Ayan Sultanov Bora Ceylan Jakob Lutz Cyuzuzo Albina Ntivuguruzwa

Is There a Cap on Human Life?

Data, Censoring & Truncation

Is there a limit to human life? While conventional wisdom may seek answers in biology, physics, or philosophy, this inquiry requires the need to transcend time. Using survival analysis, we will rely on extrapolation to determine the data beyond. We will be using the Complete IDL Database file - a dataset of 18959 centenarians and supercentenarians from 13 countries. Even though geographical components will not affect the extrapolation, we found that countries with good nutrition habits and lifestyles have a higher average. For example, Japan has a Median of 111, whereas France has the oldest participant with an age of 122.

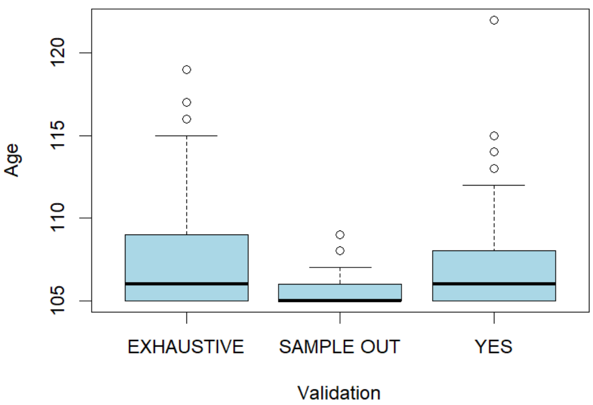
Unfortunately, our data is subject to censoring and truncation which will reduce the validity of our output. To increase the accuracy of our research we relied on truncation.

By applying a likelihood function within the interval [a, b], we are evaluating the probability of observing our data within this specific range of parameter values. This helps us identify the parameter values that best fit the observed data within the given interval

$$\delta,\ \delta = 1,\ when\ T\ is\ Observed\ and\ \delta = 0\ when\ T\ is\ censored$$

$$(t, \delta)\ For\ Observed,\ (t, 1)\ For\ Censored\ (b, 0)$$

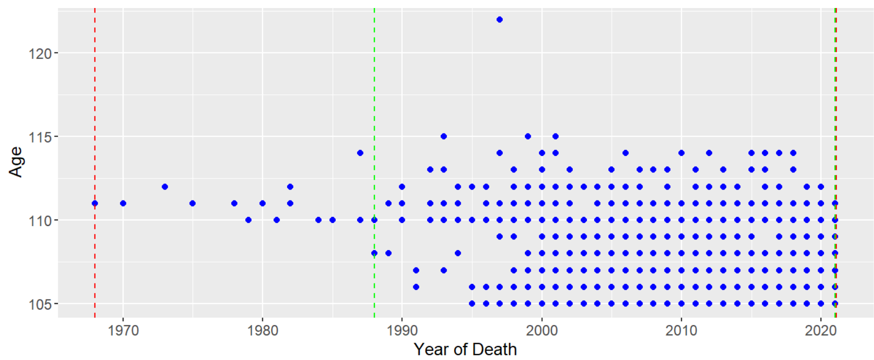
$$\frac{f(t)}{S(a)-S(b)},\ If\ a < t < b \quad \text{and} \quad \frac{h(t)\delta S(t)}{S(a)},\ t > a$$



For this project, we only considered the data with the "Yes" validation for our sample. It has a median age of 106. In order to ensure the accuracy of the extrapolation, we didn't consider the "Exhaustive" data set, which contains about half of the participants, and the Sampled Out data set. As our paper is strictly data-oriented the loss of varied age samples will not affect the result.

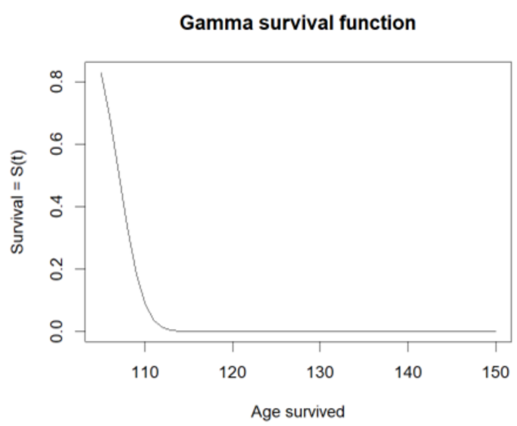
Double Interval Truncation:

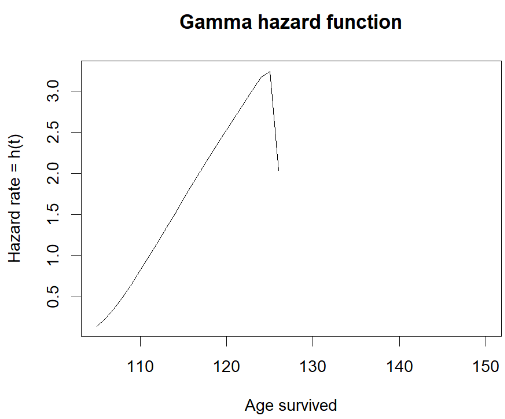
Setting the boundaries for our analysis, we identify the maximum and minimum age of death of semi-supercentenarians in our sample by a green line [1988,2021]. For supercentenarians we defined the boundary with a red line [1968, 2021].



