Swiss Table Tennis Elo Rating Simulation

Project Overview – Conceptual Brief (draft)

Context and Motivation

In Swiss table tennis, each competitive player is assigned a numerical rating called Elo. This rating can be thought of as a point score that represents a player's competitive strength.

- A player with more Elo points is considered stronger than a player with fewer points.
- The difference in Elo points between two players reflects how likely one player is to win against the other.
- When a player wins or loses a match, their Elo score changes according to a predefined rule.

In this sense, the Elo system creates a ranking hierarchy: players are ordered by Elo points, and these points evolve over time as matches are played.

While the Elo update rule is deterministic once a match result is known, real competition is inherently uncertain. Players face different opponents, match outcomes fluctuate, and a single month of competition can lead to very different ranking trajectories.

As a competitive player, I am not interested in predicting a single "most likely" future Elo value. Instead, I want to understand the range of plausible outcomes:

How much could my Elo realistically increase or decrease over a typical month of competition?

Core Idea of the Project

The key idea of the project is to treat a month of competition as a random experiment rather than a fixed trajectory.

Instead of asking: > "What will my Elo be next month?"

the project asks: > "If this month were replayed many times under realistic conditions, what distribution of Elo outcomes would we observe?"

To answer this question, the project uses **Monte Carlo simulation**: the same month is simulated thousands of times, each time with different random match outcomes. This produces a distribution of final Elo points, rather than a single predicted value.

What Is Being Simulated

The project simulates the evolution of a competitive player's Elo rating over a single month of competition, treating this month as a probabilistic experiment rather than a deterministic sequence of events.

At a high level, each simulated month follows the same logical pipeline:

1. A monthly competition structure is defined, consisting of league matches and a tournament.
2. For each match in the month, an opponent is selected according to institutional rules (league rosters or tournament pools).
3. A probability of winning is assigned to the match.
4. A win or loss outcome is randomly drawn based on this probability.
5. Elo points are updated using the official Swiss Elo update system.
6. The entire month is repeated many times using Monte Carlo simulation to generate a distribution of possible end-of-month Elo outcomes.

The goal is not to predict a single future Elo value, but to observe how Elo outcomes vary across many plausible realizations of the same competitive month.

All detailed modeling assumptions (competition format, opponent sampling rules, probability models, and parameter values) are specified explicitly in the Modeling Choices & Hypotheses section.

Why Monte Carlo Simulation Is Used

A single month of competition is subject to substantial randomness: match outcomes are uncertain, opponents vary in strength, and small differences in results can lead to different ranking trajectories.

Monte Carlo simulation addresses this uncertainty by replaying the same month thousands of times, each time with different random outcomes. This produces an **empirical distribution** of final Elo values, rather than a single point estimate.

This distribution-based perspective allows us to quantify: - expected Elo change, - downside and upside risk, - and the overall variability of ranking outcomes.

Role of Win Probability Modeling

A key component of the simulation is the assignment of winning probabilities to individual matches.

The project deliberately distinguishes between: - the simulation of match outcomes, and - the update of Elo points, which always follows the official Swiss system.

Two alternative models are used to generate match outcomes:

1. **Official Elo-based probabilities**, derived directly from the Swiss Elo formula.
2. **Machine-learning-based probabilities**, learned empirically from historical match data involving many players.

Importantly, these two models differ only in how match outcomes are generated. All Elo point updates are computed using the same official Swiss Elo update rule, ensuring that the ranking evolution remains realistic and comparable across scenarios.

What Is Compared Across Simulations

The simulation is run separately under each probability model, while holding all other elements constant.

The comparison focuses on: - how the distribution of monthly Elo outcomes changes, - whether dispersion, skewness, or tail risk differs, - and how sensitive ranking uncertainty is to probability modeling assumptions.

The project does not aim to declare one probability model "better" than the other in an absolute sense. Instead, it evaluates how different modeling choices translate into different uncertainty profiles for ranking evolution.

Output of the Project

The final outputs of the simulation are: - distributions of end-of-month Elo values, - summary statistics (mean, quantiles), - and visual comparisons between scenarios.

These outputs form the basis for interpretation, discussion of limitations, and potential extensions.

