Statistical Models for Rare Event Analysis in Competitive Swimming: A Statistical Comparison of Extreme Value Theorem, Gompretz Distribution, and Monte Carlo Simulation

Matej Dusa

Mathematics Senior Capstone Research Project

Fall 2024

**Abstract**

This paper investigates the improbability of achieving an exceptional swim time of 46.40s in the men’s 100m freestyle using statistical modeling. The analysis applies Monte Carlo simulation, the Extreme Value Theorem (EVT), and the Gompertz distribution to assess and compare the likelihood of this rare event. The findings reveal that while Monte Carlo and EVT highlight the extremity of the time, the Gompertz distribution suggests a higher probability, emphasizing the influence of different modeling assumptions on rare event prediction. These results underscore the importance of selecting appropriate statistical tools to understand performance limits in competitive sports.

**Introduction**

The pursuit of breaking record times in sports is a defining feature (Ernig, 2024). Athletes and coaches are constantly striving to push physical and technological boundaries to achieve their goals (Ernig, 2024). Analyzing the likelihood of achieving exceptionally fast swim times is not only relevant for performance limitations but it also offers insights into the limits of human capabilities within the sport (Till, 2020). Understanding the current demands of sport involves undertaking a performance-needs analysis (McGuigan, 2014). This needs analysis can include the evaluation of the physical, technical, tactical, and psychological requirements of the sport with a multitude of research available to explore these demands (Till & Baker, 2020).

However, accurately estimating the probability of extreme events occurring shows a statistical and analytical challenge (McGuian, 2020).

Knowing what tools are currently used in the field is the first step in knowing what is practical and usable for coaches in the field, where after coaches and practitioners collaborating with sport scientists can develop and implement tools that are both useable and easily administered. (McGuian,2020, p.1)

This is because using empirical methods on a real-life event produces many outside factors that are not accounted for and can skew the results (McGuian,2020). Moreover, picking the correct statistical analysis is essential for the success of the research (McGuian,2020). Mathematical insight into these events is rarely used but can offer a quantitative method of answering questions about rarity of extreme athletic performances record-breaking events (Gomes, 2019).

This research paper focuses on swimming, specifically the men’s 100m freestyle. The men’s 100m freestyle specialist don’t have a preferred body type and we can see a variety of physical features when observing the best in the world compete (Mazzili, 2019). This is very advantageous, given everyone can use their strengths in their favor (Mazzili, 2019). However, it makes the event one of the most competitive (Mazzili, 2019).

The objective of this paper is to determine the improbability of a particular swim time, specifically the newly established world record of 46.40 seconds in the men’s 100m freestyle, achieved by Pan Zhanle of The Democratic Republic of China, at the 2024 Paris Olympic Games final. To achieve this goal, I used a dataset of the 50 fastest observed performances in the 100m men’s Freestyle event. I analyzed the probability of achieving this world record time using 3 statistical methods: the Monte Carlo simulation, the Extreme Value Theorem (EVT), and the Gompertz distribution. Using these predictive analysis methods, I examine the statistical likelihood of achieving such extreme performance.

**Methodology**

The Monte Carlo simulation, the EVT, and the Gompertz distribution offer unique strengths for modeling extreme values, which are explained below. The Monte Carlo simulation provides a straightforward and computational way of estimating the minimum time based on a parameter input in a normal distribution (Bonate, 2001). 10,000 random iterations based on the previous data give us a good insight into what the expected minimum would be (Bonate, 2001).

There are multiple reasons to use the Monte Carlo simulation. Firstly, it incorporates random variability (Bonate, 2001). Monte Carlo simulation treats model parameters as random variables, which allows it to incorporate the inherent randomness and variability found in real-world systems (Bonate, 2001). This approach provides a more realistic representation of uncertainty compared to deterministic models, which assume fixed values for parameters (Bonate, 2001). Secondly, the Monte Carlo simulation is highly versatile and can be applied across various fields such as finance, engineering, and healthcare (Bonate, 2001). Lastly, the Monte Carlo simulation allows for predictive modeling (Bonate, 2001). This means that by repeatedly simulating the model with different sets of random values, Monte Carlo simulation can predict a range of possible outcomes (Bonate, 2001). This predictive capability is valuable for risk assessment and decision-making, as it provides insights into the likelihood of different scenarios occurring (Bonate, 2001).

