# Inference on the endpoint of human lifespan and its inherent statistical difficulty

Discussion on the paper by Holger Rootzén and Dmitrii Zholud

Stilian A. Stoev and Shrijita Bhattacharya

Department of Statistics, The University of Michigan
1085 S. University Ave, Ann Arbor, USA
{sstoev,shrijita}@umich.edu

April 6, 2018

**Abstract**

We offer an inference methodology for the upper endpoint of a regularly varying distribution with finite endpoint. We apply it to the IDL and GRG data sets of lifespans of supercentenarians. As in the comprehensive analysis of Rootzén and Zholud, our results underscore the effect of the data sampling scheme and censoring on the conclusions. We also quantify the statistical difficulty of distinguishing between the hypotheses of finite and infinite lifespan by providing estimates of the required sample size.

## 1 The Bible, facts, and myths

The question about whether the natural human lifespan has a hard biological limit has been of great interest since the beginning of time. Not surprisingly, therefore, the answer can be found in *The Bible, Genesis* (The Wickedness of Mankind), Ch 6:3 [1]:

*And the Lord said, My Spirit shall not always strive with man, for that he also is flesh: yet his days shall be a hundred and twenty years.*

Our colleague Ya'acov Ritov who brought this quote to our attention, explained that in the original Hebrew version, the above stipulation is gender-free and it should have applied to Jeanne Calment who lived 122 years and 164 days – the longest undisputedly documented human lifetime. The Bible itself as well as many other sources mention human lifespans of well over 120 years, but it is well-known that such data tend to be either mythical or highly unreliable.

Extreme Value Theory is the most natural statistical framework that can provide a principled answer to the question about whether or not natural human lifespan is finite. Simply put, it amounts to fitting either a Generalized Pareto Distribution model to the Peaks over Extreme Threshold, or a Generalized Extreme Value model to block-maxima of the data. Roughly speaking, the resulting index estimate (perhaps with a confidence interval) would then suggest whether the underlying distribution is of finite support or not. This recipe, however, assumes that the data are pristine, independent samples from a *fixed* population. The key challenge in applying this methodology to human lifespan data is the data itself.

Rootzén and Zholud [13] take great care in using perhaps the best available data set of super-centenarian lifespans from the International Database on Longevity [10]. They clearly demonstrate that poor understanding of the sampling scheme or selection biases, which plague many available longevity data sets, can easily lead to conflicting conclusions. Then, using a suitably defined likelihood, which accounts for the sampling scheme as well as the extremal behavior of the data, they show that the force of mortality (survival rate) of super-centenarians is likely constant with age. Therefore, there is no evidence that the current record value of the human lifespan is close to its natural biological limit if one exists at all.

We congratulate Holger Rootzén and Dmitrii Zholud on a most thorough and illuminating treatment of this important question. Their work can serve as a gold-standard of further applications of extreme value theory to data – they pay equal attention to the data, its curation, and sampling scheme, as well as to the appropriate statistical methodology.

The question about the limit of natural human lifespan is complex and confounded in many factors that bring life to an end. To best estimate the upper endpoint of the human lifespan distribution (finite or infinite), one should focus on individuals who die of natural causes and not due to a commonly preventable disease. This is essentially achieved by focusing on super-centenarians because, having passed the threshold of 110, these individuals have managed to avoid most ailments and accidents that bring demise to humans. The cause of death, however, is not indicated even in the carefully curated IDL data-base, which may in principle contribute to some bias even in the careful treatment of [13].

In contrast to [8], a large body of statistically sound research on gerontology exists where many important penultimate extreme age phenomena are discussed. Notably, the excellent study of Andersen et al [2] provides an in-depth survival analysis on the effect of ailments on age and longevity. It shows (using statistically sound Weibull regression techniques) that super-centenarians become increasingly immune to disease and ailments with age:

*"We observed a progressive delay in the age of onset of physical and cognitive function impairment, age-related diseases, and overall morbidity with increasing age. As the limit of human life span was effectively approached with supercentenarians, compression of morbidity was generally observed."*

Fries [9, 15] pioneered the "aging healthy" philosophy. He formulated the "compression of morbidity" hypothesis, which postulates that society will become increasingly more healthy in old age and the morbidity will be squeezed into shorter and shorter intervals prior to death. Other recent studies argue for a broader definition and an expansion-of-morbidity phenomenon [3]. While these studies address the important question about whether humans will "age healthily" or "age sick" the ultimate question about the limit on natural human lifespan requires solid statistical methodology, which is rooted in Extreme Value Theory.

In the rest of this note, we provide some inference methods for the endpoint of a distribution with a finite upper bound. We then apply these methods to the IDL and GRG data sets and largely confirm the detailed analysis of Rootzén and Zholud [13]. We end with a cautionary tale on the minimum sample size required to be able to draw a confident conclusion about the finiteness of human lifespan. The MATLAB code used to produce all tables and figures is given in [14].

