Models for Olympic Medal Tables
2025 MCM Problem C

Sofia Bonyadi, Aryan Dalal, Jiayi Cora Guo
Julie Quan, Shani Shaham

Spring 2025

MATH 142: Introduction to Mathematical Modeling
University of California, Los Angeles (UCLA)

Problem

In 2028, Los Angeles will host the 34th iteration of the modern Summer Olympics. For many athletes, the Olympics games is a platform to showcase their brilliance and bring pride to their nation.

As one of the major global sporting events, the Olympics games is naturally one of the most viewed and followed upon event in the world.

This talk is motivated by the curiosity to predict the medal table for the next iteration. We will also discuss medal performance by country and share who's likely to be the front-runner for the LA2028 games.

Introduction & Motivation

Task: To come up with a model for medal counts (Gold and otherwise) for each country including uncertainty intervals and measures of our model's performance.

Problems: Model type (statistical or mathematical or both)? Domain? Nonlinear Dynamics, Stochastic Process, Machine Learning? How do we evaluate uncertainty? Randomness?

Proposed Approach: Monte-Carlo Simulations

Why?: We can use Monte-Carlo Simulations to predict probabilities of a variety of presently unknown events utilizing inferential statistics. What is *inferential statistics*? We draw a random sample exhibiting the same properties of the population from which it is drawn. We then draw inferences after data analysis and reach a conclusion.

Preliminaries for Monte-Carlo Simulations

Let's begin with a quick crash-course to definitions.

Suppose there exists a sample space, Ω–this sample space contains everything that could possibly occur in our model. We denote outcomes by members of this space $\omega \in \Omega$.

Definition. *(Random Variable)*
A *random variable* is a function that sends outcomes from Ω into $\mathbb{R}$.

$$X(\omega) : \Omega \longrightarrow \mathbb{R}$$

Quick Note: A good introduction to probability theory would require a foundational understanding of measure theory. What makes *random variables* so powerful is that they're equipped with the property of being measurable.

Expectation

A random variable can be *discrete* or *continuous*. A discrete random variable takes in a countable number of distinct values. A continuous random variable takes in an uncountable infinite number of values on an interval.

Definition. *(Expected Value)*
Let X be a discrete random variable for x_i. The *expected value* of X, denoted $\mathbb{E}[X]$ is defined by

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i P(X = x_i)$$

What this means is that the expected value is the weighted average of the possible outcomes of a discrete random variable.

The fundamental idea of employing a Monte-Carlo method is to solve for an expected value.

Onwards to Monte-Carlo Simulation Method

Let X be the set of countable points. The discrete probability distribution on X is defined by the function $p(x)$ on X such that $0 \leq p(a_i) \leq 1$ and $\sum_{i=1}^{\infty} p(a_i) = 1$. Intuitively, it's simply what allows us to give probabilities to possible outcomes.

A **Monte-Carlo Method** is a computational tool to simulate random variables.

Inherently, it's an algorithm that allows us to run stochastic simulations in order to approximate outcomes.

Methodology: Vision, Assumptions, Implementation

Goal: Create a model to predict the 2028 Olympic medal tables

Central Question: Does skill matter? Is winning a medal simply a game of chance?

This question influenced our decision to build a *probabilistic*, simulation based model.

Baseline Model: Utilizing the Monte-Carlo simulation framework, we estimate future medal performance based on historical participation data. The central assumption of this model is that *all athletes in each event have an equal chance of winning a medal, regardless of nationality or historical participation*.

Why? This allows the model to reflect the intuitive idea that higher participation translates to greater medal opportunity, even under neutral skill assumptions.

General Assumptions to the Model

1. LA2028 will have the same number of events as Paris 2024.
   Why? *It's a forecasting assumption in the absence of a confirmed 2028 event schedule.*
2. For each event E_i, $i \in \mathbb{Z}$, 16 athlete slots are sampled and filled.
   Why?*It's a realistic approximation of typical Olympic event finals (e.g. Track & Field, Swimming, etc.)*
3. Countries are sampled to participate based on historical frequency.
   Why? *Countries tend to consistently enter events they've historically competed in.*
4. All events are simulated as invidiual as oppposted to team events.
   Why? *A simplifying measure for our model. Most Olympic events are generally individual.*

Baseline Model Assumption

All participants have an equal chance of winning in a single event.

Why? Removes bias from historical medal counts and emphasizes stochastic dynamics in elite competitions where competitive margins are narrow. It enables us to isolate the effect of participation volume on medal outcomes.

Estimating Participation Rate

To realistically simulate the structure of Olympic events, we first calculate the frequency with which each country has historically participated in each event.

