

Elo+Chess の層別 Lichess ベンチマーク曲線の構築

Elo+Chess 研究ノート

2026 年 5 月 31 日

Abstract

Elo+Chess ベンチマーク レポートは、大規模な層別サンプルから構築されています。Lichess.org ゲーム。このノートで使用されているベンチマーク サンプルには次のものが含まれます。483,974 1+0 バレット ゲーム、479,974 3+0 ブリッツ ゲーム、925,636 10+0 ラビッド ゲーム、660,540 の長時間の高速ゲーム。主要なゲームの種類ごとにパイプラインが拡大します すべてのゲームで、移動前と移動後のボードの状態を計算し、数百ものデータを計算します。中間変数とプレーヤーゲーム変数。これらの変数を次のように集約します。コーチングメトリクスを抽出し、それらのメトリクスを Lichess Elo バケットによって平均化します。の サンプルは生のプレーヤーに比例するのではなく、意図的に階層化されています プール: 目的は、各評価レベルでの安定した動作曲線を推定することです。自然な状態を再現するために、まばらになるバケツも含めます。rating distribution.結果は、一連の経験的なベンチマーク曲線になります。各指標ごとに、さまざまな Elo レベルのプレーヤーが特定のパフォーマンスを示す頻度を尋ねることができます。実際のゲームにおける習慣や原則。このノートでは、その方法、規模について説明します。計算といくつかの キャスリング、マイナーピースの開発、繰り返しのオープニングピースの動きと、初期のクイーンの動き。The emphasis is 方法論: Elo+Chess はチェスの原則が普遍的であるとは想定していません。すべてのスキル レベルにわたって直線的です。代わりに、ベンチマーク曲線は以下によって検査されます。レーティング範囲が広く、初心者から初級者までの領域に焦点を当てた製品です 関係が最も強く、最も解釈しやすい場所です。

1 目的

ベンチマーク システムの目標は、広範なチェスのアドバイスを、測定可能な量。「マイナーな作品を発展させなさい」というような発言は、便利ですが、マイナーピースがいくつかあるかを明らかにできると、より実用的になります。プレーヤーは通常、10手目までに成長しますが、それを同じプレーヤーと比較してどうですか 評価バケットと、その習慣が評価範囲全体でどのように変化するか。

したがって、Elo+Chess は実際のゲームからメトリクス ライブラリを構築します。それぞれの指標は、すべてのプレーヤー ゲーム行について計算され、ゲーム タイプと Elo バケットごとに要約され、レポートのベンチマーク曲線として使用されます。その後、レポートでユーザーを次の場所に配置できます。同じ曲線を表示し、ユーザーがその行動の上、下、または近くにいるかどうかを示します 同等のプレーヤー間で観察されました。

2 ソースデータと層別化

ベンチマーク ソースは、パブリック Lichess.org ゲーム データベースです。Elo+Chess が使用する 小さなサンプルではなく、ゲームの種類と評価バケットごとに大規模な層別サンプル 便利

なサンプル。現在の実稼働ランタイム アーティファクトは 4 つの主要な要素をカバーしています。ゲームのカテゴリ:

- 1+0の弾丸、
- 3+0 ブリッツ、
- 10+0 急速、
- rapid games longer than 10 minutes.

階層化の目標は、評価レベル全体にわたる幅広さです。比例ランダム サンプルはプレーヤープール全体を説明するのに効率的ですが、ベンチマーク曲線を構築するには非効率的です。最も一般的な評価バケット サンプルの大部分を占めますが、サポートが低いテールバケットでは数が少なすぎます。安定したメトリクス平均を実現するゲーム。したがって、Elo+Chess は平均的なゲームでサンプリングします エロバケツ。現在のクリーンサンプル構築では、生のゲームが最初に行われます。既知のプレーヤー、既知のレーティング、有効な結果が含まれる有効なレーティング ゲームにフィルタリングされます。そして既知の時間制御。その後、十分なゲーム数を持つプレイヤーにゲームがフィルタリングされます 一貫したスキル レベルを表すのに十分な速度と安定した評価が得られます。各プレイヤーについて、Elo+Chess はプレイヤーの速度内定格スパンを計算します ($\max(\text{Elo}) - \min(\text{Elo})$) はソース プール全体で保持されます。両方のプレイヤーが少なくとも設定された最小ゲーム数を持っている場合にのみゲームを実行し、速度内ドリフトが 300 Elo ポイントを超えないこと。ドリフトフィルターは意味がありません 通常の改善や差異を除去するため。プレイヤーの ゲームでは、暫定的なアカウントなど、実質的に異なるスキル レベルが混在することになります。復帰したプレイヤー、または期間中に複数のバケットを急速に移動したアカウント サンプリングされた期間。最後に、対象となるゲームが各平均内でランク付けされます-Elo ゲーム ID とシードの決定論的ハッシュ、および制限された数によるバケット ゲームは各バケットから保持されます。

この設計により、ベンチマーク曲線がさらに便利になります。各Eloバケット 各メトリクスの平均値を推定するのに十分な観測値に貢献し、まれに高評価または低評価のバケットが自然な形状にかき消されることはありません アクティブプレイヤーの分布。Elo+Chess で使用されるベンチマーク曲線は次のとおりです。次に、製品に最も関連する範囲に限定し、ほぼ初心者向けに説明します。初期の高度なプレイを通して。Lichess スケール レポートでは、そのプライマリ範囲 600 ~ 1600です。ユーザーが Chess.com スケールでレポートを表示すると、対応する評価バケットは、同じ基礎となる Lichess にマッピングされます。別のクロスプラットフォーム評価マッピング方法を使用したベンチマークバケット 付属の論文で説明されています: [Estimating Cross-Platform Chess Rating Mappings with モーダル回帰](#)。

3 アクティブプレイヤーの評価分布

終了時点でのLichess.orgプレイヤーのElo評価の分布を見ると 2026 年 3 月の主要なゲーム タイプ別の数とパーセンタイルは次のとおりです。この分布ビューでは、各プレイヤーは正確なゲーム タイプごとに 1 回カウントされます。対応する生データでそのプレイヤーの最新の観測レーティングを使用する 正確なゲームで少なくとも 5 つのゲームが必要な後のタイムコントロール行 タイプ。¹

¹We apply a five-game minimum because lightly active accounts disproportionately remain at the 1500 starting/provisional anchor, creating an artificial spike in all-player latest-rating distributions. Figures 1-4 このフィルターを使用した場合と使用しない場合の分布を示します。

Game type	Players	Median	75th pct.	80th pct.	% ≤ 1600	Exact 1500	1500 share
1+0 bullet	1,221,146	1,479	1,858	1,951	58.9%	4,756	0.4%
3+0 blitz	1,144,045	1,542	1,845	1,912	54.9%	2,315	0.2%
10+0 rapid	1,679,401	1,381	1,691	1,765	68.3%	3,472	0.2%
>10 rapid	668,255	1,400	1,686	1,753	68.2%	7,969	1.2%

Table 1: 最新の観察されたアクティブプレイヤーの正確な Lichess 評価分布 正確なゲーム タイプで少なくとも 5 つのゲームが必要な後のゲーム タイプ。

この分布を層別ベンチマーク サンプルと混同しないでください。このディストリビューションでは、生の状態で観察されたアクティブな Lichess プレーヤー プールについて説明します。スキャンします。次に、ベンチマーク サンプルは意図的に階層化され、Elo バケットが安定した計量曲線を十分にサポートします。

特定のゲーム タイプの Lichess.org Elo レーティングから 同等の Chess.com 評価、Elo+Chess はゲーム タイプ固有のモダルを使用します コンパニオンで推定された回帰式 [Elo+Chess rating-mapping paper](#):

$$\begin{aligned}
\hat{R}_{\text{Chess.com,bullet}} &= -531.71 + 0.9871 R_{\text{Lichess}}, \\
\hat{R}_{\text{Chess.com,blitz}} &= -551.54 + 1.0853 R_{\text{Lichess}}, \\
\hat{R}_{\text{Chess.com,10minrapid}} &= -495.04 + 1.0741 R_{\text{Lichess}}, \\
\hat{R}_{\text{Chess.com,>10rapid}} &= -369.02 + 0.9374 R_{\text{Lichess}}.
\end{aligned}$$

これらの方程式は、パーセンタイル ランクではなく、評価値をマップします。パーセンタイルランクは、特定のプレーヤー プール内で定義され、Chess.com および Lichess プレーヤー プールにはさまざまな形があります。Table 2 では、10 分間のラピッド ゲーム タイプ。最初の列には、考えられるものの小さなサンプルが示されています。Chess.com Elo 評価値。2 番目の列は対応するものを示します。2026 年 5 月 31 日現在の Chess.com に表示される Chess.com パーセンタイル ランク。71.7

10 分間の急速な Lichess.org のランクは、Lichess Elo レーティングに相当します。1,644² であり、その値は 3 列目。Chess.com-Lichess マッピング式を適用すると、1,644 になります。Lichess は、Chess.com の約 1,271 に相当します。推定値を仮定すると、マッピング式は合理的であり、Chess.com 評価間の +522 ギャップは次のとおりです。列 1 と列 4 のマッピングされた評価は、多くの数で説明できます。Chess.com プレーヤー プール内のカジュアル プレーヤーの数が Chess.com を膨らませています パーセンタイルランク。言い換えれば、Lichess.org の 72 パーセンタイルに含まれることは、Chess.com で 72 パーセンタイルに入るよりも大きな成果です。パーセンタイル ランクは、あるシステムから別のシステムにきれいにマッピングされません。実際のエロ 値は、同じプレイヤーのスコアから推定される方程式を使用してマッピングできます。2つのシステム。

Table 2 にギャップがある理由は、マッピングが 同等のプレー強度評価を推定しますが、パーセンタイルは 1 つのプラットフォーム自体の母集団内での位置。71.7パーセンタイルのプレーヤー したがって、Chess.com は自動的に Lichess の 71.7 パーセンタイルのプレーヤーにはなりません。そしてその逆も同様です。観察された数値は、Chess.com の公開データと一致しています。ラピッドプールには、より多くの低評価またはカジュアルプレイヤーの数が含まれていますが、アクティブな正確な制御 Lichess プールは経験豊富なリピート層に集中していま

²Calculated from the cumulative distribution function of the ZXQPROT3ZXQ 10+0 rapid latest-rating sample after requiring at least five games in that exact game type; see Table 1 and Figure 3.