## 2 Confidence sets for the upper endpoint of a distribution

Next we present a methodology for constructing confidence sets for the endpoint of a distribution. Our approach is based on the method of test-inversion and differs from the profile likelihood method illustrated in Figure 1 of Davison [7]. Let $X_i,\ i = 1, \ldots, n$ be iid (independent identically distributed) realizations from a distribution $F$ with an upper endpoint $x_F < \infty$. We shall suppose that $F$ belongs to the max domain of attraction of a reversed Weibull law with index $-\xi < 0$. That is,

$$\mathbb{P}(x_F - X_i < 1/z) = L(z)z^{-1/\xi}, \quad z > 0,$$

where $L$ is a slowly varying function at $\infty$ with exponent $-1/\xi$ (see, e.g., Proposition 1.13 in [12]). Equivalently, the random variables

$$Y_i = Y_i(x_F) := \frac{1}{(x_F - X_i)},\ i = 1, \ldots, n \tag{1}$$

have heavy, regularly varying right tails, i.e.,

$$\mathbb{P}(Y_i > z) = L(z)z^{-1/\xi},\ z > 0. \tag{2}$$

**The test-inversion based method.** Let $X_{(n,n)} \le X_{(n-1,n)} \le \cdots \le X_{(1,n)}$ be the order statistics of the data. Our goal is to obtain a confidence region for $x_F$. We do so by the method of test inversion. Specifically, consider the family of hypotheses testing problems:

$$T_\theta : \begin{cases} \mathcal{H}_0 : x_F = \theta \\ \mathcal{H}_a : x_F \neq \theta \end{cases}$$

for $\theta \in [\theta_0, \theta_1]$. Suppose that we have a calibrated test statistic that produces a p-value $p(\theta) = p(\theta; \mathbf{X}_n)$ having the Uniform$(0, 1)$ distribution under the null hypothesis. Then, the *method of inversion* entails that, given an $\alpha \in (0, 1)$,

$$C_\alpha := \{\theta : p(\theta) > \alpha\} \tag{3}$$

is a $100 \times (1 - \alpha)\%$ confidence region for $x_F$.

We use this test-inversion strategy to obtain a confidence interval. Observe that under the null hypothesis $\mathbf{Y}_n(\theta) := \{Y_i(\theta),\ i = 1, \ldots, n\}$ are regularly varying. Based on $\mathbf{Y}_n(\theta)$, we next define a class of statistics which are asymptotically Uniform$(0, 1)$ and independent (see (10) in Proposition 2). Applying a test of uniformity to these statistics, we obtain p-values $p(\theta)$ over the range, $\theta > X_{(1,n)}$, which by (3) yields a confidence region for $x_F$. In practice, these confidence regions depend on the choice of the test (Section 3 illustrates the dependence on order statistic parameters).

**Testing for regular variation.** We start with the ideal Pareto setting. Namely, suppose that $\mathbf{Z}_n := \{Z_i,\ i = 1, \ldots, n\}$ are iid standard $1/\xi-$Pareto distributed, i.e., $\mathbb{P}(Z_i > x) = x^{-1/\xi},\ x \ge 1,\ \xi > 0$. Recall the celebrated Hill statistics:

$$\widehat{\xi}_j(\mathbf{Z}_n) := \frac{1}{j}\sum_{i=1}^{j} \log\left(\frac{Z_{(i,n)}}{Z_{(j+1,n)}}\right) \equiv \frac{1}{j}\sum_{i=1}^{j} i \log\left(\frac{Z_{(i,n)}}{Z_{(i+1,n)}}\right),\quad j = 1, 2, \ldots, n-1, \tag{4}$$

where $Z_{(1,n)} \ge Z_{(2,n)} \ge \cdots \ge Z_{(n,n)}$ are the order statistics of the $Z_i$'s.

Consider now the *trimmed Hill statistics* introduced by Bhattacharya, et al [4]:

$$\widehat{\xi}_{k_0,j}(\mathbf{Z}_n) := \frac{1}{j}\sum_{i=k_0+1}^{k_0+j} i\log\left(\frac{Z_{(i,n)}}{Z_{(i+1,n)}}\right), \quad \text{where } 0 \le k_0 + j \le n-1 \tag{5}$$

Observe that the usual Hill statistics in (4) are recovered for $k_0 = 0$. For $1 \le k_0 \le n-2$, however, the $\widehat{\xi}_{k_0,j}$'s do not involve the top-$k_0$ order statistics of the sample. As we shall see, this will allow us to construct calibrated confidence intervals for the endpoint of a distribution, even if the data are censored, e.g., the largest $k_0$ observations are missing.

**Proposition 1** *Let $E_k = 1, 2, \ldots$ be iid exponential random variables with unit mean and let $\Gamma_k = E_1 + \cdots + E_k$ be the arrival times of a unit-rate Poisson process on $(0, \infty)$.*

(i) *We have*

$$\{\widehat{\xi}_{k_0,j}(\mathbf{Z}_n),\ 0 \le k_0,\ 1 \le j < n - k_0\} \stackrel{d}{=} \xi\left\{\frac{\Gamma_{k_0+j} - \Gamma_{k_0}}{j},\ 0 \le k_0,\ 1 \le j < n - k_0\right\}. \tag{6}$$

(ii) *Consequently, for a fixed $0 \le k_0 < n-2$, the statistics*

$$U_{k_0,j}(\mathbf{Z}_n) := \left(\frac{j\widehat{\xi}_{k_0,j}(\mathbf{Z}_n)}{(j+1)\widehat{\xi}_{k_0,j+1}(\mathbf{Z}_n)}\right)^j, \tag{7}$$

*are Uniform$(0,1)$ and independent in $j = 1, \ldots, n - k_0 - 2$.*

The proof is given the Appendix. We now turn to the general regularly varying case.

