VARIANCE IN MEDAL EXPECTANCY PREDICTIONS

Abigail Snyder
WIST Summer Fellow 2023

July 28, 2023

Monte Carlo Simulations & Medal Expectancy

This project focuses on Medal Expectancy predictions for future Olympic Games.

- Monte Carlo simulations are used to create a medal expectancy (ME) model for athlete performance in future Olympic games. The model samples from data about each athlete's performance and then runs simulations to estimate and predict how the athlete may perform in competition.
- Because the model randomly samples from a normal distributions described by mean and variance, the simulation output changes slightly with each iteration.

Problem Statement

- How much can we expect the Monte Carlo output to change between runs due to the inherent variance of the simulation?
- What is the threshold that indicates a meaningful change in ME vs model variance – 0.5%? 1%? 5%?
- What is the optimal number of simulations to run during our Monte Carlo process to converge to the “true” medal expectancy?
- Which athletes experience the largest variance in ME from run to run?

Process

1. Sport Observation

I started by making observations on individual sports, running sets of the pre-existing Monte Carlo simulations in order to identify the variance of each set of simulations, as well as the change in medal expectancy from run to run.

2. Simulation Choices

Initially, I ran 10 iterations of 100, 1000, 3500, 5000, 7500, 10,000, 20,000, and 50,000 repetitions of the Monte Carlo simulations, and then, as I began to see a pattern in the results, narrowed it to 5 simulations of 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 5000, and 7500.

3. Standard Deviation

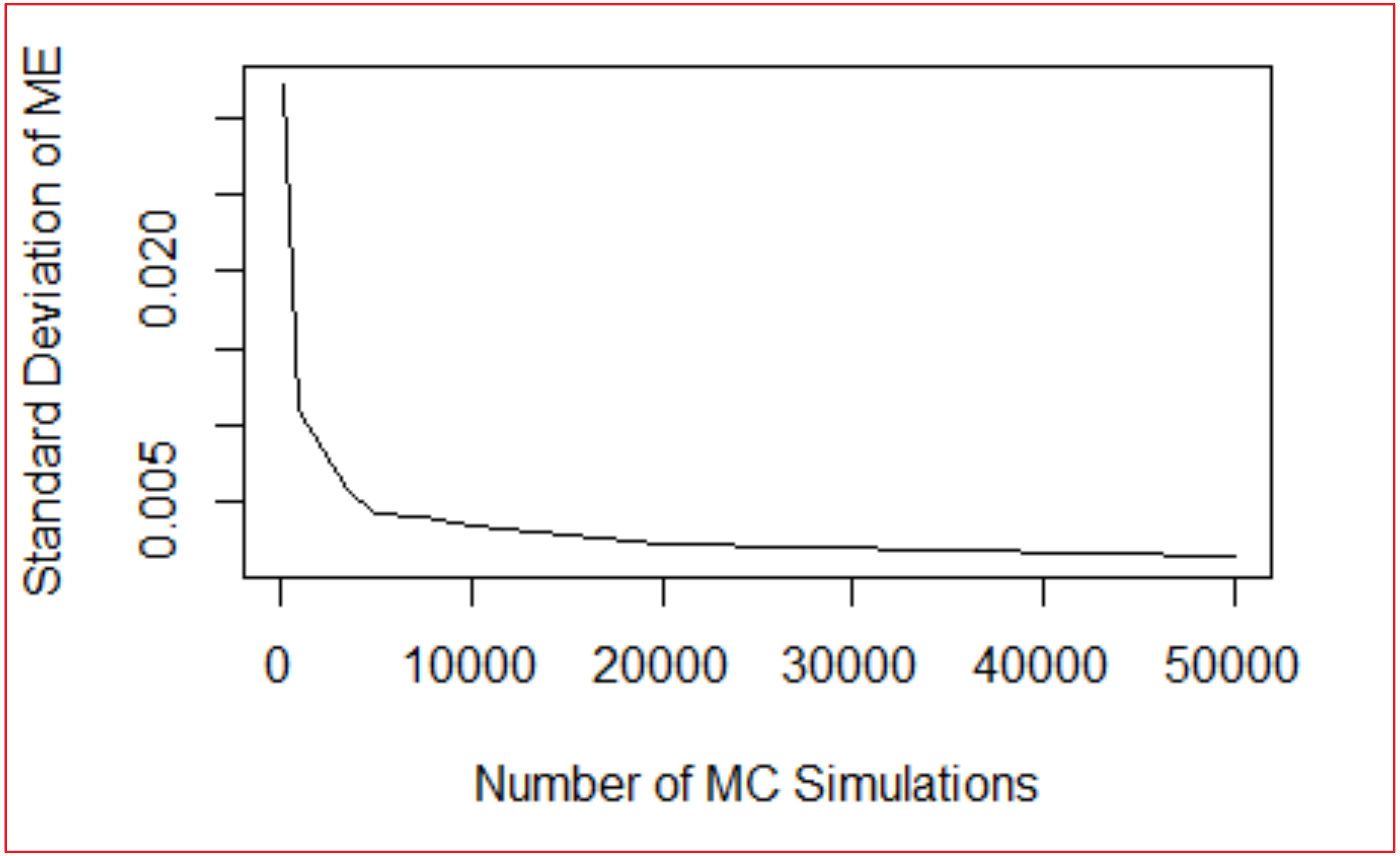
Within each set of simulations, I calculated the standard deviation of the ME for the Monte Carlo process, as well as the runtime for each set of simulations.

