An Extreme Value Analysis of Advanced Age Mortality Data

Kathryn A. Robertson
Debbie J. Dupuis, PhD
Bruce L. Jones, PhD, FSA, FCIA

Presented at The Living to 100 and Beyond Symposium
Sponsored by the Society of Actuaries
Orlando, Fla.
January 12–14, 2005

Copyright 2005 by the Society of Actuaries.
All rights reserved by the Society of Actuaries. Permission is granted to make brief excerpts for a published review. Permission is also granted to make limited numbers of copies of items in this monograph for personal, internal, classroom or other instructional use, on condition that the foregoing copyright notice is used so as to give reasonable notice of the Society's copyright. This consent for free limited copying without prior consent of the Society does not extend to making copies for general distribution, for advertising or promotional purposes, for inclusion in new collective works or for resale.

Abstract

Extreme value theory describes the behavior of random variables at extremely high or low levels. The application of extreme value theory to statistics allows us to fit models to data from the upper tail of a distribution. This paper presents a statistical analysis of advanced age mortality data, using extreme value models to quantify the upper tail of the distribution of human life spans.

Our analysis focuses on mortality data from two sources. Statistics Canada publishes the annual number of deaths in Canada, broken down by gender and age. We use the deaths data from 1949 to 1997 in our analysis. The Japanese Ministry of Health, Labour and Welfare also publishes detailed annual mortality data, including the 10 oldest reported ages at death in each year. We analyze the Japanese data over the period from 1980 to 2000.

Using the r-largest and peaks-over-threshold approaches to extreme value modeling, we fit generalized extreme value and generalized Pareto distributions to the life span data. Changes in distribution by birth cohort or over time are modeled through the use of covariates. We then evaluate the appropriateness of the fitted models, and discuss reasons for their shortcomings. Finally, we use our findings to address the existence of a finite upper bound on the life span distribution, and the behavior of the force of mortality at advanced ages.

1 Introduction

Over the last decade, several papers have been published using extreme value theory to model human mortality at extremely high ages. Aarssen and de Haan (1994) estimated a finite upper bound on the distribution of human life spans, while Galambos and Macri (2000) argued that such an upper bound could not exist. In this paper, we have used extreme value techniques to analyze Canadian and Japanese mortality data. Using the peaks-over-threshold approach, we fit generalized Pareto distributions to future lifetimes of survivors to a fixed age obtained from the Canadian data. The fitted distributions suggest that there is a finite upper bound, ω, on the distribution of human life spans. We believe that the finite estimate of ω results from the fact that our data extend only to age 100, and is not true evidence of a finite upper bound. For the Japanese data, we fit generalized extreme value distributions to the maximum age at death using the r-largest approach. While some estimates of ω are finite for the Japanese data, the confidence intervals for ω suggest that an infinite upper bound is very plausible. The Japanese parameter estimates also suggest an upward shift over time in the distribution of the annual maximum age at death.

2 Methodology

The application of extreme value theory to statistics allows us to investigate the behavior of a stochastic process at very high or very low levels. In this paper, we will use extreme value models to quantify the upper tail of the life span distribution.

Extreme value analyses depend on finite-sample approximations to several asymptotic results. In this paper, we will use the definitions and notation of Coles (2001). Suppose that $M_n = \max\{X_1, \ldots, X_n\}$ where $X_1, \ldots, X_n$ is a sequence of independent random variables having common distribution function F. Suppose, for some sequences of constants $\{a_n > 0\}$ and $\{b_n\}$,

$$P\left(\frac{M_n - b_n}{a_n} \le z\right) \to G(z) \text{ as } n \to \infty,$$

where $G(z)$ is a non-degenerate distribution function. It follows that G is a member of the generalized extreme value (GEV) family of distributions, given by the distribution function

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \qquad (1)$$

defined on $\left\{z : 1+\xi\left(\frac{z-\mu}{\sigma}\right) > 0\right\}$. There are three GEV parameters: a location parameter $-\infty < \mu < \infty$, a scale parameter $\sigma > 0$, and a shape parameter $-\infty < \xi < \infty$. When $\xi = 0$, the distribution function is the limit of (1) as $\xi \to 0$, ie.

$$G(z) = \exp\left\{-\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]\right\}$$

For large enough n, $P(M_n \le z)$ can be approximated by some member of the GEV family. That approximation is the basis of the block maxima approach to extreme value modeling. If we have a set of data where each observation is the maximum over a sufficiently large block of time, then we can fit a GEV distribution to the data and obtain estimates for μ, σ, and ξ.