Let $Q(u)$ be the tail quantile function of the $Y_i$'s in (1), i.e.,

$$Q(u) := \overline{F}_Y^{\leftarrow}(u) := \inf\{z > 0 : \overline{F}_Y(z) < u\},\ u \in (0,1)$$

By (2) and Proposition 1.5.15 in [6], we have that $Q(u) = \ell(u)u^{-\xi}$, as $u \downarrow 0$, where $\ell$ is a slowly varying function at 0 (related to $L$). Without loss of generality, we suppose that

$$Y_i = Q(V_i) = \ell(V_i)V_i^{-\xi}, \quad i = 1, \ldots, n,$$

where the $V_i$'s are independent and Uniform$(0,1)$. By the monotonicity of $Q$, and a version of the Rényi representation, for every fixed integer $k$, we have

$$\begin{aligned}(Y_{(1,n)}, \ldots, Y_{(k,n)}) &\stackrel{d}{=} \left(Q\left(\frac{\Gamma_1}{\Gamma_{n+1}}\right), \ldots, Q\left(\frac{\Gamma_k}{\Gamma_{n+1}}\right)\right) \\ &\stackrel{a.s.}{\sim} Q\left(\frac{\Gamma_{k+1}}{\Gamma_{n+1}}\right)\left(\left(\frac{\Gamma_1}{\Gamma_{k+1}}\right)^{-\xi}, \ldots, \left(\frac{\Gamma_k}{\Gamma_{k+1}}\right)^{-\xi}\right), \quad \text{as } n \to \infty.\end{aligned} \tag{8}$$

The formal proof is given in the Appendix. Therefore, as $n \to \infty$,

$$\left(\frac{Y_{(i,n)}}{Y_{(i+1,n)}},\ i = 1, 2, \ldots, k\right) \xrightarrow{d} \left(\frac{Z_{(i,k+1)}}{Z_{(i+1,k+1)}},\ i = 1, 2, \ldots, k\right), \tag{9}$$

where $Z_{(1,k+1)} \ge Z_{(2,k+1)} \ge \ldots$ are the order statistics of a standard $1/\xi$-Pareto random sample. This suggests that if one replaces the $Z_{(i,n)}$'s in (5) by the $Y_{(i,n)}$'s, the corresponding $U_{k_0,j}$ statistics in (7) would be asymptotically independent and Uniform$(0,1)$. More precisely:

**Proposition 2** *Fix the integers* $0 \leq k_0 < k$ *and let* $\mathbf{Y}_n = \{Y_i,\ i = 1, \ldots, n\}$, *where the* $Y_i$*'s are as in* (2). *Then, as* $n \to \infty$, *we have*

$$\left\{ U_{k_0,j}(\mathbf{Y}_n) := \left( \frac{j\widehat{\xi}_{k_0,j}(\mathbf{Y}_n)}{(j+1)\widehat{\xi}_{k_0,j+1}(\mathbf{Y}_n)} \right)^j,\ j = 1, \ldots, k - k_0 - 2 \right\} \xrightarrow{d} \{U_j,\ j = 1, \ldots, k - k_0 - 2\}, \tag{10}$$

*where the* $U_j$*'s are independent Uniform*$(0, 1)$.

The result follows from Proposition 1 and the Continuous Mapping Theorem when Relation (9) is applied to (5) with $n = k$.

## 3 Simulations and Data Analysis

**Understanding the testing based method.** We simulated $n = 631$ (size of the validation A (Type A) data set from [10], see also [13]) independent realizations from two models for the lifetimes of super-centenarians. The first has a finite endpoint $x_F = 122.4493$ (Jeanne Calment's age):

$$X = 110 + (x_F - 110) \times B, \quad \text{where } \mathbb{P}(B > x) = (1 - x)^{1/\xi},\ x \in (0, 1). \tag{11}$$

The second model is a shifted exponential distribution, that is

$$X = 110 + E, \quad \text{where } \mathbb{P}(E > x) = e^{-\lambda x},\ x \in (0, \infty).$$

The parameters $\xi = 0.1206$ and $\lambda = 1/1.34$ were chosen so that both models have the same mean excess lifetime over the threshold 110. Specifically, we match the estimate 1.34 of Rootzén and Zholud [13]. Figure 1 highlights the fundamental differences between the two models. The p-value maps based on the testing method (left and middle panels) were obtained as follows.

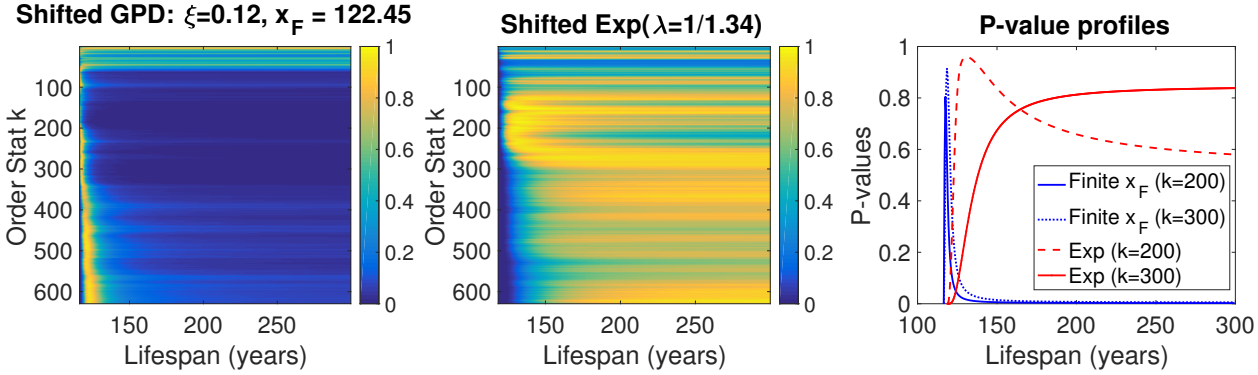


Figure 1: *Left & Middle panels:* Heatmap plots of the p-value matrices $P_{((n-2)\times T)}$ of p-values for the shifted GPD model in (11) with $x_F = 122.4493$ and $\xi = 1/8.291 = 0.1206$ and the shifted exponential model (GPD with $\xi = 0$) and $\lambda = 1/1.34$. *Right panel:* P-value profiles for the two models for several values of $k$.