On the other hand, the EVT is specifically suited for analyzing the maximum values in a dataset using the Generalized Extreme Value distribution to predict rare events. (Siergist, 2022) Extreme value theory (EVT) has emerged as one of the most important statistical disciplines in applied science. Extreme value techniques are also widely used in other disciplines, such as financial market risk assessment and telecommunications traffic prediction. EVT deals with statistical problems concerning the far tail of the probability distribution and is unique as a statistical tool since it develops models and techniques to describe the unusual event rather than the usual. Using EVT, the theoretical distribution and its population parameter that the maximum value follows are estimated from long-term observation data. (Maruyama, 2019, p. 1)

EVT offers several benefits, including the ability to analyze the most extreme parts of data, which is crucial for predicting rare events (Maruyama, 2019). EVT also provides predictive power, allowing the prediction of records in athletics, such as the 100 m, 200 m, 400 m, 4 × 100 m relay, and long jump (Maruyama, 2019). These benefits make EVT specifically applicable to assessing probabilities of extreme swim times.

The Gompertz distribution, which is commonly applied in survival, reliability or human-limitation analysis, allows for modeling events with a non-constant hazard rate, making it relevant for estimating the occurrence of high-performance athletic results (Siergist, 2022). The hazard rate, as discussed in Benjamin Gompertz’s paper, refers to the intensity of mortality or the force of mortality at a given age (Gompertz, 1825). The Gompertz curve provides a fitting for the data, showing asymptotic convergence to upper bounds, which is particularly useful for predicting threshold limits (Griffith, 2015). The Gompertz curve asymptotically approaches the upper bound without being symmetric about the inflection point, making it more suitable for this application (Griffith, 2015).

Through this comparative analysis, I evaluated the probability of achieving the target time. I also examined the efficacy of each method in the sports domain. By assessing the advantages and limitations of each statistical approach, this research helps the sports domain demonstrate the practical use of each of these methodologies.

**Data Description**

The dataset consists of 50 relevant fastest times in history in the men’s 100m freestyle in a 50m pool. The primary focus is on determining the minimum observed time (46.40s, index 1), relative to the remaining 49 times (indexed 2-50). I put together various sources of competition results, because no single resource provides a comprehensive and updated version on their own. All the data comes from cross-referencing the Kaggle dataset and USA swimming. The data from Kaggle was posted by a user with the alias “The Devastator”. The USA swimming data is collected by an automatic system that automatically pools data from swimming competitions all around the world.

After sorting the data by time (ascending), indexing the times and filtering out the irrelevant fields and records (names of athletes, place of competition, date and time, etc.), the data was set to be used in this analysis. The sample mean was 47.1659, and the standard deviation was 0.167. Figure 1 shows a line graph demonstrating the unlikely behavior of this curve. The fitted linear function is , which demonstrates the linear behavior of the fastest swim times. The point on the very left shows how the newly-set world record doesn’t fit into this equation and is very far off of what is expected. Figure 2, shows that in the whole data set of 50 fastest swim times, there is only one outlier, which is the current world record with a z-score= -3.76.

Figure 1: Line Graph of All Time Top 50 Fastest 100m Freestyle Times

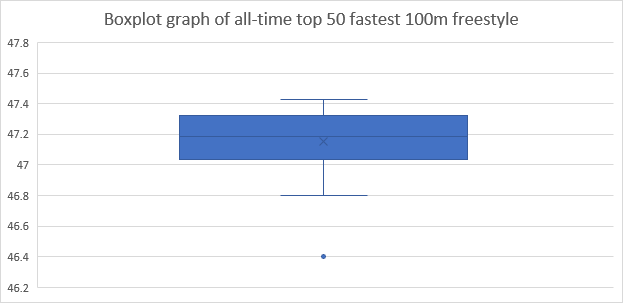


Figure 2: Boxplot of All Time Top 50 Fastest 100m Freestyle Times

**Data Analysis**

**Monte Carlo Simulation Approach**

I used the Monte Carlo simulation approach to estimate the probability of achieving time equal to or faster than 46.40s by repeatedly sampling from a probability distribution obtained from the dataset. The view on how many numbers of iterations is needed varies (Bonate,2001). Upon trying out 1000, 10 000, and 100 000 iterations, I decided to run the simulation on 10,000 iterations to achieve depth which can provide us with a trustworthy result. Using 10,000 iterations provided us with a deep enough result to account for variability in the data and also has sufficient randomness. Failure to do so will result in a simulation that doesn’t cover enough randomness and variability, which directly skews the result (Bonate,2001). For this part of the research, I used 2nd-8th place times, along with the mean and standard deviation, all from the 2024 Paris Olympic Games (which is where the world record was set). I did this to assess the probability of the event occurring at the Olympic games with the same setting. Therefore, all the athletes have the same circumstances and we minimize the outside factors.