Chess.com rating	Chess.com pct.	Same-pct. Lichess	Mapped equiv.	Gap
749	71.7%	1,644	1,271	+522
1,012	86.2%	1,871	1,515	+503
1,358	95.4%	2,095	1,755	+397
2,383	99.9%	2,557	2,251	-132

Table 2: Chess.com 10 分間の急速パーセンタイル間の比較例 アクティブな Lichess 10+0 Rapid のポイントと同じパーセンタイル ポイント 少なくとも 5 つの 10+0 ゲームが必要な場合は配布。「マッピングされた Chess.com」 equiv.” は現在の Elo+Chess を適用します 10+0 の評価スケールを同じパーセンタイルの Lichess 評価に迅速に変換します。の スケールの区別を示すために最も高い尾部の行が含まれていますが、近似された 評価ラインは、主要なコーチング範囲で最も信頼できます。

す 選手たち。重要な点は、1 パーセンタイルのスケールが「正しい」ということではなく、他は「間違っている」。それは、パーセンタイル ランクがプラットフォーム間で移植できないことです。ただし、基礎となるプレイヤーの分布が同じ形状である場合は除きます。

Note on the 1500 spike. The visible jump near 1500 は、スムージングによって作成されたものではなく、データ内に存在します。最近では 観察された評価分布、正確な 1500 の評価は 1+0 の 13.1

バレットプレイヤー、3+0 ブリッツプレイヤーの 5.6

より長い速攻プレイヤーの 17.6

評価システムにおける開始アンカーまたは暫定アンカー。

ちょうど 1500 件のスパイクは、主に、軽度アクティブなアカウントの影響によるものです。に これを確認してください。要求した後、同じ最新の評価分布を再計算しました。各プレイヤーは、正確なゲーム タイプで少なくとも 5 つのゲームを行う必要があります。スパイクはほぼ このフィルタの下では消えます: 正確な 1500 の評価は 5.6 ~ 17.6

5 試合以上の全プレイヤー分布内のプレイヤーは 0.2-1.2

分布は Table 1 で報告されます。

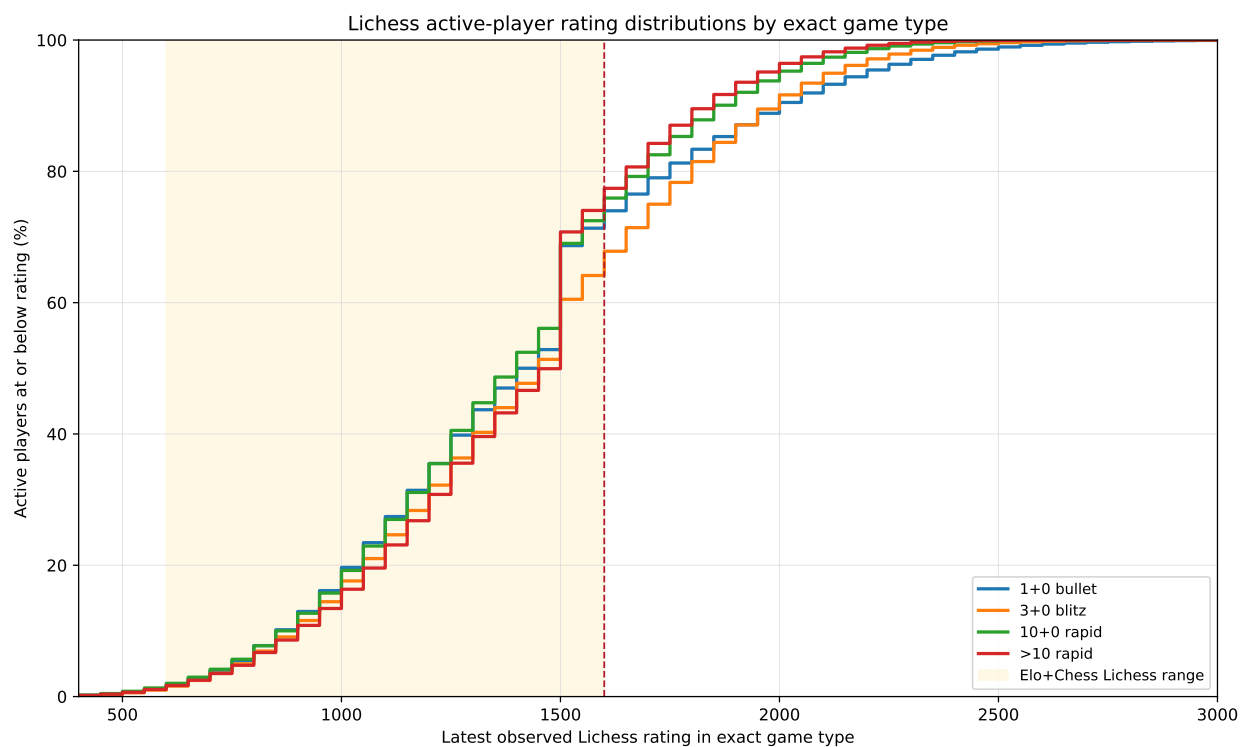


Figure 1: 正確なゲームタイプごとのアクティブプレイヤーの累積評価分布。の黄色のバンドは、メインの Elo+Chess で使用される 600 ~ 1600 Lichess 範囲をマークします。コーチングのベンチマーク。

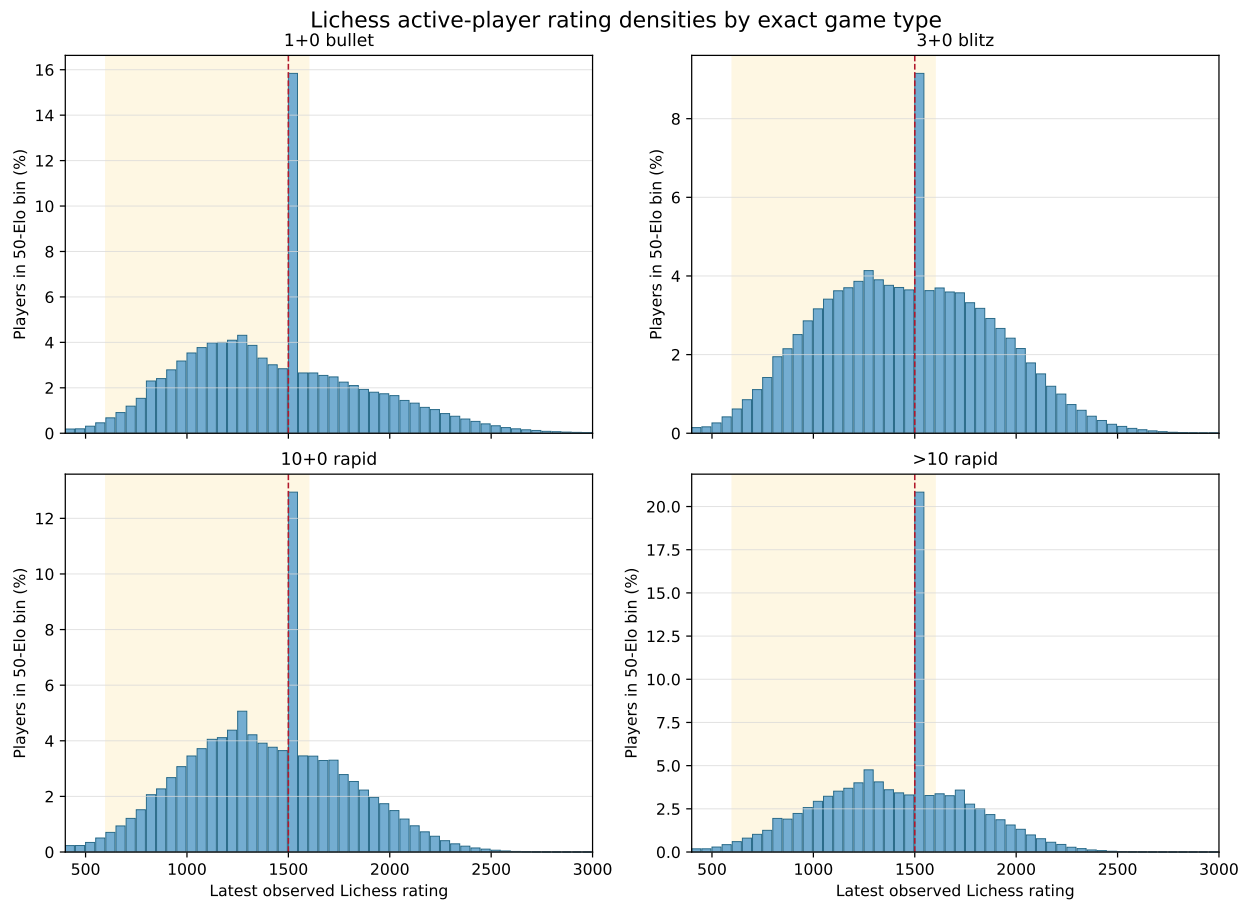


Figure 2: 同じアクティブプレイヤーの評価分布の密度ビュー。の 破線は 1500 Lichess をマークします。ここでは、生の最新評価分布が 大きな精密評価質点。黄色のバンドは 600–1600 Lichess を示します。メインの Elo+Chess メトリック レポートで使用される範囲。スパイクが最も目立つのは、多くの軽くアクティブなアカウントが正確に維持されるため、全プレイヤーの密度が高くなります。1500;図 4 は密度プロットを繰り返します 正確なゲーム タイプで少なくとも 5 つのゲームが必要な後、これにより大幅に削除されます このアーティファクト。