We compute the $U_{0,j}(\theta),\ j = 1, \ldots, k$ statistics via (10), over a grid of values $\theta(i) = x_F + ih,\ i = 1, \ldots, T$. Then, for each value $k$ of top-order statistics, we test the hypothesis that $U_{0,j}(\theta),\ j = 1, \ldots, k$ are Uniform$(0, 1)$ and obtain a p-value $P(k, i)$. We used the Anderson-Darling test but

other tailored tests may lead to better results [5]. Heatmaps of the resulting p-value matrices $P = (P(k,i))_{(n-2)\times T}$ for the finite support and exponential models are shown in the left and middle panels of Figure 1, respectively.

Recall that for a fixed $k$, the set of $\theta$'s corresponding to p-values greater than a level $\alpha$, yield a $100 \times (1-\alpha)\%$ confidence region for the maximum lifespan. These ranges can be read-off from the p-value profile plots in the right panel of Figure 1. Observe the fundamental difference between the p-value heatmaps and profiles between the two models. The p-value profiles for the finite-support model yield confidence sets with a finite upper bound, whereas those of the exponential model are monotone increasing indicating infinite support. This discussion provides some guidance on how to interpret p-value maps and profile plots for real data.

**Remarks:**

1. As for the Hill estimator of the heavy-tail index, our method requires choosing $k$. Relatively small values of $k$ yield high uncertainty (wide confidence bounds), while large values of $k$ may lead to bias since the asymptotic power-law (Pareto) scaling may not have kicked–in. The choice of $k$ is an important problem beyond the scope of this note.

2. The p-value map provides only a qualitative picture of the nature of the distribution, which should not be interpreted as the outcome of a multiple hypotheses testing problem. The test statistics in Proposition 2 are used merely as a means of obtaining confidence regions through the method of inversion (3). Specifically, once a value of $k$ has been chosen, thresholding the p-value profile yields a confidence set.

3. The behavior of the p-value profiles is similar to the log-likelihood profiles of the Generalized Pareto model, used to obtain asymptotic confidence intervals for $x_F$ in Fig. 1 of [7].

**Inference on the maximum attainable human lifespan.** Here, playing God's advocate, we assume that the lifetime distribution has a finite upper bound $x_F$ and construct confidence regions for it. We study three different data sets: (i) Type A ages from IDL; (ii) Types A and B ages from IDL, combined; (iii) The GRG data. The data was obtained from the International Database on Longevity web-site [10]. In each case, we scan the region $[122.49, 300]$ for the upper bound $x_F$ on human lifespan.

Table 1 summarizes our findings. For type A and A&B data, we find no evidence that the natural human lifespan is finite, based on the confidence regions obtained. It may appear that this conclusion is fragile. Inspired by the analysis in Figure 1 of Davison [7], we censored the type A dataset by removing the top-2 observations. If the data are treated as if no censoring is performed ($k_0 = 0$), then many of the confidence intervals obtained indicate a finite bound to human lifespan. However, if the correct value of $k_0 = 2$ is chosen, then despite the lower bounds being shifted down (due to lack of important extreme observations), the upper confidence bounds (as for the uncensored type A data) indicate no finite limit.

The GRG data leads us to the opposite conclusion, i.e., if one is content with picking $k = 100$, then with confidence at least 95% the natural human lifespan is limited to 137.47. As indicated in the careful analysis of Rootzén and Zholud, the GRG data are likely age-biased and this conclusion should not be trusted. This brief discussion corroborates the findings of [13] and [7] and underscores the crucial importance of correctly accounting for possible censoring and sampling bias in the analysis of extreme ages. Figure 2 clarifies this picture further. Note that to resolve a technical

Table 1: Confidence intervals for the maximum human lifespan obtained form the method of inversion at order statistics $k = 100$ and 200 for the Type A; A and B; GRG datasets; '–' indicates that the maximum search region of 300 was reached. See also left and middle panels in Figure 2 for other values of $k$. The last two sets of rows correspond to the censored type A dataset with top 2 observations removed (see Figure 1 of Davison [7], for analogous analysis). Observe that if censoring is not accounted for, the conclusions are reversed (see also Figure 3, below).

| Dataset | Ord Stat / Conf Levels | $k = 100$ 90% | $k = 100$ 95% | $k = 100$ 99% | $k = 200$ 90% | $k = 200$ 95% | $k = 200$ 99% |
|---|---|---|---|---|---|---|---|
| A | Lower | 128.95 | 126.65 | 124.45 | 129.95 | 127.55 | 124.95 |
| | Upper | – | – | – | – | – | – |
| A and B | Lower | 128.65 | 126.35 | 124.25 | 129.65 | 127.35 | 124.95 |
| | Upper | – | – | – | – | – | – |
| GRG | Lower | 117.37 | 117.27 | 117.17 | 117.77 | 117.57 | 117.37 |
| | Upper | 126.87 | 137.47 | – | 128.57 | 136.37 | – |
| Censored A ($k_0 = 0$) | Lower | 116.98 | 116.88 | 116.77 | 117.08 | 116.98 | 116.78 |
| | Upper | 122.78 | 128.18 | – | 153.18 | – | – |
| Censored A ($k_0 = 2$) | Lower | 119.48 | 118.48 | 117.68 | 119.78 | 119.08 | 118.18 |
| | Upper | – | – | – | – | – | – |

problem of ties among the order statistics, we dithered the data by adding to each of the lifespan observations independent random hour of death values, which are uniformly distributed on the interval $[-0.5/365.25, 0.5/356.25]$ (see Appendix B).

**Confidence regions for the maximum human lifespan.** Figure 2 shows the results from the proposed testing-based method applied to the type A and GRG data sets. The heatmap and p-value profiles for the validation A data do not support the case for a finite endpoint $x_F$. In fact, they conform rather closely to the exponential excess lifetime model in Figure 1. This finding is in agreement with the sophisticated censored likelihood approach of Rootzén and Zholud [13]. In contrast, the p-value heatmap and profiles for the GRG data are drastically different and conform with a bounded support model (see also Figure 1).