Transition to Modeling Choices & Hypotheses

The remainder of the project formalizes the elements described above.

In particular, the Modeling Choices & Hypotheses section specifies: - the exact structure of a competitive month, - opponent sampling rules, - win probability models, - and the Elo update mechanism.

The code implementation is expected to follow these specifications exactly, without introducing additional assumptions.

Research Question

How uncertain is the short-term evolution of a Swiss table tennis player's Elo ranking over a typical month of competition, and how sensitive is this uncertainty to the choice of win-probability model (official Elo-based probabilities versus a data-driven machine learning alternative)?

Modeling Choices and Hypotheses

Scope and Objective of the Model

The objective of the model is not to predict a single future Elo rating, but to characterize the **distribution of plausible short-term ranking outcomes** for a Swiss table tennis player.

The model therefore focuses on uncertainty and variability, rather than point forecasts.

The analysis is restricted to a short-term horizon of one typical month of competition, which allows the model to remain interpretable while capturing the main sources of randomness affecting ranking evolution.

Structural Assumptions on Competition Format

A month of competition is modeled as a fixed and stylized sequence of competitive events, designed to approximate a realistic competitive workload rather than replicating exact institutional formats.

The player is assumed to participate in: - **two league encounters**, each consisting of three individual one-on-one matches, and - **one national tournament**.

To ensure a consistent number of matches across simulated months, the tournament is modeled using a hypothetical placement-based structure that guarantees a fixed workload. Specifically, the tournament is assumed to involve a group stage of three matches, followed by four additional placement matches, yielding a total of seven tournament matches.

This structure does not reflect the standard Swiss tournament system, where early elimination may reduce the number of matches played. Instead, it is introduced as a modeling device to approximate the intensity of competition over a month while maintaining tractability and comparability across Monte Carlo simulations. The implications of this simplification are discussed in the limitations section.

Match Outcome Assumptions

Each match outcome is modeled as a binary random variable (win or loss).

- Match scores, set differences, and margin of victory are not considered.
- The effect of these omitted variables is implicitly absorbed into the win probability.

Match outcomes are assumed to be conditionally independent given the win probabilities, meaning that any dependence between matches operates only through changes in the player's Elo rating over time.

Opponent Sampling Assumptions

Opponents faced by the player during the month are modeled differently for league encounters and for the national tournament, reflecting the institutional structure of Swiss table tennis.

League opponents (scheduled clubs, roster-based selection)

For league play, the opposing clubs are assumed to be known in advance, as the season schedule is fixed at the beginning of the league. For each of the two scheduled league encounters, the opponent faced by the player is modeled as a random draw from the roster of the opposing club.

To reflect selection patterns in team matches, the probability that a given roster player is selected is assumed to be proportional to their observed participation rate during the season (e.g., number of matches played by that player divided by the total number of matches played by the club).

Tournament opponents (A/B pool with no rematches)

For the national tournament, eligible opponents are drawn from the most recent Swiss ranking database, restricted to players belonging to series A and B. This restriction reflects the fact that the player participates only in tournaments whose relevant competitive pools are typically within these series, and it ensures that the analysis remains focused on ranking uncertainty at a sufficiently competitive level.

In the baseline model, opponents are sampled uniformly from the eligible A/B pool. To preserve minimal tournament realism without modeling a full bracket, we impose a **no-rematch constraint**: once an opponent has been faced in the tournament, they cannot be drawn again within the same tournament simulation. This approximation captures the basic idea that tournament structures prevent repeated matchups, while remaining computationally tractable.

Win Probability Models

To simulate match outcomes within the Monte Carlo framework, the model requires an estimate of the probability that the focal player wins a given match against a specific opponent. Two alternative win-probability models are considered: an Elo-based theoretical model and a data-driven machine learning model. The purpose of this comparison is not to assert the superiority of one model over the other, but to assess how sensitive ranking uncertainty is to the choice of probability estimator.

Elo-Based Probability Model

In the first approach, win probabilities are computed using the official Swiss table tennis Elo formulation. Under this model, the probability that player i defeats player j depends solely on the difference between their Elo ratings and is given by:

$$P(i \text{ wins against } j) = \frac{1}{1 + 10^{\frac{E_j - E_i}{200}}}$$

where E_i and E_j denote the Elo ratings of players i and j, respectively.