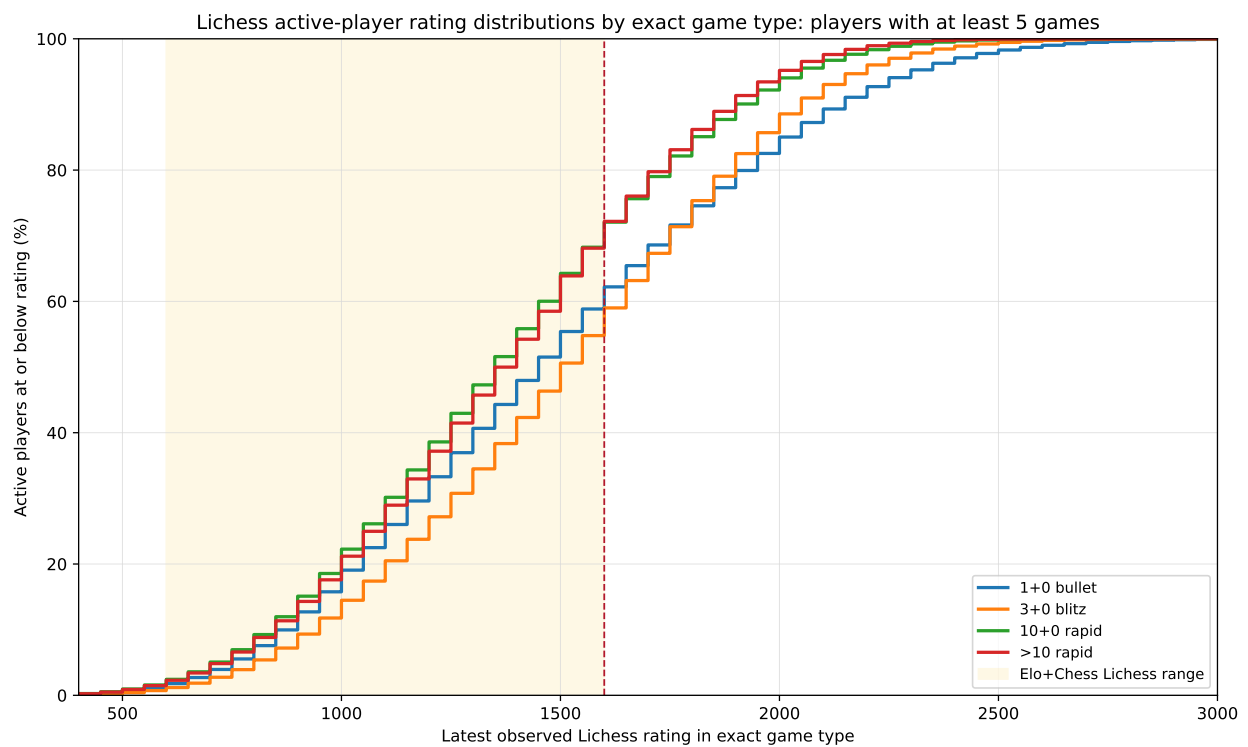


Figure 3: で要求した後の累積アクティブプレイヤー評価分布 正確なゲームタイプで少なくとも5つのゲーム。1500のジャンプは大幅に減少します。

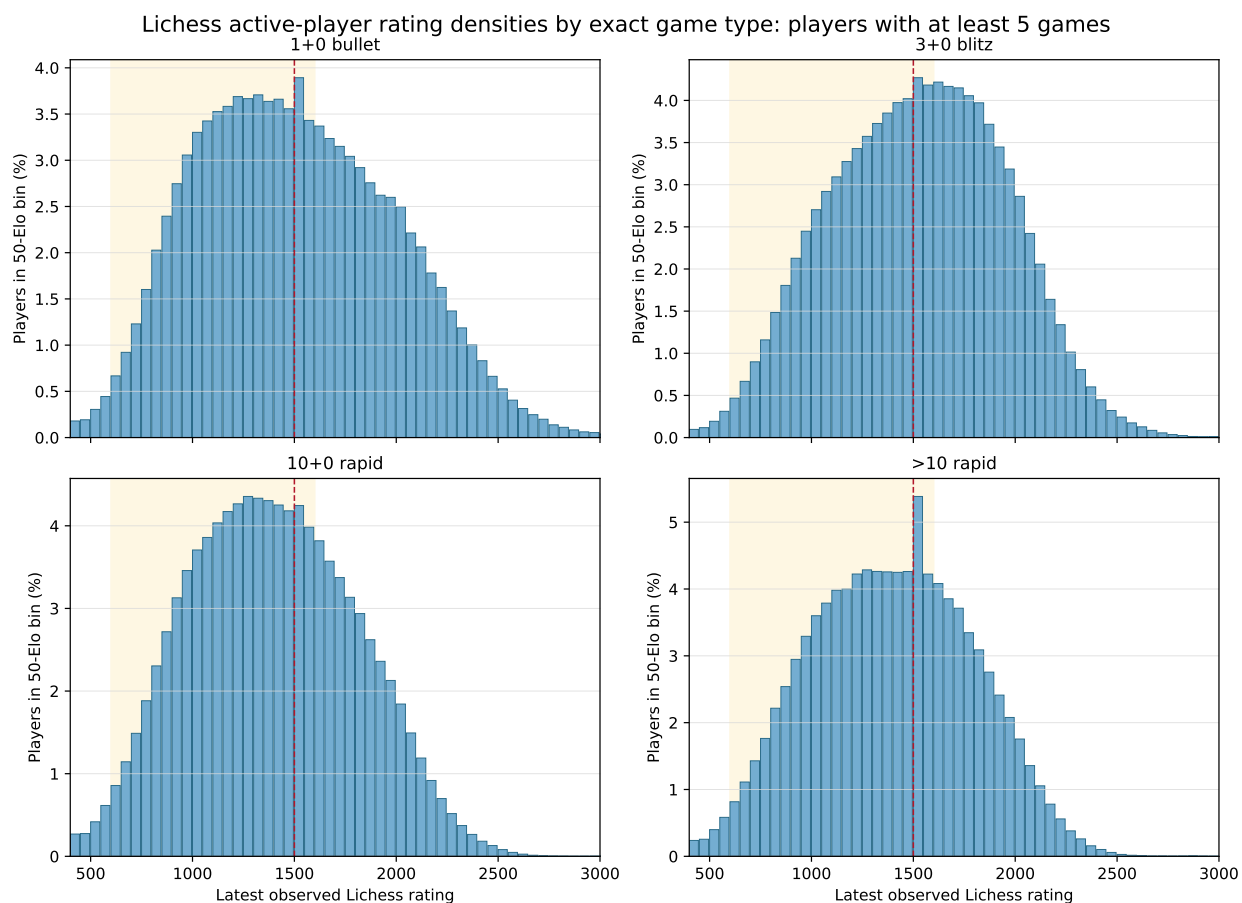


Figure 4: 正確なゲームで少なくとも 5 つのゲームが必要な後の密度ビュー タイプ。このビューは、全プレイヤーの 1500 質量点が大きいことを確認します。配布は主に、軽度アクティブなアカウントによって行われます。黄色い帯が目印です メインの Elo+Chess メトリック レポートで使用される 600 ~ 1600 Lichess 範囲。

4 現在のベンチマーク成果物の規模

計算量は意図的に大きくなります。それはプロフィールのスクレイピングではなく、小さなものです。ダウンロードされた PGN のセット、または手動で厳選された有益な例のコレクション。最初の Lichess パスは、2025 年 1 月からの完全な月次ファイルをカバーしていました。2026 年 3 月まで。非層別スキャン テーブルには 1,382,178,087 のゲームが含まれています。2,764,356,174 のプレイヤーゲーム評価行。これらの行全体で、6,623,438 個が一意です。プレイヤーのユーザー名は少なくとも 1 回は表示されます。

Speed	Games in raw scan	Player-game rows	Unique players
Bullet	519,375,423	1,038,750,846	3,077,635
Blitz	839,823,863	1,679,647,726	5,987,634
Rapid	22,978,801	45,957,602	1,535,865
All scanned speeds	1,382,178,087	2,764,356,174	6,623,438

Table 3: ベンチマーク サンプリング前の層化されていない Lichess スキャン スケール。スキャン 2025 年 1 月から 3 月までの完全な月次ゲーム履歴ファイルをカバーします。2026 年。