**The effect of censoring.** Finally, motivated by the analysis in Figure 1 of [7], we demonstrate the importance of correctly accounting for missing data. Specifically, in Figure 3, we removed the largest 2 observations from the type A data and re-did the analysis under: (i) the incorrect assumption that there are no missing data ($k_0 = 0$) – left panel; (ii) the correct assumption that $k_0 = 2$ – middle panel. The right panel shows the contrast in the p-value profiles of cases (i) and (ii). It demonstrates, how easy one can arrive at conflicting conclusions if proper care of the potential censoring of the data is not exercised.

# 4 The statistical difficulty of the problem

Here we provide an assessment of the inherent difficulty of the problem in a simplified but realistic hypothesis testing context. Namely, let $f_X(x)$, $x \geq 0$ be the probability density function of the excess lifespan of a typical super-centenarian (over age 110). Consider the hypothesis testing

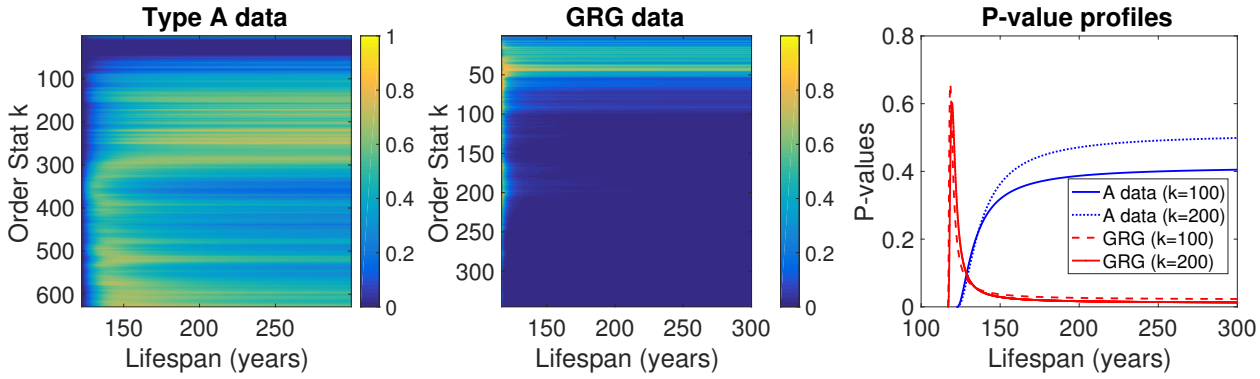


Figure 2: The layout is as in Figure 1. The statistics are based on the dithered (see the appendix for more details) type A and GRG data sets.

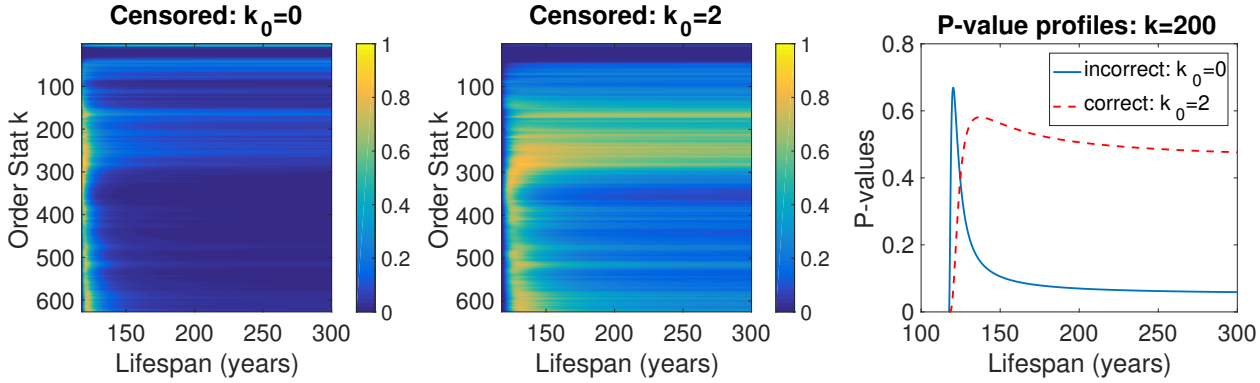


Figure 3: Type A data with the largest 2 observations deleted. The p-value heatmap in the left panel is based on the statistics $U_{0,j}$ under the incorrect assumption that there are no missing data ($k_0 = 0$); The middle panel involves the corrected statistics $U_{2,j}$ ($k_0 = 2$). The right panel shows the drastic difference between the p-value profiles.

problem:

$$\begin{cases} \mathcal{H}_\theta & : f_X(x) = p_\theta(x) := \frac{1}{\xi\theta}\left(1 - \frac{x}{\theta}\right)^{1/\xi - 1}, \ x \in (0, \theta) \\ \mathcal{H}_\infty & : f_X(x) = p_\infty(x) := \lambda e^{-\lambda x}, \ x > 0. \end{cases}$$

Based on the careful analysis of Rootzén and Zholud [13], we shall assume that the average excess lifetime of a super-centenarian is $m_e := 1.34$ and thus set $\lambda := 1/1.34$. Also, given that Jeanne Calment's excess lifespan is $\theta_0 := 122.45 - 110 = 12.45$, we shall only consider $\theta \geq \theta_0$. For a given value of $\theta$, we shall identify $\xi = \xi(\theta)$, so that

$$\mathbb{E}_\theta[X] = \frac{\theta\xi}{(\xi + 1)} = m_e = \frac{1}{\lambda} = \mathbb{E}_\infty[X].$$

That is, we *match* the means of the excess lifetimes under the null and alternative hypotheses.