This formulation is directly embedded in the Swiss ranking system and therefore serves as a natural theoretical benchmark. It assumes that the Elo rating fully summarizes player strength and that match outcomes are probabilistically determined by rating differences alone.

Machine Learning Probability Model

The second approach replaces the fixed Elo probability formula with a data-driven probability estimator learned from historical match data.

Rather than imposing a predefined functional relationship between Elo differences and win probabilities, the machine learning model infers this relationship empirically by observing past matches. Conceptually, the model learns how often players with given Elo characteristics win or lose and uses this information to estimate the probability of victory in new, unseen matchups.

To avoid individual-specific overfitting and to ensure general applicability, the machine learning model is trained on a dataset comprising historical Swiss table tennis matches involving many players, rather than only matches played by the focal player. The trained model is subsequently applied to the focal player within the Monte Carlo simulation.

Minimal Feature Specification

In line with the objective of maintaining interpretability and limiting data-engineering complexity, the machine learning model uses a minimal feature set:

- Elo rating of the focal player
- Elo rating of the opponent

No additional contextual variables (such as recent form, match importance, or competition type) are included in the baseline specification. This design choice ensures that the machine learning model remains directly comparable to the Elo-based model, as

both rely on the same informational inputs.

Model Choice and Interpretation

A **logistic regression** model is used as the baseline machine learning classifier. This choice reflects a deliberate trade-off between flexibility and transparency:

- Logistic regression outputs well-calibrated probabilities.
- Its structure closely mirrors the logistic form of the Elo model, while allowing parameters to be estimated from data rather than fixed a priori.
- The resulting probability function can be interpreted as an empirically estimated alternative to the Elo formula.

Importantly, the machine learning model is not assumed to outperform the Elo-based model by construction. Its role is to provide an empirical benchmark that captures observed match outcome frequencies in the data.

Role in the Monte Carlo Simulation

Both probability models are embedded separately into the same Monte Carlo simulation pipeline. All structural assumptions, opponent sampling procedures, and Elo update rules are held constant. Any differences observed in the resulting distributions of final Elo ratings can therefore be attributed to differences in the underlying win-probability models rather than to changes in simulation design.

Interpretation Scope

Differences between the Elo-based and ML-based simulations are interpreted as reflecting **model uncertainty** in probability estimation, not as definitive evidence of predictive superiority. The comparison highlights how sensitive short-term ranking uncertainty is to the assumptions embedded in probability modeling.

Elo Update Mechanism

After each simulated match, the player's Elo rating is updated according to the official Swiss table tennis Elo update rule. For a given match, the Elo change is defined as:

$$\Delta\text{Elo} = K \times (\text{result} - P), \quad \text{result} \in \{1, 0\}$$

where: - $K = 15$ is the Elo adjustment factor, - result equals 1 in case of a win and 0 in case of a loss, - P is the official Elo-based win probability implied by the Swiss ranking system.

The win probability P is computed using the fixed Elo ratings observed at the beginning of the month:

$$P = \frac{1}{1 + 10^{\frac{E_\text{opp} - E_\text{player}}{200}}}$$

Importantly, **within a given month, Elo ratings are held constant**. All matches are played using the player's Elo rating at the beginning of the month, and Elo changes are computed accordingly.

If the player participates in M matches during the month, the total Elo change over the month is given by the sum of individual match updates:

$$\Delta\text{Elo}_\text{month} = \sum_{m=1}^{M} \Delta\text{Elo}_m$$

The player's Elo rating at the end of the month is therefore:

$$\text{Elo}_\text{end} = \text{Elo}_\text{start} + \Delta\text{Elo}_\text{month}$$

where Elo_start denotes the Elo rating at the beginning of the month.

In the Monte Carlo simulation, match outcomes are generated using either the Elo-based probability model or the machine-learning-based probability model described earlier. However, **Elo point updates always follow the official Swiss Elo update rule**, ensuring that simulated ranking changes remain fully consistent with the real-world ranking system.

Monte Carlo Assumptions

Uncertainty in ranking evolution is modeled using Monte Carlo simulation.

