1712304 cases for the pre-flop.

To convert the gross amount to the net gain or loss, we subtract the total amount invested from the current pot equity: Net = (AIE * $potsize$) - $invested$. For example, if the pot size after the turn betting is 10 sb (5 sb from each player), and 11 out of 44 outcomes will win, zero will tie, and 33 will lose (written as +11 =0 -33), then our current net pot equity is Net = 0.25 * 10 - 5 = -2.50 bets. All further discussion of equity will be in terms of net pot equity.[12]

In general, AIE is an unrealistic measure of true equity because it does not consider the effects of future betting. For example, a *pure drawing hand* like **P1: 3♡-2♡ P2: A♠-T♠ Board: K♠-T♡-7♢** **A♡**, has AIE = 0.2045 (9/44) for P1, but has very favourable implied odds, since it stands to win several bets if a heart lands, and need not lose any additional bets if it does not improve. Conversely, a hand that will be mediocre in most of the outcomes will have reverse implied odds, and the true equity will be less than the AIE metric would indicate. In cases where a player may be forced to fold to a future bet, using AIE can result in a gross over-estimate of hand value.

In contrast, *roll-out equity* (ROE) is based on an explicit enumeration of each future outcome, accounting for the betting in each case according to some standard policy. A simple estimate might assume exactly one bet per round contributed by each player. A better measure would account for folding of the worst hands, and raising of the best hands. The tools we will develop in the following sections establish a more refined betting policy that can be used for all future betting rounds, providing a much more accurate estimate of the true equity of each situation. We refer to the basic implementation as AIE DIVAT, whereas the well-informed roll-outs give rise to the superior ROE DIVAT system.

3.2 DIVAT Overview

Poker decisions are extremely context-sensitive. Each decision depends on every decision that preceded it, throughout the history of previous games, and especially with respect to the previous actions in the current game. The scope of the DIVAT analysis is a single complete betting round, encompassing all of the player decisions made between chance events. Since it is essential to keep the frame of reference as simple as possible, we simply ignore much of the relevant context of each situation. This is not as serious a limiting factor as it might seem, as will be shown in section 4.

In particular, each player's betting actions (and responses to the opponent) in previous betting rounds are treated as being essentially information-free, and no inferences are made with respect to their likely holdings. Since the EHR metric is based on unweighted hand rank calculations, the distribution of opponent hands is implicitly assumed to be uniform. In a sense, each new round is taken as an independent event, with no memory of how the current situation arose, apart from the current size of the pot. However, the implications of bets and raises within a single betting round *are* taken into account. By largely ignoring the history of how each situation arose, we lose some relevant context, and possibly introduce error to the assessment. Fortunately, this does not completely undermine the analysis, because those errors apply to both players in roughly equal proportions, and thus the

[11] With the normal (*table stakes*) rules of poker, a player is said to go *all-in* when they commit their last chips to the pot. The player cannot be forced to fold, and is eligible to win the portion of the pot they have contributed to if they win the *showdown*. (If more than one active player remains, betting continues toward a *side pot*). Hence, from the perspective of the all-in player, there is no future betting, and the AIE is an accurate calculation of their true equity.

[12] Note that measures of equity require the perfect knowledge hindsight view to assess the fraction of future outcomes that are favourable. In contrast, measures of hand strength are with respect to a player's imperfect information view, and are used for realistic assessment, such as establishing baseline betting sequences.

effects largely cancel out.

The core component of the Ignorant Value Assessment Tool is the DIVAT *baseline betting sequence*. The baseline sequence is determined by *folding policies* and *betting policies*, applied symmetrically to both players. Each of the DIVAT policies is designed to use the simple metrics, while also compensating for some of their liabilities.

To understand DIVAT on an intuitive level, we need to consider the implications of deviations between the baseline sequence and the sequence that actually took place in the game. The baseline sequence represents a sane line of play between two "honest players", free of past context, and devoid of any deceptive plays. If the total number of bets that actually went into the pot was not equal to the baseline amount, then the *net difference in the resulting equities* can be measured.

For example, when Alfred decided to try for a *check-raise trap* on the flop, he was taking a calculated risk in the hope of a higher overall payoff. If his deceptive play is successful, then more money will go into the pot than in the baseline, and he is awarded the equity difference on that extra money. If the ploy fails, then he will receive a penalty based on the lost equity (which happens to be the same amount in this case). Thus, his judgement in attempting a non-standard play is immediately endorsed or criticized from the net difference in equities.[13] Although the obvious limitations in measurement and scope can produce a somewhat uninformed view of the ideal play in some cases, the net difference in equities is a highly meaningful characteristic most of the time.

3.3 DIVAT Details and Parameter Settings

The DIVAT *folding policy* is used to decide when a player should fold to an opponent's bet. The folding policy is based on a *quasi-equilibrium* strategy, motivated by game-theoretic equilibrium strategies. Since the 7cHR metric is in the range zero to one, and is roughly uniformly distributed across that range on the river, the 7cHR is compared directly to the game-theoretic optimal frequency for folding. The game-theoretic equilibrium fold frequency is an invariant that depends only on the size of the bet in relation to the size of the pot. If the bet size is bs and the pot size is ps at the beginning of the river betting round, and the opponent then bets, the optimal fold frequency is $bs/(ps + bs)$. For example, if the pot size on the river is 8 sb and the bet size is 2 sb, the DIVAT policy would fold all hands with 7cHR below $2/(8+2) = 0.20$, as an estimate of the bottom 20% of all hands.[14]

Prior to the river round, this simple method is not applicable. First, some weak hands may be worth calling for their *draw potential*. One can imagine a situation where it is correct to call a bet with *any* hand, because the pot is much larger than the size of the bet (giving extremely high *pot odds*). Clearly it is not correct in principle to fold at the same rate when there are still cards to be dealt. Secondly, as an average of future outcomes, 7cHR does not have a uniform distribution. Very low values (*e.g.*, below 7cHR = 0.15 on the flop) do not occur, due to the ubiquitous possibility of improvement. Very high values are also under-represented, being vulnerable to a certain fraction of unlucky outcomes. Thirdly, the 7cHR metric is not especially well-balanced with respect to the extra expense of mediocre hands (reverse implied odds), as mentioned previously.

To account for these factors, we add an offset to the game-theoretic folding frequency, and use that as a fold threshold for the EHR metric. Thus the formula for pre-river fold policies is: Fold Threshold = $bs/(ps + bs) + \textit{offset}$. The offsets were first estimated on a theoretical basis, then verified and tuned empirically. The empirical data based on millions of simulated outcomes is omitted in the interest of brevity. On the turn, the normal offset is +0.10, so the fold threshold in the previous example would be increased to hands below 7cHR = 0.30. On the flop, the normal offset is +0.075, thus hands would be folded below 7cHR = 0.275 when the initial pot is four bets in size. Prior to the flop, all hands have sufficient draw odds to see the three-card flop, so no offset is required.[15]

The DIVAT *betting policy* is used to decide how many bets and raises a hand is worth. It is strictly a *bet-for-value* policy, meaning that all bets and raises are intended to be positive expected value actions, with no deceptive plays (*i.e.*, no bluffing with a weak hand, nor trapping with a strong one). Betting in direct proportion to hand strength is a very poor poker strategy in general, because it conveys far too much reliable information to the opponent. Nevertheless, it serves as a reasonable guideline of how many bets each player should wager in a given situation,

[13] Since the net difference is relative, there is actually no distinction between a bonus for one player and a penalty for the other. The assessment is symmetric and zero-sum in this regard.

[14] Folding more often than this would be immediately exploitable by an opponent who always bluffs, showing a net profit overall. Conversely, folding less often than the equilibrium value would be exploitable by betting with slightly weaker hands. Employing the game-theoretic optimal fold policy makes a player indifferent to the opponent's strategy.

[15] The pre-flop offset could be negative, but it would make little difference in practice, since almost all cases will proceed to the flop.

	Fold Offset	Make1	Make2	Make3	Make4
Pre-flop	0.000	0.580	0.825	0.930	0.965
Flop	0.075	0.580	0.825	0.930	0.965
Turn	0.100	0.580	0.825	0.930	0.965
River	0.000	0.640	0.850	0.940	0.970

Table 1: Standard (moderate) DIVAT settings.

all else being equal.

Although bluffing is an absolutely essential component of any equilibrium strategy, the benefits of bluffing are exhibited as a side-effect, increasing the profitability of strong hands. The actual bluffing plays themselves have a net EV of zero against an equilibrium strategy (neither gaining nor losing equity). Since the DIVAT fold policies and betting policies are applied in an oblivious manner, the highly predictable behavior of the bet-for-value policy is not being exploited. In effect, the deceptive *information hiding* plays are taken as a given, and all players simply employ the ideal bet-for-value policy without consideration to the opponent's possible counter-strategies. Since the DIVAT policy provides a realistic estimate of how many bets should be invested by each player, and is applied in an unbiased fashion to all players, the weighted gains and losses in hindsight EV are a low-variance estimate of the long-term differentials in decision quality.

As an example, if Player 1 holds a hand of value EHR = 0.70 and Player 2 holds an EHR = 0.90 hand, then the bet-for-value betting sequence for the round would be bet-Raise-call (bRc), indicating that each player will normally invest two bets on that betting round. In the reverse case, with P1 = 0.90 and P2 = 0.70, the expected sequence would be bet-Call (bC), an investment of one bet each, because the second player does not have a strong enough hand for a legitimate raise. This demonstrates a natural asymmetry of the game, and reflects the inherent advantage enjoyed by the player in second position.