To further minimize the effect of outside factors, such as lane distribution, water temperature, geographical location, or random chance, I ran the Monte Carlo simulation again, with the standard deviation from the 2020 Tokyo Olympics, while keeping the mean from 2024 Olympic Games (Bonate,2002). This way, I could minimize the effect of the outside factors and further maximize randomness in the sample.

In table 1, you can see the times used in both of the Monte Carlo simulation approaches. The index is the placement in the final, the fields show which competition yielded each sample result, and the records show the actual times achieved at each competition. Further, the table also shows the mean and standard deviation for each of these fields. I added the results from the 2020 and 2016 Olympic Games to show the rarity of the Olympic winner being so far ahead from the rest of the field. The red cell shows the standard deviation from the 2020 Olympic Games. This cell is highlighted because it will be later used in the calculation process.

|  |  |  |  |
| --- | --- | --- | --- |
| Index | OLY2024 | OLY2020 | OLY2016 |
| 1 | 46.4 | 47.02 | 47.58 |
| 2 | 47.48 | 47.08 | 47.8 |
| 3 | 47.49 | 47.44 | 47.86 |
| 4 | 47.5 | 47.72 | 47.88 |
| 5 | 47.71 | 47.82 | 48.01 |
| 6 | 47.8 | 47.86 | 48.02 |
| 7 | 47.96 | 48.04 | 48.12 |
| 8 | 47.98 | 48.1 | 48.41 |
| Mean 2-8 | 47.70285714 | 47.72285714 | 48.01428571 |
| SD 2-8 | 0.219295409 | 0.356918024 | 0.206386138 |

The results of my Monte Carlo simulations are found below in the Results section.

**Extreme Value Theorem**

I applied the EVT to estimate the probability of observing extreme swim times using the GEV distribution. The EVT is specifically used to determine the theoretical maxima or minima within a dataset, making it relevant for this topic question (Maruyama, 2019). The theoretical maxima and minima refer to the largest or smallest values observed in a dataset or within blocks of data, often analyzed to predict the behavior of extreme events in the tail of a distribution (Maruyama, 2019). EVT assumes that, under certain regularity conditions, the distribution of maxima converges to one of the three types of limiting distributions: the Gumbel, Fréchet, or Weibull distributions, collectively represented by the Generalized Extreme Value distribution (Maruyama, 2019). “The generalized extreme-value distribution combines the three types of extreme value distributions (Gumbel, Fréchet, and Weibull) into a single family, allowing for a unified approach to modeling extreme events” (Hosking, 1985, p. 252).

The cumulative distribution function (CDF) of the GEV distribution is:

Where:

* x is the variable of interest (e.g., swim time),
* μ is the location parameter,
* σ is the scale parameter, and
* ξ is the shape parameter.

In this approach, I followed with these steps:

1. I used the python library genextreme in scipy.stats, and programmed a machine learning model to fit my data into the GEV distribution.
2. I used the swim times from indices 2 to 50 to accurately estimate the shape, location, and scale parameters of the GEV model.
3. I evaluated the cumulative distribution function (CDF) of the GEV at 46.40s, providing the probability of observing a swim time at or below this value.

The results of my EVT calculations are found below in the Results section.

**Gompertz Distribution Approach**

Just like the EVT, the Gompertz distribution also uses a probability density function fitting to assess the likeliness of an event occurring (Griffith, 2015). The distribution’s flexibility in handling non-constant hazard rates makes it a valuable tool for rare event analysis in sports performance (Griffith, 2015). When this data is fitted, this convex parabola shaped probability distribution function assumes an asymptote where the probability becomes zero, which can be interpreted as “the human limit” based on the fitted data (Griffith, 2015).

Where:

* x is the variable of interest,
* η is the scale parameter, and
* b is the growth rate parameter.

In this approach:

1. I fitted the Gompertz distribution to the dataset, excluding the swim time at index 1.
2. I examined the asymptotic limit of the Gompertz distribution to assess the theoretical boundary for swim times.
3. I calculated the probability of observing a swim time of 46.40s using the CDF of the fitted Gompertz distribution.