運用ベンチマークのアーティファクトは、生のスキャンよりもはるかに小さいため、それらはメトリック構築のために選択されます。まだ大きいです: ベンチマーク ステージは、数百にわたる移動状態の拡張および集約パイプラインです。ゲームの種類ごとに数千のゲーム。

Game type	Games	Player-game rows	Unique players	Move states	Feature columns
1+0 bullet	483,974	967,948	119,732	22,985,272	277
3+0 blitz	479,974	959,948	129,398	28,138,904	277
10+0 rapid	925,636	1,851,272	229,235	57,554,339	277
>10 rapid	660,540	1,321,080	133,661	38,580,214	277

Table 4: 調整後の層別ベンチマーク アーティファクトのスケール。共通の 277 列のランタイム機能スキーマ。あ「プレイヤーとゲームの列」は、1 人のプレイヤーの視点から見た 1 つのゲームを意味するため、ほとんどのゲームは 2 つのプレイヤー ゲーム行を生成します。

3+0 ブリッツだけでも、現在の製品ベンチマークでは 479,974 ゲームが使用されます。129,398 人のユニークなプレイヤーと 28,138,904 の異なる移動状態行が関与します。の 整列されたブリッツ プレイヤー ゲーム機能テーブルには、277 の派生変数が含まれています。の 迅速な機能構築のアーティファクトは、移動状態の段階ではさらに広範囲になります。フィーチャ構築中に使用される高速移動状態テーブルには 168 が含まれます。277 列のプレイヤー ゲーム フィーチャーに集約される前の move-state 列 テーブル。4 つのゲーム タイプすべてのフロントエンド ランタイム ベンチマーク アーティファクトは次のとおりです。次に、同じ 277 列のプレイヤー ゲーム機能スキーマに合わせて調整します。戦術分類のブレイクアウト変数を拡張しました。これらの特徴テーブルから、レポートは、最も解釈しやすいコーチング指標を選択して表示します。

これらのカウントは、結果を解釈するために重要です。ベンチマーク曲線 手動でレビューされた数百のゲームからの推定値ではありません。代わりに、彼らは、評価対象となる数千万の州と 4 つ以上の理事会の出身者 現在の 4 つの製品版ゲーム カテゴリ全体で 100 万人のプレイヤー ゲームの行。

5 パイプラインの概要

ベンチマーク建設パイプラインには 6 つの主要なステップがあります。

1. 対象のゲーム タイプとレーティング階層でレーティングされた Lichess ゲームを選択します。
2. 各ゲームを手ごとに解析し、手前と手後を保存します 基板の状態。
3. ボード、ムーブ、 プレイヤーの視点と対戦相手の視点。
4. 移動状態変数をプレーヤーゲーム機能変数に集約します。
5. プレーヤーゲームの変数を明確なレポート指標に変換します。 解釈の方向性: 通常は高いほど優れており、通常は低いほど より良い、または説明的なもの。
6. Elo バケットごとに各メトリクスを平均し、結果を手動で検査します 関係。

多くのレポート指標は直接的なものではないため、この構造は重要です。 PGN ヘッダーに表示されます。位置を再構築し、決定する必要があります。 どの駒がどの開始マスから移動したか、移動前と移動前を比較 移動後のマテリアルと構造、およびその結果を集計します。 レベルをプレーヤー ゲーム レベルに移動します。

6 中間変数

中間変数は、ボード上で何が変更されたかを説明するように設計されています。 そしてレポートプレイヤーの観点からそれがなぜ重要なのかを説明します。 例としては次のものが挙げられます。

- 手番、プレイヤーの手番、動かす側、手番の色、プレイヤー 色;
- from-square、to-square、移動駒タイプ、捕捉駒タイプ、および 移動がキャプチャ、チェック、キャスル、プロモーション、またはメイトであるかどうか。
- 移動前と移動後のマテリアルバランス。
- マイナーピースが指定された手数だけ開始マスから離れたかどうか。
- 連続したプレイヤーの動きで同じ駒が動いたかどうか 開口部。
- クイーンが手番 10 より前に動いたかどうか。
- キャスリング側、キャスリングの手数、およびルークが接続されたかどうか キャスル後。
- ルーズピース、ハンギングピース、ダブルポーン、孤立ポーン、パス ポーンと後方ポーン。
- rook ファイルのステータス、オープン ファイル、ハーフオープン ファイル、およびキング セーフティ 特徴;
- 戦術的機會、実行された戦術、物的圧力戦術、 保証されたフォローアップ利益と実現利益。

正確な中間セットはビルド段階によって異なります。実行時レポートデータベースが保持するレポートを迅速に提供するために必要な列のみを提供し、完全な機能を構築します。アーティファクトには、より広範な移動状態テーブルと、構築に使用されるフィーチャ パッチが含まれます。最後のプレイヤー ゲーム機能の行。

7 開放原理メトリクスの例

開始原則は、ユーザーにとって簡単のため、有用な例です。理解しやすく、移動状態の計算に簡単に接続できます。

7.1 Minor Pieces Developed by Move 10

このメトリクスは、プレイヤーの 4 人の騎士と司教のうち何人が離脱したかをカウントします。プレイヤーの10手目までにホームマスに移動します。途中の計算 移動者の色、移動する駒の種類、開始マス、およびプレイヤーの移動を使用します 番号。白の場合、ホームマスは b1、g1、c1、f1 です。黒人にとって、彼らは、b8、g8、c8、f8。各プレイヤー ゲームについて、Elo+Chess は個別のホームをカウントします。プレイヤーが移動した騎士と司教の間のマス目は 10 です。

これは、通常、高いほど良い指標です。すべての未成年者がそうであるとは主張しません 駒は常に 10 手目で動かなければなりません、発展途上のプレイヤーの間では経験的に 関係は明らかです: より高い Elo バケットに到達したプレイヤーは成長する傾向があります もっとマイナーな部分を以前に。

7.2 Repeated Piece Moves in the Opening

このメトリクスは、代わりに同じ駒を再度移動するのに費やされた余分な初期の手をカウントします。新しいものを開発するという。中間計算では以前の計算が使用されます。目的地のマス目、現在の出発地マス目、およびプレイヤーの移動番号。もし 前のプレイヤーの移動で同じ駒が移動し、時間内に再び移動されます。プレイヤーが最初に 10 回移動すると、反復移動回数が増加します。

これは、通常は低いほど良い指標です。無駄な開栓温度を測定します。簡単な方法。いくつかの繰り返しの動きは戦術的に正当化されますが、規模が大きくなると、バケット カーブは一般的な傾向を捉えています: ターゲットでより強いプレイヤー レベルでは、同じピースを繰り返して移動する初期の手が少なくなります。

7.3 Queen Touched Before Move 10

このメトリクスは、プレイヤーが 10 移動する前のクイーンの移動をカウントします。計算には移動駒の種類とプレイヤーの移動番号が使用されます。動いている部分の場合 がクイーンでプレイヤーの手数が最大10の場合、カウントが増加します。

これは、Elo+Chess ターゲット範囲にとっても、通常は低い方が良いです。早い クイーンの動きは材料を獲得したり、譲歩を強制したりする可能性があります、多くの発展途上のプレイヤー 女王を連れ出すのが早すぎる、女王への攻撃でテンポを失い、遅れてしまう マイナーピースの開発とキングの安全性。

7.4 Castling and King Safety

キャスリングメトリクスは、キャスリングフラグ、キャスリングサイド、キャスリング移動番号、およびキャスリング後のボードの状態。例としては、プレイヤーが城にいるかどうか、プレイヤーが手番 10 でキャスルしたかどうか、プレイヤーがキングサイドにキャスルしたか、またはクイーンサイド、キャスリング後にルークが接続されているかどうか、およびルークファイルがあるかどうか。キャスリング後は開いているか半分開いていました。

これらの指標はすべて同じように解釈されるわけではありません。キャスル完成と通常、ターゲット範囲内では、移動 10 によるキャスリングの方が高いほど優れています。クイーンサイドキャスリングまたはオープン ファイルへのキャスリングは、よりコンテキストに依存する場合があります。だからこそ、Elo+Chess は道徳ルールを単にハードコーディングするものではありません。バケットを構築します 曲線を描いて、どの関係がコーチングに十分安定しているかを検査します。

8 バケット曲線と手動検査

プレイヤー ゲーム メトリクスが計算された後、Elo+Chess はそれらを Lichess Elo でグループ化します。バケットを検索し、バケットの平均を計算します。バケット平均は生の経験値です。そのメトリクスのベンチマーク ポイント。傾向線はレポートで使用され、Elo スタイルのグレードをユーザーの値に割り当てます。

曲線は使用前に手動で検査されます。メトリクスの方が便利です。バケット平均が評価範囲全体で一貫した関係を示す場合、製品によって提供されます。メトリクスは、ノイズが多く、平坦で、高度な場合にはあまり役に立ちません。非単調であるか、ユーザーに説明するのが難しい特殊なケースによって引き起こされます。

この手動検査ステップは、Elo+Chess 上部より上で特に重要です。カットオフ。より強いプレイヤーは、いつオープニングの原則を破るべきかを知っており、それを実行します。効果的に。初心者から初級者までの範囲を超えると、単純な習慣と評価の間では、平坦になったり形が変わったりすることがよくあります。製品 したがって、経験的關係が強い範囲を強調します。そして教育的にも有益です。