- The same month of competition is simulated repeatedly.
- Each simulation uses independent random draws for opponent selection and match outcomes.
- The collection of final Elo ratings across simulations defines an empirical distribution of ranking outcomes.

The number of simulations is chosen sufficiently large to ensure stable estimates of distributional properties such as quantiles and variance.

Simplifications and Excluded Factors

Several real-world factors are intentionally excluded from the model:

- Physical condition, fatigue, and injuries
- Psychological effects
- Strategic behavior and learning during tournaments
- Long-term form dynamics

These exclusions are motivated by data limitations and by the desire to keep the model transparent and interpretable. Their potential impact is discussed qualitatively in the limitations section.

Interpretation Framework

The resulting Elo distributions are interpreted as **risk profiles**, analogous to return distributions in financial modeling.

Differences between distributions obtained under different probability models reflect model uncertainty rather than purely random noise.

Methodology (Conceptual Algorithm Without Code)

This section describes the simulation procedure used to generate the distribution of end-of-month Elo ratings.

The methodology translates the modeling choices and hypotheses into a precise algorithmic sequence, without introducing additional assumptions or implementation-specific details.

Step 1 – Initialization

At the beginning of each simulation run:

- A focal player is defined with an initial Elo rating observed at the start of the month.
- This Elo rating remains fixed throughout the entire simulated month and serves as the reference rating for all probability and Elo update calculations.
- All model parameters (competition structure, opponent pools, probability model choice) are fixed and identical across simulation runs.

As a result, all variability in simulated outcomes arises solely from randomness in opponent selection and match outcomes.

Step 2 – Definition of the Competitive Month

Each simulation run represents one stylized month of competition.

The competitive month consists of a fixed sequence of matches, including:

- two league encounters, each composed of three individual matches, and
- one national tournament with a fixed total number of matches.

The tournament structure is designed to guarantee a constant workload across simulations and includes institutional constraints such as the absence of rematches within the same tournament.

The player is assumed to participate in all scheduled matches.

Step 3 – Opponent Selection

For each match in the simulated month, an opponent is selected according to predefined rules:

- **League matches**: opponents are drawn from the rosters of scheduled opposing clubs, with selection probabilities reflecting observed participation frequencies.
- **Tournament matches**: opponents are sampled from an eligible pool of A/B-ranked players, subject to a no-rematch constraint within the tournament.

Opponent Elo ratings are treated as known and fixed during the month.

Step 4 – Win Probability Assignment

Once an opponent is selected, a probability of winning the match is assigned.

Two alternative probability models are considered:

- **Elo-based probabilities**, computed using the official Swiss Elo formula and the fixed initial Elo rating.
- **Machine-learning-based probabilities**, estimated from historical match data and applied using the same Elo inputs.

In both cases, win probabilities depend on the fixed initial Elo rating of the player and the opponent's Elo rating.

Step 5 – Match Outcome Simulation

Given the assigned win probability:

- The match outcome is simulated as a binary random variable (win or loss).
- Outcomes are generated using independent random draws across matches, conditional on the assigned probabilities.

This step captures the intrinsic randomness of competitive match outcomes.

Step 6 – Elo Update Calculation

After each simulated match:

- An Elo change is computed using the official Swiss Elo update rule.
- The expected score used in the update formula is always computed using the player's initial Elo rating and the opponent's Elo rating.

Although Elo changes are calculated at match level, these changes do not affect subsequent match probabilities within the same month.

The total Elo change over the month is obtained by summing individual match updates.

Step 7 – End-of-Month Elo Recording

At the end of the simulated month:

- The player's final Elo rating is computed as the initial Elo plus the total accumulated Elo change.
- This final Elo value represents one possible realization of the month under the given probability model.

Step 8 – Monte Carlo Repetition

Steps 1–7 are repeated a large number of times.

Each repetition represents an alternative but plausible realization of the same competitive month, differing only through random opponent selection and match outcomes.

This procedure generates an **empirical distribution** of end-of-month Elo ratings.

Step 9 – Aggregation and Analysis

From the simulated distribution:

- summary statistics (mean, variance, quantiles) are computed,
- distributional properties such as dispersion and tail behavior are analyzed,
- results are compared across probability models.

No individual simulation run is interpreted as a prediction. Only aggregate distributional features are used for interpretation.