The *Make1 threshold* is the strength needed to warrant making the first bet of a betting round. The *Make2 threshold* specifies the strength required to raise after the opponent bets. The *Make3 threshold* is the strength required for a re-raise. The *Make4 threshold* governs re-re-raises (called *capping* the betting, because no more raises are allowed).[16] The betting thresholds are derived from the equilibrium points ensuring positive expectation. On the river, the betting thresholds must account for the fact that the opponent will only call with a hand that has some minimum value. A bet cannot earn a profit if the opponent folds, and the hand must win more than half of the time *when it is called* to have positive expectation. For example, if the opponent will call at the game-theoretic optimal frequency, then the range above that point is the relevant interval. A Make1 threshold of 0.64 approximately corresponds to a policy of betting with the top 41.4% of holdings in that interval.[17] Prior to the river, there is tangible value in forcing the opponent to correctly fold a hand that has a small chance of winning (sometimes called the *fold equity*). We handle this case by considering the appropriate interval to be the entire window from zero to one, lowering the thresholds by the corresponding ratio. A Make1 threshold of 0.58 approximately corresponds to a policy of betting the top 41.4% of all possible holdings.

The standard betting thresholds for each round of play are listed in Table 1.(Billings and Kan (2006))

3.4 Basic AIE DIVAT Analysis of the Example Game

We now re-visit the example game presented in section 2.1 to illustrate the quantitative DIVAT assessment of all the decisions made by each player. This analysis uses a basic version of DIVAT, based on immediate hand rank, seven-card hand rank, and all-in equity. Although roll-out equity is clearly superior (being better-informed and provably unbiased), the all-in equity is simpler to compute and easier to follow. Table 2 presents the pertinent values in the analysis of each round of betting.

In the first round of play, Betty chose to make a deceptive play by raising with a hand that is unlikely to be the best. She may not have done this with the expectation of Alfred folding immediately, but the raise could create a

[16] Some Limit games permit a fifth level, or even unlimited raises in two-player (*heads-up*) competition. This does not greatly affect the analysis, as situations with more raises are increasingly uncommon, and can be handled if necessary, even for the infinite case.

[17] The zero EV equilibrium points for a 4-bet maximum are derived recursively. The fractions corresponding to the Make1 - Make4 thresholds are the top 12/29, 5/12, 2/5, and 1/2 of the interval, respectively. For unlimited raises, the zero EV equilibrium point is the top $\sqrt{2} - 1$ at all levels.

Board: <none>			
	IHR	7cHR	EHR
P1: A♣-K♣	0.9376	0.6704	0.9376
P2: 7♡−6♡	0.1604	0.4537	0.4537
AIE = 0.6036 (+1029832 =7525 -674947)			
LFAT change = +0.2073 (of 1712304)			
DIVAT Baseline = SlCrC	Actual Seq = SlRrC		
Round EV = +0.2073 [SlCrC]			
Actual Equity = +0.6218 [SlRrC]			
Baseline Equity = +0.4145 [SlCrC]			
DIVAT Difference = +0.2073 sb			

Preflop Analysis

Board: <K♠-5♡-3♢>			
	IHR	7cHR	EHR
P1: A♣-K♣	0.9685	0.8687	0.9685
P2: 7♡−6♡	0.0634	0.3798	0.2216
AIE = 0.7636 (+756 =0 -234 of 990)			
LFAT change = +0.9600			
DIVAT Baseline = bC	Actual Seq = kBrC		
Round EV = +0.5273 [bC]			
Actual Equity = +2.6364 [kBrC]			
Baseline Equity = +2.1091 [bC]			
DIVAT Difference = +0.5273 sb			

Flop Analysis

Board: <K♠-5♡-3♢ T♣ >			
	IHR	7cHR	EHR
P1: A♣-K♣	0.9411	0.8902	0.9411
P2: 7♡−6♡	0.0662	0.2146	0.2146
AIE = 0.9091 (+40 =0 -4 of 44)			
LFAT change = +1.4545			
DIVAT Baseline = bF	Actual Seq = bC		
Round EV = 0.0000 [bF]			
Fold Equity = 0.9091 [bF]			
Actual Equity = +5.7273 [bC]			
Baseline Equity = +5.0000 [bF]			
DIVAT Difference = +0.7273 sb			

Turn Analysis

Board: <K♠-5♡-3♢ T♣ 4♡ >			
	IHR	7cHR	EHR
P1: A♣-K♣	0.8576	0.8576	0.8576
P2: 7♡−6♡	0.9955	0.9955	0.9955
AIE = 0.0000 (+0 =0 -1 of 1)			
LFAT change = -12.7273			
DIVAT Baseline = bRc	Actual Seq = bRc		
Round EV = -4.0000 [bRc]			
Actual Equity = -11.0000 [bRc]			
Baseline Equity = -11.0000 [bRc]			
DIVAT Difference = +0.0000 sb			

River Analysis

Table 2: Round-By-Round Analysis of the Example Hand

misrepresentation that will have lasting effects much later in the game. Moreover, if she never raised with weaker hands, she would potentially be conveying too much information to her opponent. If Alfred had only called the raise, then there would be no net difference in equity compared to the DIVAT baseline, since the same amount of money would be invested by the two players. In the actual game sequence, Alfred correctly re-raised, and is thus awarded a skill credit for the round, at the expense of Betty. Since Alfred's hand will win about 60% of the future outcomes, he earns a small fraction of the two extra bets that went into the pot (Net = 0.60 * 2 - 1 = +0.20 sb).[18]

The flop was favourable for Alfred, giving him a strong hand, and increasing his chance of winning a showdown to 76%. He tried for a *check-raise trap*, which was successful, netting him another 0.53 sb in equity. If Betty had simply checked, she would have earned that amount instead of Alfred, and his gamble would have back-fired.

The turn brings another good card for Alfred, increasing his expected share of the pot to 91%. Since his check-raise on the flop may have revealed the true strength of his hand, he opts for the straight-forward play of betting. Now Betty is faced with a difficult decision between calling with a weak hand, or folding when the pot is relatively large. Betty might have supposed that there was still a chance that Alfred had no pair, and that she could perhaps win if a **6** or **7** landed. The necessity of making informed guesses is the very essence of poker. In this particular case, that belief would be wrong, and folding would have had a higher expected return for Betty. The magnitude of the resulting misplay is accurately reflected by the DIVAT Difference for the round. [19] When a fold is involved, the DIVAT Difference is not the same as the baseline EV for the round. Here Betty loses a net of 1.64 sb from the actions of the round (bet-Call), but calling only loses 0.73 sb more than folding. The difference is due to the *fold equity* that Betty refused to abandon (1/11 of the pot prior to the betting), which gives her partial compensation for the calling error.

[18] In reality, the equity difference is not that large, because Betty will usually not continue with the hand when it does not connect with the flop, but stands to win several extra bets in the future if the flop is favourable. These positive implied odds are reflected in the roll-out equity, which indicates that Alfred's advantage in this situation is actually much smaller, as shown in Table 3.

[19] We again see the effects of favourable implied odds when a **4** lands, which out-weigh the *reverse implied odds* when a **6** or **7** lands. The slightly better prospects after the calling error are captured by the roll-out equity analysis.

A♣-K♣ vs 7♡-6♡ | K♠-5♡-3♢ T♣ 4♡
Betting sequence: SlRrC/kBrC/bC/bRc

Round	LFAT	RndEV	DIVATdiff
Pre-flop	+0.207	+0.207	+0.207
Flop	+0.960	+0.527	+0.527
Turn	+1.455	+0.909	+0.727
River	-12.727	-4.000	0.000

Total DIVAT Difference = +1.462 sb

Example Hand AIE DIVAT Summary

A♣-K♣ vs 7♡-6♡ | K♠-5♡-3♢ T♣ 4♡
Betting sequence: SlRrC/kBrC/bC/bRc

Round	LFAT	RndEV	DIVATdiff
Pre-flop	+0.207	+0.029	+0.029
Flop	+0.960	+0.523	+0.523
Turn	+1.455	+0.909	+0.636
River	-12.727	-4.000	0.000

Total DIVAT Difference = +1.189 sb

Example Hand ROE DIVAT Summary

Table 3: Full-Game Analysis of the Example Hand (AIE and ROE)

The perfect **4**♡ for Betty on the river erases all of the good luck Alfred enjoyed up to that point. The LFAT swing indicates that the card cost him 12.73 sb, based solely on his pot equity after the turn betting. However, the damage is actually much greater, because he naturally continues to bet his strong hand and walks into a raise, losing an additional 4 sb.

Here we witness the dramatic difference between the perfect knowledge hindsight view of EVAT, and the more realistic DIVAT view. The EVAT baseline sequence for the final round of betting would be check-Bet-fold, which is quite absurd with respect to the imperfect information reality. The DIVAT baseline of bet-Raise-call is much more reflective of real play. Since Alfred did not compound the problem with a re-raise, he loses only the expected amount from the river betting, and thus does not receive any penalty. If he had chosen to play check-Bet-call (perhaps based on a telling mannerism by Betty), he would in fact *gain* in the view of DIVAT, because he lost less than was expected.

Not only is the DIVAT baseline more reasonable, it is also clear that the EVAT analysis is statistically biased, because it is preferential toward one particular *style* of play over another, apart from EV considerations. As we have seen, it is simply impossible to play the river betting round without frequent misplays, with respect to the EVAT perfect information baseline. This means that a player who employs a conservative style (folding in neutral or marginal EV situations) will be viewed more favourably than a player with a liberal style (calling in those same situations), simply because the conservative player gets to the river round with a marginal hand less often. In fact, if one wanted to maximize the EVAT opinion of one's play, it would be wise to sacrifice slightly positive expectation situations on the flop and turn, simply to avoid being measured against an impossible standard on the river. The irrational EVAT baseline has the same undesirable effect on earlier rounds, but it is most pronounced on the river, where the effect is not dampened by averaging over future outcomes. In contrast, the DIVAT baseline provides one standard way to play the hand for *all* cases, without regard to the opponent's actual (hidden) holding.