Occasionally, we have data on several high order statistics for each block, and not just the block maxima. This extra information can aid us in obtaining better estimates of the GEV parameters μ, σ, and ξ in (1). Suppose there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P\left(\frac{M_n - b_n}{a_n} \le z\right) \to G(z) \text{ as } n \to \infty,$$

for some μ, σ, and ξ. Then the joint limiting distribution of

$$\tilde{M}_n^{(r)} = \left(\frac{M_n^{(1)} - b_n}{a_n}, \ldots, \frac{M_n^{(r)} - b_n}{a_n}\right)$$

as $n \to \infty$ has probability density function,

$$f(z^{(1)}, \ldots, z^{(r)}) = \exp\left\{-\left[1+\xi\left(\frac{z^{(r)}-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}\prod_{k=1}^{r}\sigma^{-1}\left[1+\xi\left(\frac{z^{(k)}-\mu}{\sigma}\right)\right]^{-1/\xi-1} \qquad (2)$$

on $\left\{z^{(k)} : 1+\xi\left(\frac{z^{(k)}-\mu}{\sigma}\right) > 0\right\}$ for $k = 1, \ldots, r$, where $z^{(r)} \le \cdots \le z^{(1)}$.

As well as describing the behavior of the maximum of a sequence of random variables, extreme value theory can be used to characterize the distribution above some large threshold u. If, for large n, $P(M_n \le z) \approx G(z)$ for some μ, σ, and ξ, then for sufficiently large u, the distribution function of $X - u$, given $X > u$, is:

$$P(X - u \le y \mid X > u) \approx H(y) = 1 - \left(1+\xi\frac{y}{\tilde{\sigma}}\right)^{-1/\xi} \qquad (3)$$

on $\left\{y : 1+\xi\frac{u}{\tilde{\sigma}} > 0\right\}$ where $\tilde{\sigma} = \sigma + \xi(u-\mu)$. $X - u \mid X > u$ is referred to as an exceedance of the threshold u. The family of distributions given by (3) is called

the generalized Pareto (GP) family. Again, the $\xi = 0$ case is interpreted as the limit,

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right),$$

an exponential distribution function. If we have observed all of the values above some high threshold u, we can use the peaks-over-threshold approach to fit a generalized Pareto distribution to the exceedances of u, and obtain estimates of $\tilde{\sigma}$ and ξ. If we have instead fit a GEV distribution to a set of data using the block maxima or r-largest approach, then we can find the corresponding GP parameters, given a high threshold value, u.

For mortality data, $H(y)$ can be interpreted as follows: Let X represent the time-to-death random variable for a person aged 0. Then, for some high age u, $H(y)$ represents the probability that the person will die before age $u + y$, given survival to age u, ie. ${}_y q_u$. Expressions for yearly mortality rates q_x and forces of mortality μ_x can be derived from the GP distribution. For a high age $x \geq u$,

$$q_x = 1 - \left[1 + \frac{\xi}{\tilde{\sigma} + \xi(x-u)}\right]^{-1/\xi}$$

and

$$\mu_x = \frac{1}{\tilde{\sigma} + \xi(x-u)}.$$

The existence of a finite upper bound, ω, on the life span distribution can also be addressed via the GP parameters. When $\xi < 0$, there is a finite upper bound of $\omega = u - \frac{\sigma}{\xi}$. When $\xi \geq 0$, there is no finite upper bound. Thus, we can test for the existence of a finite ω by testing the hypothesis $\xi < 0$.