Under $\mathcal{H}_\theta$, the maximum excess lifespan $\theta$ is finite. Observe that as $\theta$ grows, distinguishing between $\mathcal{H}_\theta$ and $\mathcal{H}_\infty$, becomes increasingly more difficult since $p_\theta \to p_\infty$. Our goal is to quantify the sample size necessary for a satisfactory answer to the problem.

Suppose that we have $n$ independent realizations from the excess lifetime distribution. Every (possibly randomized) testing procedure can be identified with a function $\phi : [0,\infty)^n \to [0,1]$, such that, given data values $\mathbf{x} = (x_1, \ldots, x_n) \in [0,\infty)^n$, the hypothesis $\mathcal{H}_\theta$ is rejected with probability $\phi(\mathbf{x})$ and otherwise $\mathcal{H}_\infty$ is rejected with probability $(1 - \phi(\mathbf{x}))$. The overall testing error can then be measured as the sum of the type I and type II errors:

$$\mathcal{E}_n(\phi) := \int_{[0,\infty)^n} \phi(\mathbf{x}) p_\theta(\mathbf{x}) d\mathbf{x} + \int_{[0,\infty)^n} (1 - \phi(\mathbf{x})) p_\infty(\mathbf{x}) d\mathbf{x},$$

where $p_\theta(\mathbf{x}) = \prod_{i=1}^n p_\theta(x_i)$.

For any fixed $\theta$ and $n$, the optimal testing error is then obtained by minimizing over all possible testing procedures $\phi$:

$$\mathcal{E}_n^{\text{opt}}(p_\theta, p_\infty) = \inf_{0 \le \phi \le 1} \mathcal{E}_n(\phi) = \int_{[0,\infty)^n} (p_\theta \wedge p_\infty)(\mathbf{x}) d\mathbf{x}.$$

The latter is the so-called *testing affinity* between the distributions $p_\theta$ and $p_\infty$, which is in fact achieved by the Neyman-Pearson likelihood ratio-based test.

Given a value of $\theta$, by inverting the testing affinity, one can in principle compute the minimum sample size $n(\theta)$, required for the optimal (Neyman-Pearson) test to be able to achieve a specified testing error $\delta = \mathcal{E}_n^{\text{opt}}(p_\theta, p_\infty) \in (0,1)$. We provide instead lower and upper bounds on the minimum sample size, based on the easier to compute *Hellinger affinity*:

$$\rho_n(p_\theta, p_\infty) := \int_{[0,\infty)^n} \sqrt{p_\theta(\mathbf{x}) p_\infty(\mathbf{x})} d\mathbf{x} = \rho^n(\theta),$$

where

$$\rho(\theta) = \int_0^\infty \sqrt{p_\theta(x) p_\infty(x)} dx = \sqrt{\frac{\lambda}{\theta\xi}} \int_0^\theta (1 - x/\theta)^{(1-\xi)/2\xi} e^{-\lambda x/2} dx. \tag{12}$$

By the Le Cam testing inequalities, we obtain (see, e.g., page 44 in [11]),

$$1 - \sqrt{1 - \rho^{2n}(\theta)} \le \mathcal{E}_n^{\text{opt}}(p_\theta, p_\infty) \le \rho^n(\theta). \tag{13}$$

This yields the bounds

$$\frac{\log(1 - (1-\delta)^2)}{2\log(\rho(\theta))} \le n(\theta) \le \frac{\log(\delta)}{\log(\rho(\theta))},$$

where the integral (12) can be computed numerically.

Table 2 illustrates the difficulty of the testing problem. Specifically, for the current sample size of $n = 631$ observations from the type A dataset, the conservative bound on the combined type I and type II error is about 0.28. More precisely, for this sample size, the lower and upper bounds on the optimal (testing affinity) error based on (13) are $[0.0388, 0.2757]$. In reality, the testing error is perhaps closer to the lower bound. This simple analysis shows that the available data may not be sufficient to give a very confident answer to the question of whether or not the human lifespan is finite. Furthermore, these results provide some guidance for future studies of the problem based on larger samples and record values of $\theta$.

Table 2: Lower and upper bounds for the minimum sample size required to achieve testing error $\delta$, as a function of $\theta$ based on Hellinger affinity.

| | $\theta = 122.45$ | | $\theta = 130$ | | $\theta = 140$ | | $\theta = 150$ | |
|---|---|---|---|---|---|---|---|---|
| $\delta = 0.28$ | 179 | 624 | 528 | 1,838 | 1,276 | 4,445 | 2,349 | 8,186 |
| $\delta = 0.10$ | 407 | 1,128 | 1,199 | 3,325 | 2,899 | 8,039 | 5,340 | 14,807 |
| $\delta = 0.05$ | 570 | 1,467 | 1,681 | 4,326 | 4,064 | 10,460 | 7,485 | 19,264 |
| $\delta = 0.01$ | 959 | 2,256 | 2,828 | 6,650 | 6,838 | 16,079 | 12,594 | 29,614 |

# 5 Discussion

Reducing the question about the endpoint of a distribution to a single number does not tell the full story behind the data. Here, we proposed a methodology for constructing confidence regions for the upper endpoint of a distribution in the domain of attraction of a reversed Weibull law. The methodology is likely asymptotically sub-optimal to parametric techniques such as generalized likelihood ratio tests or profile likelihood-based confidence intervals. Nevertheless, our methods are likely more robust to departures from the asymptotic parametric model, they can handle missing extreme observations, and provide also a graphical device that aide the understanding of the extremal behavior of the data. Our analysis of lifespans of super-centenarians confirms the conclusions of Rootzén and Zholud and underscores the importance of correctly accounting for missing extremes and sampling bias. We ended with a cautionary tale on the sample size required to distinguish between the finite and infinite lifespan hypotheses.