From athlete data (1992-2024), we count the number of appearances by each country in each event and normalize this by the number of Olympic Games in the period. This gives us

$$\textbf{AvgParticipation}_{E,\mathcal{C}} = \frac{\#\text{ Athletes from Country } \mathcal{C} \text{ in Event } E}{\#\text{ Olympic Games since 1992}} \quad (1)$$

Denote $P \in \mathbb{R}^{E\times\mathcal{C}}$ the participation matrix storing these values. For each event E, the row P_E is the probability distribution over countries and is used to sample which countries would send athletes to the event E.

Uniform Medal Assignment

Given $P \in \mathbb{R}^{E \times \mathcal{C}}$, let $\textbf{AvgParticipation}_{E,\mathcal{C}} < 0.3$ denote the threshold to permit removal of countries with weak participation.

We simulate $n = 16$ athlete slots with replacement that is, a country may appear multiple times in a single event (allows us to reflect countries with deeper delegations to have more chances to win).

Events lacking sufficient medal data are excluded from simulations.

From the sampled athlete slots, we require *at least* 3 unique countries to ensure models can be awarded properly.

Select 3 unique countries at random to receive Gold, Silver and Bronze medals for that event.

Selected winners recorded in medal matrix, M.

Monte-Carlo Averaging

The above process simlates one cycle of Olympic Games. We repeat the simulation for count $N = 1000$. The expected number of medals of type $m_{\text{type}} \in \{\text{Gold, Silver, Bronze}\}$ for country $\mathcal{C}$ is given by

$$\textbf{ExpectedMedals}_{\mathcal{C},m_{\text{type}}} = \frac{\#\text{ Medals of } m_{\text{type}} \text{ won by } \mathcal{C}}{N}$$

Final result is an expected medal table by total medals per country.

Model Improvement

Thus far, what we have described has been for a baseline model. This is a neutral starting point allowing us to isolate the impact of participation volume.

Unfortunately, the baseline model lacks predictive power for historically strong nations. A reasonable guess would be that nations with strong existing infrastructure would be amongst the top in the medal count table.

This motivates our development of an improved model that incorporates historical medal outcomes into performance-based weighting.

Additional Assumptions

1. Historical medal performance informs current medal probabilities.
   Why? *Countries with strong sporting infrastructure tend to remain strong over time.*
2. Countries with no historical medals still have a non-zero chance of winning.
   Why? *Stochastic dynamics and constant improvement of infrastructure.*
3. Countries cannot send more than 3 athletes per event in the simulation.
   Why? *Following Paris 2024, there exists a maximum of three athletes per eligibility of the NOC.*
4. Host nation receives a Multiplier boost in participation and performance probability.
   Why? *Reflects the "host effect" phenomenon.*

Model Enhancements

Decaying Weight:

$$\omega_y := e^{-\alpha(2028-y)} \qquad\qquad \alpha = 0.15$$

For each $\mathcal{C}$, event E, and medal type m_{type}

$$\textbf{WeightedMedals}_{\mathcal{C},E,m_{\textbf{type}}} := \sum_{y;\ \text{multiples of 4}}^{2024} \textbf{MedalsWon}_{\mathcal{C},E,m_{\textbf{type}}} \cdot \omega_y$$

Let $M_{\text{gold}}, M_{\text{silver}}, M_{\text{bronze}}$ denote medal probability matrices where the ith row is an event and jth column is a country. Athletes sampled from participation matrix P.

Model Enhancements

For each event, medals are assigned by weighted probability. For $\mathcal{C}$ in event E, probability of winning a medal is given by

$$P\left(\mathcal{C} \text{ wins medal}\right) \propto \left(\textbf{WeightedMedals}_{\mathcal{C},E,m_{\textbf{type}}} \cdot \textbf{Slots}_{\mathcal{C}}\right)^2$$

For "Host Effect" Phenomenon,

$$P_{\mathcal{C},E} \leftarrow P_{\mathcal{C},E} \qquad\qquad \text{if } \mathcal{C} = \texttt{United States}$$

$P_{\mathcal{C},E}$ denotes each $\mathcal{C}$'s probability of being sampled for a given event E. If the host country is eligible to win a medal, its weight in the medal matrix is also scaled by 1.05.

$$\textbf{WeightedMedals}_{\texttt{USA}} \leftarrow \textbf{WeightedMedals}_{\texttt{USA}} \cdot 1.05$$

To ensure that countries with no prior medal history are not excluded from winning, a small baseline value of $\varepsilon = 10^{-6}$ is added to each entry in the medal probability matrices.

Results and Analysis: LA2028 Projected Medal Table

Table 1: Data with 95% prediction intervals (Integer point estimate, Decimal bounds).