All of the above results have been given in terms of stationary sequences of random variables. They can be adapted for use with data from non-stationary sequences, as well. In a non-stationary sequence, characteristics of the stochastic process change with changes in some related random variable. For example, the distribution of life spans might shift upward over time. Let t be a covariate, for example, the year of death or the birth cohort. The GEV or GP parameters can be expressed as functions of t. We will investigate changes in the distribution of extreme life spans over time through the use of such covariate functions.

Applications of extreme value theory often deal with environmental or structural data. For applications of the GEV model, see Robinson & Tawn (1995) and Tawn (1992). Smith (1986) provides an example of the r-largest approach. Davison & Smith (1990) used the peaks-over-threshold approach to model

several sets of environmental data. Coles (2001) provided numerous applications of extreme value theory, using the block maxima, r-largest, and peaks-over-threshold approaches.

Extreme value theory has been applied to actuarial science, as well. Embrechts *et al*. (1999) used a peaks-over-threshold approach to model industrial fire insurance claims data. They fit GP distributions to the exceedances of high threshold claim amounts u. The exceedance amounts corresponded to reinsurance claims under an excess-of-loss treaty with retention level u.

3 The Data

Statistics Canada publishes annual numbers of deaths from the Canadian population. The data are provided for males and females, and are broken down by individual ages from 0 to 99 and grouped over age 100. Doray (2002) investigated the fit of several mortality models to the Statistics Canada deaths data from 1949 to 1997. In his analysis, Doray constructed five cohort life tables from the Canadian data. The cohorts cover the birth periods 1869-72, 1873-77, 1878-82, 1883-87 and 1888-92. Each life table is provided for males and females separately, and encompasses ages 80-99 and 100+. We have used Doray's cohort life tables in our extreme value analysis of the Canadian data. For ease of notation, we refer to the cohorts as Cohort 1 (birth years 1869-72) through Cohort 5 (birth years 1888-92). Figure 1 shows the empirical mortality rates for the Canadian data, based on Doray's life tables.

The Japanese Ministry of Health and Welfare has published detailed mortality statistics since 1899. Robine *et al*. (2002) examine several features of the Japanese data, including the maximum and tenth largest ages at death in each year. In our analysis, we looked at the ten largest ages at death for each year.

Figure 2 displays the largest, second largest, and tenth largest ages at death over the period from 1899 to 2000. The data for 1944 to 1946 are not available. The maximum ages attained in the earlier years are often higher than those attained later in time. In 1872, a Japanese national birth registration system was implemented. Due to possible age misreporting, Robine *et al*. (2002) advise that the reported ages of people born before 1872 may not be accurate. We therefore chose to base our analysis on the data from 1980 to 2000. In 1986, the maximum age at death for a male was 120, considerably higher than any other age at death from 1980 to 2000, and fourteen years older than the second oldest age at death in that year. To attain age 120 in 1986, the person in question would have been born prior to 1872. We treated this point as a potential outlier, performing our analysis with it included and also with it replaced by a more plausible value.

4 Results

4.1 Canadian Data

For the Canadian data, we were not given the maximum age at death in each year. Based on the life table structure of the data, we decided to take the peaks-over-threshold approach in our analysis. In doing so, we assumed that the life spans of people within a given cohort and gender were independent and identically distributed random variables. Using maximum likelihood, we fit GP distributions over a range of thresholds.

Let u be a high threshold age, and let X_y be the age at death random variable for a life in Cohort y, $y = 1,\ldots,5$. Let H_u^y be the distribution function for $X_y - u$, given that $X_y > u$. The peaks over threshold approach states that, for $z > 0$,

$$P(X_y \le u + z \mid X_y > u) \approx H_u^y(z) = 1 - \left(1 + \frac{\xi_y z}{\sigma_y}\right)^{-1/\xi_y}$$

where $\sigma_y > 0$ and ξ_y are the GP parameters associated with Cohort y.

