Somersaulting to Gold:
Optimizing USA's Artistic Gymnastics Team Selection Strategy

Emily Zhao, Elyse McFalls, Young Jun, Evan Kim

2023-11-20

I. Introduction

Gymnastics has swiftly risen to become one of America's cherished pastimes, with recent polls revealing it to be the most anticipated Olympic event among Americans.[1] This heightened enthusiasm is largely attributed to the historical success the US has enjoyed in the Olympics, particularly in women's artistic gymnastics. Dominant performances by teams like the Final Five in 2016 and athletes like Simone Biles (2011 - current) have brought immense prominence to the sport and its athletes.[2] While the US men's artistic gymnastics team may not be as dominant, they have consistently secured top-five placements in team events at the Olympics and have bolstered decorated athletes like Khoi Young (2018 - current) in recent years.[3]

We expect that millions of Americans are anticipating the gymnastics events at the 2024 Olympic Games in Paris, France.[4] This anticipation is coupled with the expectation of US dominance at the Olympics, mirroring the success we have seen in the past. One may ask then: How can we ensure that the US reigns supreme? The answer lies in assembling formidable teams capable of winning the most medals.

Historically, US Olympic team rosters are shaped by the results of the US Olympic Team Trials. The trials will be hosted in June, 2024, for this upcoming cycle.[5] Here, each invited athlete's performance determines their spot on the team. Athletes are required to complete all apparatuses twice (4 for women, 6 for men), and scores from each event (8 for women, 12 for men) are combined to generate a final score. The selection committee then uses a combination of the competition results and the teams' needs to decide on the final rosters. In the last Olympic cycle, the top two performers for women and the top performer for men at the trials secured automatic spots, while the remaining slots were determined by the selection committee.[6] Given this meticulous process, it will be some time before the US Olympic gymnastics teams are unveiled, and even longer before we can assess how they will perform on the big stage. The selection committee also faces the arduous task of determining the optimal team makeup. This season, the challenge is particularly difficult for the women's team given the abundance of strong candidates.[7]

In this paper, we present a simulation-based approach to determine the best US Olympic men and women's gymnastic teams for the 2024 Paris Olympics. Our method utilizes historical data on each athlete's performances to simulate pseudo-Olympic data to see how well the US will fare under different team makeups. We imagine that our algorithmic approach would be useful supplementary material for the selection committee to consider when finalizing the roster. Moreover, the results of these simulations may also be enjoyed by fans of the sport who are interested in gaining some insight into what to expect from the Olympics given the athletes' past scores.

II. Methodology

Data Collection:

We were originally provided two data sets from UCSAS: the 2020 Tokyo Olympics women's results and the 2022-2023 season's men and women's results from various competitions. Given the dynamic nature of gymnastics, where athletes' careers are typically short and marked by rapid changes, the period between two Olympic cycles is crucial. It often sees the retirement of established gymnasts and the rise of new talents, significantly altering the sport's competitive hierarchy. (Atikovic et al., 2017)[8] Therefore, we have elected to only consider the 2022-2023 season's results as relevant data as it more accurately reflects the current competitive landscape and athlete performance trends. In addition, via web scraping and research, we were able to add athletes, both US and non-US based, that were not in the provided data set but are worth considering in our simulations.

Any scores that were missing or recorded as "0" were dropped from the input data for the simulation as we assumed the athlete had chosen not to compete in the event. We also extensively cleaned the data for inconsistencies, mis-typed, and changed names for each athlete, while also checking for athletes who switched country allegiances.

Motivation for Kernel Density Estimation (KDE):

Our refined data set shows the distribution of scores for each athlete-apparatus combination over the most recent two years. This observation suggests that each combination possesses a unique, true underlying distribution of scores, justifying the use of density estimation. Our method of choice is Kernel Density Estimation (KDE), as opposed to piece-wise histogram density estimation, given the near-continuous nature of gymnastics scores. It offers a significant advantage by centering a smooth kernel function at each data point and aggregating these to form an overall density estimate.

