Predictive Analytics for Modeling Threshold Life Tables

Min Ji, Mostafa Aminzadeh, Min Deng
Towson University, Maryland, USA

Advances in Predictive Analytics Conference
Waterloo, December 2, 2017
Outline

1. Introduction
2. Threshold life tables
3. Maximum likelihood estimation
4. Bayesian inference
5. Concluding remarks
The generalized Pareto distribution

- The peaks-over-threshold (POT) approach in the extreme value theory describes the behavior of large observations which exceed some high threshold;

- The generalized Pareto distribution (GPD) provides a unifying approach to the distribution of threshold excesses;

- In actuarial science literature, the (GPD) has been regarded as a promising approach to the modeling of high age mortality;
Excess distribution

For a large value $u$, the conditional random variable over the threshold $u$ is defined to be $Y = X - u \mid X > u$, with the excess distribution expressed as

$$F_u(y) = \Pr\{X - u \leq y \mid X > u\} = \frac{F(u + y) - F(u)}{1 - F(u)}, \quad (1)$$

for $0 \leq y < x_F - u$, where $x_F < \infty$ is the right endpoint of $F$ if it exists.
The excess distribution $F_u(y)$ over some suitably high threshold $u$ can be well approximated by a generalized Pareto distribution given by

$$G(y) = \begin{cases} 
1 - (1 + \frac{\gamma y}{\sigma})^{-1/\gamma}, & \text{if } \gamma \neq 0 \\
1 - \exp\left(-\frac{y}{\sigma}\right), & \text{if } \gamma = 0
\end{cases}$$

(2)

defined on $\{y : y > 0 \text{ and } (1 + \gamma y / \sigma) > 0\}$. If $\gamma < 0$ the distribution of threshold excesses has an upper bound of $-\sigma / \gamma$; if $\gamma \geq 0$ the distribution has no upper limit. The case of $\gamma = 0$ is interpreted by taking the limit $\gamma \to 0$, leading to an exponential distribution with parameter $1 / \sigma$. 

The general Pareto distribution
Research question

- A threshold life table has been proposed by Li et al. (2008), using the Gompertz law connected with the GPD;

- There are difficulties in parameter estimation for the threshold life tables;

- Some research used a predetermined threshold;

- Thresholds were set at integer ages;

- Cannot guarantee smooth death probabilities over ages.
Research objectives

- To achieve a smooth threshold life table
- To use Bayesian approach to estimate the parameters
- To derive predictive density of lifetime distribution and other quantities of interest
Some notations

Define the following notations for life time distribution

- $X$: the age at death random variable, assumed to be continuous.
- $f(x)$: the probability density function of $X$.
- $F(x)$: the distribution function of $X$.
- $S(x) = 1 - F(x)$: the survival function.
- $\mu(x) = f(x)/S(x)$: the force of mortality, which is called the failure rate in credibility theory.
- $d(x)$: the number of death at age $x$.
- $E(x)$: the amount of exposure-to-risk at age $x$. 

Threshold life tables

The threshold life table proposed by Li et al. (2008) is mathematically expressed as follows:

\[
F(x) = \begin{cases} 
1 - \exp \left( - \frac{B}{\ln c} (c^x - 1) \right) & \text{if } x \leq u \\
F(u) + (1 - F(u)) \left( 1 - \left(1 + \frac{\gamma(x-u)}{\sigma}\right)^{-1/\gamma} \right) & \text{if } x > u 
\end{cases}
\]

where \( F(x) \) is the distribution function of the lifetime random variable.
Propose two constraints

1. To ensure $F(x)$ is continuous and differentiable at $x = u$, we need
   \[ \sigma = \frac{1}{Bc^u}, \]
2. To ensure $f(x)$ is differentiable at $x = u$.
   \[ \gamma = -\frac{\ln c}{Bc^u}, \]

the probability density function becomes
\[
f(x) = \begin{cases} 
\exp \left(-\frac{B}{\ln c} (c^x - 1)\right) Bc^x & \text{if } x \leq u \\
\frac{Bc^u}{\ln c} f(u) \left(1 - (x - u) \ln c\right)^{\frac{Bc^u}{\ln c} - 1} & \text{if } x > u
\end{cases}.
\]
Properties of the proposed threshold life table

- Five parameters are reduced to three free parameters $B$, $c$, and $u$;

- The shape and scale of the GDP distribution is implicitly determined by the threshold and the force of mortality at the threshold;

- The second constraint indicates $\gamma < 0$, which means the GPD in Equation (3) has a right end point $-\sigma/\gamma$;

- The limiting age of human beings after imposing the two constraints is $u - \sigma/\gamma = u + \frac{1}{\ln c}$. 
The likelihood function

- Use the US mortality data in year 2015 at ages 65 to 105: \(d_x\) and \(E_x\), for \(x = 65, 66, 67, \ldots, 105\).