For Cohort y, let ℓ_x^y be the number of survivors to age x. Doray's life tables give ℓ_x^y for $x = 80,\ldots,100$. The number of deaths between age x and age $x+1$ is therefore ℓ_x^y - ℓ_{x+1}^y. The likelihood contribution for each age $x = 80,\ldots,99$ is the probability of dying between age x and age $x+1$, raised to the number of deaths, or

$$\left[H_u^y(x-u+1) - H_u^y(x-u)\right]^{\ell_x^y - \ell_{x+1}^y}.$$

The likelihood contribution for the survivors to age 100 is the probability of survival to age 100, raised to the number of survivors, or

$$\left[1 - H_u^y(100-u)\right]^{\ell_{100}^y}.$$

The resulting likelihood function for cohort y is

$$L_y(\sigma_y, \xi_y) = \left[\prod_{x=u}^{99}\left(H_u^y(x-u+1) - H_u^y(x-u)\right)^{\ell_x^y - \ell_{x+1}^y}\right]\left[1 - H_u^y(100-u)\right]^{\ell_{100}^y} \qquad (4)$$

with corresponding log-likelihood function,

$$\begin{aligned} l_y(\sigma_y, \xi_y) &= \log(L_y(\sigma_y, \xi_y)) \\ &= \sum_{x=u}^{99}(\ell_x^y - \ell_{x+1}^y)\log\left[H_u^y(x-u+1) - H_u^y(x-u)\right] + \ell_{100}^y \log\left[1 - H_u^y(100-u)\right] \end{aligned} \qquad (5)$$

Combining the likelihood contributions for each cohort, the overall log-likelihood function for a threshold of u is:

$$l(\sigma, \xi) = \sum_{y=1}^{5}\left[l_y(\sigma_y, \xi_y)\right]$$

where $\sigma = (\sigma_1, \ldots, \sigma_5)$ and $\xi = (\xi_1, \ldots, \xi_5)$. Using the R language (R Development Core Team, 2004), we maximized $l(\sigma, \xi)$ with respect to σ and ξ to obtain maximum likelihood estimates of the parameters. We defined the log-likelihood function within R, and used the function optim() to perform the maximization.

The first model we fit allowed σ_y and ξ_y to differ for each cohort and gender combination. We repeated the maximization for integer thresholds from 80 to 98. The resulting parameter estimates are given in Table 1, for several thresholds.

Figure 3 displays the maximum likelihood estimates of $\log(\sigma_y)$ versus y for the Canadian data, over several different thresholds. For each threshold, there appears to be a linear relationship between the estimate of $\log(\sigma_y)$ and y. Based on that observation, we fitted a second GP model to the Canadian data, with

$$\sigma_y = \exp[\beta_0 + \beta_1 (y-1)], y = 1, \ldots, 5.$$

Again, we placed no restrictions on the behavior of ξ_y. The parameter estimates for this model are shown in Tables 2 and 3, for several threshold choices.

In order to obtain the GP maximum likelihood parameter estimates, we can take u to be any integer between 80 and 98. If the exceedances above some threshold u_0 follow a GP distribution, then the exceedances over any higher threshold are also GP. We used several diagnostic tools to select the minimum threshold u_0 over which the Canadian data exceedances are GP.

Suppose the number of deaths in Cohort y at age $x \geq u_0$ is given by d_x^y. The observed number of deaths is given by

$$d_x^{y,obs} = \ell_x^y - \ell_{x+1}^y.$$

Under the GP assumption, the expected number of deaths is

$$d_x^{y,\exp} = \ell_{u_0}^y \left[\frac{H_{u_0}^y (x+1-u_0) - H_{u_0}^y (x-u_0)}{1 - H_{u_0}^y (x-u_0)} \right].$$

To test the goodness of fit of the GP distribution to exceedances of u_0, we used the χ^2 test statistic,

$$\chi^2 = \sum_{y=1}^{5} \left[\sum_{x=u}^{99} \frac{\left(d_x^{y,obs} - d_x^{y,\exp}\right)^2}{d_x^{y,\exp}} + \frac{\left(\ell_{100}^{y,obs} - \ell_{100}^{y,\exp}\right)^2}{\ell_{100}^{y,\exp}} \right]$$

For the male data, according to the χ^2 statistics, the fit of the GP distribution was not acceptable at the 5% level for thresholds below $u = 92$. For $u \geq 92$, the χ^2 test provided no evidence against the GP distribution. The χ^2 tests of the female data suggested that a GP distribution was appropriate for exceedances of

thresholds $u \geq 94$. We will therefore take u_0 to be 92 for males and 94 for females. The parameter estimates for those thresholds are listed in Table 4.