# 6 Acknowledgements

We thank Thomas Mikosch, Holger Rootzén, and Dmitrii Zholud for the opportunity to engage in a stimulating research discussion. We are also grateful to Ya'acov Ritov for enlightening discussions on Statistics, Philosophy, and The Bible. SS and SB were partially funded by the NSF grant DMS-1462368.

# A Proofs

**Proof of Proposition 1.** Observe that $Z_i = V_i^{-\xi}$, where $V_i,\ i = 1, \ldots, n$ are iid Uniform$(0, 1)$. A version of the Rényi representation entails that

$$(V_{(n,n)}, V_{(n-1,n)}, \ldots, V_{(1,n)}) \stackrel{d}{=} \left( \frac{\Gamma_1}{\Gamma_{n+1}}, \frac{\Gamma_2}{\Gamma_{n+1}}, \ldots, \frac{\Gamma_n}{\Gamma_{n+1}} \right).$$

Using this representation and the fact that $Z_{(i,n)} = V_{(n-i+1,n)}^{-\xi}$, from (4), we obtain

$$\{\widehat{\xi}_{k_0,k}(n),\ 0 \le k_0 < k < n\} \stackrel{d}{=} \left\{ -\frac{\xi}{k - k_0} \sum_{i=k_0+1}^{k} i \log \left( \frac{\Gamma_i}{\Gamma_{i+1}} \right),\ 0 \le k_0 < k < n \right\}. \qquad (14)$$

Note that the random variables $\Gamma_i/\Gamma_{i+1},\ i=1,\dots,n-1$ are independent. Indeed, this follows from the fact that for all $k$,

$$\left(\frac{\Gamma_1}{\Gamma_k},\dots,\frac{\Gamma_{k-1}}{\Gamma_k}\right) \quad \text{and} \quad \Gamma_k$$

are independent. Furthermore, $\Gamma_i/\Gamma_{i+1}$ has the Beta$(i,1)$ distribution and hence $W_i := (\Gamma_i/\Gamma_{i+1})^i$ is Uniform$(0,1)$. This implies that

$$-i\log\left(\frac{\Gamma_i}{\Gamma_{i+1}}\right) = -\log(W_i),\ i=1,\dots,n-1,$$

are iid standard exponential, which in view of (14) yields (6).

By the so-established part (i), for a fixed $k_0$, we have that

$$\{(k-k_0)\widehat{\xi}_{k_0,k}(n),\ k=k_0+1,\dots,n-1\} \stackrel{d}{=} \xi\,\{\Gamma_{k-k_0},\ k=k_0+1,\dots,n-1\}\,.$$

Therefore, for the statistics defined in (7), we obtain

$$\begin{aligned}\{U_{k_0,k}(n),\ k=k_0+1,\dots,n-2\} &= \left\{\left(\frac{(k-k_0)\widehat{\xi}_{k_0,k}(n)}{(k-k_0+1)\widehat{\xi}_{k_0,k}(n)}\right)^{k-k_0},\ k=k_0+1,\dots,n-1\right\}\\ &\stackrel{d}{=} \left\{\left(\frac{\Gamma_{k-k_0}}{\Gamma_{k-k_0+1}}\right)^{k-k_0},\ k=k_0+1,\dots,n-2\right\}.\end{aligned}$$

As argued above, the random variables $(\Gamma_i/\Gamma_{i+1})^i,\ i=1,2,\dots$ are iid Uniform$(0,1)$, which proves (7). □

**Proof of Relation** (8)**.** We have

$$\begin{aligned}\left(Q\left(\frac{\Gamma_1}{\Gamma_{n+1}}\right),\dots,Q\left(\frac{\Gamma_k}{\Gamma_{n+1}}\right)\right) &= Q\left(\frac{\Gamma_{k+1}}{\Gamma_{n+1}}\right)\left(Q\left(\frac{\Gamma_1}{\Gamma_{n+1}}\right)\Big/Q\left(\frac{\Gamma_{k+1}}{\Gamma_{n+1}}\right),\dots,Q\left(\frac{\Gamma_k}{\Gamma_{n+1}}\right)\Big/Q\left(\frac{\Gamma_{k+1}}{\Gamma_{n+1}}\right)\right)\\ &= Q\left(\frac{\Gamma_{k+1}}{\Gamma_{n+1}}\right)\left(\frac{\ell(\Gamma_1/\Gamma_{n+1})}{\ell(\Gamma_{k+1}/\Gamma_{n+1})}\left(\frac{\Gamma_1}{\Gamma_{k+1}}\right)^{-\xi},\cdots,\frac{\ell(\Gamma_k/\Gamma_{n+1})}{\ell(\Gamma_{k+1}/\Gamma_{n+1})}\left(\frac{\Gamma_k}{\Gamma_{k+1}}\right)^{-\xi}\right).\end{aligned}$$

By the Strong Law of Large Numbers, we have $\Gamma_{k+1}/\Gamma_{n+1} \xrightarrow{a.s.} 0$, and hence by the slow variation property of $\ell$, for all fixed $k$ and $i=1,\dots,k$, we have

$$\frac{\ell(\Gamma_i/\Gamma_{n+1})}{\ell(\Gamma_{k+1}/\Gamma_{n+1})} = \frac{\ell((\Gamma_i/\Gamma_{k+1})(\Gamma_{k+1}/\Gamma_{n+1}))}{\ell(\Gamma_{k+1}/\Gamma_{n+1})} \xrightarrow{a.s.} 1.$$

This yields (8).

# B The need for dithering

Table 3 shows that each of the three longevity data sets involves a fair number of identical ages.