Country	Gold	Silver	Bronze	Total
USA	36 (36.0–36.4)	39 (39.1–39.5)	30 (29.9–30.3)	106 (105.2–105.9)
CHN	20 (20.3–20.6)	15 (15.2–15.5)	14 (13.8–14.1)	50 (49.4–50.1)
GBR	12 (11.8–12.1)	11 (11.3–11.5)	17 (17.3–17.6)	41 (40.5–41.0)
AUS	13 (12.4–12.7)	14 (13.5–13.8)	14 (13.6–13.9)	40 (39.7–40.2)
ITA	7 (7.2– 7.4)	10 (10.1–10.4)	13 (12.6–12.9)	30 (30.1–30.5)
GER	9 (8.4– 8.7)	7 (7.4– 7.6)	7 (7.3– 7.5)	23 (23.2–23.6)
JPN	8 (8.3– 8.5)	7 (7.2– 7.5)	8 (7.4– 7.7)	23 (23.1–23.5)
FRA	5 (5.2– 5.5)	9 (8.6– 8.9)	7 (7.0– 7.2)	21 (21.0–21.5)
NED	7 (7.3– 7.6)	7 (6.5– 6.7)	7 (6.4– 6.6)	21 (20.4–20.8)
CAN	7 (6.9– 7.1)	5 (4.8– 5.0)	6 (5.7– 5.9)	18 (17.5–17.9)

Scaling and Rounding

Due to missing data and fewer events in the 2024 Paris Olympics, our initial simulation totaled only 243 medals per type. To accurately represent the 356 events expected for the 2028 Olympics, we applied scaling and rounding methods:

1. **Scaling:** Medal counts were proportionally adjusted to sum up to approximately 356 medals per type.
2. **Minimum Medal Guarantee:** Each country with a positive expected medal count received at least one medal.
3. **Rounding:** Medal counts were rounded to the nearest integer, minimizing discrepancies and preserving original proportions.

This approach produces a realistic and representative medal distribution for the 2028 Olympics.

Top 50 Countries by Total Medal Count

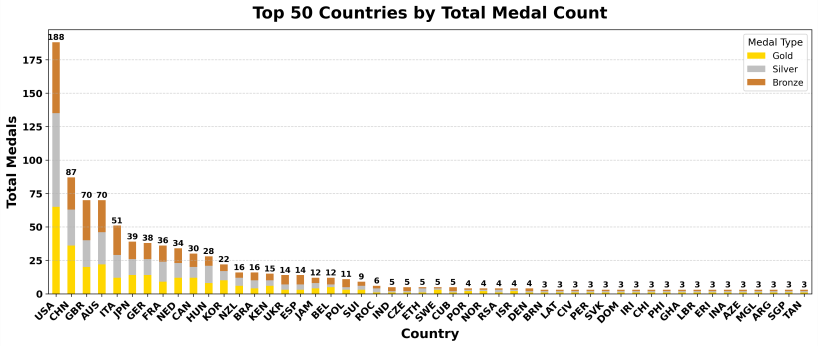

Figure 1: Medal Count corresponding to Table 1.

Change Relative to Paris 2024

Instead of raw counts, the following figure compares each country's *share* of the overall medal pool in Paris 2024 (denoted in blue) with its simulated share for LA2028 (denoted in orange)

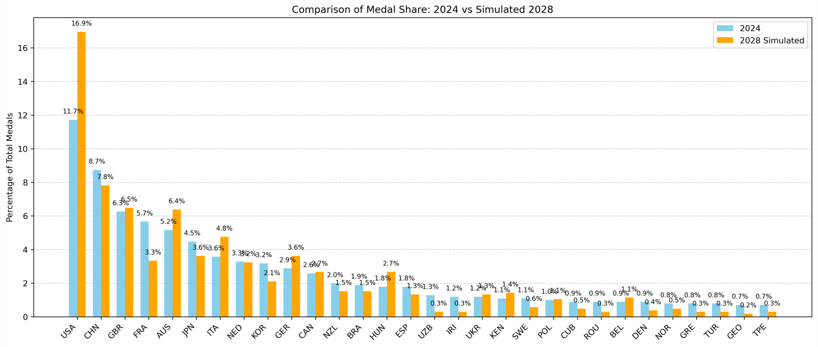

Figure 2: Share of total medals: Paris 2024 results (blue) versus our LA2028 simulation (orange). Labels give the exact percentage of all medals won.

First-Time Medal Probabilities

Only a handful of countries with no prior Olympic podium finish are forecast to breakthrough in LA2028. Liberia leads with +7.1%. Guinea, Maldives and Mali sit at 0.2%. All other candidates lie below < 0.2%.