Figure 4 displays PP and QQ plots for the fitted GP distributions for males and females, Cohort 1. There are very few points on the diagnostic plots, since our data only went up to age 100, but there are no obvious problems with the existing points. The diagnostic plots for Cohorts 2 through 5 (not shown) did not highlight any problems with the fits, either. Figures 5 and 6 compare the fitted and observed mortality rates, q_x, for $x \geq u_0$. Again, the GP distributions appear to provide a good fit to the observed data.

The GP distributions fit to exceedances over 92 for males and 94 for females had negative shape parameters for each cohort. As discussed in Section 2, when $\xi_y < 0$, the lifespan distribution for Cohort y has a finite upper bound, $\omega_y = u - \frac{\sigma_y}{\xi_y}$. Figure 7 displays the estimates of ω_y for each gender/cohort combination (the starred values), along with 95% profile confidence intervals. Most of the estimates fall around 110-112 years. To find the profile confidence interval for ω_y, we first wrote the GP parameter σ_y as

$$\sigma_y = \xi_y (u - \omega_y)$$

We substituted that expression for σ_y into the log-likelihood function given by (5), so that l_y was a function of ω_y and ξ_y. We defined the profile log-likelihood for ω_y as

$$l_{y,p}(\omega_y) = \max_{\xi_y} \{ l_y(\omega_y, \xi_y) \}.$$

A 95 percent profile confidence interval for ω_y is given by

$$\left\{ \omega_y : 2\left[l_y(\hat{\omega}_y, \hat{\xi}_y) - l_{y,p}(\omega_y) \right] \leq \chi^2_{1,0.05} \right\}$$

where $\hat{\omega}_y$ and $\hat{\xi}_y$ are the maximum likelihood estimates of ω_y and ξ_y, and where $\chi^2_{1,0.05}$ is the 95 percent quantile of the χ^2 distribution, on one degree of freedom.

4.2 Japanese Data

For the Japanese data, we were provided with the ten highest ages at death in each year, for males and females. Based on the structure of the available data, we decided to take an r-largest approach to modeling Japanese mortality. In order to fit a GEV using the r-largest likelihood function, based on the pdf in (2), we needed exact observations. The Japanese data is interval censored: for each death, we know only that it occurred in a one-year interval from some age x to age $x+1$. We approximated the exact ages at death by assuming that deaths are uniformly distributed over the year. To obtain "exact"

ages, we added a random amount between zero and one to each of the integer ages.

The r-largest likelihood function for one gender is:

$$L = \prod_{t=1980}^{2000} \left\{ \exp\left\{ -\left[1+\xi\left(\frac{z_t^{(r)} - \mu(t)}{\sigma(t)} \right) \right]^{-1/\xi} \right\} \prod_{k=1}^{r} \sigma^{-1} \left[1+\xi\left(\frac{z_t^{(k)} - \mu(t)}{\sigma(t)} \right) \right]^{-1/\xi - 1} \right\} \tag{6}$$

where $\mu(t)$ and $\sigma(t)$ are functions of the year of death, t, and $z_t^{(r)} \le \cdots \le z_t^{(1)}$ are the r oldest ages at death in year t. For the Japanese data, $r = 10$. To maximize the r-largest likelihood and obtain parameter estimates, we used the R function rlarg.fit() in the package ismev.

The likelihood function in (6) assumes that the number of deaths from the Japanese population was the same in each year of data. In reality, the number of deaths likely increased over time, which would lead to different values for μ and σ in each year. By setting μ and σ to be functions of t, we have allowed for a change in population size. However, it will not be possible to distinguish between the effect of a change in the number of deaths and actual mortality improvements.

Figure 2 suggests that the ten oldest deaths are shifting upward over time. To account for that shift, we first let

$$\mu(t) = \mu_0 + \mu_1 t^*$$

where

$$t^* = \frac{t - 1980}{20}$$

is the standardized year of death, $0 \le t^* \le 1$. The maximum likelihood estimates for this model are given in Table 5.