While there are over twenty packages in R that assist with density estimation, varying in both theoretical backdrop and computational performance, we employ the `KernSmooth::bkde()` function, renowned for its speed, accuracy, and robust maintenance. Developed by Matt Wand and Brian Ripley in 2010, `KernSmooth::bkde()` implements the linear binning technique, which distributes data mass between grid points for enhanced accuracy by utilizing "weighted bin counts instead of individual data points". (Deng et al., 2011)[9]

Developing the Simulation:

For the purpose of our analysis, we have fine-tuned the `KernSmooth::bkde()` function to our data set's specific characteristics. We selected the Gaussian kernel for its efficacy in creating a smooth, continuous approximation of the underlying score distributions, while deliberately forgoing the scale normalization which is typically associated with canonical kernels to maintain the natural scale of gymnastics scores.[10] Also, the `bandwidth` is set as 0.25 all throughout, which yields the optimal empirical balance between over-smoothing and under-smoothing for our aggregated data set. (See Appendix, Figure 1)

Furthermore, for the range of scores we should estimate for each athlete-apparatus combination, we incorporated a `nugget` metric to allow extra variability for athlete-apparatus combinations with fewer than four observations. For example, if the minimum score for an athlete-apparatus combination is initially 4.3, and the nugget adjustment brings it down to 2.3, we re-calibrate the lower limit of the range from 4.3 to 2.3 before proceeding with the kernel density estimation, while ensuring that the lower limit never reaches below 0, a logical baseline for gymnastics scores. This facet provides a buffer that accommodates the potential for higher variability in sparser data. Our methodology is grounded in the principle that past performance can be a predictor of future results, but with a recognition of the stochastic nature of athletic performance. The nugget is calibrated using the standard deviation of all scores for a given apparatus, rewarding consistent high-level performance over time rather than exceptional one-off performances.

Incorporating our fine-tuned `KernSmooth` function with our own simulation algorithm, we conducted a total of 1,000 iterations for the following: 4 women's individual apparatuses — Balance Beam (BB), Floor Exercise (FX), Uneven Bars (UB), Vault (VT) — and 6 men's individual apparatuses — Horizontal Bar (HB), Floor Exercise (FX), Parallel Bars (PB), Pommel Horse (PH), Still Rings (SR), and Vault (VT). By conducting a total of 1,000 iterations of simulated Olympic events separately for both men and women, we generated a robust collection of data sets reflecting the potential range of scores for each athlete-apparatus combination, assuming all athletes compete in the events they have historic data for. The aggregation of these results facilitated a comprehensive analysis of the top performers for each apparatus and potential team compositions for both the US and other team-qualifying countries.

Weighted Individual Medal Contributions:

In order to determine the medal potential of each US gymnast, we calculated an average weighted medal score for each event. We assigned gold a value of 3, silver a value of 2, bronze a value of 1, and anything lower a value of 0. For each simulation, we assigned each competitor a weighted score based on their ranking and then calculated the average weighted score across all simulations. If a gymnast did not have historical data for a specific event, their simulated results would return 0, subsequently removing them from contention for that specific event. When two gymnasts happen to earn the same score within one of the top three spots, they would both be rewarded the same medal score. This average weighted medal score is relative to the performance of other gymnasts and therefore is a better representation of a gymnast's potential on the team than their score. Then, across simulations of each event, we found each competitor's average weighted medal score. By conducting this weighted medal calculation for each women's and men's event, we can determine the most competitive gymnasts by event and also determine how frequent and valuable their medal contributions are.

This calculation averages the gymnast's range of medal placements in an event to a single value. For instance, if a gymnast scores only one bronze throughout the 1,000 simulations of an event, their average weighted medal score would be 0.001. Meanwhile, a gymnast who consistently earns silver or gold throughout the 1,000 simulations will obtain a weighted average medal score between 2 and 3. This process allows us to identify which individual events Team USA has the highest chance of winning and whether there are particular events that may benefit from a "specialist", a gymnast who may only excel at one particular event.

We used this calculation for all four apparatus for women, all six apparatus for men, and the individual all-around event for both women and men. The individual all-around event weighted score calculation is derived from the same simulation data as the individual events. The individual women's and men's all-around scores are calculated by taking the sum of the athlete's performance in each simulation for all 4 and 6 respective events. Then, the same ranking and medal scoring process for each simulation is applied to the summed all-around scores.

