

POKER AI · EVALUATION · BENCHMARK DESIGN

What Does It Mean for a Poker Bot to Be Good?

Ranking, Exploitability and Benchmark Design in Poker AI

Henry De Courcy Thompson, FullHouse Research

ABSTRACT

This note examines a deceptively difficult question in poker AI: what should it mean for one agent to be *better* than another? Tournament profit is intuitive, but it is field-dependent and can conceal severe exploitability. Equilibrium-based notions offer stronger guarantees in heads-up poker, yet become less decisive in multiplayer settings. Drawing on the Annual Computer Poker Competition, Local Best Response analysis and population-evaluation work, the note argues that combining these properties into a single weighted score merely relocates subjective judgement. A more interpretable approach is to evaluate expected value over a deliberately designed distribution of strategic environments, while making the remaining benchmark choices explicit and defensible.

Keywords — poker AI, agent evaluation, exploitability, benchmark design, Nash equilibrium

Introduction

When considering the role of poker hackathon leader, one of my first considerations was how to effectively rank the contestants in a fair and representative way. Having read about frontier AI benchmark development and the risks of overfitting, this struck me as a potentially significant nuance, which would have to be handled carefully. The following is the outcome of my research into poker AI ranking and benchmarking, and the challenges associated with it.

The naive benchmark

We begin with the naivest possible approach, which a novice might propose: simply have the agents play a 'realistic' poker game, maybe $\sim 1,000$ random hands, and rank them according to their profit or loss. I hope that the problem is immediately obvious to most readers; anyone who's played knows that poker is statistically very noisy. Over a small or medium sample size, card distribution, a few large pots and some good luck will overwhelm marginal strategic differences. A weaker agent could easily finish above a stronger one through variance alone.

Fortunately, the solutions to this are very intuitive: run far more hands, implementing 'duplicate poker'. For each set of dealt cards, the cards are fixed at the corresponding table position and the agents are rotated until each agent has played each hand. This slashes the card-related variance, as we can compare agents in the same spot. (Bard et al., 2013)

So variance is one problem, but not the most troubling as it is quantifiable and therefore can be adjusted for intelligently. But even if we could eliminate variance completely, we run into a more ambiguous problem. What exactly should we be measuring to assess performance?

Assuming minimal variance, the large-simulation method may now seem quite attractive. Indeed, various early agentic poker competitions used this methodology to decide winners and to benchmark frontier models. To determine the winner of a competition or tournament, this is often sufficient, however for a more general benchmark or performance metric it has many vulnerabilities.

Benchmarking the field

For one, this measurement is intrinsically dependent on the field that the agent is playing against, and its strategy distribution. If many submissions are variants of the same strategy, or in general if the strategy distribution of the field isn't representative of the target population, this risks skewing the measurement towards a certain strategy. [Balduzzi et al. \(2018\)](#) identify redundancy invariance as a desirable property of an evaluation system, and propose Nash averaging as one way of achieving it in their agent-vs-agent setting.

A closed tournament only needs to identify the strongest agent relative to its entrants. A benchmark claiming to measure general poker ability has a harder job: its result should remain meaningful beyond the particular field on which it was measured. Furthermore, a limited field does not stress-test exploitability at all. A sophisticated strategy that beats other serious agents may be vulnerable to an automatic all-in agent, but such a naive agent is unlikely to enter the field of course.

Exploitability: strong against whom?

This brings us to our first case study, courtesy of [Lisý and Bowling \(2017\)](#). In this paper, LBR (Local Best Response) is proposed as a cheap way of quantifying exploitability, giving a lower bound on how much a best-response opponent could win. Across the four ACPC agents they evaluated, including Slumbot 2016 and Act1 2016, they established with 97.5% confidence a lower bound on exploitability above 3.18 BB/hand (big blinds per hand). Against a worst-case opponent, even an auto-folding agent would lose at least four times less than these sophisticated strategies. Even an LBR restricted to fold/call could profit against all four.

Naturally, if no adversaries find these exploits, the agents are safe and continue winning, but Lisý and Bowling showed that their strategic dominance was surprisingly precarious. This reinforces the notion that although field-relative performance may be sufficient to decide a closed tournament winner, it should not be the only metric considered for a general benchmark.

Ranking rules define the winner

Even within tournaments, there are various issues and nuances to consider when declaring a winner. The ACPC (Annual Computer Poker Competition) ran into such a hurdle when it investigated total bankroll versus instant runoff methods. [\(Bard et al., 2013\)](#) Total bankroll rewards total expected value (EV) against the whole field, which can be skewed by being best at exploiting the weakest agents, for example. Instant runoff ranks the field, eliminates the weakest agent and re-ranks, rewarding robustness against stronger opponents more. Both have merits and real-world correspondence, but can produce different rankings from the same set of agents because they reward different notions of 'best'. Organisers must therefore decide exactly what they want to measure and reward, as rankings are sensitive to various valid metrics.