In Table 5, the parameter estimates are given for three groups: males, including the possible outlier in 1986, males, with the possible outlier replaced by a more plausible value, and females. To find a plausible replacement for the possible male outlier, we began by fitting a GEV distribution to the original data, using the 10-largest approach. We calculated the conditional expected value for the maximum age at death in 1986, given that the person in question was born in or before 1872, and would therefore be at least 114 years old in 1986. We replaced the original value of 120 with the expected value, fit a GEV to the modified data, and repeated the procedure until the estimate for the maximum in 1986 converged. The final replacement for the outlier was 115.07.

The Japanese female data showed some evidence that the variance was changing over time. We investigated that possibility by allowing σ to vary with the year of death. The second r-largest model we fit set

$$\mu(t) = \mu_0 + \mu_1 t^*$$

and

$$\sigma(t) = \exp\{\sigma_0 + \sigma_1 t^*\}$$

The maximum likelihood estimates for this model are given in Table 6.

For both sets of male data, allowing σ to vary over time did not significantly improve the fitted models. The likelihood ratio statistics comparing the constant σ model to the second model were close to zero. For the female data, the addition of a fifth parameter did significantly improve the fit of the GEV distribution. In the remaining analysis, we have used the constant σ model for the male data, and the time-varying σ model for the female data.

To evaluate the goodness of fit of the fitted GEV models, we considered the corresponding GP distributions. We chose to look at exceedances of $u = 108$ for males, and $u = 111$ for females. Those thresholds lie above the highest tenth oldest ages at death, allowing us to be sure that we have all of the deaths above the threshold age. For each year t, the GP parameters were ξ, the shape parameter from the GEV, and

$$\tilde{\sigma}(t) = \sigma(t) + \xi(u - \mu(t)).$$

Let $Y_t \sim GP(\tilde{\sigma}(t), \xi)$ be an exceedance of u in year t. Since $\tilde{\sigma}(t)$ varies over time, the distribution of exceedances is different from year to year. We based the PP and QQ plots in Figures 8, 9, and 10 on the standardized variables,

$$\tilde{Y}_t = \xi^{-1} \log\left\{1 + \xi\left(\frac{Y_t - u}{\tilde{\sigma}(t)}\right)\right\}.$$

By rearranging variables, we can show that

$$P\left(\tilde{Y}_t \le x\right) = P\left(Y_t \le u + \frac{\tilde{\sigma}(t)}{\xi}\left(e^{\xi x} - 1\right)\right) = 1 - e^{-x}.$$

Thus, $\tilde{Y}_t$ has a standard exponential distribution.

Since the estimated shape parameters for the females and the males (with the outlier replaced) are negative, the fitted distributions have finite upper bounds, ω. We used a parametric bootstrap to obtain lower 95 percent confidence bounds for ω for the most recent year of data, 2000. We began by resampling 21 years of data from each of the fitted GEV distributions. Using the 10-largest approach, we fit a GEV distribution to the resampled data. We then used the new fitted parameters to estimate ω, setting $\hat{\omega} = \infty$ if the estimate of ξ was non-negative. Repeating those steps 2000 times, we obtained a set of estimates of ω. We took the 100th smallest value from the set of estimates to be a lower 95-percent confidence bound for ω.

For the female data, $\hat{\omega} = 124.21$, with a lower 95-percent confidence bound of 115.51. For males, with the outlier replaced, $\hat{\omega} = 135.56$, with a lower 95-percent confidence bound of 118.17. When the 1986 outlier is included in the

male data, the estimate of ξ is positive, leading to an infinite estimate of ω. However, it is still possible to construct a lower 95% confidence bound using the parametric bootstrap. The confidence bound in this case is 147.11.

Table 7 gives estimates of the mortality rates, q_x, and the force of mortality, μ_x, from the fitted GEV models. The estimates are calculated using the values of $\mu(t)$ and $\sigma(t)$ for the most recent year, 2000.