Weighted Team Medal Contributions:

To obtain the weighted team medal contributions, we first filtered for the countries that qualified for a full team. Then, from the individual event simulations, we selected the top three athletes in each apparatus per country to generate a group of top-performing women and men. From this group, we created all possible combinations of 5 athletes from the groups of top performers per country. This left us with all of the possible teams for all of the qualified countries given our criteria. Following the team all-around rules and using the top three scores for each event, we generated all of the scores for each team.[11] The team with the highest score for a given country was selected for each simulation and ranked against the other countries' best teams to determine medal placements for the simulation. Similar to the weighted individual medal contributions, teams that receive gold are assigned a weight of 3, silver placements are assigned a weight of 2, bronze placements are assigned a weight of 1, and anything else is assigned a weight of 0. The average of a team's placement is their expected medal contribution given that their team was selected as one of the best US teams at least once. Teams that are not selected as the best US team in any simulation will have an expected medal contribution score of 0. The rationale behind this is that we want to prioritize teams that are more consistently the best. A team that has a high score and/or high expected medal count but only reigns as the best U.S. team once or twice is likely to be a fluke given the nature of our simulations. Moreover, teams that do not appear as the best US team in any simulation likely will not be the best US team for the team event.

For men's, the other qualifying teams include Canada, China, Germany, Great Britain, Italy, Japan, Spain, Switzerland, Netherlands, Turkey, and Ukraine. For women's, the other qualifying teams include Australia, Brazil, Canada, China, France, Great Britain, Italy, Japan, Netherlands , Romania, and South Korea.[12] All these teams had at least five athletes qualify under our initial threshold criteria except for the men's Netherlands team which only had four athletes qualify under our criteria. After further research on the historic performance of the Netherlands men's team, we believed that they are not in contention for a team all-around medal and therefore removed their team from the team simulations.[13]

Assumptions:

We applied a series of assumptions to reduce noise in the data, computational demand, and simplify the selection process. Firstly, we decided to establish a performance cutoff for the competitors inputted into the simulation function. We removed any athlete who has not placed in the top 5 for any competition in the 2022-2023 seasons, as we assumed that their lack of data from the recent seasons or their relatively sub-par performance indicated that they likely would not make an Olympic team. This criteria is supported by the 2020 Team USA artistic gymnastics team selections as all members of the selected teams had placed in the top 5 of a competition from 2019 to 2020. This criteria allows for flexibility as even if a gymnast is a "specialist", their top 5 strong performance in a singular event will be sufficient for including them in the data set. However, individual athletes who already qualified for the 2024 Olympics were not bound to this cutoff.[12]

In the selection of our team, we also assume that these gymnasts will be performing at their optimal condition. In our selection process, we have also removed some athletes which have either announced their retirement or have suffered a recent significant injury. One notable gymnast who was removed from the data set due to this criteria is Suni Lee. At her peak performance, gymnastics fans believe she is a strong contender on both uneven bars and balance beam.[14] However, in April 2023, she announced that she was withdrawing from collegiate gymnastics and had been diagnosed with an incurable kidney disease. The uncertainties surrounding her disease and its progression over the next year would potentially make her chance of performing well volatile, so we could not confidently place her in our recommendation given her medical condition unknowns. For the rest of the athletes, we assume that gymnasts will be able to participate in multiple events throughout the Olympics competition and will also not get hurt or injured from now until the Olympics.

In the simulation of results, we grouped all historic vault scores, either from team or the first or second round of individual vaults, all into a common vault category, unifying labels such as "VT1" and "VT2" into simply "VT". We also applied the assumption that a gymnast's performance in each round — qualifying, final, all-around qualifying, all-around final, team qualifying, and team final — would all follow a similar kernel density distribution and therefore can be aggregated into one cumulative apparatus category. To test the assumption of similar scores by round, we inputted all data from our constrained data set to conduct three-way ANOVA tests by athlete, round stage (final or qualifying), round type (all around, individual, or team), and all interactions with score as the response. We used a significance level of 1.25% for women and 0.83% for men to give each test a Type I error rate of 5% with Bonferroni's correction.[15]

For both men and women, the effect of athlete on score was unsurprisingly a significant effect (p-value: <0.001), controlling for all else. For women, only the balance beam apparatus showed sufficient evidence for a difference in score by round type (team, all around, individual) with a p-value of 0.011, controlling for all else. For men, the interaction between athlete and round type was significant for the floor apparatus (p-value: <0.001). This finding supports the notion that the effect of round type on floor scores is contingent on the athlete or vice versa. Given the overwhelming insignificant outcomes for the round stage and round type main effects along with their interactions at our specified significance levels, we moved forward with aggregating all of the results for each apparatus. The figures below show the average score for each apparatus by round for the top scorers to further demonstrate similarities across rounds. (See Appendix - Figure 2, 3)

III. Results

Individual Events - Top USA Women A total of 180 female athletes were each simulated 4,000 times to generate 1,000 performances in each of the four apparatus events in women's gymnastics. After conducting four simulations reporting for the four events, we generated an aggregated group of seven athletes where all athletes ranked in the Top 3 Americans for any event. In the calculations of success, we used the metric "weighted medal score" to represent their weighted medal contribution to Team USA. As mentioned earlier, bronze represents 1, silver 2, and gold 3. This score does not indicate how many medals each athlete earns but rather the "value" of their medals earned.