The round-by-round LFAT transitions cannot easily be combined to determine a net effect on the game as a whole. They cannot simply be summed, since instances of early good luck can be wiped out by a final reversal, as in the example game. The LFAT value can, however, provide some relevant context when interpreting the round-by-round analyses.

The DIVAT analyses of each round are treated as being essentially independent of each other, and can be summed to give the net difference for the complete game. Indeed, the sum is generally more meaningful than the individual rounds in isolation, because a player may make a deliberate misplay on an early betting round in order to induce larger errors on the turn or river. For example, a player may *slow-play* by only calling with a strong hand on the flop, intending to raise on the turn when the bet size is larger. Conversely, a player may *raise for a free-card* with a drawing hand, intending to check on the turn after the opponent checks. All else being equal, the success or failure of the tactic will be reflected in the total DIVAT Difference for the game.

Table 3 gives a summary for the complete example game. The corresponding summary for the better-informed and theoretically sound ROE DIVAT is given for direct comparison. The ROE DIVAT procedure has been formally proven to be statistically unbiased by Zinkevich *et al.* (2006).

Overall, we can conclude that Alfred did better than expected with the cards and situations that arose in this game. In hindsight, he made better choices than Betty, worth a total of about +1.46 small bets in terms of long-term expected value. The fact that Alfred actually lost 11 sb on the hand is not particularly relevant, and only serves to occlude the most important considerations.

Of course, this is only a single game, and as previously noted, Alfred's hand was somewhat easier to play well than Betty's. The non-independence of consecutive games should also be kept in mind. In the wider scope, Betty's play might have been entirely appropriate, and might earn some compensation in future games. Long-term plans taking effect over many games are also captured by the running total of DIVAT Differences over the course of a match. Given many examples over a series of games, the accumulated DIVAT Differences provide a much more accurate, low-variance estimate of the eventual win rate for one player over another.

3.5 Smoothing and Using Equilibrium Baselines

Further refinement to DIVAT can be obtained by averaging the results of several runs, using different parameter settings for each. For this purpose, we defined nine sets of folding parameters that reflect styles of play ranging from very loose to very tight; and defined nine sets of betting parameters that reflect styles of play ranging from very aggressive to very conservative. In our experiments, we normally run a sweep of five settings from loose/aggressive to tight/conservative.[20]

This provides a natural *smoothing* of the DIVAT Differences over the short-term. For example, a situation might arise that is close to the borderline between bet-Call and bet-Raise-call. Over the five separate runs, the baseline might be bet-Call four times and bet-Raise-call once. Thus instead of a single threshold with a jump from 2.0 bets to 4.0 bets, we have a smoother transition with five smaller steps, yielding a jump of 2.4 bets on average.[21]

In effect, the average of several runs is a hedge, being less committal in situations where the best line of play is more debatable. This can be especially useful over the range of fold thresholds, because the difference in EV between folding and continuing with the hand can be very large (since the whole pot is at stake, rather than fractions of a bet). However, over the course of many games the average of several runs will simply converge on the DIVAT Difference line for the median (moderate style) parameter settings. Thus, smoothing is primarily useful for short matches.

In imperfect information games where an equilibrium solution is known (or can be accurately estimated), the ideas behind DIVAT can be used in a manner analogous to smoothing. For example, in the case of poker, suppose we have a weak hand on the final round of betting, which would be checked in the DIVAT baseline. If we know that the bluff frequency in this situation should be 10%, then the corresponding baselines for betting would be computed, and would contribute 10% of the total weight for actions made with a weak hand. The pertinent regions of the strategy space are not continuous (unlike smoothing, which was done over a continuous range of values), but that is of no consequence for determining the average EV outcome for a given situation. Thus knowing an equilibrium strategy provides us with an appropriate way to compute a weighted average of possible actions (or action sequences).

We have not attempted to use a full quasi-equilibrium strategy to obtain a collection of DIVAT baselines to be averaged. However, it remains an interesting possibility for future work.

4. EXPERIMENTS

The Ignorant Value Assessment Tool is based on relatively simple methods for hand strength measurement, betting policies, and folding policies. During development, each method was empirically tested and refined iteratively, since the components are mutually dependent to some degree.

For example, the folding policy was tuned to be consistent with the strengths and liabilities of the simplistic IHR and 7cHR hand strength metrics. In the case of 7cHR, it was observed that reverse implied odds were not

[20] Although the fold parameters and betting parameters could be varied independently, the number of runs would grow multiplicatively, without much benefit. For the same reason, we keep the parameter settings for each of the four betting rounds consistent, and apply the same settings for the actions of both players, rather than mixing them. The intention is to aim for a range of distinct perspectives over a span of reasonable styles, rather than an exhaustive enumeration of many possibilities.

[21] The weight of each run need not be equal. Since the median value is deemed to be the most reasonable, the weights for five runs could be 1-2-3-2-1 instead of 1-1-1-1-1; or could follow the binomial distribution 1-4-6-4-1 to reflect a Gaussian dispersion of styles over the specified range.

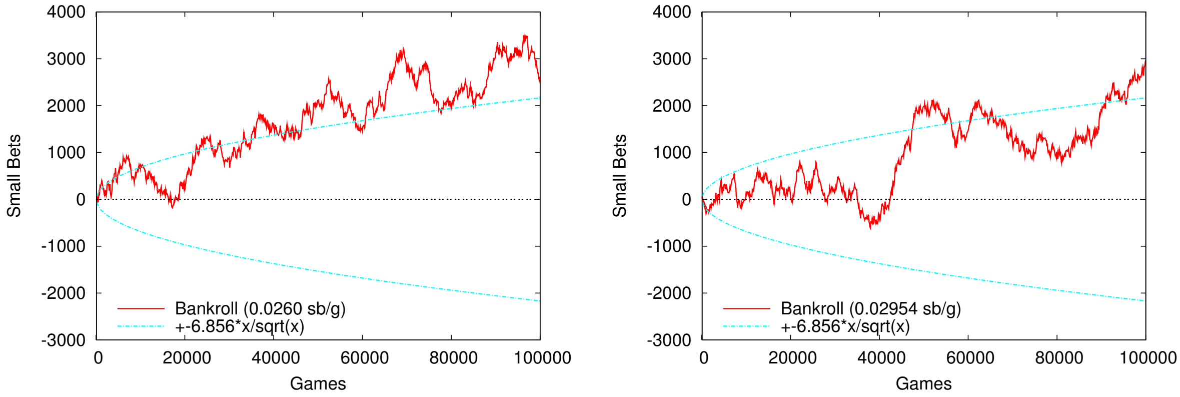


Figure 1: High variance in two Always_Call *vs* Always_Raise matches.

being given enough consideration, resulting in folding policies that were too liberal when cards were still to be dealt. The correction was to add an offset to the game-theoretic threshold value.[22] The value of this offset was determined empirically using roll-out simulations to find the best practical setting for achieving a neutral outcome, thereby estimating the game-theoretic equilibrium point. The experimentally derived offset was consistent with the predicted offset based on expert calculations and intuition.

Each series of experiments was repeated on at least seven different matches, involving a wide range of playing styles and skill levels. This is necessary because the conditions of the match dictate the frequency and type of situations that will arise. The computer players ranged from simple rule-based algorithms to the strongest known game-theoretic and adaptive programs. The human players ranged from intermediate strength to world-class experts.

To make the presentation of experimental results easier to follow, we will limit ourselves to only five types of matches. The first type is between two extremely simple players: Always_Call and Always_Raise. Although both algorithms are extremely weak as poker players, there are several advantages to studying this contest, because it holds many variables constant. The exact betting sequence is always known: if Always_Raise is the first player, the betting sequence is bet-Call (bC) for every round; if Always_Call is the first player, the betting sequence is check-Bet-call (kBc) for every round. There is never a fold, so every betting round is involved in every game. The final pot size is always 14 sb, and the net outcome for one player is either +7 sb, -7 sb, or zero (a tie will occur approximately 4.06% of the time). The long-term expected value for each player is exactly zero, since neither player exhibits superior skill. The outcome of each game is similar to coin-flipping, with occasional ties. The natural variance can easily be determined for these conditions. Simulation experiments were run for 25 million games, showing a variance of 47.010, for a standard deviation of ± 6.856 small bets per game (sb/g). The empirical variance for each 100,000-game match agreed, also showing ± 6.856 sb/g.

This gives us a frame of reference for the natural fluctuations on the amount won or lost during the match (the "money line", labelled as "Bankroll"). The ± 6.856*x/sqrt(x) guide-curves indicate the one standard deviation boundary on the total outcome after x games. The 95% confidence interval would correspond to roughly two standard deviations. The variance in real matches strongly depends on the styles of the two players involved, but ± 6 sb/g is typical.[23] When complete data is available, we use the actual measured variance during the match for more accurate guide-curves.

Figure 1 shows two separate matches of 100,000 games between Always_Call and Always_Raise. The money line in these two matches nicely illustrates just how pernicious the problem of variance can be. In the first match, there is every appearance that one player is outplaying the other, by an average of more than +0.025 small bets per game (sb/g). Although there are some fluctuations along the way, they are small enough to give the impression of a steady increase. This is extremely misleading, since we know that the long-term average is exactly zero. With the aid of the guide-curves, we can see that the money line strays outside the boundary of

[22] Actually, the linear adjustment involved both a multiplier and an offset ($y = mx + b$), but a multiplier of 1.0 was found to be sufficient, leaving only the constant offset as a correction factor.