The results of my calculations for the Gompertz distribution are found below in the Results section.

**Results**

**Monte Carlo Simulation**

In the first attempt, I used the mean and standard deviation from 2024 Olympic Games. The minimum observed time was 46.90s, which is above the set world record. In the second attempt, using the 2024 Olympic Games mean and 2020 Olympic Games standard deviation, the minimum randomly sampled time was 46.27s. A time below the objective value occurred only once in 10,000 iterations, giving us a probability of 0.0001.

**Extreme Value Theorem**

Applying the EVT through the GEV distribution model provided an analytical approach to estimating the probability of achieving a swim time at or below 46.40s. The parameters estimated for the Generalized Extreme Value (using the Python’s library scipy.stats.genectreme) distribution were:

* **Shape parameter (ξ):** 0.6040125060311626
* **Location parameter (μ):** 47.13371710730888
* **Scale parameter (σ):** 0.18696174939365043

These parameters define the behavior of extreme swim times in the dataset. The shape parameter is positive, indicating a Frechet-type distribution with a heavy tail. This shows that extreme swim times (like exceptionally fast performances) have a slight likelihood of occurring, which is typical when analyzing athletic records (Maruyama, 2019). The location parameter centers the distribution at approximately 47.134s, representing the value which extreme swim times cluster around when graphing their probability density. Finally, the scale parameter of 0.187 indicates moderate variability in the extreme values, meaning that the fastest swim times are expected to deviate within a relatively narrow range around the central value. Together, these parameters are used to precisely model extreme swim times and the calculation of probabilities for rare events such as the 46.40s swim time.

Using these parameters, the cumulative distribution function of the GEV distribution at 46.40s showed a probability of 0.000567. This suggests that a swim time of 46.40s represents a rare event within the distribution of observed swim times. The fitted GEV distribution is visualized in Figure 3, with the swim time of 46.40s marked, illustrating its extremity relative to other observed values.

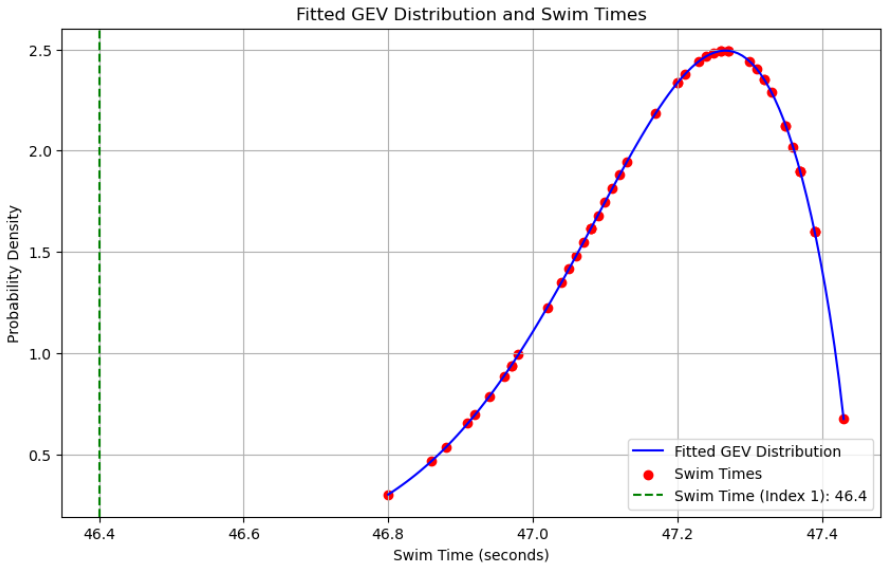


Figure 3: Fitted GEV Distribution Function

**Gompertz distribution**

The probability of achieving a swim time at or below 46.40s, calculated by the CDF of the Gompertz distribution, was found to be 0.498757 (or 49.88%). The asymptote of the Gompertz distribution is found as 2.094373. This asymptote theoretically represents an upper boundary for the probability density function, indicating the point beyond which the probability would stop increasing (Gompertz, 1825). This probability is relatively high (compared to Monte Carlo simulation and EVT), meaning that within the fitted Gompertz distribution, a swim time of 46.40s is not as improbable as might be expected based on the previous extreme event analyses. However, this probability is also close to 50%, suggesting that a time this low is on the borderline of what is typical versus what is extreme. Furthermore, the probability is not extremely low. The position of the 46.40s swim time so close to the left side of the distribution suggests it’s still on the more extreme end of performance but not entirely beyond what’s expected while examining the Gompertz distribution. (Figure 4)