In the table below, each event column has three yellow-highlighted cells, indicating the athletes with the top 3 medal scores, and one blue-highlighted cell, indicating the fourth highest medal score. Simon Biles emerged from the simulation as having a top 3 medal score in all five events (BB: 1.515, FX: 2.747, UB: 0.064, VT: 1.94, AA: 2.962). Shilese Jones posted top 3 medal scores for UB, VT, and AA and also fourth place scores in BB and FX (BB: 0.034, FX: 0.195, UB: 0.567, VT: 0.260, AA: 0.595). For the rest of the gymnasts, they either placed in the top 3 scores for one or two events. (See Figure 4 below)

From the total weighted medal scores, assuming each athlete competes in all events, Biles emerges as the clear strongest contributor to the weighted medal scores (9.228). Konnor McClain posts the second highest total weighted medal score (2.016). The gap between Biles and the rest of the women candidates is wide, expectedly as most gymnasts only show weighted medal scores above 0.1 for one or two events.

IV. Discussion

Limitations:

Extensions:

V. Appendix

Figure 1. Below is an example of the KDE density estimation for a specific athlete-apparatus combination based on scores from the 2022-2023 seasons, using our fine-tuned `KernSmooth::bkde()` function. The red dots represent the scores from the data set, and the blue line estimates the underlying distribution, using the Gaussian family as a parameter.

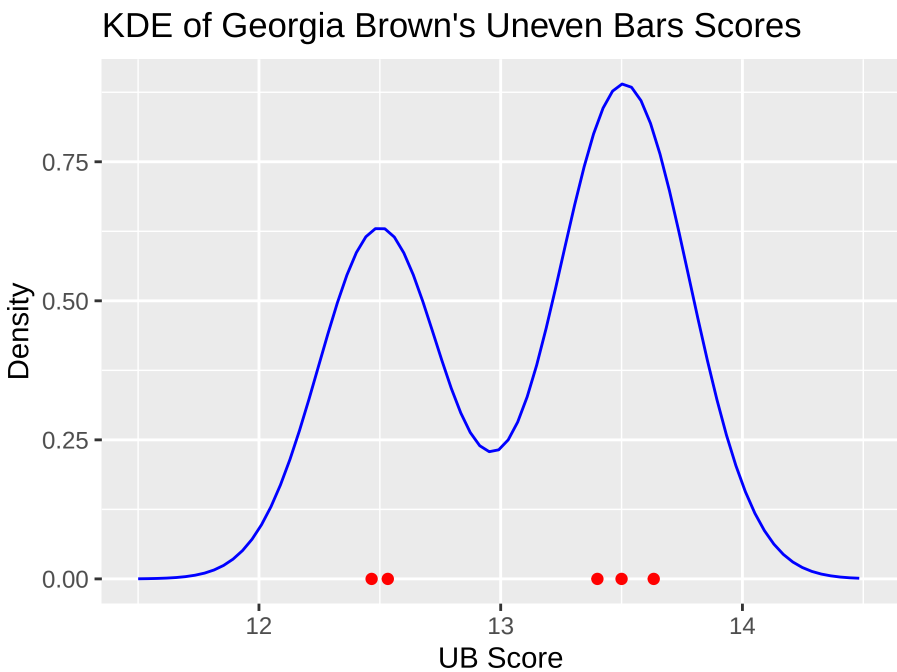

Figure 2. Below is a table summary of the p-values for Three-way ANOVA for women, where each row represents a particular apparatus. "Round", as written in the table below, is short for Round Stage (final vs. qual); "Type" is short for Round Type (team, all-around, individual); The symbol 'x' indicates an interaction term. As expected, every test showed a significant relationship between athlete and score, controlling all else, indicating that at least one athlete has a different score than all the others for each apparatus. (*Significance level of 0.0125 was set to give a Type I error rate of 5% with Bonferroni's correction)

Apparatus	Athlete	Stage	Type	Athlete x Stage	Athlete x Type	Stage x Type
Balance Beam	<0.001*	0.734	0.011*	0.943	0.753	0.161
Uneven Bars	<0.001*	0.022	0.021	0.897	0.025	0.257
Floor Exercises	<0.001*	0.801	0.418	0.990	0.998	0.985
Vault	<0.001*	0.030	0.328	1.000	1.000	0.723