9 フルレンジの開口曲線の例

図 5 は、4 つの 3+0 ブリッツ開始原理曲線を示しています。灰色の点は、Lichess の定格範囲が広いことを示します。青いポイントと青いトレンド 行は、メインの Elo+Chess コーチングで使用される 600 ~ 1600 の Lichess 範囲を示しています。ベンチマーク。赤い破線は上限カットオフを示します。

同じパターンがフルレンジの 10+0 ラピッドおよびより長いラピッドでも見られます。人工物。中心となるモデリングの決定は変更されていません。経験に基づいたモデルを使用します。コーチングの主張を安定した範囲で行い、単純な習慣曲線であるかのように振る舞うことを避ける 永遠に同じように情報を提供し続けます。

同じパターンが Table 5 に数値的に表示されます。坂道は 100 Elo ポイントごとのメトリック値の変化として報告されます。キャスリング、マイナー作品 展開、繰り返しの駒の動き、初期のクイーンの動きはすべて明確に示されています 600 ~ 1600 の範囲での移動。1600 度を超えると、傾斜ははるかに小さくなります。これは、評価の高いプレイヤーがブレイクするという考えと一致しています。基本原則をより選択的に、より効果的に適用します。2 つ目のメカニズムは、飽和。習慣によっては、指標が実質的な上限または甘い値に達する場合があります。

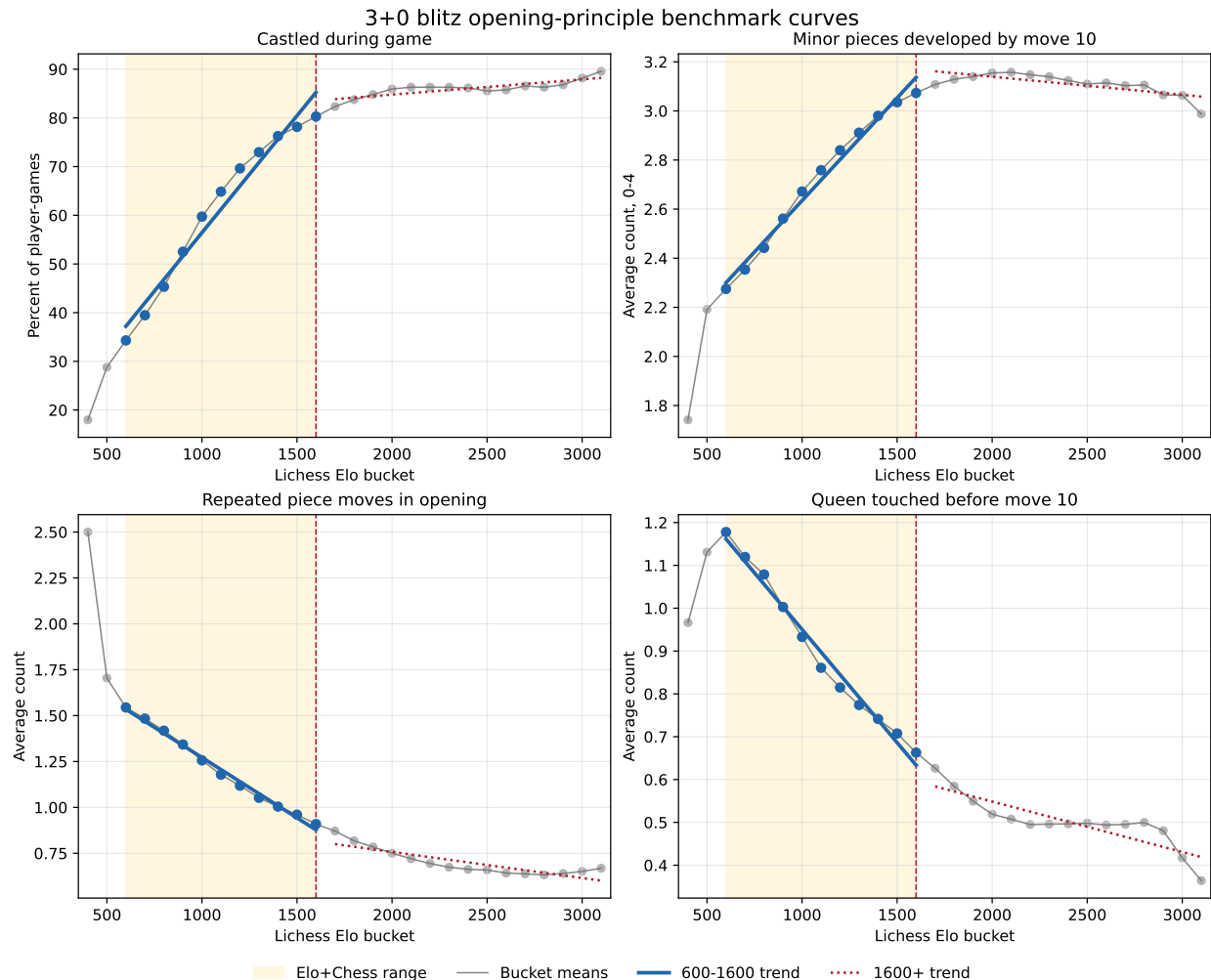


Figure 5: 3+0 ブリッツの開始原理ベンチマーク カーブを選択しました。の Elo+Chess コーチング範囲では、人間関係が最も強く、最も役立ちます。上限カットオフを超えると、いくつかの関係が実質的に平坦になります。

す。スポット: 強いプレイヤーはすでにその習慣を身につけているので、そうする理由はありません。急上昇を続けるカーブ。多ければ多いほど良いとは限りません その点。したがって、より強いプレイヤーがブレークする可能性があるため、データは両方とも平坦になります。具体的な立場でのルールと、多くの基本的な習慣がすでに浸透しているため、有効な動作範囲。

10 上限カットオフが重要な理由

Elo+Chess の対象ユーザーは、初心者から初級者までのプレイヤーです。これらのために プレイヤーの基本的な習慣はレーティングと強く関連しています: 駒の開発、繰り返しの初期の駒の動きを避ける、不必要な初期のクイーンの動きを避ける、城 時間内にルークを接続し、材料を管理し、駒が緩んだままになることを避けます。ぶら下がっている。

評価が高くなると、同じ表面上の習慣が診断されにくくなります。強い センターが閉鎖されているため、プレイヤーはキャスリングを遅らせる可能性があります。クイーンを早めに移

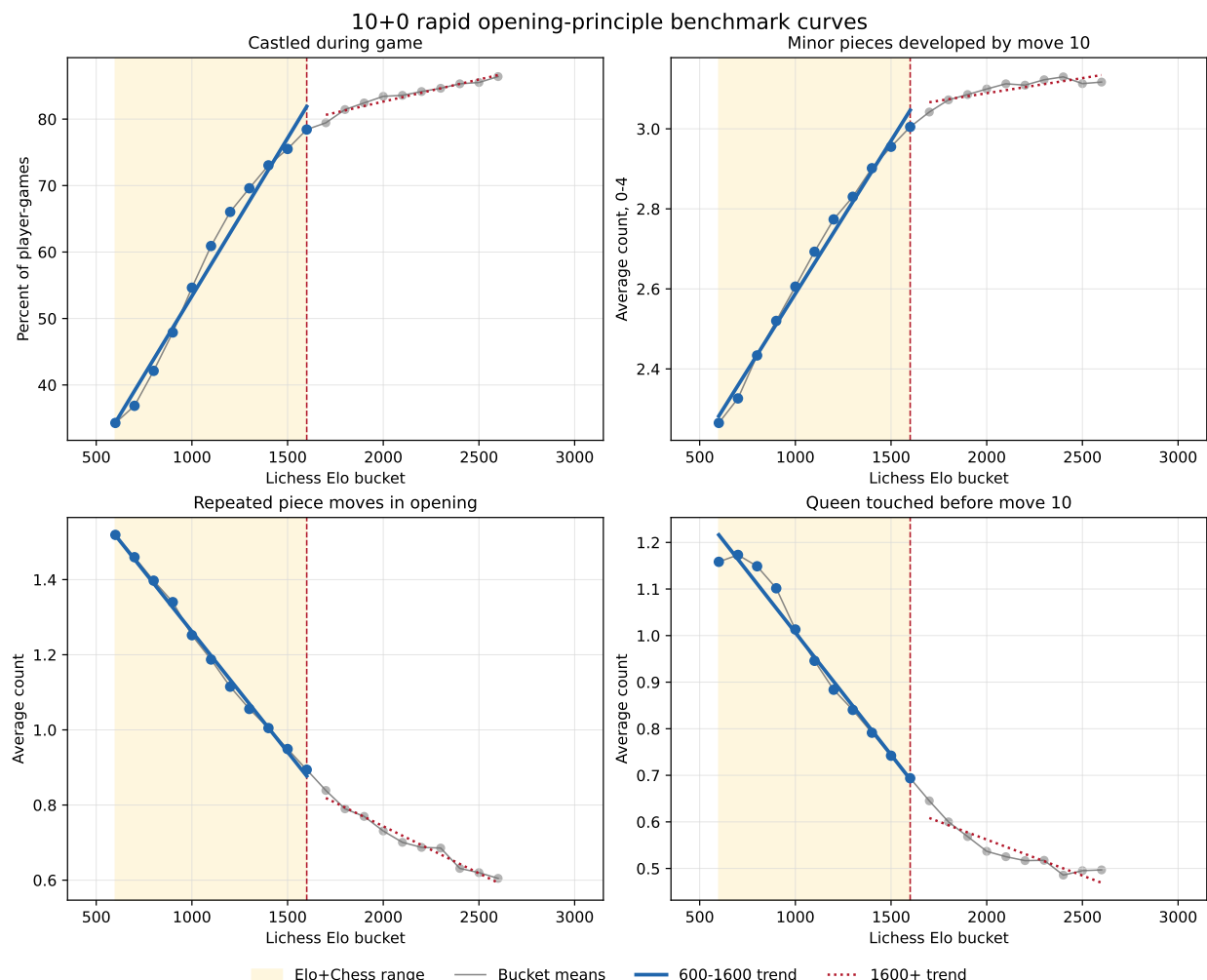


Figure 6: 10+0 のラピッド ゲーム用に選択された開始原理ベンチマーク カーブ。の Elo+Chess コーチング範囲では、人間関係が最も強く、最も役立ちます。上限カットオフを超えると、いくつかの関係が実質的に平坦になります。