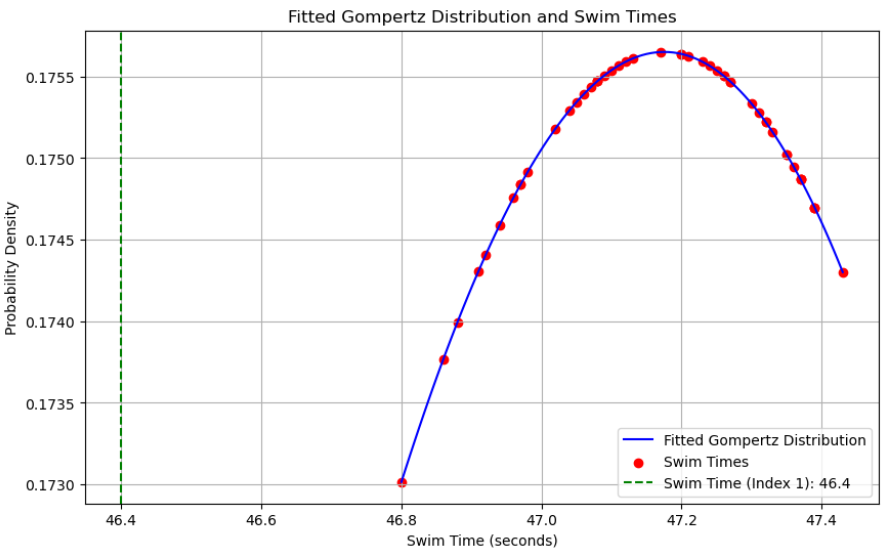


Figure 4: Fitted Gompertz Distribution Function

**Comparison of Methods**

Each method provided a unique probability estimate for achieving a swim time of 46.40s:

* **Monte Carlo Simulation Probability:** 0.0001
* **EVT (GEV Distribution) Probability:** 0.000567
* **Gompertz Distribution Probability:** 0.498757

Overall, the first two methods prove the rarity of achieving a swim time of 46.4 seconds in the men’s 100m freestyle. The Monte Carlo simulation made it seem like it’s almost impossible. The EVT also suggested this event is highly unlikely.

All these methods vary in levels of probability based on their underlying assumptions. The Monte Carlo simulation, which directly reflects the observed distribution, yielded a relatively empirical probability, while EVT and Gompertz approaches provided theoretical analysis.

Theory and statistical models constitute distinct entities built on different information, the behavior of economic agents, and statistical systematic information, respectively. This constitutes a necessary condition for the statistical model to be used as an unprejudiced witness on the basis of whose testimony the empirical adequacy of the theory model can be assessed. The theory influences the choice of an appropriate statistical model in two ways. First, the theory determines the choice of the observed data of interest. Although the choice of the observed data is theory laden, once chosen, the data acquire an objective existence which is theory free. (Spanos, 2019, p. 19)

Empirical analysis is data-driven and directly reflects the observed distribution of swim times (Bonate, 2001). It uses simulating random samples based on the dataset to estimate probabilities, relying only on the patterns and variability present in the given data (Bonate, 2001). In this case, the Monte Carlo simulation produces an estimate of how frequently a swim time like 46.40s could occur based on repeated resampling, ignoring the type of distribution.

Theoretical analysis uses predefined mathematical models, like the GEV or Gompertz distribution, to analyze data (Hoskins, 1985). These models are based on statistical principles and involve assumptions about how extreme events or certain patterns behave (Hoskins, 1985). For example, the EVT uses a specific type of distribution to estimate probabilities, even for events beyond what has been observed in the data (Maruyama, 2022).

The Gompertz distribution, on the other hand, uses a model that accounts for varying likelihoods of rare events (Grammaticos, 2022).

“The Gompertz distribution is a continuous probability distribution often used to model the distribution of human mortality and other biological phenomena. It is characterized by its asymmetry and its ability to model the increasing rate of failure over time (Grammaticos, 2022, p. 10)”

Theoretical approaches are helpful for making predictions outside the range of the data and for understanding the bigger picture, but their accuracy depends on how well the chosen model matches reality (Spanos, 2019).