4. Elbow Curve Calculation

I also calculated the elbow of the standard deviation curve in order to mathematically identify the precise number of simulations at which the standard deviation began to level.

OBSERVATIONS

- There is a decrease in variance as the number of simulations increases.
- At some point, that curve levels out and becomes insignificant.
- The find_curve_elbow function from the R pathviewr package, finds the “elbow” in bivariate data, enabling us to mathematically identify this point.



Sport 1: Standard Deviation of
Medal Expectancy for the 10 Athletes with the Highest
Medal Expectancy vs. Number of Monte Carlo Simulations

ELBOW CURVE FOR THE 10 ATHLETES WITH HIGHEST MEDAL EXPECTANCY

Sport		Optimal Number of Simulations as identified by the Elbow Curve	Resulting Standard Deviation of Athlete Medal Expectancy
Sport 1	MC	1500	0.0075
Sport 2	MC	1000	0.0094
Sport 3	MC	1000	0.0097
Sport 4	H2H	1000	0.011
Sport 5	H2H	2000	0.007
Sport 6	H2H	1000	0.007
Sport 7	H2H	1000	0.006

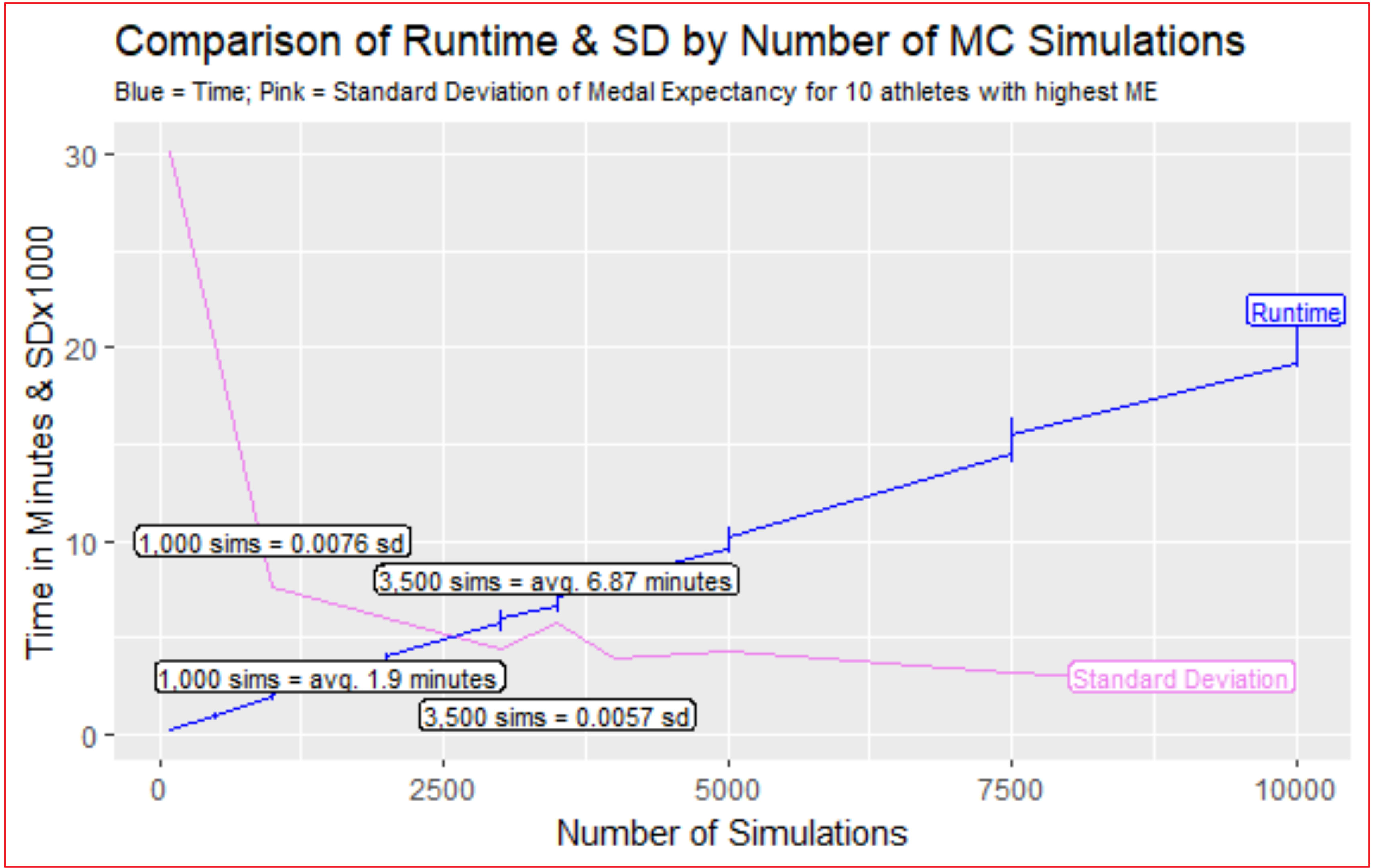
RUNTIME

The runtime of the Monte Carlo process proved to increase linearly as the number of simulations increased. The exact runtime varied between sports, likely as a result of the number of teams, athletes, and competitions involved in any given sport.