Methodological Clarification

The simulation does not aim to predict future results. It addresses the following counterfactual question:

> Given a fixed competitive month and fixed Elo reference ratings, how much variability in end-of-month Elo outcomes arises purely from randomness and probability modeling assumptions?

Why This Methodology Is Consistent

- It respects the institutional rule that Elo ratings are updated only after the month.
- It ensures full comparability across probability models.
- It provides a clean additive structure for interpreting Elo risk.

Expected Results and Interpretation

This section outlines the qualitative patterns and distributional features that are expected to emerge from the Monte Carlo simulations, prior to observing any numerical results. These expectations are derived directly from the structure of the model, the competition format, and the probabilistic assumptions described in the previous sections.

The purpose of this section is not to anticipate specific numerical values, but to clarify the mechanisms through which uncertainty arises and to define interpretable benchmarks for the subsequent empirical results.

Expected Shape of the Elo Distribution

Given that the monthly Elo change is modeled as the sum of multiple independent match-level updates, the distribution of end-of-month Elo ratings is expected to be:

- approximately unimodal,
- centered around a mean close to the expected Elo change implied by the win probabilities,
- with dispersion increasing as the number of matches increases.

However, because individual match outcomes are binary and Elo updates are asymmetric when win probabilities differ from 50%, the resulting distribution may deviate from strict normality. In particular, mild skewness may arise when the player is systematically favored or unfavored against typical opponents.

Expected Role of Match Volume and Competition Structure

The simulated month includes a relatively large number of matches, combining league encounters and a tournament.

As a result:

- the variance of the monthly Elo change is expected to be substantially larger than the variance of a single match outcome,
- extreme positive or negative outcomes become plausible even if the expected Elo change is moderate.

The tournament component, which introduces a broader and less predictable opponent pool, is expected to contribute disproportionately to the tails of the distribution, increasing both upside and downside risk relative to league play alone.

Expected Impact of Fixed Elo Within the Month

Because the Elo rating is held fixed throughout the month:

- win probabilities do not adapt to interim success or failure,
- early wins do not mechanically increase the probability of later wins, and vice versa.

This absence of intra-month feedback implies that the total Elo change is an additive process driven by independent realizations. Consequently, the distribution of outcomes reflects pure outcome uncertainty rather than dynamic momentum effects.

This design choice is expected to yield cleaner and more interpretable distributions, at the cost of ignoring short-term psychological or form-related dynamics.

Expected Comparison Between Probability Models

When comparing simulations based on the Elo-based probability model and the machine-learning-based probability model, several qualitative differences may emerge.

If both models produce similar average win probabilities across matchups, the mean end-of-month Elo change is expected to be comparable. However, differences may arise in higher-order distributional properties:

- dispersion may differ if one model assigns more extreme probabilities,
- tail risk may increase if the machine learning model captures nonlinearities or asymmetries not present in the Elo formula,
- skewness may differ if one model systematically over- or underestimates the player's winning chances against specific opponent segments.

Importantly, any such differences are interpreted as reflecting **model uncertainty** in probability estimation, not as definitive evidence of predictive superiority.

Expected Sensitivity to Opponent Sampling

Because opponents are sampled probabilistically rather than deterministically, the distribution of Elo outcomes is expected to be sensitive to the realized mix of opponent strengths.

Simulations in which the player faces an unusually strong or weak set of opponents are expected to populate the lower or upper tails of the distribution, respectively. This mechanism highlights that ranking uncertainty arises not only from match outcomes but also from the stochastic nature of competition exposure.

Interpretation Framework

The simulated Elo distributions are interpreted as **risk profiles**, analogous to return distributions in financial contexts.

In this interpretation:

- the mean Elo change represents expected performance,
- the dispersion captures ranking volatility,
- the lower tail reflects downside risk,
- the upper tail reflects upside potential.

The comparison across probability models therefore informs us how sensitive perceived ranking risk is to modeling assumptions, rather than providing a single forecasted outcome.

Role of This Section

This section establishes clear expectations against which the empirical simulation results can be evaluated. Deviations between observed results and these expectations will be interpreted as informative signals about the interaction between competition structure, probability modeling, and stochastic variability.