[23] Many games will end early when one player folds, but those frequent small net outcomes (n) are offset by the occasional large outcomes, which carry greater weight toward total (n^2) variance.

one standard deviation, but is still easily within the realm of possibility. In other words, this is not a severely abnormal outcome, and it is not statistically significant at the 95% confidence interval.

In the second match (generated with an independent random number generator, due to our own suspicions), we again see a high-variance event with a potentially misleading conclusion.[24] Another feature to notice about this second match is the dramatic rise from about -650 sb at 38,000 games, to almost +2000 sb by 48,000 games. During that 10,000-game span (which is much longer than most competitive matches) one player demonstrates an enormous win rate of +0.25 sb/g that is entirely due to random noise.

The second type of match we will examine is a self-play match, with the game-theoretic equilibrium-based program PSOPTI4 playing both sides. This player is *stationary* (*i.e.*, uses a strategy that is randomized but static, is oblivious to the opponent, and does no learning or adaptation), so clearly the long-term expectation is exactly zero. However, unlike the ALWAYS_CALL *vs* ALWAYS_RAISE match, the play is realistic – the player folds bad hands, raises good hands, bluffs, and attempts to trap the opponent.

The third match type is between PSOPTI4 and PSOPTI6. The newer incarnation plays a substantially different style, but loses to the older version in head-to-head competition.[25] Both players are static, so learning effects are eliminated. Since we want to compare the DIVAT estimate to the long-term EV, this match and the self-play match were extended to 400,000 games, concatenating the first 100,000-game series above with three new 100,000-game series (each generated with a different random seed). The same series of cards was then used to play a *duplicate match*, with the players occupying the reverse seats (note that the self-play match is self-duplicate). Thus 800,000 distinct games were played between PSOPTI4 and PSOPTI6, with each 400,000-game match being directly comparable to the other. The hand-by-hand results were averaged to obtain the *Duplicate Money Average* and the *Duplicate DIVAT Average* lower-variance estimators, for comparison with the single-sided DIVAT.

The fourth type of match we examine here is a man versus machine match between world-class expert "thecount" (Gautam Rao), and the first game-theoretic equilibrium-based program PSOPTI1. This match was featured in our 2003 paper (Billings *et al.* (2003)). The 7030-game contest again featured large swings, which we now know were almost entirely attributable to stochastic noise, with each player having alternating phases of extreme luck.

A match involving real players is significantly different from a match played under the sanitized conditions of the laboratory. Here we witness a clash between radically different approaches to the game. Top-flight experts change their style rapidly and frequently, as the game conditions dictate. The human player avoids being predictable, explores new tactics, and learns over time how best to exploit the opponent. In this match, "thecount" started with a hyper-aggressive style that is highly effective against most human opponents, but was counter-indicated against the program. After shifting to a more patient approach, he enjoyed more success. Much of that was due to fortuitous cards and situations that coincidentally occurred at around the same time; but the DIVAT analysis is able to extract signal from the wash of noise, revealing the overall *learning curve* and the positive impact of that major shift in style.

The final match we will examine is between PSOPTI4 and the adaptive program VEXBOT. VEXBOT is the strongest poker program to date, having defeated every computer opponent it has faced, and often provides a serious threat to top-flight human players (Billings (2003), Billings *et al.* (2004)). However, the learning systems embedded in the VEXBOT architecture are slow and imperfect. They often require many thousands of games to train, and frequently lead to local minima that are well below the maximum exploitation level of the given opponent. The DIVAT analysis reveals much more about the changes of VEXBOT over time, and again shows how misleading the basic money line can be.

The experiments shown in this section address: (1) the correctness and variance-reduction properties of DIVAT, (2) the ability to reveal learning curves and changes in style for *non-stationary* players, (3) the robustness of the DIVAT Difference line, and (4) the usefulness of the round-by-round analysis to gain extra insights into match results. Many other experiments were conducted during the development of the DIVAT system. In the interest of brevity and focus, we do not show experimental results pertaining to: (1) system components and parameter

[24] These matches were not selected after the fact – they were the first two matches generated.

[25] Note that this does **not** mean that PSOPTI4 is a better player than PSOPTI6. The result of any one match-up is not necessarily indicative of overall ability. By analogy, consider a pseudo-optimal Rock-Paper-Scissors player that chooses Rock 30% of the time, Paper 40%, and Scissors 30%. The maximum exploitation best response (+0.10 units per game) is achieved by Always_Scissors, but that is one of the worst possible strategies in general. Poker results are highly non-transitive (*e.g.*, A beats B beats C beats A) in practice. Nor is there a simple cycle – it is a complex *wreath* of dominance relationships between styles.

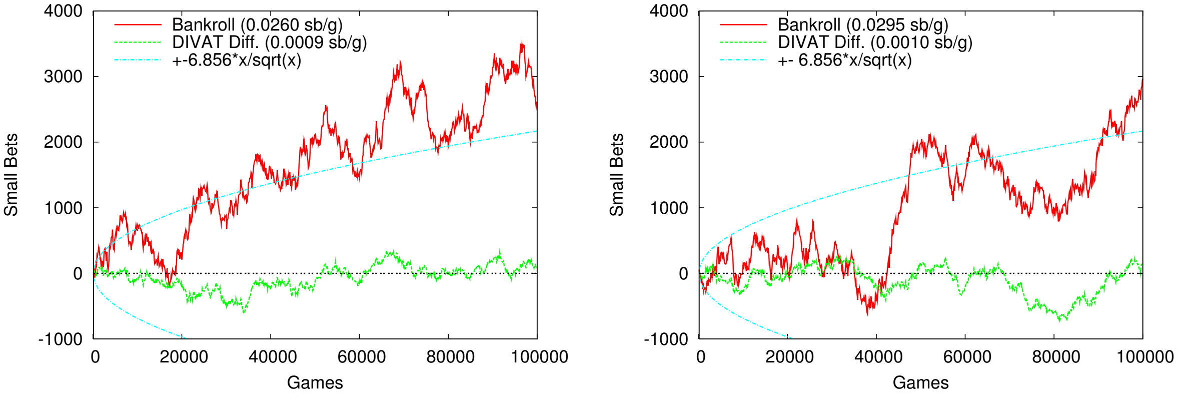


Figure 2: DIVAT Difference analysis of two Always_Call *vs* Always_Raise matches.

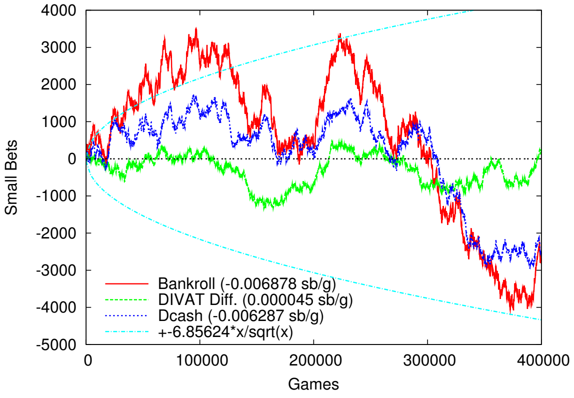


Figure 3: DIVAT Difference analysis and DIVAT Cash baseline for an extended Always_Call *vs* Always_Raise match.

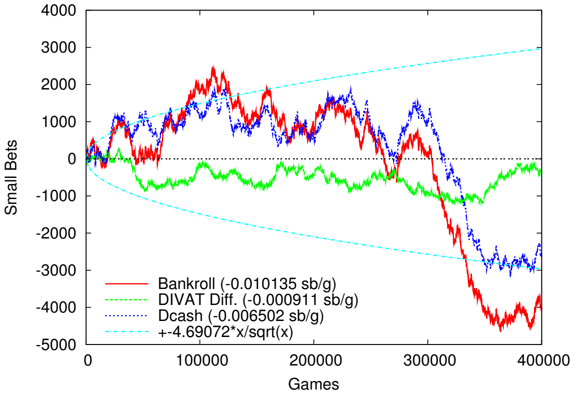


Figure 4: DIVAT Difference analysis of PsOpti4 self-play match.

tuning (other than robustness), (2) the significant statistical bias of the (perfect information) EVAT view, (3) other reduced-variance estimators arising out of the DIVAT analysis (such as *Money minus Dcash*), (4) asymmetric assignments of credit and blame to each player (such as *RedGreen points*), (5) comparison of ROE and AIE formulations of DIVAT, and (6) *smoothing* by averaging over a range of parameter settings.

4.1 Correctness and Variance Reduction

We now present a series of experiments to verify empirically that the DIVAT Difference is an unbiased estimate of the long-term expected value between two players. We will measure the overall reduction in variance in each case.

4.1.1 Always_Call versus Always_Raise Match

Figure 2 shows the ROE DIVAT Difference line for the two 100,000-game matches between Always_Call and Always_Raise. In each case, the DIVAT Difference line hovers near the zero line, showing much better accuracy than the money line, and showing no apparent bias in favour of one player over the other. The standard deviation for the money line is ± 6.856 sb/g in each case (as expected), whereas the measured standard deviation for the DIVAT Difference lines are ± 2.934 sb/g and ± 2.917 sb/g, respectively, for an overall reduction in variance by a factor of 5.50.

Figure 3 shows the results after extending the first match for an additional 300,000 games. Following the steady climb over the first 100,000 games, the money line displays a natural regression toward the mean, but later shows a relentless down-swing, reaching about -1.4 standard deviations. This graphically demonstrates that the full statistical interval is indeed used over the course of normal stochastic events.