Figure 3. Below is a table summary of the p-values for Three-way ANOVA for men, where each row represents a particular apparatus. "Round", as written in the table below, is short for Round Stage (final vs. qual); "Type" is short for Round Type (team, all-around, individual); The symbol 'x' indicates an interaction term. As expected, every test showed a significant relationship between athlete and score, controlling all else, indicating that at least one athlete has a different score than all the others for each apparatus. (*Significance level of 0.0083 was set to give a Type I error rate of 5% with Bonferroni's correction)

Apparatus	Athlete	Stage	Type	Athlete x Stage	Athlete x Type	Stage x Type
Horizontal Bar	<0.001*	0.911	0.930	0.997	0.746	0.912
Parallel Bars	<0.001*	0.680	0.721	0.161	0.708	0.810
Floor	<0.001*	0.812	0.091	0.182	<0.001*	0.109
Still Rings	<0.001*	0.185	0.113	0.836	0.053	0.992
Pommel Horse	<0.001*	0.015	0.101	0.711	0.186	0.114
Vault	<0.001*	0.039	0.047	0.954	0.999	0.576

VI. Sources Cited

1. "Poll: Gymnastics Is America's Most Highly Anticipated Olympic Sport". SGB Media. 2021. https://sgbonline.com/poll-gymnastics-is-americans-most-highly-anticipated-olympic-sport/
2. "How one adjustment helped USA Gymnastics become a power and led to Rio gold". Sports Illustrated. 2016. https://www.si.com/olympics/2016/08/10/usa-womens-gymnastics-system-rio-olympics-martha-karolyi.
3. United States men's national artistic gymnastics team. Wikipedia. https://en.wikipedia.org/wiki/United_States_men%27s_national_artistic_gymnastics_team
4. Paris 2024 - Artistic Gymnastics. https://www.paris2024.org/en/sport/artistic-gymnastics/
5. "Minneapolis, Minnesota to host 2024 US Olympic Team Trials – Gymnastics, become Gymnastics City USA 2024". USA Gymnastics. 2023. https://usagym.org/minneapolis-minnesota-to-host-2024-u-s-olympic-team-trials-gymnastics-become-gymnastics-city-usa-2024/
6. "What to know about the US Olympic gymnastics trials". The Seattle Times. 2021. https://www.seattletimes.com/sports/olympics/what-to-know-about-the-u-s-olympic-gymnastics-trials/
7. "Picking a strong US gymnastics team is easy. Whom to leave out is hard". The Washington Post. 2023. https://www.washingtonpost.com/sports/olympics/2023/08/29/us-gymnastics-team-world-championships/
8. Atikovic, A., Kalinski, S. D., Cuk, I. Age trends in artistic gymnastic across world championships and the olympic games from 2003 to 2016. Science of Gymnastics Journal. 2017. https://www.researchgate.net/publication/320335292_Age_trends_in_artistic_gymnastic_across_world_championships_and_the_olympic_games_from_2003_to_2016
9. Deng, H., Wickham, H. Density estimation in R. 2011. https://vita.had.co.nz/papers/density-estimation.pdf
10. Wand, M., Moler, Cleve., Ripley, B. "Package KernSmooth". CRAN. 2023. https://cran.r-project.org/web/packages/KernSmooth/KernSmooth.pdf
11. "Gymnastics 101: Competition format". NBC Olympics. 2021. https://www.nbcolympics.com/news/gymnastics-101-competition-format
12. "Who Qualified to Paris 2024?". The Gymternet. 2023. https://thegymter.net/2023/10/09/who-qualified-to-paris-2024/
13. "Worlds Team Preview #8 - The Netherlands". Reddit, r/Gymnastics. 2023. https://www.reddit.com/r/Gymnastics/comments/16nqb9r/worlds_team_preview_8_the_netherlands/
14. "Olympic gymnastics champion Suni Lee began pointing to 2024 long ago. It could be a difficult road". AP News. 2023. https://apnews.com/article/sunisa-lee-paris-olympics-gymnastics-95c3857ec9338170792baca56ffe6c07
15. "ANOVA and the Bonferroni Correction". Towards Data Science. 2021. https://towardsdatascience.com/anova-vs-bonferroni-correction-c8573936a64e?gi=ffb12a01219d