5 Discussion

5.1 Canadian Data

Our analysis of the Canadian data resulted in negative estimates of the shape parameter for both the male and female data. Those estimates, in turn, led to finite estimates of the upper bound, ω, on the life span distribution. Despite the apparent statistical evidence, we do not believe that there is truly a finite upper bound on the distribution of human life spans. If there was, then any person reaching age ω minus one day would have a probability of one of dying on that one day. Such a situation seems highly unlikely. Even if there were a finite ω, the values that we estimated, mostly around 110-112 years, seem far too low.

Since the peaks-over-threshold approach produced unexpected results, we examined the assumptions that we made in fitting the GP models. The peaks-over-threshold approach requires that the threshold u be set high enough that the distribution of exceedances of u can be approximated by a GP distribution. If we knew the underlying distribution from which the data originated, then we could determine theoretically what the limiting distribution of exceedances should be. We could also gauge how far we must move the threshold into the tail of the distribution for the GP approximation to be valid. Doray (2002) looked at the Gompertz and Kannisto distributions as possibilities for modeling the distribution of human life spans. Both of those distributions fall into the "domain of attraction" of the $\xi = 0$ subset of extreme value distributions. In other words, the limiting distribution of exceedances from true Gompertz or Kannisto data should have $\xi = 0$, and therefore an infinite upper bound.

We examined the Kannisto distribution fit by Doray (2002), using maximum likelihood, to the Canadian male data from Cohort 1. We found the "best-fitting" generalized Pareto distribution to exceedances from that Kannisto by minimizing the squared difference between the quantile functions. Let

$$H_u(x) = 1 - \left[1 + \frac{\xi(x-u)}{\sigma}\right]^{-1/\xi}$$

be the generalized Pareto distribution function for exceedances of u with parameters σ and ξ, and let $h(x)$ be the corresponding density function. Let $F_X(x)$ be the distribution function of the Kannisto distribution, fit by maximum likelihood to the Canadian male data, Cohort 1. Let

$$F_{X|X>u}(x) = \frac{F_X(x) - F_X(u)}{1 - F_X(u)}$$

be the associated conditional distribution for values greater than u. For a series of thresholds, u, we minimized

$$\int_0^1 \left[F_{X|X>u}^{-1}(p) - H_u^{-1}(p)\right]^2 h_u(H_u^{-1}(p))dp$$

with respect to the GP parameters, σ and ξ. We then examined QQ plots comparing the quantiles of the conditional Kannisto distribution to those of the best-fitting GP distribution. There was considerable curvature in the QQ plots for thresholds below 110. Since our data only allows us to look at thresholds up to $u = 98$, it is not surprising that we did not see the results we would expect for the upper bounds. Although the diagnostic plots in Figure 4 show no obvious problems with the GP distribution, our choices of u were too low for the results to be valid. However, we did not have data at sufficiently high ages to detect this. To obtain credible results from the peaks-over-threshold approach, we would need more detailed data above age 100, and a considerably higher threshold age, u.

For low thresholds, as we have in the Canadian data, finite estimates of ω can actually be expected. Suppose the true distribution is a Gompertz distribution. It is possible to show that, as we move the threshold further into the tail of the distribution, the GP distribution that best approximates the true Gompertz distribution over the relevant range of values has a ξ that approaches zero from the left hand side. This is not surprising since only GPs with $\xi < 0$ will yield an increasing force of mortality.

5.2 Japanese Data

The maximum likelihood estimates of the GEV parameters for the Japanese analysis reflect several features of the data. For both males and females, the maximum likelihood estimates of μ_1 are positive, leading to an increasing $\mu(t)$. Given the plots of the data in Figure 2, the positive estimates of μ_1 are not surprising. The data plots show an increasing trend in the ten largest order statistics from 1980 to 2000. Less clear is the reason behind that increase. One possibility is that mortality has improved from 1980 to 2000, leading to a higher average value for the maximum age at death. However, over the same time

period, the Japanese population has grown. The larger pool of lives may be allowing us to see observations further into the right tail of the distribution than we did in previous years.

For the female data, the estimate of σ_1 was also positive, leading to an increasing function $\sigma(t)$. From the plot of the female data in Figure 2, the variance of the oldest age at death appears to increase slightly from 1980 to 2000. The positive estimate of σ_1 reflects that increase.