Between the two theoretical analyses, the EVT approach suggested a strong unlikelihood of the event occurring. Unlike the Gompertz distribution, which showed that while it seems that the probability should be very low, the probability is not extreme at all. This result provides context but differs from results obtained using other methods, which highlight the extremity of this performance more strongly.

**Conclusion**

This study examined the improbability of achieving a 46.40s swim time in the men’s 100m freestyle event by employing three distinct statistical approaches: the Monte Carlo simulation, the EVT, and the Gompertz distribution. Each method provided unique insights into the likelihood of extreme swim performances, revealing the strengths and limitations of different approaches to rare event prediction.

The Monte Carlo simulation yielded a low probability for the 46.40s time, reflecting the empirical rarity of such an outcome based on random sampling from the observed data distribution. Similarly, the EVT approach, through the Generalized Extreme Value (GEV) distribution, modeled the probability of minimum values in the dataset and confirmed that this swim time lies on the extreme end, reinforcing its rarity. These two methods suggest that a time as fast as 46.40s is highly unlikely within the context of typical competitive performances.

However, the Gompertz distribution presented a different perspective, yielding a probability close to 50% for a swim time of 46.40s. This relatively high probability indicates that, within the fitted Gompertz model, a swim time of 46.40s does not lie as far outside the realm of expected performance as might be inferred from the other methods. The Gompertz model’s non-constant hazard rate allows for a broader range of expected values, suggesting that extreme performances, while rare, are more plausible within this framework (Tjørve, 2017).

"An extreme event is an occurrence of a value of a random variable that lies outside the range of typical values, often at the tails of its probability distribution (de Haan, 2006, chapter 3)." Based on my findings listed above, 46.40s was outside the range of typical values. Therefore, I can conclude that there is sufficient evidence to show that 46.40s 100m freestyle is a mathematically extreme event.

The varying results across the methods highlight the importance of selecting the appropriate statistical model when analyzing rare events in sports performance. While Monte Carlo simulation and EVT capture the empirical and theoretical extremity of the swim time, the Gompertz distribution provides an alternative view, suggesting that exceptional performances may occur with moderate probability. This highlights the need for careful model selection and comparison in sports analytics, particularly when interpreting the likelihood of record-breaking performances.

A larger dataset can significantly impact an analysis by increasing the reliability and accuracy of the results. It reduces the impact of outliers, improves the power of statistical tests, and provides a more comprehensive representation of the population, leading to more generalizable conclusions (Frankfort-Nachmias, 2020). However, it also increases the complexity of data handling and may require more advanced computational resources (Frankfort-Nachmias, 2020). The results might differ when using a larger sample. There are sources with more than 50 fastest swim times, but these data tables lack updating and are not up to date.

Future research could further provide additional support for the use of these approaches, potentially adding other statistical models. Future research could further validate and expand upon these approaches by incorporating additional statistical models, such as Bayesian inference or machine learning techniques, to refine predictions and analyze non-linear trends (Baker et al., 2020). Longitudinal studies could be conducted to track performance trends over decades, accounting for factors such as advancements in training methods, technology, and human physiology (Baker et al., 2020). This would provide a more comprehensive view of performance limits in competitive swimming and other sports.

This study contributes to sports analytics by highlighting the differences in rare event prediction, offering a foundation for more informed analysis of extreme athletic performances. Understanding these differences is critical as it allows analysts, coaches, and sports scientists to identify the factors contributing to outlier performances and assess their likelihood with greater accuracy (Till, 2020). Such insights are essential for optimizing training programs, developing predictive models for talent identification, and setting realistic benchmarks for both individual athletes and teams (Till, 2020).

**References**

Bonate, P. L. (2001). A brief introduction to Monte Carlo simulation. *Clinical Pharmacokinetics, 40*(1), 15-22. <https://doi.org/10.2165/00003088-200140010-00002>

de Haan, L., & Ferreira, A. (2006). Extreme Value Theory: An Introduction. Springer.

Emig, T., & Adam, G. (2024). Evolution of world running record performances for men and women: physiological characteristics. *Frontiers in Physiology*. https://doi.org/10.3389/fphys.2024.1372092

Frankfort-Nachmias, C., & Leon-Guerrero, A. (2020). *Social statistics for a diverse society* (9th ed.). SAGE Publications.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, *115*, 513-583. https://doi.org/10.1098/rstl.1825.0026

Grammaticos, B., Meloun, J., & Purdy, J. G. (2022). *Distribution of performances and scoring in athletics*. Mathematics and Sports, 3(1), 1-20.