Gamma distribution

In our pursuit to find a cap on human life, we used the Gamma distribution function. Using R studio we obtained the maximum likelihood parameters of Gamma distribution that fit the IDL dataset.

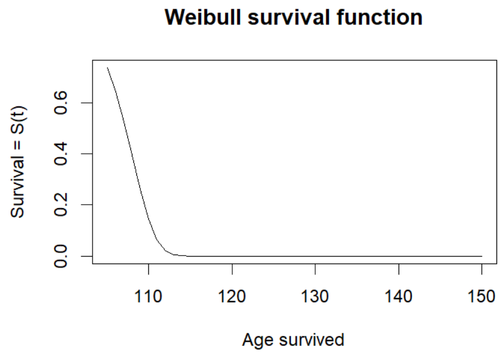


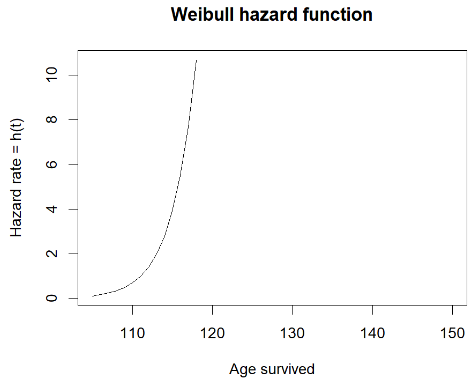


The hazard function is increasing for this model, until it reaches the maximum hazard rate at the age of 124. After this the hazard rate decreases until the age of 126 years, after which R defines other rate values as "infinite".

Weibull distribution

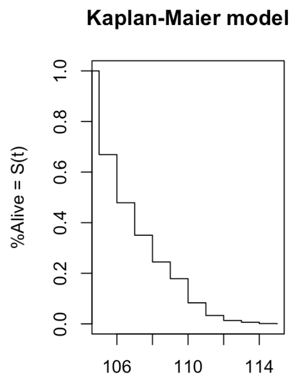
Alternatively, we tried to fit the IDL database to the Weibull distribution. We estimated the parameters of the Weibull distribution using maximum likelihood estimation based on the age at death data.

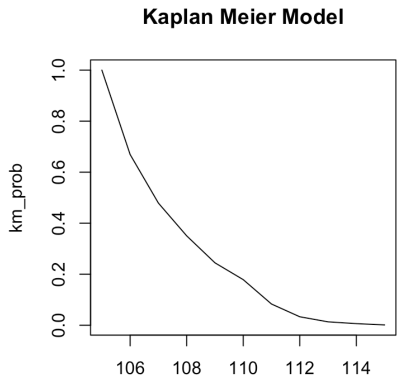




Kaplan Maier Estimates

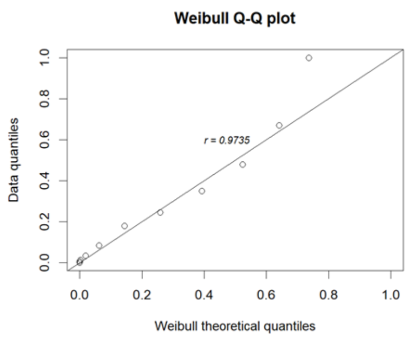
It is not possible to extrapolate beyond data by using a nonparametric model. However, nonparametric models can be used to compare the parametric models we may use. In order to assess the parametric models, we have used the Kaplan Meier model (or product limit estimator), which is a nonparametric model that models the survival function. Both of the graphs below illustrate the age and Kaplan Meier probabilities.

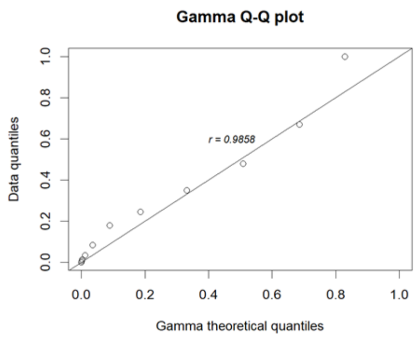




Quantile to quantile comparison

By using the Kaplan Maier model and a quantile-quantile plot we compared the degree of fitness of the Weibull distribution and the Gamma distribution survival functions to the dataset. Gamma distribution fits the dataset slightly better with a correlation coefficient of 0.9858 compared to a correlation efficient of 0.9735 of the Weibull distribution.





Conclusion

In this report, we investigated whether there is a limit on human life or not. We selected Gamma and Weibull distributions as the best fitting distributions for our dataset by estimating the correlation between the model quantiles and the Kaplan Meier model, which is a common nonparametric estimation model used to estimate the survival function.

In our analysis, we witness a steady decline of both survival functions as age increases, signifying the inevitability of mortality. At the age of 127 years, R studio classifies the values of the survival functions as zero. However, due to the nature of these mathematical functions, they never truly reach zero but dwindle infinitesimally close to zero as they asymptotically approach it. This finding underscores the complex interplay between statistical models and the existential reality of mortality.

Furthermore, the hazard rates of both distributions increase until 124 years, indicating an increasing risk of mortality as age progresses. After 126 years however, the hazard rates are defined as “infinite” by R studio, meaning the values are astronomically high. This finding signifies that the values surpass numerical computation capabilities, suggesting an enormous increase in the probability of death. After 126 years the probability of death reaches a value so high, that it is practically impossible to survive after passing this threshold.

While our models provide valuable insights into the limits of human life, they also underscore the harsh reality of mortality's inexorable grip on our existence.