Sport	Average Runtime (in minutes) for 1000 simulations
Sport 1	1.6
Sport 2	1.89
Sport 3	7.2
Sport 4	1.94
Sport 5	2.16
Sport 6	4.97
Sport 7`	5.94

RUNTIME

In this graph, which tracks runtime and standard deviation for the 10 athletes with the highest ME within this sport, the minimal gains in accuracy (defined here as a decrease in expected variance) between 1000 and 3500 simulations are noted.



IMPROVEMENTS IN ACCURACY & TIME COST

In order to specify the improvement in accuracy as the number of simulations increases, as well as the cost of the increased runtime, I created comparison tables for several of the sports.

The first column lists the number of simulations, followed by the average standard deviation in medal probability and average runtime. The last two columns indicate the change from one row to the next in runtime and standard deviation as a percentage.

Sport A: Change in Standard Deviation & Runtime				
sims	sd	time	time_pct_change	sd_pct_change
100	0.0061	0.22	NA	NA
500	0.003	1.07	395.14	-59.15
1000	0.002	2.15	101.37	-32.49
1500	0.0017	3.23	49.46	-14.7
2000	0.0014	4.34	34.38	-18.45
2500	0.0013	5.42	24.85	-2.99
3000	0.0011	6.61	21.9	-16.36
3500	0.001	7.67	16.14	-11.61
4000	0.0011	8.81	14.77	7.43
5000	0.009	11.1	26.07	-19.5
7500	0.008	16.61	49.67	-6.65

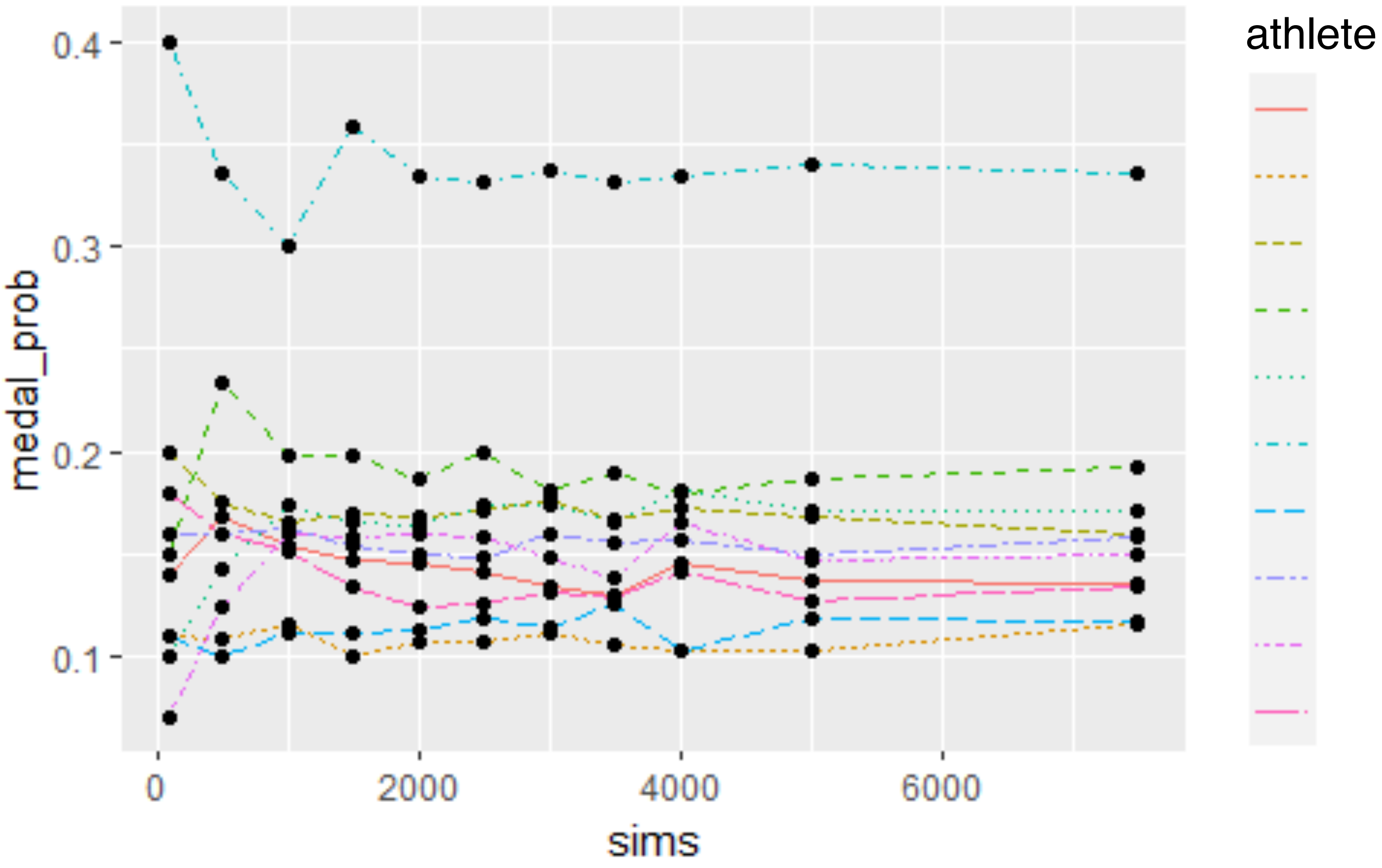
IMPROVEMENTS IN ACCURACY & TIME COST

Sport B: Change in Standard Deviation & Runtime				
sims	sd	time	time_pct_change	sd_pct_change
100	0.0073	0.6	NA	NA
500	0.0035	3.02	403.58	-51.71
1000	0.0024	5.94	97.09	-33.24
1500	0.0018	8.74	47.13	-22.08
2000	0.0014	11.53	31.84	-23.75
2500	0.0013	14.12	22.5	-8.21
3000	0.0015	17.08	20.94	18.26
3500	0.0012	20.14	17.92	-18.85
4000	0.0012	23.31	15.74	-1.58
5000	0.0009	29.29	25.65	-24.94
7500	0.0006	43.29	47.7	-31.76

OPTIMAL NUMBER OF MONTE CARLO SIMULATIONS

- Mathematically, the optimal number of simulations, as determined by the find_elbow_curve function, varies somewhat between sports.
- Given that the average elbow of the curve is approximately 1,800, and that the improvement in expected variance ranges from 27% to 35% between 1,000 and 2,000 simulations, my recommendation would be to **run each sport's Monte Carlo simulation 2,000 times to optimize the results.**
- The caveat to this is that the approximate runtime per sport is between 3-4 minutes, which, for 950 sports is approximately 63 hours: **close to double the current estimated runtime** (at 1,000 simulations).

