ACTUARIAL RESEARCH CLEARING HOUSE
1993 VOL. 1

An Alternative Method For Least-Squares Fitting of Parametric Survival Models

Anita Singh
Lockheed Engineering
Las Vegas, Nevada

Rohan Dalpatadu
Department of Mathematics
University of Nevada, Las Vegas

Malwane Ananda
Department of Mathematics
University of Nevada, Las Vegas

Abstract

The problem of estimation of parameters of the two well-known three parameter models of force of mortality, viz., Weibull, and Makeham models, is considered in this paper. The method of least-squares, commonly used for estimation of parameters of these models, involves log-transformation of the force of mortality data. We propose an efficient numerical scheme for solving the problem. The proposed method involves fitting a series of linear regressions, and finding the least-squares estimates of all the parameters by the method of Golden Section search for computer minimization. Some examples of the proposed method will be given to compare the existing least-squares method with the proposed procedure.

Introduction

The force of mortality $\mu(t)$ is defined as

$$\mu(t) = \frac{f(t)}{1-F(t)}, \tag{1}$$

where $f(t)$ and $F(t)$ are the probability density function and cumulative distribution function respectively. Most general forms of $\mu(t)$ involve three parameters. Some of the commonly used forms of $\mu(t)$ are as follows:

1. Makeham form: $\mu(t) = A + Bc^t$, $B > 0$, $c > 1$, $A > -B$, $t \geq 0$. (2)
2. Weibull three parameter form: $\mu(t) = k(t-\delta)^n$, $k > 0$, $\delta > 0$, $n > 0$, $t \geq \delta$ (3)

   or

$$\mu(t) = \left(\frac{t-\delta}{\theta}\right)^{\beta}, \ \beta > 0, \ \delta > 0, \ \theta > 0, \ t \geq \delta \tag{4}$$

Given an ordered sample

$$t_1 < t_2 \cdots < t_n$$

of size n from a population with mortality rate $\mu(t)$, we are interested in estimating unknown parameters.

Commonly used methods of estimation of the unknown parameters are:

(a) a sequence of Taylor series approximations for the Makeham model.

(b) a trial and error type method to estimate δ and then a log transformation followed by linear regression for the Weibull model.

The proposed method is a direct least-squares estimate for the unknown parameters. In the proposed method we fix one parameter and find the least square estimators for the remaining parameters. Finally we find the fixed parameter by minimizing the sum of squared error (SSE).

In the next section we give a brief discussion about the existing methods for both cases and in the following section we describe the proposed method.

Existing Methods

1. $\mu(t) = A + Bc^t$. Let A_0, B_0, and c_0 be initial estimates. Then

$$\mu(t) \approx A_0 + B_0 c_0^t + (A - A_0) + (B - B_0)c_0^t + (c - c_0)B_0 t c_0^{t-1}.$$

Now obtain the least-squares estimates for $A - A_0, B - B_0$, and $c - c_0$. Define $A_1 = A_0 + (\hat{A} - A_0)$, $B_1 = B_0 + (\hat{B} - b_0)$, $c_1 = c_0 + (\hat{c} - c_0)$. and obtain the least-squares estimates for $A - A_1$, $B - B_1$, and $c - c_1$ using the model

$$\mu(t) = A_1 + B_1 c_1^{\,t} + (A - A_1) + (B - B_1)c^t{}_1 + (c - c_1)B_1 + c_1^{\,t-1}.$$

The sequences $\{A_k\}$, $\{B_k\}$, and $\{c_k\}$ converge to the estimates of A, B, and c. (This is Nesselle's method. For a full description, see Survival Models and their Estimation by Dick London.)

2. $\mu(t) = k(t-\delta)^n$ or $\mu(t) = \left(\frac{t-\delta}{\theta}\right)^\beta$. Using a log transformation we obtain

$$\log[\mu(t)] = \log\left[\frac{f(t)}{1-F(t)}\right] = \log k + n \cdot \log(t-\delta)$$

or

$$\log[\mu(t)] = -\beta \log\theta + \beta\log(t-\delta) = \delta + \beta \cdot \log(t-\delta)$$

The method is to keep fitting straight lines to $(\log(t_i - \delta), \log[\mu(t_i)])$ data for various values of δ, until a reasonable fit is found.

We can use this method for the Makeham model by fitting straight lines to $(c^{t_i}, \mu(t))$ for various values of c, until a reasonable fit is found.