There appears to be a positive correlation between the money line and the DIVAT Difference line (particularly over the last half of the second 100,000-game match, and in the middle regions of the 400,000-game series). This is expected, because the two measures are not completely independent. In general, strong hands are easier to play correctly than weak hands, because the EV consequences of an incorrect raise (or lack thereof) are generally smaller than the EV consequences of an incorrect fold (or lack thereof).

In this match, ALWAYS_RAISE makes frequent errors by betting instead of checking with weak hands. The ALWAYS_CALL player makes frequent errors by checking instead of betting strong hands (in first position), but those errors are effectively forgiven when the opponent bets. The ALWAYS_CALL errors of calling instead of raising with especially strong hands are distinct, as are the large EV errors of calling instead of folding with very weak hands. We know that the weighted sum of all misplays will exactly balance out, because the overall EV difference between these two players is zero.

Also in Figure 3, we show the DIVAT Cash (Dcash) baseline, which is the amount that would be won or lost if both players followed the DIVAT baseline policy for the entire game. Since Dcash is a reasonable estimate of the normal outcomes, we can see that *Money minus Dcash* would be a decent lower-variance estimator of the difference in skill, and is certainly much better than the money line alone. However, this estimate has a higher variance than the DIVAT Difference, is more strongly dependent on the stochastic luck of the card outcomes, and can only be applied to complete games. The Dcash line is also biased in favour of styles that are similar to the (overly honest) DIVAT baseline, whereas the ROE DIVAT Difference is a provably unbiased estimator.

The identical sequence of 400,000 deals was used for subsequent experiments, to eliminate sampling differences. We include the Dcash line in those graphs to provide a common frame of reference.

4.1.2 PSOPTI4 Self-play Match

Figure 4 shows the results for the PSOPTI4 self-play match. Here the Bankroll line represents the actual amount of money won or lost (in sb) when this particular player plays both sides of each hand. We can observe that the self-play money line is much closer to the Dcash line than it was for the essentially random ALWAYS_CALL *vs* ALWAYS_RAISE match, because the play is much more realistic. Nevertheless, the final Bankroll difference is about 35% greater in magnitude. This is also to be expected, because the baseline reflects a highly conservative "boring" style of play. PSOPTI4 has a relatively low-variance style itself, as evidenced by the measured standard deviation of "only" ± 4.691 sb/g. The measured standard deviation of the *Money minus Dcash* estimator is ± 2.825 sb/g in this match. The measured standard deviation of the DIVAT Difference is ± 1.863 sb/g, which is a 6.34-fold reduction in variance from the self-play money line.[26]

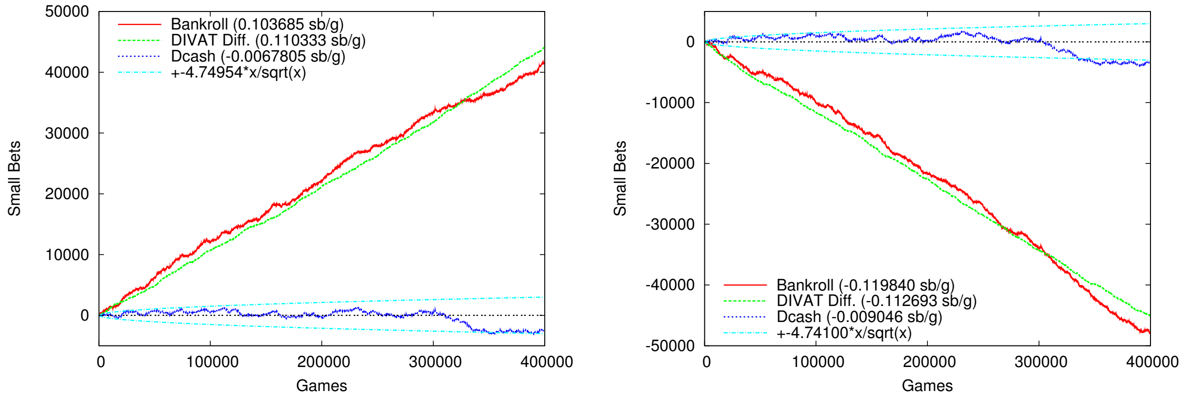
In practical terms, this means that routine 40,000-game matches could be replaced with 6000-game matches having comparable variance. More to the point, the same 40,000-game matches can discern between two players who are nearly three times closer to each other in skill level, instead of requiring matches of more than a quarter million games. With 95% confidence, a 40,000-game match distinguishes differences of roughly 0.06 sb/g, which represents a substantial difference in skill (such as a professional player against a mediocre player). Being able to discern a difference of 0.02 sb/g (such as the difference between a good player and a very good player) is much more useful in practice.

4.1.3 PSOPTI4 versus PSOPTI6 Duplicate Match

Figure 5 shows the results for the duplicate match between PSOPTI4 and PSOPTI6 over the same 400,000 deals. All graphs are shown from the perspective of the player in the North seat,[27] so PSOPTI4 wins both sides of this

[26] The DIVAT Difference line does not cling tightly to the zero line in this graph, but the meaningful indicator of EV correctness and low variance is simply the levelness (*i.e.*, horizontal flatness) of the line. Once it drifts away from zero, it should persist at the same constant difference. Obviously the DIVAT Difference line is much flatter than the money line overall.

[27] The North seat still alternates between first and second position (large and small blind). By convention, the first named player is North, having the big blind on all odd numbered hands and the small blind on all even numbered hands.



(a) PsOpti4 (North) *vs* PsOpti6 (South) (b) PsOpti6 (North) *vs* PsOpti4 (South)

Figure 5: DIVAT Difference analysis of PsOpti4 *vs* PsOpti6 duplicate match.

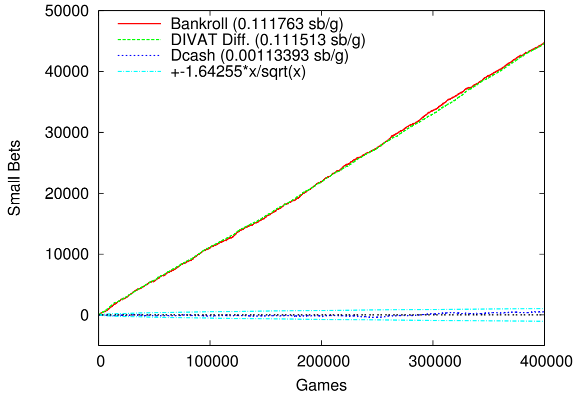


Figure 6: Duplicate Money Average and Duplicate DIVAT Average.

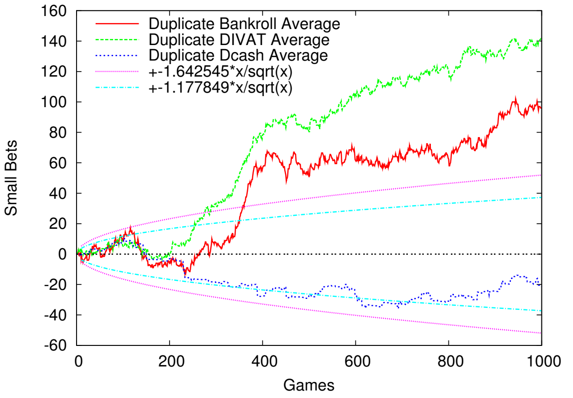


Figure 7: Duplicate Money Average and Duplicate DIVAT Average over the first 1000 games.

dual by a convincing margin of +0.112 sb/g. In both passes, the money line dips below the DIVAT Difference line at about the 300,000-game mark, because of the huge swing of luck in South's favour at that stage in the match.

The Bankroll difference is not a very accurate measure of skill difference even after 400,000 games. On the North side of the cards PsOpti4 wins at +0.104 sb/g, compared to +0.120 on the (stronger) South side; whereas the DIVAT estimate is within ± 0.0012 sb/g in either case. Using the Dcash line as a correction of the money line would considerably improve the accuracy. However, the DIVAT Difference is strictly more powerful for predicting future outcomes.

The measured standard deviations for Bankroll are ± 4.750 and ± 4.747 sb/g respectively. For *Money minus Dcash* they are ± 2.948 and ± 2.945 sb/g. For the DIVAT Difference, they are ± 1.929 and ± 1.932 sb/g respectively, for an overall reduction in variance by a factor of 6.05.

Note that the measurements for the two halves of this duplicate match are very consistent, because 400,000 games is sufficiently large for the number and variety of opportunities to be reasonably well balanced. Over a shorter duration, one side could enjoy more good opportunities, while facing relatively few difficult situations.

Figure 6 shows the results after combining the outcomes of each North and South pair of duplicate games. Both the Duplicate Money Average and the Duplicate DIVAT Average are good low-variance predictors of future outcomes, and match each other almost exactly over the long-term.[28] The measured standard deviation for Duplicate Money Average is ± 1.643, for a reduction in the normal variance between these two (fairly conservative) players

[28] The duplicate Dcash line shown in the figure is not exactly zero because of the way it is computed. In cases where a player folds, the roll-out equity is computed by applying the DIVAT baseline to all future chance outcomes, and taking the average. Thus if one player folds while the other continues with the hand, their Dcash lines will usually be slightly different.

by a factor of 8.36. The measured standard deviation for Duplicate DIVAT Average is ± 1.178, for a 16.25-fold reduction in variance, making it the lowest variance estimator we currently have.

Figure 7 zooms in on the first 1000 games of the pairwise duplicate games in Figure 6. Here we can see the limits of the discrimination power for each metric. The guide-curves show the one standard deviation bound for the Duplicate Money Average (at ± 1.643), and the narrower bound for the Duplicate DIVAT Average (at ± 1.178).