- Assume the number of deaths at age \(x\) follows a binomial distribution.

- The likelihood function of a threshold life table fitting to the period mortality data is

\[
L(B, c, \gamma, \sigma, u) = \prod_{x=65}^{105} \left( \frac{E_x}{d_x} \right) q_x d_x (1 - q_x)^{E_x - d_x} \tag{5}
\]

where \(q_x = \Pr(X < x + 1|X > x) = 1 - S(x + 1)/S(x)\) is the probabilities of dying between ages \(x\) and \(x + 1\).
Introduction
Threshold life tables
Maximum likelihood estimation
Bayesian inference
Concluding remarks

Model 1: two constraint

The likelihood function becomes

\[
L(B, c, u) = \prod_{x=m-1}^{x=m+1} \left( \frac{E_x}{d_x} \right) (1 - e^{-Bc^x \frac{c-1}{\ln c}}) d_x e^{-Bc^x \frac{c-1}{\ln c} (E_x - d_x)} \\
\times \left( \frac{E_m}{d_m} \right) (1 - e^{-Bc^u \frac{c^m}{\ln c}} (1 - (m + 1 - u) \ln c) \frac{Bc^u}{\ln c})^d_m \\
\times \left( e^{-Bc^u \frac{c^m}{\ln c}} (1 - (m + 1 - u) \ln c) \frac{Bc^u}{\ln c} \right)^{E_m-d_m} \\
\times \prod_{x=m+1}^{105} \left( \frac{E_x}{d_x} \right) (1 - (1 - \frac{\ln c}{1 - (x - u) \ln c} \frac{Bc^u}{\ln c}) d_x (1 - \frac{\ln c}{1 - (x - u) \ln c} \frac{Bc^u}{\ln c} (E_x - d_x)),
\]

where \([\cdot]\) denotes the greatest integer function and \(m = \lfloor u \rfloor\). For any search number \(m\), the boundaries are \(m \leq u < m + 1\) and \(c \leq e^{\frac{1}{106-m}}\) respectively.
Model 2: no constraint

The likelihood function can be written as

\[
L(B, c, u, \gamma, \sigma) = \prod_{x=65}^{m-1} \left( \frac{E_x}{d_x} \right) \left( 1 - e^{-Bc^x \frac{c-1}{\ln c}} \right) d_x \left( e^{-Bc^x \frac{c-1}{\ln c}} (E_x - d_x) \right)
\times \left( \frac{E_m}{d_m} \right) \left( 1 - e^{-B \ln c (cu - cx)} (1 + \frac{\gamma}{\sigma} (m + 1 - u)) \right)^{\frac{-1}{\gamma}} \left( 1 + \frac{\gamma}{\sigma + \gamma(x - u)} \right) \frac{1}{\gamma} (E_x - d_x)
\]

where \( m = \lfloor u \rfloor \), and \( \gamma(106 - u) + \sigma > 0 \) if \( \gamma \) is negative.
Model comparison

To compare the results of fitting threshold life tables to the mortality data, we consider the following three models:

- Model 0: the Gompertz law
- Model 1: the threshold life table with two constraints
- Model 2: the threshold life table without constraints
### Maximum likelihood estimates

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td><strong>0</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>1.85E-05</td>
<td>1.91E-05</td>
</tr>
<tr>
<td><strong>c</strong></td>
<td>1.1062</td>
<td>1.1058</td>
</tr>
<tr>
<td><strong>u</strong></td>
<td>100.003</td>
<td>93</td>
</tr>
<tr>
<td><strong>γ</strong></td>
<td>-0.256</td>
<td>-0.252</td>
</tr>
<tr>
<td><strong>σ</strong></td>
<td>3.917</td>
<td></td>
</tr>
<tr>
<td><strong>ω</strong></td>
<td>109.95</td>
<td>108.28</td>
</tr>
<tr>
<td><strong>Logl</strong></td>
<td>-1745.34</td>
<td>-1736.96</td>
</tr>
<tr>
<td><strong>BIC</strong></td>
<td>3498.11</td>
<td>3481.35</td>
</tr>
</tbody>
</table>

**Table:** Parameter estimates of the three mortality curves
## Likelihood ratio test

<table>
<thead>
<tr>
<th>Sex</th>
<th>Null</th>
<th>Alternative</th>
<th>$\chi^2$ statistics</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 0</td>
<td>Model 1</td>
<td>16.76</td>
<td>4.2E-05</td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>Model 2</td>
<td>771.99</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 0</td>
<td>Model 1</td>
<td>23.15</td>
<td>1.5E-06</td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>Model 2</td>
<td>287.50</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Likelihood ratio tests for model selection
Figure: Raw mortality probabilities, the fitted Gompertz law and the fitted threshold life table with constraints, for males (left) and females (right)
Graphic comparison

