A Matter of Survival: Using Censored Regression to Analyze Athlete Peak Performances

David Awosoga* and Matthew Chow+

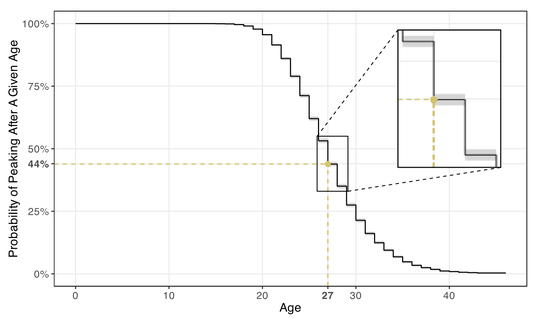
*Faculty of Mathematics, +Faculty of Arts

Introduction

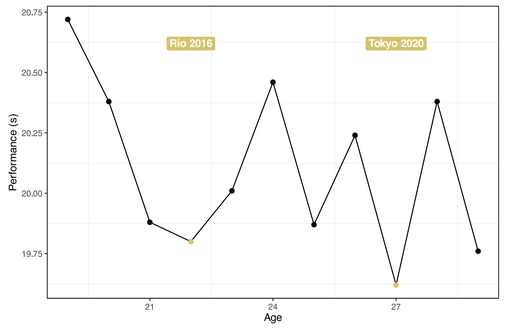
Athlete performances generally follow a parabolic trajectory, ascending to reach a "peak" at some point during their career and trending downwards for the remainder of it. In competitive sports, an athlete's peak is constrained almost exclusively by their age[1]. This has massive implications for competitors in Athletics, where the Olympic Games are their most prolific competition but occur only once every 4 years.

Objective

Since over 70% of competitors in Athletics participate at exactly one Olympic Games[2], understanding when peak performances occur can provide valuable information to maximise an athlete's opportunity to qualify. Identifying a set of covariates that can predict the age of peak performance is essential to quantifying the impact of physiological and event-specific patterns that contribute to the age of an athlete's peak.



A Kaplan-Meier curve showing the probability of peaking after a given age.



The career progression of Andre De Grasse in the 200 m.

Data

- Complete event results from the past seven Olympic Games – Atlanta 1996 to Tokyo 2020 – were paired with individual career progression data for every athlete who competed within this span.
- Athletes who competed in multiple individual events were separated into independent observations.

Exploratory Data Analysis

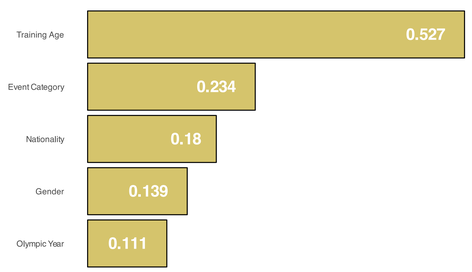
- We use survival analysis to identify the "time until an athlete peaks". Athletes who had a performance after December 2022 are labeled as "active" and therefore censored.
- The average age of participation at the Olympic Games in Athletics is just under 27 years old. Unlike other sports such as gymnastics[3], this has been very consistent at each Games over the past 25 years.
- This age coincides with the median peak age of Athletics competitors, which was found to be 27 years old.

Keeping Up With the Times

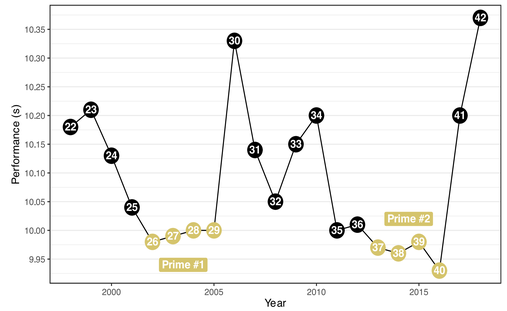
- There are some indications of a relationship between the relative peaks in athlete performance and the years of Olympic Games.
- These preliminary results match our intuition about the incentive that the Games provide, as the difference between an athlete's proximity from their career best in Olympic years versus non-Olympic years is statistically significant.

Predictive Modeling Results

- Five features were chosen in the modeling framework: gender, nationality, event type, an indicator for Olympic year, and training age.
- Training age is included to estimate the number of years that an athlete has trained at an "elite" level.
- Most analyses investigate athlete peak ages using summary statistics and their confidence intervals[4], but here we used an oblique random survival forest[5] to predict peak age given the selected covariates.



Variable Importance



The career progression of Kim Collins in the 100 m from age 22.

Case Study: Kim Collins

- Residual analysis identifies athletes who have a peak performance earlier or later than expected.
- Five-time Olympian Kim Collins of St. Kitts and Nevis had his career-best 100 m performance of 9.93 seconds at age 40, 12 years after his predicted peak age of 28.
- Collins' career progression is noteworthy because he seemingly experienced *two* primes, one between the ages of 26 – 29 and another from the ages of 37 – 40.

Future Extensions

- Model predictions could be improved by including additional variables to capture how an athlete qualifies for an Olympic Games.
- More rigorous treatment is necessary for athletes who compete in multiple individual events.
- Further analysis of the rate of progression and regression around an athlete's peak could help identify the window of their prime.

UNIVERSITY OF WATERLOO

UWAGGS

Read the paper here!

References:

1. Toussaint, J. (2012) Bien vieillir: à quelles conditions? Gérontologie et Société , 35(HS1), 41–53.
2. OlyMADMen. Olympedia – Main Page. https://www.olympedia.org. Accessed: 2024-04-21.
3. Mike Davis. A History of the Average Age in Women's Gymnastics. https://themedalcount.com/2020/01/18/a-history-of-the-average-age-in-womens-gymnastics/. Accessed: 2024-03-28
4. Arc Centre of Excellence in Population Ageing Research. Peak Performance Age in Sports. https://cepar.edu.au/sites/default/files/peak-performance-age-sport.pdf. Accessed: 2024-03-28
5. Jaeger, B. C., Welden, S., Lenoir, K., Speiser, J. L., Segar, M. W., Pandey, A. and Pajewski, N. M. (2024) Accelerated and interpretable oblique random survival forests. Journal of Computational and Graphical Statistics, 33(1), 192–207.

Acknowledgements:
This work was made possible via collaboration within the University of Waterloo Analytics Group for Games and Sports (UWAGGS). A special thank you to Rithika Silva for his extensive support acquiring the data.