Griffith, Marie (2015) "Modeling the Development of World Records in Track and Field," McNair Scholars Journal: Vol. 19: Iss. 1, Article 9. Available at: <https://scholarworks.gvsu.edu/mcnair/vol19/iss1/9>

Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, *27*(3), 251-261.

McGuigan, M. R. (2014). *Performance analysis in sport: A review and critique*. Journal of Sports Sciences, 32(10), 901-914.

McGuigan, H., Hassmén, P., Rosic, N., & Stevens, C. J. (2020). Training monitoring methods used in the field by coaches and practitioners: A systematic review. International Journal of Sports Science & Coaching, 15(3), 439-451. https://doi.org/10.1177/1747954120913172

Maruyama, F. (2022) Ultimate Olympics Records in Athletics Using Extreme Value Theory. *Open Journal of Applied Sciences*, **12**, 541-554. doi: [10.4236/ojapps.2022.124038](https://doi.org/10.4236/ojapps.2022.124038)

Mazzilli F. (2019). Body Height and Swimming Performance in 50 and 100 m Freestyle Olympic and World Championship Swimming Events: 1908 - 2016. *Journal of human kinetics*, *66*, 205–213. https://doi.org/10.2478/hukin-2018-0068

Siegrist, K. (2022a) “PROBABILITY, MATHEMATICAL STATISTICS, AND STOCHASTIC PROCESSES”, University of Alabama in Huntsville, <https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/05%3A_Special_Distributions/5.34%3A_The_Gompertz_Distribution>

Siegrist, K. (2022b) “PROBABILITY, MATHEMATICAL STATISTICS, AND STOCHASTIC PROCESSES”, University of Alabama in Huntsville, <https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/05:_Special_Distributions/5.30:_The_Extreme_Value_Distribution>

Spanos, A. (2019). Probability theory and statistical inference: Empirical modeling with observational data. Cambridge University Press.

Takata Gomes, D., & Henriques-Rodrigues, L. (2019). Swimming performance index based on extreme value theory. *International Journal of Sports Science & Coaching, 14*(1), 51-62.

Till, K., & Baker, J. (2020). Challenges and [Possible] Solutions to Optimizing Talent Identification and Development in Sport. *Frontiers in Psychology*, 11, 664. <https://doi.org/10.3389/fpsyg.2020.00664>

Tjørve, K. M. C., & Tjørve, E. (2017). [The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family](https://edgeservices.bing.com/edgesvc/chat?udsframed=1&form=SHORUN&clientscopes=chat,noheader,udsedgeshop,channelstable,ntpquery,devtoolsapi,udsinwin11,udsdlpconsent,udsarefresh,cspgrd,&shellsig=8eb5c7e5246ad34baf9cc8ad7197fdf55b3c2384&setlang=en-US&darkschemeovr=1&udsps=0&udspp=0#sjevt%7CDiscover.Chat.SydneyClickPageCitation%7Cadpclick%7C0%7C58ad0658-90cc-45a1-97cd-2970b3ea0f8e)[1](https://edgeservices.bing.com/edgesvc/chat?udsframed=1&form=SHORUN&clientscopes=chat,noheader,udsedgeshop,channelstable,ntpquery,devtoolsapi,udsinwin11,udsdlpconsent,udsarefresh,cspgrd,&shellsig=8eb5c7e5246ad34baf9cc8ad7197fdf55b3c2384&setlang=en-US&darkschemeovr=1&udsps=0&udspp=0#sjevt%7CDiscover.Chat.SydneyClickPageCitation%7Cadpclick%7C0%7C58ad0658-90cc-45a1-97cd-2970b3ea0f8e). *PLoS ONE, 12*(6), e0178691. [https://doi.org/10.1371/journal.pone.0178691](https://edgeservices.bing.com/edgesvc/chat?udsframed=1&form=SHORUN&clientscopes=chat,noheader,udsedgeshop,channelstable,ntpquery,devtoolsapi,udsinwin11,udsdlpconsent,udsarefresh,cspgrd,&shellsig=8eb5c7e5246ad34baf9cc8ad7197fdf55b3c2384&setlang=en-US&darkschemeovr=1&udsps=0&udspp=0#sjevt%7CDiscover.Chat.SydneyClickPageCitation%7Cadpclick%7C1%7C58ad0658-90cc-45a1-97cd-2970b3ea0f8e)