動してください 局面に具体的な戦術が含まれているため、または同じ駒を2回動かす それはテンポを勝ち取ったり、構造的な譲歩を強いたりするからである。まだ生々しい習慣 意味はあるが、習慣と評価の関係はより深くなる 条件付き。これが、Elo+Chess がメインのコーチングに上限カットオフを使用する理由です。主張と、その範囲を超えるユーザーに密接な関係が警告される理由 原則的な習慣と評価の間で問題が発生します。

11 両面戦術メトリクスとパーセンタイル目標

一部のメトリクスは、開始原理の例とは逆の形状をしています。フォーク イベントには本質的に両面性があるため、これらは役立つ例です。選手 ピアよりも有用なフォークを見つけて変換するのは良いことですが、より強いため、変換されたフォークの人口率はレーティングとともに低下します 対戦相手はフォークの機会を減らし、戦術的な脅威から守ります もっと積極的に。

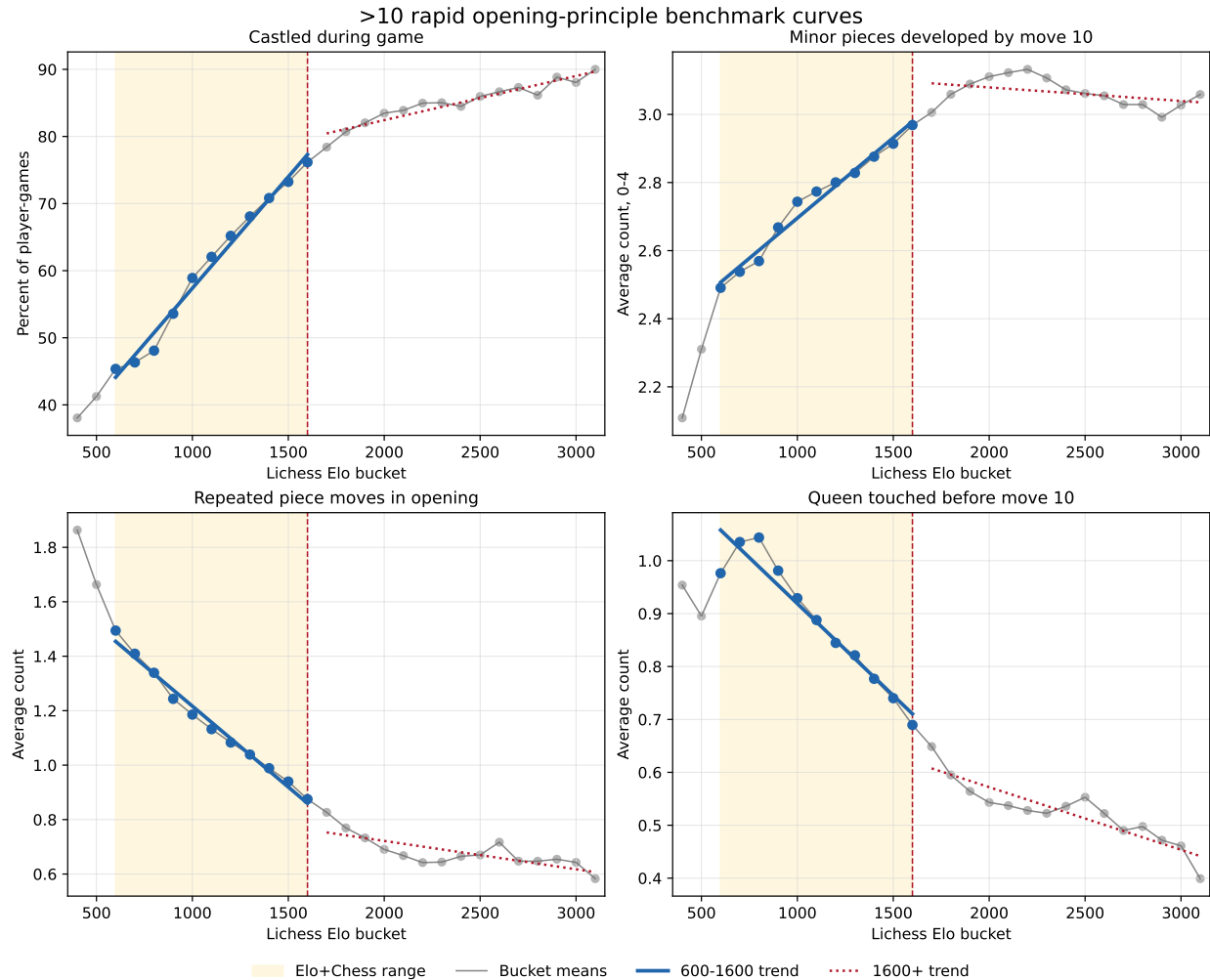


Figure 7: 長時間のラピッドゲーム向けに選択された開始原理のベンチマーク曲線。

図 8 はこの効果を示しています。両方ともフォークを変換しました レポートプレイヤーと相手によるコンバートフォークの頻度が低くなります 評価が上がります。それは、フォークの実行が悪いという意味ではありません。それは生の出来事を意味します レートは相手のエラー率とゲームのテクスチャによって部分的に制御されます。メトリクスの場合 したがって、このように、Elo+Chess はピア履歴パーセンタイルを内部で使用します。世界的なバケット平均の傾向だけに依存するのではなく、プレイヤー自身の Elo バケットを評価します。

これらのパーセンタイル ターゲットを構築するために、Elo+Chess はゲーム タイプごとにプレイヤーをサンプリングし、Elo バケットは、サンプリングされた各プレイヤーの最近の適切なゲームを使用します。の現在のターゲット セットは 600 から 1600 までの 11 バケットを使用し、1 つあたり 200 人のプレイヤーがいます。バケット、および各プロダクションのプレイヤーあたり最大 50 件の最近の対象となるゲーム ゲームタイプ。これにより、ゲーム タイプごとに 2,200 のプレイヤー サンプルが得られ、約 バケット内パーセンタイルのゲーム タイプごとに 110,000 のプレイヤー ゲーム行 配布物。

図 9 は、結果のターゲットを示しています 3+0 ブリッツにおけるサンプルプレイヤー換算フォークレートの分布。それぞれのボックスは、全体のバケット平均ではなく、サンプリングされたプレイヤー全体のバケット内分布 個人的なゲーム。レポート UI では、ユーザーの値が関

Game type	Metric	600–1600 slope per 100 Elo	1600+ slope per 100 Elo
1+0 bullet	Castled during game	4.137	0.550
	Minor pieces developed by move 10	0.063	0.009
	Repeated piece moves in opening	-0.051	-0.026
	Queen touched before move 10	-0.040	-0.011
3+0 blitz	Castled during game	4.804	0.313
	Minor pieces developed by move 10	0.084	-0.007
	Repeated piece moves in opening	-0.066	-0.014
	Queen touched before move 10	-0.053	-0.012
10+0 rapid	Castled during game	4.753	0.665
	Minor pieces developed by move 10	0.076	0.008
	Repeated piece moves in opening	-0.064	-0.025
	Queen touched before move 10	-0.052	-0.015
>10 minute rapid	Castled during game	3.320	0.659
	Minor pieces developed by move 10	0.047	-0.004
	Repeated piece moves in opening	-0.059	-0.010
	Queen touched before move 10	-0.035	-0.012

Table 5: 4 つの例の開口原理に対するバケット カーブの傾きの近似値 プロダクションゲームタイプ全体のメトリクス。

Game type	Elo buckets	Players per bucket	Games per player	Player samples
1+0 bullet	11	200	49–50	2200
3+0 blitz	11	200	27–50	2200
10+0 rapid	11	200	37–50	2200
>10 minute rapid	11	200	47–50	2200

Table 6: 現在のピア履歴パーセンタイル ターゲット セット。これらは バケット内分布を推定するために使用されるプレーヤー履歴サンプル パーセンタイル形式のメトリクス。

連する値と比較されます。ユーザーのマップされた Elo バケットのボックス。これは、次のような発言の根拠です。プレーヤーのコンバートされたフォーク率が他のプレーヤーと比較して高いか低いか 人口全体のイベント発生率がレーティングとともに低下した場合でも、同じレベルになります。