After 200 games, neither technique has detected a difference in skill between the players, with a net difference close to zero. It is fair to say that PsOpti6 had some "luck", in that the most telling skill differences were not yet exposed during that phase.

After 300 games, the Duplicate DIVAT Average is beginning to favour PsOpti4 (about +1.5 standard deviations), whereas the Duplicate Money Average is still inconclusive. There is sharp increase in both metrics between 300 and 400 games, where PsOpti4 is able to exhibit its superior skill quite a bit faster than usual. The non-zero duplicate Dcash line gives us a hint that some of that difference may be due to the folding behaviors of the two players (either PsOpti6 folds too easily to bluffs, or does not fold correctly when it is behind).

After 400 games, the Duplicate DIVAT Average is at about +3.4 standard deviations, and would conclude with more than 99% confidence that PsOpti4 is the better player heads-up against PsOpti6. The Duplicate Money Average cannot make the same conclusion even at the 95% confidence level. The same statements are still true after 1000 games.

This example happened to be quite favourable to the DIVAT method. In general, to reach the 95% confidence interval for this highly unbalanced (+0.112 sb/g) contest, the Duplicate Money Average would require about 870 games, while the Duplicate DIVAT Average would need about 450 games. For a live match, the normal single-sided DIVAT Difference would require about 1200 games, whereas the simple Bankroll difference would need about 7260 games on average.

A natural question is whether the two approaches are somewhat orthogonal to each other, and could be combined for an even sharper resolution. Unfortunately, we can easily deduce that there must be a substantial amount of overlap between the two techniques. As mentioned previously, the single-sided DIVAT baseline is especially useful for discerning the "normal" outcome when a strong second-best hand loses many bets to an even stronger hand. The Duplicate Money Average achieves a similar neutralization, because both players are given the opportunity to play the weak side and strong side of that situation.

Moreover, if one player gained on the DIVAT scale by losing less on the weaker side, that superior skill could also be reflected in the Duplicate Money Average by showing a net profit over the pair of duplicate games. In the case of a successful or unsuccessful bluff, we can see that the player is given full credit or full blame with either the DIVAT measure or the duplicate result.

For a pair of duplicate games, noise is introduced into the result after the actions of the two players diverge. In particular, when one player folds while the other continues with the hand, all chance outcomes from that point forward are subject to the usual effects of stochastic noise.[29] Directly measuring the loss or gain of equity from each decision yields a more stable estimate.

4.2 Learning Curves for Non-stationary Players

We now show how the DIVAT analysis can provide powerful insights into the changing behavior of non-stationary players, which includes virtually all strong human players, and the most advanced programs.

4.2.1 The 2003 "thecount" versus PsOpti1 Match

Figure 8 compares the money line to the ROE DIVAT Difference line for the 2003 match between "thecount" and PsOpti1.[30] We observe a similar improvement from ± 5.756 sb/g standard deviation for the money line to ± 2.109 sb/g for the DIVAT Difference, for a 7.45-fold reduction in variance. The *Money minus Dcash* estimate

[29] We see a similar source of variance in the *Money minus Dcash* estimator, whenever the assumed fold policy differs from actual events.

[30] Note that a complete-knowledge log is required for this analysis, including all cards folded by either player. The match was played on our online poker server, which maintains complete-knowledge logs of all games played.

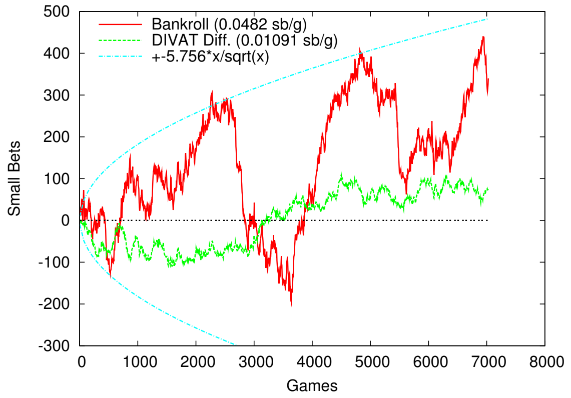


Figure 8: DIVAT Difference analysis of "thecount" *vs* PsOpti1.

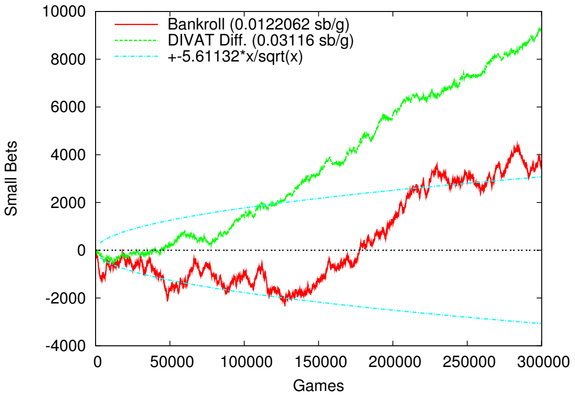


Figure 9: DIVAT Difference analysis of Vexbot *vs* PsOpti4.

had a standard deviation of ± 3.417 sb/g.

Overall, the DIVAT Difference line indicates that "thecount" exhibited a small advantage against PsOpti1 over the course of this match. However, it also suggests that the actual win rate should have been less than one quarter of what the money line indicates. In view of the high noise element, and given the estimated difference of only +0.01 sb/g, it is not difficult to imagine that the final result could have been dominated by short-term luck.

If we trust the DIVAT analysis, it tells a very different story of how the match progressed. First, it appears that PsOpti1 held a slight advantage in play early in the match. This is consistent with the comments "thecount" made after the match. He used an extremely aggressive style, which is effective against most human opponents (who can be intimidated), but turned out to be counter-indicated against this program. PsOpti1 won many games during that phase by calling to the end with relatively weak hands, frequently beating a bluff.

The human expert then went on an extended winning streak, but the DIVAT line suggests that the play was actually close to break-even during that phase. The dramatic collapse that occurred before game 3000 was almost entirely due to bad luck. However, that turnaround did cause "thecount" to stop playing, and do a complete reassessment of the opponent. He changed tactics at this point of the match, toward a more conservative style.

The DIVAT analysis indicates that this was a good decision, despite the fact that he continued to lose due to bad luck. Then the cards broke in favour of the human again, but his true advantage does not appear to have changed by much. Toward the end of the match, the two players appear to be playing roughly on par with each other. Regardless of whether this is a perfectly accurate synopsis of the true long-term expected values, one point is irrefutable: that it is almost impossible to discern the truth from the money line alone.

Interestingly, the DIVAT Difference line also appears to reveal the learning curve of "thecount" as the match progresses. The general upward bend of the DIVAT win rate suggests that the human expert was continuously adapting to the program's style, and learning how to exploit certain weaknesses. PsOpti1 is not an easy opponent to learn against, because it employs a healthy mixture of deceptive plays (bluffing and trapping). Nevertheless, it is a static strategy that is oblivious to the opponent, and is vulnerable to systematic probing and increasing exploitation rates over time.[31] The exposition of learning curves is one of several unplanned bonuses from the DIVAT analysis technique.

4.2.2 The Vexbot versus PsOpti4 Match

Figure 9 shows the DIVAT Difference line for the match between Vexbot and PsOpti4. The measured standard deviation is ± 5.611 sb/g for the money line, ± 3.671 sb/g for *Money minus Dcash*, and ± 2.696 sb/g for the DIVAT Difference, reducing variance by a factor of 4.33.

In this experiment, the adaptive program took a particularly long time to find a winning counter-strategy, and

[31] As a point of reference, the first author's win rate was approximately +0.3 sb/g against PsOpti1, after extensive experience. Against subsequent pseudo-optimal computer opponents, the win rate has been in excess of +0.4 sb/g.

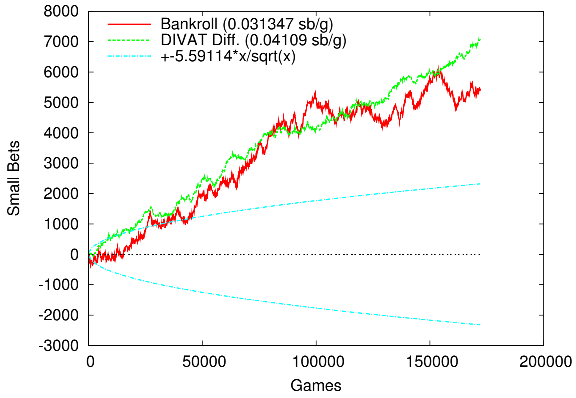


Figure 10: DIVAT Difference analysis over the last 170,000 games.

the strategy it finally adopted secured only a modest win rate. In other runs, VEXBOT discovered a much more effective exploitation much more quickly (often after only a few thousand hands).[32]

After 300,000 games, the money line is inconclusive (a little over one standard deviation), and close to meaningless (having come from minus one standard deviation, which could be due to normal drift, as we have seen in previous matches). In contrast, the DIVAT analysis is quite certain that VEXBOT is exhibiting an advantage. Moreover, the DIVAT line indicates that VEXBOT found the counter-strategy after about 80,000 games, whereas the money line does not start to run parallel until about 130,000 games. The 50,000-game lag phase is yet another demonstration of the misleading nature of high-variance stochastic outcomes.

Figure 10 ignores the early learning phase of VEXBOT by removing the first 130,000 games of the match. This re-calibrated view overlays the two lines, and shows that the win rate was over +0.03 sb/g from that point forward, while the DIVAT total was over +0.04 sb/g.[33] Furthermore, there is evidence of a reversal at about 80,000 games (210,000 games in total), where VEXBOT appears to slip into a less effective counter-strategy. VEXBOT continues to learn throughout a match, albeit at a decreasing rate because of "inertia" from the full match history. We can see from the slope of the DIVAT line that VEXBOT was winning at about +0.05 sb/g during the middle stages, but fell off to about +0.033 sb.g over the final 90,000 games.