In the above methods, the force of mortality $\mu(t)$ is first estimated (from the given data set) for each $i = 1, 2, \cdots, n$. One way of estimating μ is as follows:

If $U_i = F(t_i)$ is the ith ordered observation from a sample of size n, then $E(U_i) = \frac{i}{n+1}$ and $F(t_i)$ can be estimated using

$$\hat{F}(t_i) = \frac{i}{n+1}$$

and therefore $\mu(t_i)$ can be estimated by

$$\hat{\mu}(t_i) = \frac{\hat{F}(t_{i+1}) - \hat{F}(t_i)}{(t_{i+1} - t_i)(1 - \hat{F}(t_i))}$$

$$= \frac{1}{(t_{i+1} - t_i)(n+1-i)}$$

Another way of estimating $\mu(t_i)$ is by using a more robust estimator of $F(t)$.

$$\hat{F}(t_i) = \frac{i-.3}{n+.4}$$

and the estimated value of $\mu(t_i)$ is

$$\hat{\mu}(t_i) = \frac{1}{(t_{i+1} - t_i)(n+.7-i)}.$$

Proposed Method

Consider

$$\log\left[\frac{1}{1-F(t)}\right] = \int \mu(t)dt$$

$$y(t) = \int \mu(t)dt \text{ where } y(t) = \log\left[\frac{1}{1-F(t)}\right].$$

By fixing one parameter, we convert $\int \mu(t)dt$ in to linear form and then find the other unknown parameters by least square method. Finally, we find the fixed parameter by minimizing the SSE.

One can find an estimator

$$\hat{y}(t_i) = \log\frac{n+1}{n+1-i}$$

of y(t) by replacing $F(t)$ by $\hat{F}(t) = \frac{i}{n+1}$, or a more robust estimator

$$\hat{y}(t_i) = \log\left[\frac{n+.4}{n+.7-i}\right]$$

by replacing $F(t)$ by

$$\hat{F}(t_i) = \frac{i-.3}{n+.4}.$$

1. For the Makeham model:

$$\log\left[\frac{1}{1-F(t)}\right] = At + \frac{B(c^t - 1)}{\log c}$$

If we first fix c, then this is in linear form

$$y(t_i) = AX_1 + BX_2$$

where

$$y(t) = \log\left[\frac{1}{1-F(t)}\right],\ X_1 = t \text{ and } X_2 = \frac{c^t - 1}{\log c}.$$

The following algorithm is used to find the estimators of A, B, and c.

Step 1: Find an initial search interval [L,H] for c.

Step 2. For fixed $c \in [L,H]$ fit a linear regression line to the points $(y(t_i), x_1(t_i), x_2(t_i))$, $i = 1,2,\cdots n$ and obtain least square estimators $\hat{A}_c$ and $\hat{B}_c$.

Step 3. Compute $SSE(c) = \sum\left[y(t_i) - \hat{A}_c X_1 - \hat{B}_c X_2\right]^2$.

Step 4. Use a computer search minimization to find c_0, which minimizes SSE(c). We have used the golden section search procedure for this step and details of this procedure are given later.

The estimators are $\hat{c} = c_0$, $\hat{A} = \hat{A}_{c_0}$ and $\hat{B} = \hat{B}_{c_0}$.

2. For the Weibull model:

First consider the parametrization given in (3). Then

$$\log\left[\frac{1}{1-F(t)}\right] = \frac{k(t-\delta)^{n+1}}{(n+1)}$$

$$\log\left[\log\left(\frac{1}{1-F(t)}\right)\right] = \log\left(\frac{k}{n+1}\right) + (n+1)\ \log(t-\delta).$$

This is in the form

$$y(t_i) = A + BX_1$$

where

$$
\begin{aligned}
y(t_i) &= \log\left[\log\left(\frac{1}{1-F(t)}\right)\right] \\
A &= \log\left(\frac{k}{n+1}\right) \\
B &= n+1 \\
X_1 &= \log(t-\delta).
\end{aligned}
$$

The following procedure is used to obtain estimators for n, k and δ.

Step 1: Find an initial search interval $[L,H]$ for δ.

Step 2: For fixed $\delta \in [L,H]$, fit a straight line to the points

$(y(t_i),\ x_1(t_i))$; $i = 1,2,\cdots n$ and obtain least square estimators $\hat{A}_\delta$ and $\hat{B}_\delta$.