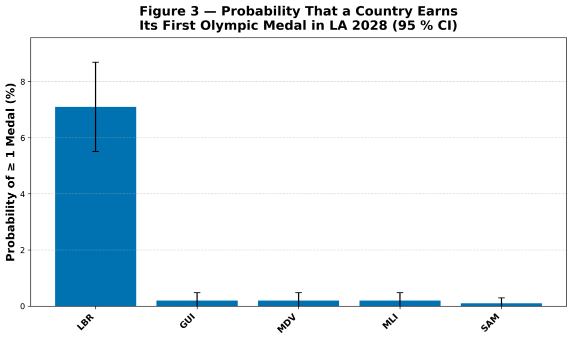

Figure 3: Error bars denote 95 % prediction intervals from 1000 Monte-Carlo runs.

Home-Country Bias

What if we didn't have a home-country bias? We run the Monte-Carlo simulations again with **HostMultiplier** := 1 and compare against our 1.05 multiplier.

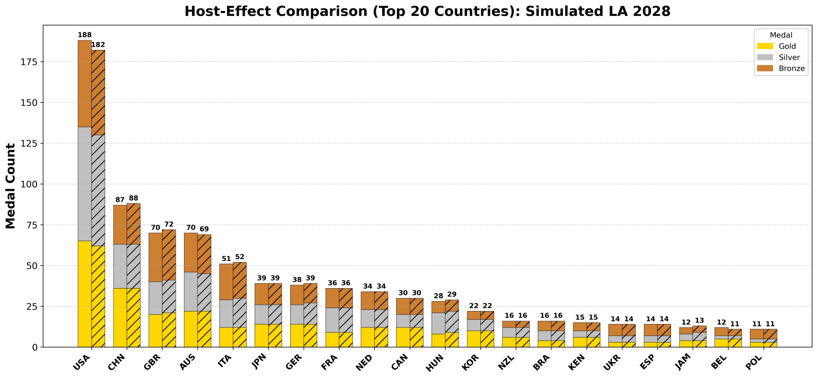

Figure 4: Solid Bars: 1.05 Multiplier; Hatched Bars = 1.00 Multiplier.

Our Model is Not Perfect

While our model provides generally acceptable results into LA2028, there exists a non-zero finite amount of limitations and variables that were not incorporated.

While predictions primarily rely on historical data, a **better prediction** will arise from incorporating both present and historical states.

Our model heavily relies on simply historical patterns in medal wins but non-sporting metric such as GDP, Economic Development and recent athletic performance are not assumed.

Countries with no past data or current geopolitical conflicts are not accurately predicted. Reliance on historical data inherently favors countries that have won medals in the past. Adding $\varepsilon > 0$ showed little to no effect on outcomes.

Improvements

An initial plan included to decompose results by sport in order to pinpoint each country's core strengths. In practice, this would require mapping each of the 287 simulated events and recomputing shares. Given the extent of the task, we decided to leave sport-level aggregation and this is an avenue for improvement.

A sophisticated treatment of event structures would certainly benefit in providing a more accurate prediction. One of our assumptions were to assume both team and individual events as only individual. This is clearly an assumption we can forgo to improve results.

A New Model: Multiple Linear Regression; Markov-Chain Monte-Carlo Simulations; Graph Learning

Conclusion

What we have achieved so far: We have been successful in employing inferential statistics through the Monte-Carlo framework to provide a prediction that's in-line with what the median viewer would expect.

What can we do better: Our modeling process involved significant measures–to reduce computational cost and decrease model complexity. However, a revised approach with changing our assumptions can provide insight into whether our assumptions are reasonable or not.

What we would like to do in the future: Excellence can seldom be quantified–as such, we rely heavily on statistical frameworks to guide inference. However, using statistical analysis as a guide, we can deploy different models to target specific sporting phenomenon through *regression*, *optimization*, *game theory*, etc.

Acknowledgement

Sofia B., Aryan D., Jiayi C. G., Julie Q. and Shani S. would like to thank *Professor of Mathematics, Mason A. Porter* and *Graduate Teaching Assistant, Lyu Haoyang* for their inputs during the project timeline.

They were an invaluable resource during our modeling process.

All external resources used up to and including past attempts at modeling medal prediction for the Olympics games have been referenced in the report.

Thank you for a wonderful MATH 142!

$$142! = \Gamma(143) = \int_0^\infty t^{142} e^{-t}\, dt, \quad \Re(z) > 0$$
$$\approx 2.695 \times 10^{245}$$

This document was typesetted with LaTeX