4.3 Robustness of DIVAT Difference

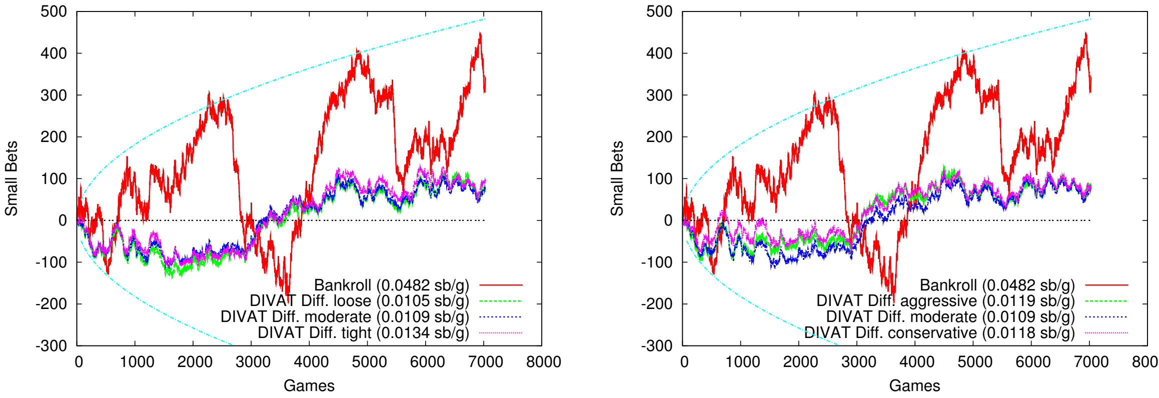
Over the course of many experiments, it was observed that even radical changes to the underlying policies and parameters had relatively little effect on the DIVAT Difference line. This indicates that it is a robust measurement – it is not overly sensitive to the exact construction of each component, nor to the precise settings of parameters.

Figure 11(a) shows three runs of DIVAT, holding all parameters constant except for the fold policy on the turn. These are varied to reflect a loose style (calling with rather weak hands that may or may not have any chance of winning), a tight style (surrendering in all borderline cases), and a normal well-balanced style between the two extremes. Simulations of 100,000 complete roll-out turn scenarios indicate that each increment increases the fold frequency by approximately 6.9% absolute (*e.g.*, for a pot-size of 4 sb: from 36.3% folds for tight, to 29.3% folds for normal, to 22.5% folds for loose).

Figure 11(b) shows three runs of DIVAT, holding all parameters constant except for the betting and raising thresholds on the turn. These are varied to reflect an aggressive style (having rather low standards for betting and raising), a conservative style (requiring a very solid holding), and a normal well-balanced style between the two extremes. Simulations of 100,000 complete roll-out turn scenarios indicate that each increment decreases the bet frequency by approximately 7.2% absolute (*e.g.*, from 45.0% betting for aggressive, to 37.2% betting for normal, to 30.5% betting for conservative).

[32] This illustrates another limitation of the duplicate match system. VEXBOT is capable of displaying many different personalities, from wildly aggressive and over-optimistic to passive and sullen. Since the sequence of cards on the opposing side provides a very different learning experience, there is no way of predicting which VEXBOT will show up in each run, thus nullifying some of the beneficial effects of opportunity equalization.

[33] Note that we have committed the statistical crime of selecting our endpoints, but only to serve an illustrative purpose.



(a) Varying DIVAT Turn Fold Offsets (+0.05, +0.10, +0.15)

(b) Varying DIVAT Turn Bet Thresholds (Make1 0.52, 0.58, 0.64)

Figure 11: Varying DIVAT parameters.

The lines are all close together, so clearly these settings did not have a significant impact on the overall DIVAT Difference in this match. This is not too surprising, in view of how the metric is being used. Even if the absolute measurements are somewhat inaccurate in certain cases, the same objective standards are always being applied equally to both players. To put it more colloquially, even a “crooked measuring stick” can be used to compare the relative lengths of two objects (and using a hockey stick or a driftwood walking stick could work equally well for that particular purpose).

If certain kinds of inaccuracies are common, they will tend to cancel each other out fairly quickly. Relatively infrequent opportunities might not balance out as quickly, but those sources of inaccuracy might not occur at all over the short-term, or might not have much impact on the overall score in any case.

Similar robustness results were observed for broad ranges of all parameters, over a large variety of match conditions. Even under extreme settings, the overall assessment appears to degrade gracefully. For example, there is relatively little change in the DIVAT Difference line until using fold policies that are obviously much too liberal, or much too restrictive (well outside the normal behavior of strong players). In some cases the settings might be inconsistent with the actual styles of the players involved, such as very aggressive betting parameters for a match between very conservative players, but this does not appear to lead to serious consequences or suspicious evaluations.

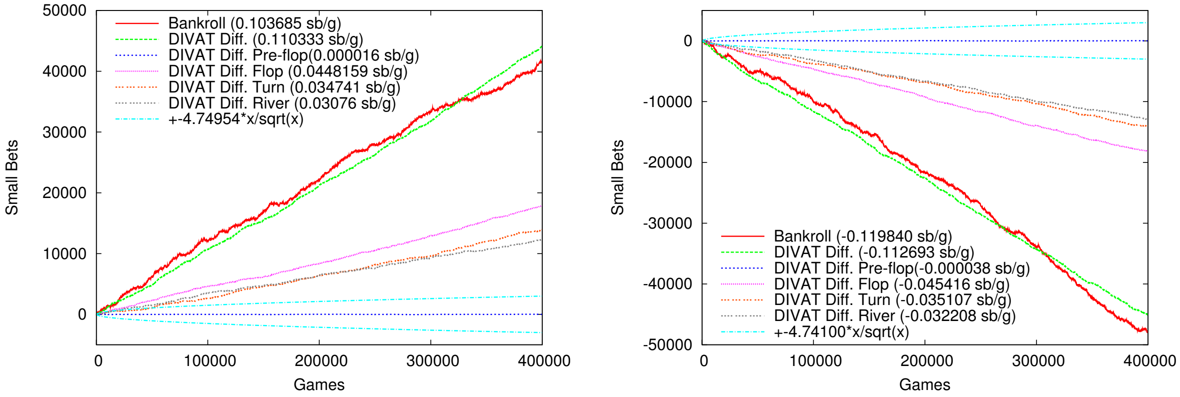
As previously mentioned, it has been shown that roll-out equity DIVAT is a statistically unbiased estimator of long-term expected value. Although this is certainly a desirable theoretical property, it is not absolutely essential for the assessment tool to have practical value. We have seen that stochastic noise can have disastrous consequences on evaluation, occluding the true averages even for very long matches. In practice, an assessment technique that has a lot of variance reduction power can be highly valuable, even if it cannot be formally proven to be unbiased. By design, the DIVAT Difference is largely insensitive to the fluctuations incurred from stochastic outcomes, imbuing it with considerable power for variance reduction, regardless of the precise formulation.

4.4 Round-By-Round DIVAT Analysis

The DIVAT system is designed for round-by-round analysis of each game, with the sum of all rounds characterizing the game as a whole. However, viewing the cumulative results for each round independently can be very enlightening.

Figure 12 shows the breakdown of DIVAT Differences for the pre-flop, flop, turn, and river in the PsOpti4 *vs* PsOpti6 duplicate match. The difference in pre-flop skill is shown to be minuscule, which is correct because both programs use the same (randomized) pre-flop expert system. The remaining rounds all contribute roughly equal portions to the total DIVAT Difference, meaning that PsOpti4 consistently outplayed PsOpti6 in every phase of the game.

The EV magnitude of decisions on the flop is generally much smaller than decisions in the last two rounds,



(a) PsOpti4 (North) *vs* PsOpti6 (South) (b) PsOpti6 (North) *vs* PsOpti4 (South)

Figure 12: Round-by-round analysis of PsOpti4 *vs* PsOpti6 duplicate match.

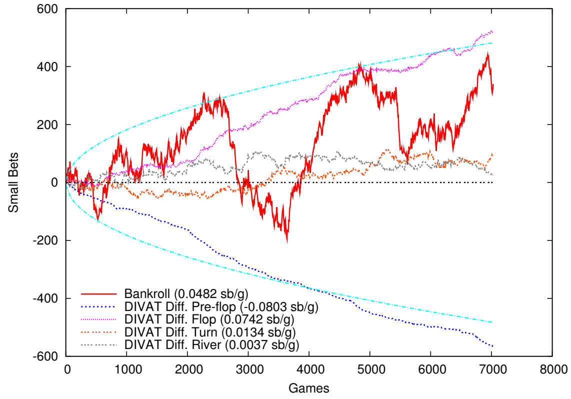


Figure 13: Round-by-round analysis of "thecount" *vs* PsOpti1.

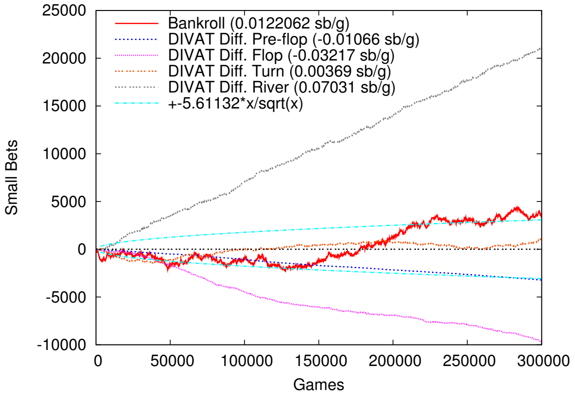


Figure 14: Round-by-round analysis of Vexbot *vs* PsOpti4.

because (1) the bet size is half as much (so the *pot odds* are roughly double), (2) pot equities are usually closer to 50%, meaning less EV is at stake, and (3) large implied odds further equalize the true equities. On the other hand, the flop occurs more frequently than the later rounds.[34] Similarly, decisions on the river usually count for a full big bet in equity, which is somewhat larger than decisions on the turn, but they are also somewhat less frequent.

Figure 13 shows the round-by-round breakdown for the match between "thecount" and PsOpti1, with some fascinating revelations. Of immediate note is the pre-flop DIVAT Difference line, which indicates that "thecount" was being badly outplayed before the flop. The loss rate of -0.08 sb/g is substantial, exceeding the standard per game win rate for most professional players. This is quite surprising, because the EV magnitude of pre-flop decisions is normally very small,[35] making it the least important round in two-player Limit Hold'em.

In reviewing the match log, the reason for the difference became apparent: "thecount" was folding too frequently before the flop (about 16% of the time). By refusing to fight with weak hands, he was sacrificing the 0.5 sb small blind (or 1.0 sb big blind) too often, incurring a large net loss. With this kind of objective analysis, even a good player can identify weaknesses or "leaks" in their strategy, and make appropriate adjustments.

On the flop betting round, the situation is reversed, with "thecount" holding a large edge in play. Although it has not been verified, it is possible that PsOpti1 was surrendering excessive amounts of equity by folding too easily after the flop. It is also possible that "thecount" was exhibiting a superior understanding of certain strategic aspects, such as the effects of implied odds. There could also be some compensation for his pre-flop selectivity, in that he is more likely to have the better hand when both players connect with the flop. Whatever the reasons,

[34] The graphs show the total (absolute) effect of each round, rather than the average (relative) magnitude of the differences.

[35] Most of the time one player is no more than a 60-40 favourite, and the small fraction of a bet in EV advantage is further diminished by the effects of implied odds.

the DIVAT analysis reveals a huge disparity in play, thereby identifying an area of weakness for researchers to investigate.

On the turn and river rounds, "thecount" maintained a small DIVAT advantage. Given the styles of the two players (folding many hands on the pre-flop and on the flop), the turn and river rounds occurred less often than usual, thus having a smaller effect on the final outcome. Again of interest is the fact that "thecount" appeared to improve his results on the turn as the match progressed, as he learned more about the opponent's predictable style and weaknesses. In comparison, he did not appear to find additional ways to exploit the opponent on the river round. The full-game DIVAT summary simply indicates a slight playing advantage for "thecount"; but the round-by-round analysis goes well beyond that, providing many additional insights into the reasons for the differences in equity.

Figure 14 shows the round-by-round breakdown for the match between VEXBOT and PSOPTI4, again with some surprising insights. VEXBOT held a negligible advantage from play on the turn, and was actually losing equity on the pre-flop and (especially) on the flop. However, those losses were more than offset by huge equity gains on the river.

To an expert observing the match, the meaning of this in terms of poker strategy is clear. VEXBOT was employing a hyper-aggressive "fast" style on the flop, putting in many raises and re-raises. This creates a false impression of strength, at a moderate expense in terms of EV, because the relative differences in pot equity on the flop are small. That false impression was then exploited with follow-up bluffs on the river, when PSOPTI4 folded too often based on its implicit beliefs about the strength of the opponent's hand. A casual observer might conclude that VEXBOT is a "maniac" based on its play on the flop, but the DIVAT analysis clearly shows that there is a method to its madness.

Moreover, it is evident that VEXBOT discovered this imbalance in the strategy of PSOPTI4 very early in the match, since the river DIVAT score maintains the same slope from the beginning. The change at 80,000 games was in fact due to a shift in style on the flop, where VEXBOT eventually learned that it did not need to put in as many extra raises to set-up the same exploitative plays on the river.

5. CONCLUSIONS

Stochasticity is a major impediment to the accurate assessment of player skill in poker. The simple money outcome of a short-term match is highly unreliable – many thousands of games are required to obtain statistically significant results. Some variance-reduction methods are available, such as duplicate tournaments, but they do not eliminate the problem of assessment in normal live play.

The Ignorant Value Assessment Tool provides an accurate unbiased estimate of the long-term expected value between two players. DIVAT is a practical system that provides a significant reduction in variance, thus extracting signal from the noise. DIVAT uses hindsight analysis to quantify the difference in value between the players' actual actions and a standard benchmark betting sequence. The comparison sequence is based on game-theoretic invariant frequencies and quasi-equilibrium policies, reflecting an appropriate amount to be invested by each player in the given situation. Although much of the relevant context is largely ignored (the previous rounds of betting in particular), the most important aspects are estimated adequately enough for the system to be highly effective in practice.

DIVAT is versatile in its uses, and promises to be an important tool for all researchers working in the domain of poker. The results from a match can be broken down in many ways and analyzed with DIVAT. For example, the player position can be isolated (separating the games played as the first player from those as the second player). We have also used a variation of the tool, called *runtime DIVAT*, to provide more meaningful feedback during matches (where special consideration must be given for the biases caused by the partial observability in live games). There are numerous other uses for the DIVAT analysis technique that have not been addressed here, in the interest of brevity.

The first extension to DIVAT will likely be for multi-player games of Limit Hold'em. In principle, this generalization should be much easier than other two-player to multi-player generalizations (such as game-theoretic Nash equilibria, or imperfect information game-tree search approaches). However, there will invariably be some theoretical and practical issues that will need to be resolved.

Developing DIVAT systems for other betting structures (*e.g.*, Pot Limit, No Limit), and other poker variants (*e.g.*, Omaha, 7-card Stud) should be fairly straight-forward. Imperfect information games with a known equilibrium strategy can employ similar methods, using a weighted mixture of baselines corresponding to the relative weights of actions in each mixed strategy.

Applying the general methods to other imperfect information domains is feasible in principle, but might encounter new obstacles due to fundamental differences in the structure of the game-trees. For example, games in which the chance nodes and decision nodes are intertwined and not easily separable could be problematic.

Acknowledgments

This research was supported by the Izaak Walton Killam Memorial Scholarship Foundation, the Alberta Heritage Scholarship Fund, the Alberta Informatics Circle of Research Excellence (iCORE), the Alberta Ingenuity Center for Machine Learning (AICML), and the Natural Sciences and Engineering Research Council of Canada (NSERC-PGS).

6. REFERENCES

Billings, D. (1995). Computer Poker. M.Sc. thesis, Department of Computing Science, University of Alberta.

Billings, D. (2003). Vexbot wins Poker Tournament. *International Computer Games Association Journal*, Vol. 26, No. 4, p. 281.

Billings, D. and Bard, N. (2005). Compact Pre-computed Hold'em Tables. Unpublished.

Billings, D., Burch, N., Davidson, A., Schauenberg, T., Holte, R., Schaeffer, J., and Szafron, D. (2003). Approximating Game-Theoretic Optimal Strategies for Full-scale Poker. *The Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, IJCAI/AAAI'03*, pp. 661–668.

Billings, D., Davidson, A., Schaeffer, J., and Szafron, D. (2002). The Challenge of Poker. *Artificial Intelligence*, Vol. 134, Nos. 1–2, pp. 201–240.

Billings, D., Davidson, A., Schauenberg, T., Burch, N., Bowling, M., Holte, R., Schaeffer, J., and Szafron, D. (2004). Game-Tree Search with Adaptation in Stochastic Imperfect-Information Games. *Computers and Games: 4th International Conference, CG'04, Ramat-Gan, Israel, July 5-7, 2004. Revised Papers* (eds. H. J. van den Herik, Y. Björnsson, and N. Netanyahu), Vol. 3846 of *Lecture Notes in Computer Science*, pp. 21–34, Springer-Verlag GmbH.

Billings, D. and Kan, M. (2006). Development of a Tool for the Direct Assessment of Poker Decisions. Technical Report TR06-07, University of Alberta Department of Computing Science.

Frank, I. and Basin, D. A. (2001). A Theoretical and Empirical Investigation of Search in Imperfect Information Games. *Theoretical Computer Science*, Vol. 252, No. 11, pp. 217–256.

Ginsberg, M. (2001). GIB: Imperfect Information in a Computationally Challenging Game. *Journal of Artificial Intelligence Research*, Vol. 14, pp. 303–358.

Sheppard, B. (2002a). *Toward Perfection of Scrabble Play.* Ph.D. thesis, Computer Science, University of Maastricht.

Sheppard, B. (2002b). World-championship-caliber Scrabble. *Artificial Intelligence*, Vol. 134, Nos. 1–2, pp. 241–275.

Sklansky, D. (1992). *The Theory of Poker.* Two Plus Two Publishing.

Tesauro, G. (1995). Temporal Difference Learning and TD_Gammon. *Communications of the ACM*, Vol. 38, No. 3, pp. 58–68.

Tesauro, G. (2002). Programming Backgammon Using Self-Teaching Neural Nets. *Artificial Intelligence*, Vol. 134, Nos. 1–2, pp. 181–199.

Wolfe, D. (2002). Distinguishing Gamblers from Investors at the Blackjack Table. *Computers and Games 2002* (eds. J. Schaeffer, M. Müller, and Y. Björnsson), LNCS 2883, pp. 1–10, Springer-Verlag.

Zinkevich, M., Bowling, M., Bard, N., Kan, M., and Billings, D. (2006). Optimal Unbiased Estimators for Evaluating Agent Performance. *American Association of Artificial Intelligence National Conference, AAAI'06.* To appear.