Step 3. Compute

$$\mathrm{SSE}(\delta) = \sum\left(y(t_i) - \hat{A}_\delta - \hat{B}_\delta X_1\right)^2$$

Step 4. Use the golden section search procedure to find δ_0, which minimizes $\mathrm{SSE}(\delta)$.

Step 5. Find $\hat{A}_{\delta_0}$ and $\hat{B}_{\delta_0}$.

Estimators for the parameters δ, n, and k are

$$\hat{\delta} = \delta_0, \quad \hat{n} = \hat{B}_{\delta_0} - 1$$

$$\hat{k} = \hat{B}_{\delta_0}\, e^{\hat{A}\delta_0}.$$

Now consider the second parametrization given in (4). Then

$$\log\left[\frac{1}{1-F(t)}\right] = \frac{1}{\theta^\beta}\frac{(t-\delta)^{\beta+1}}{(\beta+1)}$$

$$\log\left[\log\left(\frac{1}{1-F(t)}\right)\right] = -\log\left[\theta^\beta(\beta+1)\right] + (\beta+1)\log(t-\delta).$$

For given δ, this is in the linear form and we can find the estimators for δ, θ and β by minimizing SSE(δ) as before.

The golden section search procedure is described in the next section and some examples are given in the following seciton.

The Golden Section Search Procedure

Let $f(x)$ be continuous on $[a,b]$ and suppose $f(x)$ has a unique minimum at $c \in [a,b]$.

Step 1: Let $h = b - a$, $x_1 = b - \gamma h$, and $x_2 = a + \gamma h$, where $1/\gamma$ is the golden ratio, i.e., the positive root of $x^2 + x - 1 = 0$, i.e., $\gamma = \frac{\sqrt{5}-1}{2}$.

Step 2: If $f(x_1) < f(x_2)$, then set $b = x_2$; else set $a = x_1$.

Step 3: Set $h = b - a$, $x_1 = b - \gamma h$, and $x_2 = a + \gamma h$. (It is easily seen that only one of x_1 and x_2 is a new value.) If $x_2 - x_1 <$ "tolerance", then stop; else go to Step 2.

Solution: The last computed value of x (x_1 or x_2) is the value c at which f attains its minimum.

Case Examples

1. The first example is from Kapur & Lamberson's Reliability in Engineering Design. The model is a 3-parameter Weibull model $\mu(t) = \left(\frac{t-\delta}{\theta}\right)^\beta$.

   The data set is {22000, 25000, 30000, 33000, 35000, 63000, 104000}. The following estimators were obtained using the two methods:

	Trial and Error Method	Proposed Method
$\hat{\delta}$	19600	20584
$\hat{\beta}$	0.840	0.739
$\hat{\theta}$	44100	26192
R^2	0.9810	0.9864
SSE	0.1244	0.0893

The proposed method gives a "better fit" than the trial and error method.

2. This example uses simulated data from a Weibull model with parameters $\delta = 0,\ \beta = 1,\ \theta = 1$. The data set is {0.0441, 0.3140, 0.3547, 0.4373, 0.8177, 1.0554, 1.2532, 1.4686, 1.6342, 2.0658}. The proposed method gave the following estimates:

$$\hat{\delta} = 0,\ \hat{\beta} = .8360,\ \hat{\theta} = 1.1570,\ R^2 = 0.9196,\ \text{SSE} = 0.7251$$

3. This example uses simulated data from a Weibull model with parameters $\delta = 2,\ \beta = 1.2,\ \theta = 2$. The data set is {2.1951, 2.3140, 2.3390, 2.3519, 2.4708, 2.5637, 2.6144, 2.6330, 2.8768, 2.9590, 3.1709, 3.4930, 3.4995, 3.7692, 3.8461, 4.7498, 5.4524, 6.0464, 6.7954, 6.8012}. The proposed method gave the following estimates:

$$\hat{\delta} = 1.6944,\ \hat{\beta} = 1.3945,\ \hat{\theta} = 2.1790,\ R^2 = 0.8987,\ \text{SSE} = 2.2883.$$

References

1. London, Dick; Survival Models and Their Estimates, Abington, Connecticut, ACTEX Publications 1988.

2. Kapur, K.C. and Lamberson, L.R.; Reliability in Engineering Design, New York, John Wiley and Sons 1977.