Multiplayer poker complicates equilibrium

Heads-up poker has one major theoretical luxury: as a two-player zero-sum game, an exact Nash-equilibrium strategy guarantees at least the value of the game in expectation, regardless of how the opponent plays. Unfortunately for game theorists (or fortunately, for the masochists, deep learning enthusiasts, or heuristics/abstraction enjoyers), multiplayer poker loses this unusually strong safety property. As [Brown and Sandholm \(2019\)](#) observe in their Pluribus paper, Nash equilibrium has important shortcomings outside the two-player zero-sum setting, with no equally convincing replacement solution concept. Pluribus was therefore ultimately judged empirically: could it consistently beat strong human opponents?

Why not combine the metrics?

At this point, a seemingly clean solution presents itself. If field performance and exploitability measure different but desirable properties, why not simply combine them?

Let P denote an agent's performance against the evaluated field and X its estimated exploitability. We could define some composite score

$$S = \alpha P - \beta X$$

and rank the agents accordingly. An agent would therefore be rewarded both for winning against the field and for avoiding obvious strategic weaknesses.

This is attractive because it seems to resolve the ambiguity while still giving us the single ordering that a competition ultimately requires. Unfortunately, it mostly shifts the problem elsewhere.

The first issue is the choice of weights. How much additional field performance should compensate for additional exploitability? If one bot wins an extra 1 BB/100 against the competition field but can be exploited for an additional 0.5 BB/100 by a targeted adversary, which should we prefer? There is no natural exchange rate between these quantities. Choosing α and β therefore embeds our own judgment and biases into the objective notion of poker skill.

Normalisation creates the same problem in disguise. Field performance and exploitability may occupy very different numerical ranges, so the method used to standardise them can introduce the same biases.

More importantly, once the formula is published, it becomes the contestants' optimisation target. Rather than simply attempting to build a strong poker agent, contestants are incentivised to maximise the particular function

$$\alpha P - \beta X.$$

This is not necessarily disastrous, and it's very possible that contestants will define a similar objective function or benchmark themselves and try to maximise it during development. Every competition metric creates incentives. But if our aim is to measure general poker ability, constructing an arbitrary exchange rate between fundamentally different properties seems unsatisfying, and could lead to overfitting to these parameters rather than aiming to simply build the strongest poker bot.

Designing the evaluation environment

This suggests a cleaner approach. Rather than complicating the scoring formula with implicit biases, we can make the evaluation environment more representative.

Let D_{eval} denote a distribution of evaluation scenarios, including table composition, positions, stack sizes and more. We measure how much the agent expects to win against a representative distribution of strategies.

For agent i ,

$$P_i = \mathbb{E}_{x \sim D_{\text{eval}}} [EV(i \mid x)].$$

The submitted field may form part of D_{eval} , but should be augmented with fixed benchmark agents, strategically distinct archetypes and limited adversarial strategies to encourage robustness.

There is a further complication: if the exact composition of D_{eval} is fixed and known, agents can be tuned specifically against the benchmark rather than for broader strategic strength. Where generalisation matters, some combination of held-out opponents, refreshed reference agents or undisclosed evaluation scenarios may therefore be useful.

Notice that this is still not perfectly objective, but rather shifts the bias yet again. Choosing which strategies to include, how to cluster similar agents and how much weight each group receives all require human judgement. In this configuration, the benefit is that the bias is more explicit and justifiable than defining arbitrary exchange rates or coefficients.

Another benefit is that this benchmark can also be stress-tested. If small, reasonable changes to D_{eval} repeatedly change the winner, then the ranking itself is fragile.

So the goal is not to eliminate judgement, but to make it explicit, stable and defensible.

Conclusion

The statistical difficulty in ranking poker, which is initially more obvious, is insignificant compared to the judgmental ranking decisions. Field EV and exploitability capture different properties. A bot can dominate a particular field while remaining strategically fragile, while another can be much harder to exploit but less effective at extracting value from weaker opponents.

A weighted composite appears to solve this, but mostly replaces one judgement call with another through arbitrary weights and normalisation. A more interpretable approach is to keep the primary metric simple and design the evaluation distribution carefully, which is easier to make explicit and stable under stress-testing.

There is no completely neutral definition of poker skill. Any ranking system encodes choices about what kinds of strength matter. The objective is not to find a perfect formula, but to design an experiment in which the strategies we consider strong are the ones most likely to rank highly.

The objective is not to find a perfect formula, but to design an experiment in which the strategies we consider strong are the ones most likely to rank highly.

References

Sources cited throughout this note.

Re-evaluating evaluation

D. Balduzzi, K. Tuyls, J. Pérolat, and T. Graepel. *Advances in Neural Information Processing Systems 31*, pp. 3272–3283, 2018.
<https://arxiv.org/abs/1806.02643>.

The Annual Computer Poker Competition

N. Bard, J. Hawkin, J. Rubin, and M. Zinkevich. *AI Magazine*, 34(2):112–114, 2013.
doi: 10.1609/aimag.v34i2.2474.

Superhuman AI for multiplayer poker

N. Brown and T. Sandholm. *Science*, 365(6456):885–890, 2019.
doi: 10.1126/science.aay2400.

Equilibrium approximation quality of current no-limit poker bots

V. Lisý and M. Bowling. In *AAAI Workshop on Computer Poker and Imperfect Information Games*, pp. 361–366, 2017.
<https://arxiv.org/abs/1612.07547>.