The mortality rates in Table 7 increase fairly steadily with age for both females and males, with the 1986 outlier replaced. The forces of mortality for those two groups increase sharply as the ages approach the estimated upper bounds. When the 1986 outlier is included in the analysis, the force of mortality and mortality rates decrease slightly by age.

Despite the finite estimates of the upper bounds, ω, for females and males with the 1986 outlier replaced, it is still very plausible that the upper bounds are actually infinite. In Section 4.2, we found lower 95-percent confidence bounds for ω using a parametric bootstrap. In doing so, we obtained a set of estimates of ω. For the female data, 307 of the 2000 estimates were infinite. For the male data, with the outlier replaced, the number of infinite estimates was 1189. With the 1986 male outlier included, the initial estimate of ω was infinite, as were 1769 of the parametric bootstrap estimates. Based on those counts, two-sided 95% confidence intervals for ω would extend to ∞ for all three of the data sets.

The PP and QQ plots for the r-largest fits are shown in Figures 8, 9 and 10. The points on the female plots lie very close to the unit diagonal, and show no major difficulties. The possible outlier from 1986 shows up clearly on the QQ plot in Figure 8. In Figure 9, the maximum age at death in 1986 lies closer to the QQ plot unit diagonal than in Figure 8, but it is still obvious. Our restriction that the replacement value be greater than 114 has kept the point significantly higher than the rest of the data in the 1980-2000 interval.

6 Conclusion

Extreme value theory provides useful tools for analyzing data at very high levels. Unfortunately, our Canadian mortality data did not extend far enough into the tail of the life span distribution for the extreme value approximations to be valid. To effectively
use the extreme value techniques, we would need more detailed information about mortality after age 100. For the Japanese data, the GEV distribution fit by an r-largest approach provided a reasonable fit. For that data set, confidence intervals for ω supported our belief that there is no finite upper bound on the distribution of human life spans.

7 Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada in the form of a Postgraduate Scholarship to the first author and research grants to the second and third authors. We would like to acknowledge NUPRI, the Nihon University Population Research Institute, for providing the Japanese mortality data.

References

1. Aarssen, K., and L. de Haan. 1994. On the Maximal Life Span of Humans. *Mathematical Population Studies* 4(4): 259-281.

2. Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.

3. Davison, A.C., and R.L. Smith. 1990. Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society, Series B (Methodological)* 52(3): 393-442.

4. Doray, L. G. 2002. Living to age 100 in Canada in 2000. *Proceedings of the Society of Actuaries' Living to 100 and Beyond: Survival at Advanced Ages Symposium, January 17-18 2002*. Online at http://www.soa.org/ccm/content/research-publications/research-projects/living-to-100-and-beyond-survival-at-advanced-ages/.

5. Dupuis, D.J. 1997. Extreme-Value Theory Based on the r Largest Events: A Robust Approach. *Journal of Hydrology* 200: 253-264.

6. Embrechts, P., S. I. Resnick, Gennady Samorodnitsky. 1999. Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal* 3(2): 30-41.

7. Galambos, J. and N. Macri. 2000. The Life Length of Humans Does Not Have a Limit. *Journal of Applied Statistical Science* 9(4): 253-264.

8. R Development Core Team. 2004. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online at http://www.R-project.org.

9. Robine, J., Y. Saito, and C. Jagger. 2002. Living and dying beyond age 100 in Japan. *Proceedings of the Society of Actuaries' Living to 100 and Beyond: Survival at Advanced Ages Symposium, January 17-18, 2002*. Online at http://www.soa.org/ccm/content/research-publications/research-projects/living-to-100-and-beyond-survival-at-advanced-ages/.

10. Robinson, M. E., J. A. Tawn. 1997. Statistics for Extreme Sea Currents. *Applied Statistics* 46(2): 183-205.

11. Smith, R.L. 1986. Extreme value theory based on the *r* largest annual events. *Journal of Hydrology* 86: 27-43.

12. Tawn, J. A. 1992. Estimating probabilities of extreme sea-levels. *Applied Statistics* 41(1): 77-93.