**Figure:** Raw mortality probabilities and the with and fitted threshold life table without constraints, for males (left) and females (right)
Prior distributions

• Assume the parameters $B$, $c$, and $u$ of the threshold life table with smoothness constraints are independent;

• Three independent prior distributions are

\[
\begin{align*}
  f(B) &= \frac{1}{\lambda} e^{-\frac{B}{\lambda}}, \quad B > 0, \lambda > 0, \\
  f(c) &= \frac{\theta}{(1 - c_m^{-\theta})c^{\theta+1}}, \quad 1 < c \leq c_m, \theta > 0, \\
  f(u) &\propto (u - m)^{\alpha-1}(m + 1 - u)^{\beta-1}, \quad m \leq u < m + 1, \alpha > 0, \beta > 0,
\end{align*}
\]

where $\lambda$, $\theta$, $\alpha$, and $\beta$ are hyper parameters.
Joint posterior probability

- Let \( \mathbf{E} \) and \( \mathbf{d} \) be the vector with elements \( E_x \) and \( d_x \) respectively, for \( x = 65, 66, \ldots, m - 1, m, m + 1, \ldots, 105 \);

- Assuming the counts of death at different ages are independent random variables;

- The joint posterior probability density function of the parameters \( B, c, u \) can be written as

\[
p(B, c, u | \mathbf{d}) \propto f(B)f(c)f(u)L(B, c, u) \\
\propto \theta(u - m)^{\alpha - 1}(m + 1 - u)^{\beta - 1} \frac{\lambda c^{\theta + 1}}{\lambda c^{\theta + 1}} e^{-\lambda/B} L(B, c, u),
\]

where \( L(B, c, u) \) the likelihood function in Equations (6)
Conditional posteriors

The conditional posteriors can be written as

\begin{align*}
  f_1(B|c, u, E, d) &\propto e^{-\frac{B}{\lambda}} f(B, c, u, E, d), \\
  f_2(c|B, u, E, d) &\propto \frac{1}{c^{\theta+1}} f(B, c, u, E, d), \\
  f_3(u|B, c, E, d) &\propto (u - m)^{\alpha-1} (m + 1 - u)^{\beta-1} f(B, c, u, E, d),
\end{align*}

where

\begin{equation*}
f(B, c, u, E, d) = \prod_{x=m+1}^{105} \left( 1 - \frac{\ln c}{1-(x-u)\ln c} \right)^{Bc_u} \left( 1 - \frac{\ln c}{1-(x-u)\ln c} \right)^{Bc_u} (E_x-d_x)
\end{equation*}
Following Aminzadeh (2013), we use the Markov chain Monte Carlo (MCMC) method based on the Metropolis & Hastings algorithm to generate samples;

At each iteration of which all parameters are updated simultaneously;

Propose to use the following proposal distributions for the parameters $B$, $c$, and $u$.

- The proposal distribution for $B$ is $\text{EXP}(B_{\text{mle}})$, an exponential distribution with mean equal to mle of $B$.
- The proposal distribution for $c$ is $N(c_{\text{mle}}, 0.005^3)$, a normal distribution with mean equal to mle of $c$.
- The proposal distribution for $u$ is $N(u_{\text{mle}}, 0.1^2)$, a normal distribution with mean equal to mle of $u$.
Figure: Markov chain Monte Carlo samples
Simulation results

<table>
<thead>
<tr>
<th></th>
<th>True value</th>
<th>MCMC</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>MSE</td>
</tr>
<tr>
<td>$B$</td>
<td>1.915E-05</td>
<td>1.906E-05</td>
<td>9.53E-12</td>
</tr>
<tr>
<td>$c$</td>
<td>1.1058</td>
<td>1.1060</td>
<td>4.648E-06</td>
</tr>
<tr>
<td>$u$</td>
<td>100.003</td>
<td>99.91</td>
<td>0.104</td>
</tr>
</tbody>
</table>

Table: Compare the MCMC method and maximum likelihood estimation based on 100 sets of simulated numbers of death data
Conclusions

- Threshold life table with two constraints is favorable model
- The MCMC method performs better than the maximum likelihood estimation
- Predictive analysis of the threshold life tables enables Bayesian inference of the life time distribution, limiting age, and other quantities of interest
Introduction

Threshold life tables

Maximum likelihood estimation

Bayesian inference

Concluding remarks

Selected references:


