NHL Poisson Goal Prediction Model: Full Technical Overview

This report provides a detailed, multi-page technical summary of a Poisson-based statistical model designed to estimate NHL scoring distributions, win probabilities, and fair no-vig betting odds. It is intended for sportsbook analysts, data science hiring managers, and advanced portfolio reviewers.

1. Introduction

This document outlines the complete methodology behind a Poisson-based model for NHL goal prediction. The goal is to model how many goals each team scores, derive a total goals distribution, estimate win probabilities in regulation, and convert those probabilities into fair American odds. This approach represents one of the core baseline statistical models used in sports analytics.

2. Data Inputs

The model begins with team-level scoring metrics: - Goals For per game (GF/G) - Goals Against per game (GA/G) - League average GF/G - League average GA/G These inputs allow us to construct an understanding of how strong a team's offense and defense are relative to league norms.

3. Attack and Defense Strength Calculation

Each team is evaluated relative to the league using: attack_factor = team_GF_per_game / league_avg_GF defense_factor = team_GA_per_game / league_avg_GA An attack factor above 1 means above-average offensive production. A defense factor above 1 indicates allowing more goals than average (weaker defense).

4. Expected Goals (Lambda)

Expected goals for each team are calculated as: lambda_home = league_avg_GF * attack_home * defense_away * 1.05 lambda_away = league_avg_GF * attack_away * defense_home * 0.95 The constants 1.05 and 0.95 approximate home-ice advantage effects.

5. Poisson Goal Distributions

Using the Poisson distribution: P(k goals) = exp(-lambda) * lambda^k / k! We generate goal probabilities for k = 0 to 10. Values above 10 are negligible and therefore truncated.

6. Total Goals Distribution via Convolution

To compute the probability of combined goals, we convolve the home and away distributions. This produces a full distribution for total goals from 0 to approximately 20. This distribution powers Over/Under line probabilities and fair totals pricing.

7. Regulation Win Probabilities

The joint probability matrix P(home_goals = i, away_goals = j) allows us to compute: P(Home win) = sum(i > j) P(Away win) = sum(j > i) P(Tie) = sum(i = j) These values are foundational for pricing moneylines and related markets.

8. Over/Under Market Modeling

Given a betting line L (e.g., 6.5 goals): P(Over L) = sum(Total > L) P(Under L) = sum(Total < L) These probabilities are later transformed into fair odds.

9. Fair American Odds (No Vig)

Fair odds are computed using: If p > 0.5: odds = -100 * p / (1 - p) If p < 0.5: odds = 100 * (1 - p) / p These odds represent neutral bookmaker pricing without margin.

10. Example: New York Rangers vs New Jersey Devils

lambda_home = 3.952 lambda_away = 2.930 Over/Under 6.5: P(Over 6.5) = 0.531 Fair odds: Over -113 Regulation win probabilities: Home = 0.576 Away = 0.279 Tie = 0.145 Moneyline (fair): New York Rangers: -184 New Jersey Devils: +184

11. Limitations

- Poisson independence assumption - No player-level adjustments - No goalie metrics included - No rest/travel factors - Team strengths are deterministic

12. Future Improvements

- Fit attack/defense parameters using maximum likelihood - Integrate goalie analytics (GSAA, xG_saved) - Use the Dixon–Coles time-adjusted Poisson model - Add Monte Carlo simulation for more detailed projections - Incorporate lineup, rest-day, and injury adjustments