Sport B:
Number of Monte Carlo Simulations vs. Medal Probability for the 10 Athletes with the Highest Medal Expectancy



Hypothesis Testing

All Athletes: 0.005

- **Null hypothesis:** the mean of the standard deviation is 0.005 or greater
- **Alternative hypothesis:** the mean of the standard deviation is less than 0.005

```
        One Sample t-test

data:  sport_data_comp$sd.all
t = -14.715, df = 31, p-value = 7.823e-16
alternative hypothesis: true mean is less than 0.005
95 percent confidence interval:
        -Inf 0.002304907
sample estimates:
  mean of x
0.001953939
```

Hypothesis Testing

Top 10 ME Athletes: 0.005

- **Null hypothesis:** the mean of the standard deviation is 0.005 or greater
- **Alternative hypothesis:** the mean of the standard deviation is less than 0.005

```
        One Sample t-test

data:  sport_data_comp$sd.top10
t = 9.7866, df = 31, p-value = 1
alternative hypothesis: true mean is less than 0.005
95 percent confidence interval:
        -Inf 0.007986178
sample estimates:
  mean of x
0.007545222
```

Top 10 ME Athletes: 0.01

- **Null hypothesis:** the mean of the standard deviation is 0.01 or greater
- **Alternative hypothesis:** the mean of the standard deviation is less than 0.01

```
        One Sample t-test

data:  sport_data_comp$sd.top10
t = -9.4389, df = 31, p-value = 6.248e-11
alternative hypothesis: true mean is less than 0.01
95 percent confidence interval:
        -Inf 0.007986178
sample estimates:
  mean of x
0.007545222
```

10 Sports with the highest Standard Deviation among athletes with the highest Medal Expectancy

sport	sims	sd.all	avg_medal_prob.all	sd.top10	avg_medal_prob.top10
Sport 1	2000	0.0012541462	0.020000000	0.010359893	0.21369
Sport 2	2000	0.0052313536	0.114285714	0.009994469	0.29296
Sport 3	2000	0.0022475097	0.038961039	0.009555360	0.24632
Sport 4	2000	0.0016189759	0.043478261	0.009331968	0.29098
Sport 5	2000	0.0017648159	0.052631579	0.009038772	0.29606
Sport 6	2000	0.0014361557	0.015625000	0.008830693	0.19214
Sport 7	2000	0.0007832864	0.010169492	0.008752627	0.21969
Sport 8	2000	0.0028623662	0.088235294	0.008572936	0.28822
Sport 9	2000	0.0009901050	0.019480519	0.008430037	0.23983
Sport 10	2000	0.0037958700	0.130434783	0.008396800	0.29924

Athletes with the highest variance

sport	athlete_team_highest_sd	country_highest_sd	sd_highest_sd	avg_medal_prob_highest_sd
Sport 1	Athlete/Team 1	CAN	0.019049278	0.019480519
Sport 2	Athlete/Team 2	GER	0.018045082	0.038961039
Sport 3	Athlete/Team 3	CRO	0.017732033	0.130434783
Sport 4	Athlete/Team 4	ESP	0.017592612	0.043478261
Sport 5	Athlete/Team 5	GBR	0.016517415	0.090909091
Sport 6	Athlete/Team 6	CUB	0.016507574	0.114285714
Sport 7	Athlete/Team 7	USA	0.016328656	0.014218009
Sport 8	Athlete/Team 8	POL	0.015809016	0.022058824
Sport 9	Athlete/Team 9	CHN	0.015610093	0.013761468
Sport 10	Athlete/Team 10	ESP	0.015182226	0.020000000

Conclusions

1. **2000 Simulations**

   The balance between runtime and optimizing accuracy seems to be 2000 simulations

2. **Level of Significance**

   At 2,000 runs, when considering all athletes, a .5% change in ME can be considered a meaningful change. For the athletes with the highest ME, the variance may be up to 1%.

3. **Athlete Variance**

   There was no clear differentiating factor between those athletes who experienced less variance and those who experienced more variance run-to-run

Further Questions

1. Specific Sports or Athletes?

Are there sports or athletes that experience greater variance based on the data that is used in the Monte Carlo medal prediction process?

2. Contributing Factors?

What factors lend themselves to greater variance in medal expectancy predictions?

Q & A