This *digitization effect* is due to the fact that human lifetimes are reported as integer number of days. Indeed, the period of 12 years and 164 days (the excess lifetime of Jeanne Calment) over the super-centenarian threshold of 110 years involves (approximately) $m = 4,547$ days (assuming

Table 3: Sample sizes and corresponding numbers of unique numerical values for each data set.

| Data set | $A$ | $A\&B$ | $GRG$ |
|---|---|---|---|
| Sample size | 631 | 668 | 347 |
| Distinct values | 478 | 499 | 313 |

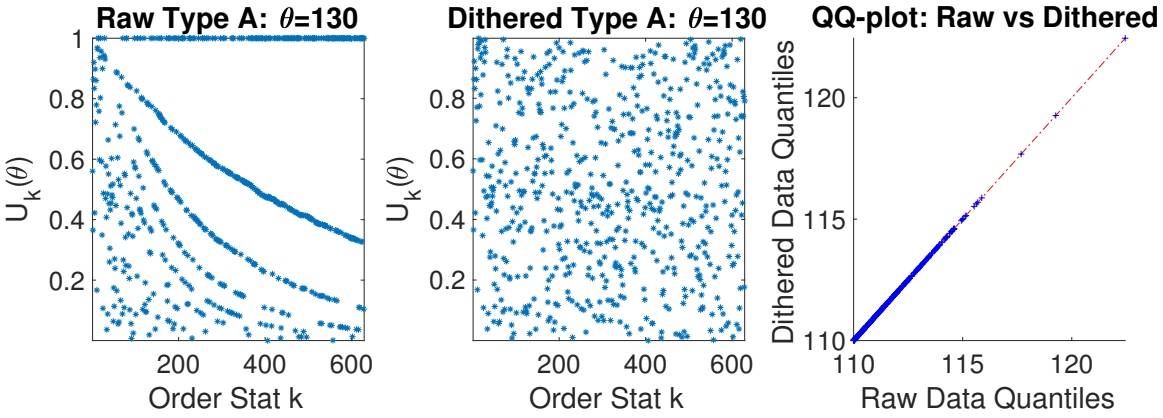


Figure 4: Scatter-plots of the $U_{0,j}$ statistics (10) based on the raw (left panel) and dithered (middle panel) type A data. The right panel shows a quantile-quantile plot of the raw versus the dithered data.

the year has 365.25 days). If one samples uniformly and at random $n = 631$ excess ages (in integer number of days) from $\{1, \ldots, m\}$, then the expected number of distinct ages in this sample is $m \times (1 - (m-1)^n/m^n) \approx 589.23$. The non-uniform excess age distribution leads to fewer distinct values but this ball-park computation explains the source of the seemingly odd digitization effect. The digitization effect in-of-itself does not indicate issues with the sampling scheme, but leads to many ties among the order statistics which affect the empirical distribution of the $U_{0,j}$'s (Figure 4). Before we can apply the proposed methodology, we need to fix this problem. We do so with *dithering*, i.e., to each lifetime $X_i$, we add a random time-of-day when the person departed. Formally, we consider the dithered sample $X_i^* := X_i + \Delta_i$, $i = 1, \ldots, n$, where the $\Delta_i$'s are independent and uniformly distributed in the interval $[-0.5/365.25, +0.5/365.25]$. Such dithering has virtually no effect on the distribution of the excess lifetimes but it eliminates the large number of ties among the order statistics and corrects for the odd digitization effect on the scatter-plots of the $U_{0,j}$-statistics (see Figure 4).

## References


[1] *The Holy Bible*. King James Version, American Edition. `https://www.bible.com/bible/547/GEN.6.kjva`.

[2] Stacy L. Andersen, Paola Sebastiani, Daniel A. Dworkis, Lori Feldman, and Thomas T. Perls. Health span approximates life span among many supercentenarians: Compression of morbidity at the approximate limit of life span. *The Journals of Gerontology: Series A*, 67A(4):395–405, 2012.

[3] H. Beltrán-Sánchez, F. Razak, and S.V. Subramanian. Going beyond the disability-based

morbidity definition in the compression of morbidity framework. *Global Health Action*, 7, 2014.

[4] S. Bhattacharya, M. Kallitsis, and S. Stoev. Trimming the Hill estimator: robustness, optimality and adaptivity. *ArXiv e-prints*, May 2017.

[5] Peter J. Bickel, Ya'acov Ritov, and Thomas M. Stoker. Tailor-made tests for goodness of fit to semiparametric hypotheses. *Ann. Statist.*, 34(2):721–741, 2006.

[6] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, 1987.

[7] A.C. Davison. 'The life of man, solitary, poor, nasty, brutish, and short' Discussion of the paper by Rootzén and Zholud. *Extremes*, 2018. In this volume.

[8] X. Dong, B. Milholland, and J. Vijg. Evidence of a limit to human lifespan. *Nature*, 538:257–259.

[9] J.F. Fries. Aging, natural death, and the compression of morbidity. *New England Journal of Medicine*, 303(3):130–135, 1980.

[10] IDL. International Database on Longevity, 2018. `http://supercentenarians.org/DataBase`.

[11] Lucien Le Cam and Grace Lo Yang. *Asymptotics in Statistics*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 2000. Some basic concepts.

[12] S. I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, 1987.

[13] H. Rootzén and D. Zholud. Human life is unlimited – but short. *Extremes*, 2018. Discussion paper.

[14] S. Stoev and Bhattacharya. Matlab code accompanying the paper: 'inference on the endpoint of human lifespan and its inherent statistical difficulty', 2018. `http://hdl.handle.net/2027.42/142999`.

[15] A. Swartz. James Fries: Healthy Aging Pioneer. *American Journal of Public Health*, 98(7):1163–1166, 2008.