**Documentation of used Python libraries:**

SciPy Community. (2024). *SciPy documentation: Generalized extreme value distribution*. SciPy. Retrieved December 2, 2024, from <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.genextreme.html>

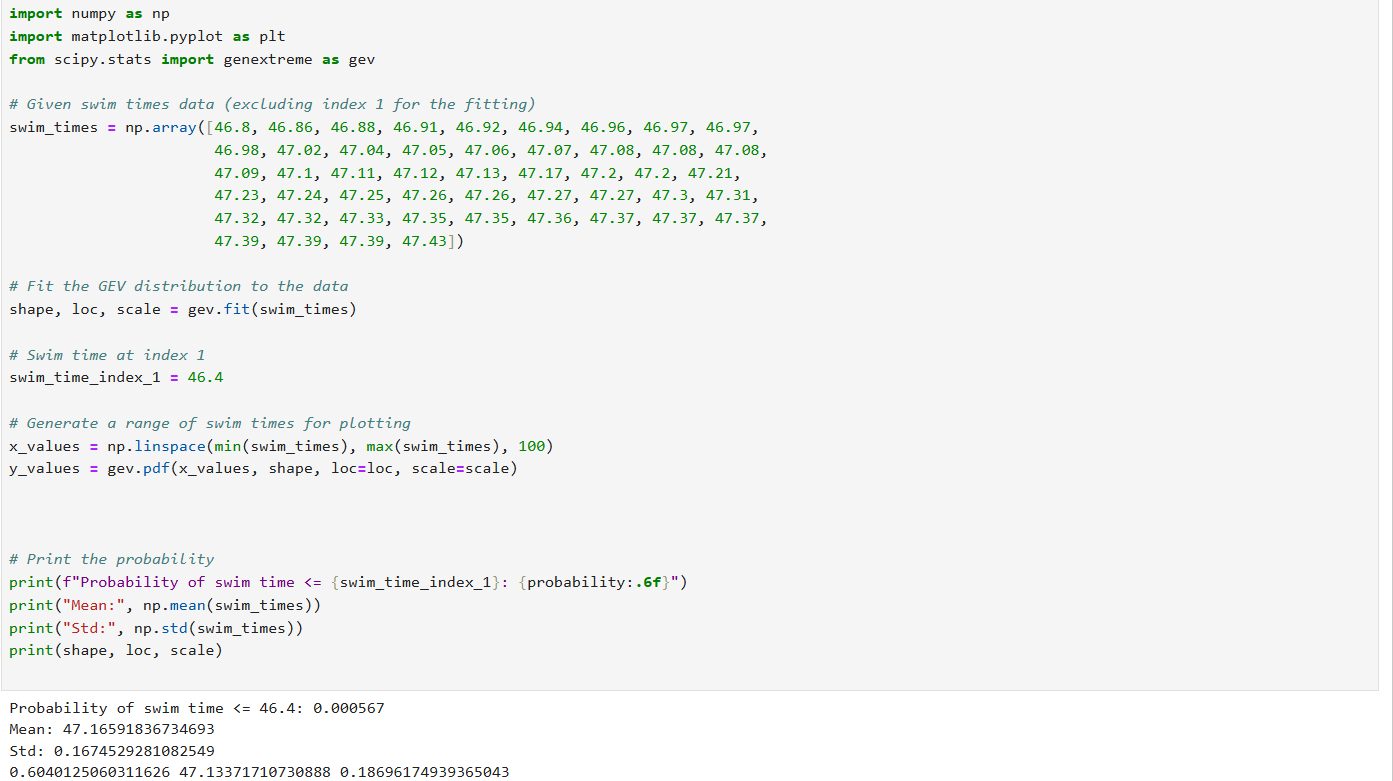
SciPy Community. (2024). *SciPy documentation: Gompertz distribution*. SciPy. Retrieved December 2, 2024, from <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gompertz.html>

**Appendix A**

**Programming Scripts**

**Python**

Below you can find the Python 3 script ran in Jupyter Notebook, which was used to fit and assess the parameters and probability for the EVT.



Here, you can find the Python 3 script ran in Jupyter Notebook, which was used to fit and assess the asymptote and probability for the Gomperz distribution.