12 レプリケーションコスト

この方法は原理的には再現可能ですが、軽量ではありません。再構築 ベンチマーク曲線には次のものがが必要です。

- 大規模な Lichess 月次ゲーム履歴ファイルへのアクセス。
- ゲームの種類、評価バケット、資格による層別フィルタリング ルール。
- 拡張可能な合法的な手パーサーとボード状態評価ツール 数千万の移動状態。
- オープニング、マテリアル、ポーン構造、キングセーフティ、ルークの機能ロジック アクティビティ、戦術、時間管理、ゲームの結果。
- 移動状態からプレイヤーのゲーム機能までの集約。

Two-sided converted-fork rates by Elo bucket

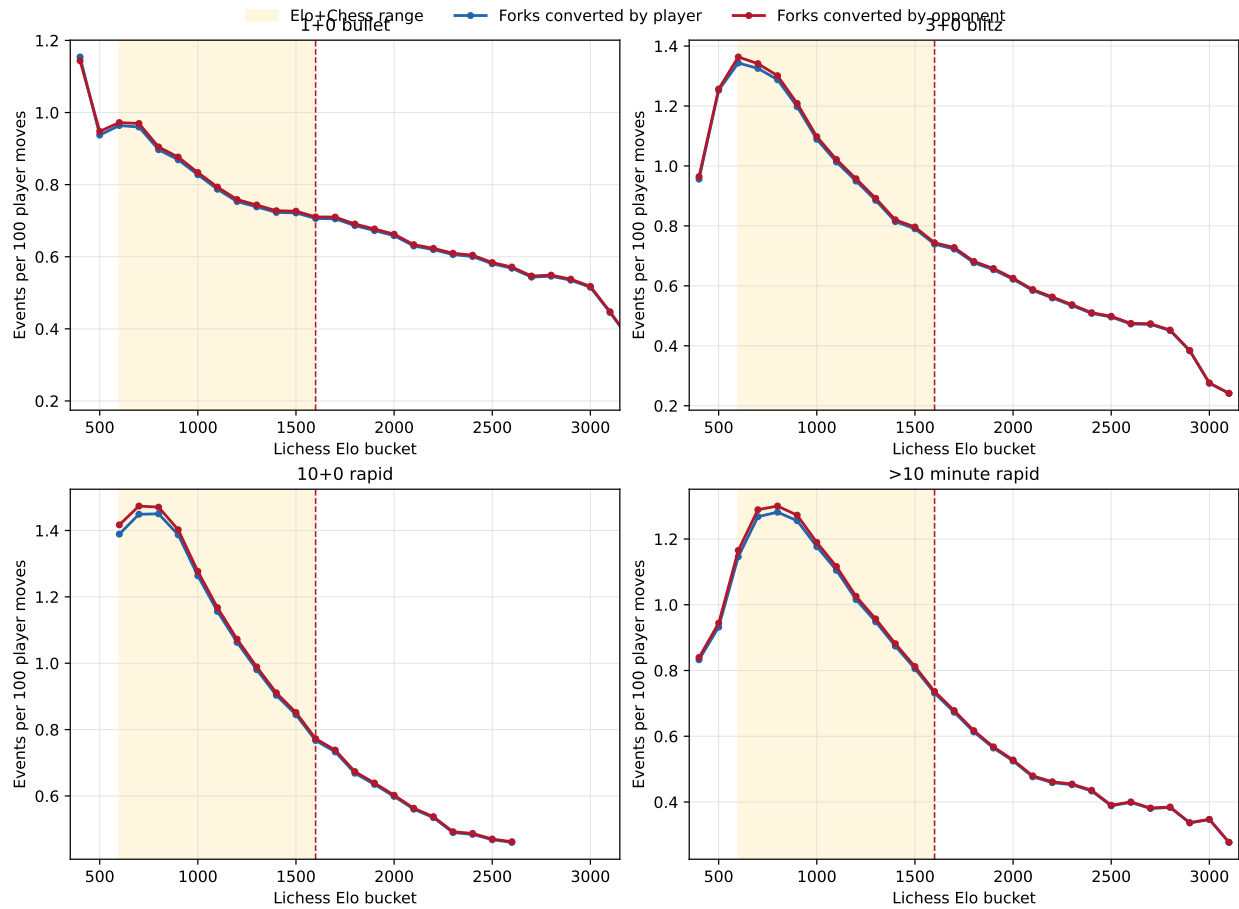


Figure 8: Elo バケットによる両面変換フォーク レート。変換されたフォークイベント 双方のディフェンスが良く、クリーンフォークが少ないため、評価が高くなる 変換が可能です。

- すべてのレポート指標のバケットレベルの概要。
- すべての候補ベンチマーク曲線を使用する前に手動でレビューします。 コーチングレポートです。

3+0 ブリッツの場合、これはほぼ 50 万試合、ほぼ 100 万試合を意味します。プレイヤー ゲームの行、および 2,800 万を超える移動状態の行。四つにまたがってベンチマーク ゲーム タイプ、現在のベンチマーク アーティファクトには 2.55 を超えるものが含まれています 100 万 ゲーム、510 万以上のプレイヤーゲーム行、147 以上 100 万行の移動状態行。スケールは価値の中心です。ベンチマーク: プレイされたゲームで実際に何が起きたかをレポートでユーザーに表示できます 一般的なチェスのスローガンだけに頼るのではなく、自分のレベルに近いプレイヤーによって行われます。

13 制限事項

ベンチマーク曲線は経験的な記述であり、因果関係の証明ではありません。選手 バケット平均と一致するだけでは評価は得られません。メトリクスも 複雑なポジションの要約が不完全: 初

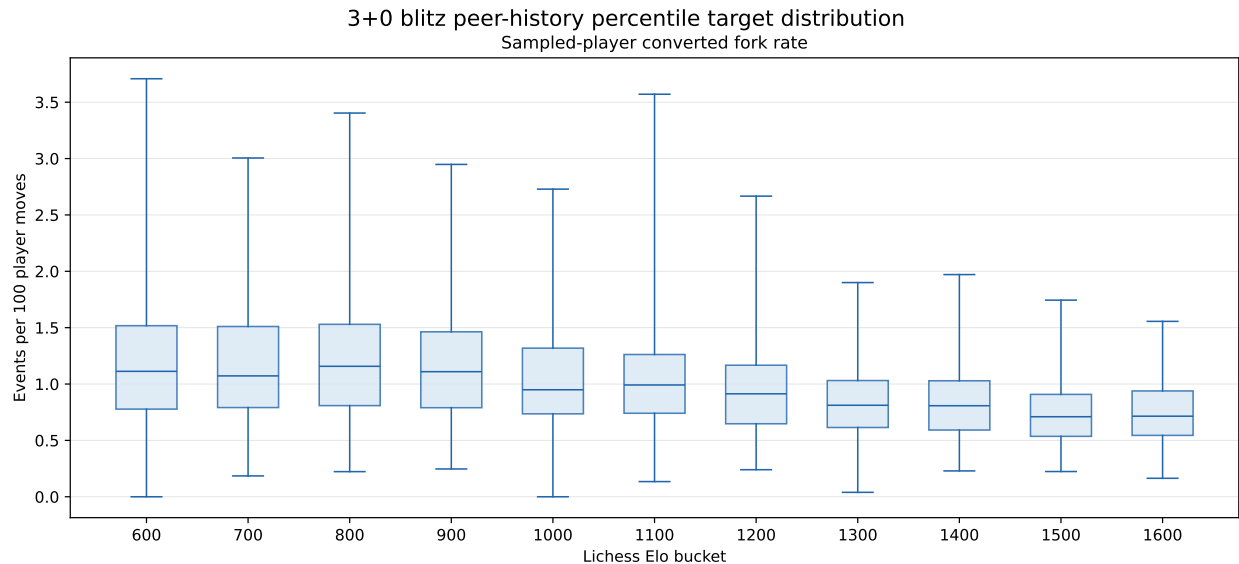


Figure 9: サンプル プレーヤーの 3+0 ブリッツ ピア履歴ターゲット分布の例 換算されたフォークレート。パーセンタイルは Elo バケット内で計算されるため、レポートは プレーヤーを、ほぼ同様の戦術的機会に直面している仲間と比較する 料金。

期のクイーンの動きは優れている可能性がある ある立場では無謀であり、別の立場では無謀です。曲線の目的は次のとおりです。繰り返される習慣を特定し、それを同僚の行動と比較する。計算やコーチングの判断を置き換えます。

曲線も Lichess スケールの曲線です。ユーザーが Chess.com ゲームを持ち込む場合 履歴、Elo+Chess は、関連する評価バケットを Lichess ベンチマークにマッピングします 別のクロスプラットフォーム マッピング モデルを使用してスケールします。そのマッピング層は Lichess-to-Chess.com に関するコンパニオン Elo+Chess 研究ノートに記載されています。評価変換: [Estimating Cross-Platform Chess Rating Mappings with モーダル回帰](#)。このマッピングはパーセンタイルを保持するものとして解釈しないでください。ランク。2つの評価間の推定同等の評価を保持します。システム。パーセンタイル ランクはアクティブ プレーヤーであるためプラットフォーム固有のままです 分布が異なります。

14 生産レポートグラフィックス

プロダクション レポートは、ベンチマーク アーティファクトをメトリック レベルのパネルに変換します。完全なビルド パイプラインを知らなくても読むことができます。数字 10 and 11 ライブ Elo+Chess レポート デザインの 3 つの例を示します。

図 10の開口原理パネル 通常のベンチマーク曲線のプレゼンテーションを使用します。各カードは次から始まります メトリック ファミリーとメトリック名を入力すると、短い解釈ルールが得られます。の 右上隅の大きなグレードは、生のグレードから意図的に分離されています。メトリック値: ユーザーが直面する質問「この習慣はどのように評価されるのか」に答えます。一方、以下の行には測定値が保存されています。4つの概要行には、生涯、過去 10 回、勝数、および損失が表示されます。これにより、単なる装飾ではなく診断パネル。生涯と最後の10が一致する場合、おそらく習慣は安定しているでしょう。勝ちと負けが異なる場合、指標は次のようになります。単純な習慣ではなく、結果に依存するか、ゲームの状態に関連付けられます。

チャート領域には、抑制された暗い背景、金色のトレンドライン、および色が使用されています。4つのスライスのマーカー。傾向の注釈は、次のことを明示的に示します。ベンチマーク関係はEloとともに上昇または下降します。これは重要な理由です。マイナーピースの開発など、一部の指標は高いほど良いものですが、初期のクイーン移動など、低いほうが良いものもあります。したがって、カードは、ユーザーに傾斜だけから方向を推測するように要求しません。決勝戦「これが意味すること」と「コーチングのヒント」の行でベンチマーク結果を翻訳 経験的なコンテキストを維持しながら、具体的なチェスの指示に変換します。

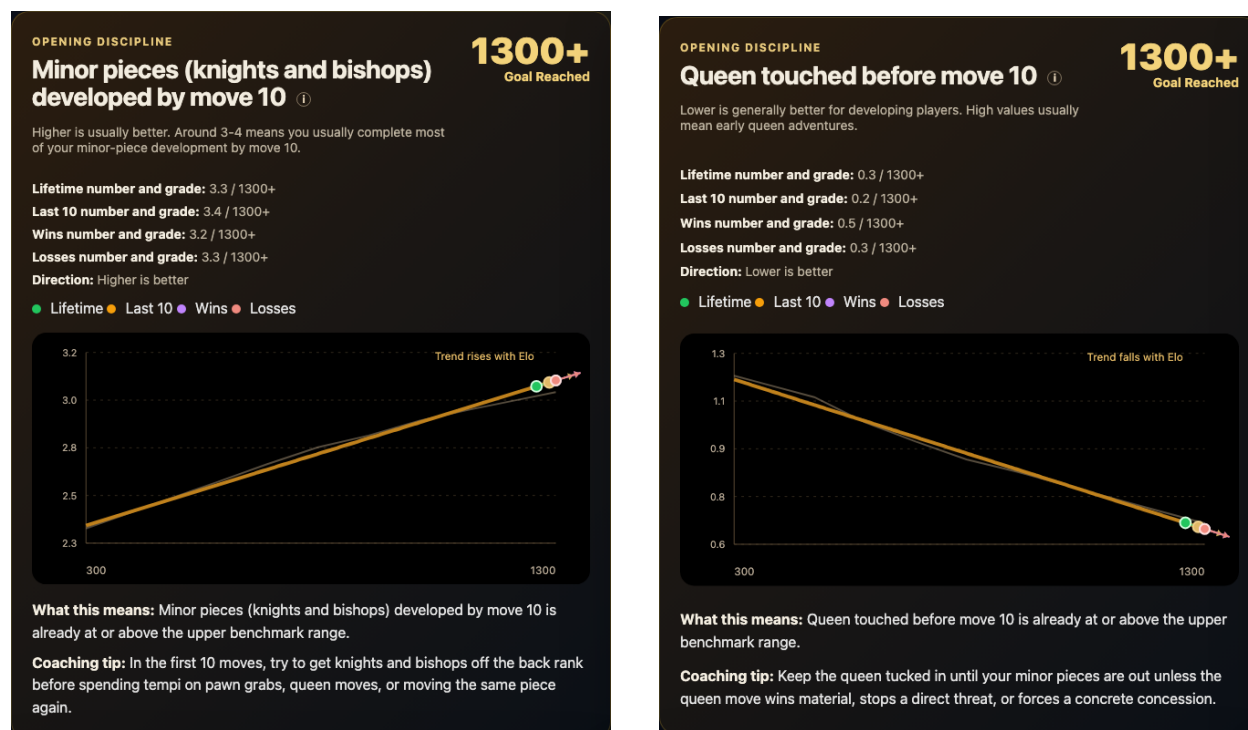


Figure 10: 生産開始方式のパネル。同じ視覚的な文法で、高いほど良い指標、マイナーピースの開発、低いほど良い指標、初期のクイーンの動き。両方のパネルで生涯、過去10回、勝利数、負け数が表示されます。そのため、レポートでは安定した習慣と最近の習慣、または結果に依存した習慣を区別できます。パターン。

図11はパーセンタイルスタイルを示します。生のレートが適切に表されないメトリクスに使用されるプレゼンテーション。単一の世界的な傾向。ウィニングフォークは戦術的な例です: 人口率変換されたフォークの数は、プレイヤーのスキルと相手のエラー率の両方によって変化します。のしたがって、カードはユーザーの価値をピアプレイヤーの履歴と比較します。マッピングされたEloバケット。箱ひげ図には、ピア分布が表示されます。四分位範囲、中央値、およびひげ。ユーザーのマーカーが上にプロットされます。同じスケールで、生の値とパーセンタイルランクの両方がラベル付けされています。隣接比較バケットはミュータされた形式で表示されたままになるため、ユーザーは、ピアバケットはローカルであり、ユニバーサルスケールではありません。

この設計により、パーセンタイルメトリックが監査可能になります。ユーザーには生の値が表示されます。パーセンタイル、比較に使用されるピアバケット、および変換されたChess.comからLichessへのバケットのマッピング。同じ生涯、過去10回、勝利数、損失行は保持されるため、戦術パーセンタイルスコアを引き続き確認できます。直近の影響と結果の依存性について。

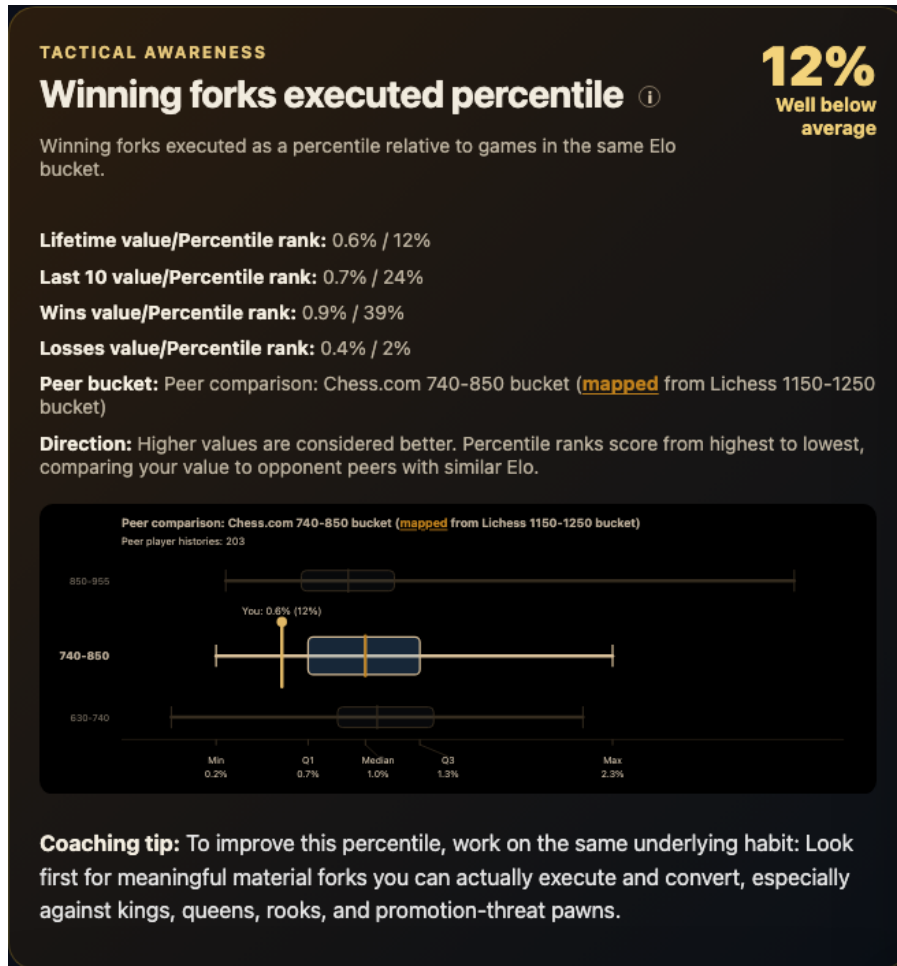


Figure 11: 実行された勝利フォークの生産パーセンタイル パネル。箱ひげ図 は、レポートで使用する同じバケットのピア分布を示しています。 マーカーは、ユーザーの生の値とパーセンタイル ランクを示します。

15 結論

Elo+Chess ベンチマーク レポートは、大規模な移動状態の計算から構築されています。層別の Lichess サンプル。各メトリクスは、具体的なボード状態のセットとして始まります。および移動状態変数は、プレイヤー ゲームの特徴となり、平均化されます。Elo バケットによって経験的なベンチマーク曲線を形成します。オープニング原則 このアプローチがなぜ役立つのかを例で示します。単純なチェスの習慣には強い、ターゲット範囲内の評価との目に見える関係があり、上限カットオフを超えると、関係は平坦になるか、より条件的になることがよくあります。これが、Elo+Chess がコーチングの主張を評価に焦点を当てている主な理由です データがそれらを最も明確に裏付ける範囲。