New solution and its impact on Increasing, Decreasing and Bathtub shaped failure rate model

Mahmudul Alam
Disaster Prevention Research Institute, RECP
Kyoto University, Gokasho
Uji-shi, Kyoto 611-0011
JAPAN
mmalamstat@gmail.com
m.alam@aw2.ecs.kyoto-u.ac.jp

Abstract

The maximum likelihood equations of IDB distribution can't be solved analytically. Solutions of the MLE equations can be obtained numerically. But the problem is to detect the initial value of the parameters to solve the nonlinear MLE equations. A technique is developed for formulating the initial value of the parameter to solve the MLE equations of IDB distribution. Considering all the facts, though IDB distribution is not suitable for graduating mortality data but the model can graduate mortality data of Bangladesh in the age range of 0-8 very well.

Key words: IDB distribution, Mortality, MLE, Estimation.

Introduction

In general, mortality data shows very high frequency at and near zero age group with a sharp decline within a few years of life, then becoming more or less constant for a couple of years and again rises steadily resulting to a more or less bathtub shaped curve. The exponential power life testing, and increasing, decreasing and bathtub shaped failure rate (IDB) distribution both belonging to the mixed failure rate group have the property to produce bathtub shaped hazards as well as frequency distribution under certain conditions. The primary aim of the study is to fit mortality data from different sources including Bangladesh to IDB model and to verify the degree of fitness.

Fitting probability distributions to data faces problems for estimating parameters when the underlying distribution is IDB. There are several statistical methods that can give estimation of model parameters (method of moments, quantile or percentile, least squares (Zhang et al., 2006), SCEM-UA algorithm (Gong, 2006), etc). However, in general, these methods are not as efficient as the MLE (Zhang and Xie, 2007). Two problems are encountered here with this method. The first is that all the mortality data are grouped into age interval. The second is the choice of initial solution. The pattern of maximum likelihood function and maximum likelihood estimation are different with group data from that of ungrouped one (Lawless, 1982) as can be seen in the Methodology and materials section. Improper selection of the initial values may not help solving the iterative nature of likelihood parameters. Aim of the study is to search for an appropriate primary initial solution to solve the maximum likelihood equations for group data.

There are a number of studies dealing with models for bathtub-shaped failure rate. For example, Gaver and Acar (1979) proposed a model that has a broad range of flexibility-constant hazard to bathtub hazard under different conditions, a flexible bathtub hazard model for non-repairable systems with uncensored data (Jaisingh *et al*., 1987), a lifetime distribution with an upside-down bathtub-shaped hazard function (Dimitrakopoulou *et al*., 2007) and so on. Mudholkar and Srivastava (1993) introduced an exponentiated Weibull distribution. Xie and Lai (1996) gave another additive model with bathtub-shaped failure rate. The parameters of this model can be estimated using graphical method. An additive Burr XII model of four parameters is studied in Wang (2000). There are also several other studies that investigated the combination of two Weibull distributions for bathtub shape failure rate function, such as the competing risk model, multiplicate model, and sectional model given in Chen (2000). Graphical methods and representations of the mixed Weibull distributions were discussed in Jiang and Murthy (1995, 1999), Jiang and Kececioglu (1992) and recently Lai, et.al (2001).

Increasing, Decreasing, Constant and Bathtub-shaped failure rate distribution (IDB)

In a study of how estimation errors and model assumptions affect the costs in the optimal preventive maintenance problem, it soon became evident that for this purpose the established distributions were not always feasible. Distributions with one or two parameters like Weibull distribution impose very strong restrictions on the data. This is well illustrated by their inability to produce bathtub curves. On the other hand, more flexible distributions usually have five or more parameters (Jaisingh *et al*., 1987), which seem to make the study of estimation and optimization from small samples a rather hopeless numerical task (Hjorth, 1980).

A second idea leading towards the distribution was the lack of physical motivation for various bathtub models. While the Weibull distribution, the extreme value distribution, the normal distribution, and the whole class of IFR (increasing failure rate) distributions have some sort of physical motivations, the burn in argument of bathtub does not seem to have much relevance in many situations where such distributions arise. The reason for this seems to be that a probabilistic selection argument, and not a physical one, is the most relevant explanation of the bathtub curve. For practical interest with lifetime data subject to increasing age, an increasing failure rate of each individual was the natural starting point. These arguments, together with considerations of mathematical simplicity, led us to search for a new distribution, and with capacity to also describe bathtub curves. The distribution to be studied is defined by the survival function

$$S(t) = (1+\tfrac{t}{\lambda})^{-\beta} e^{\frac{\delta t^2}{2}} \qquad (1)$$

The hazard function can be written as

$$h(t) = \frac{\beta}{t+\lambda} + \delta t = \frac{\theta}{1+\gamma t} + \delta t \qquad (2)$$

where, $\gamma = 1/\lambda$ and $\theta = \beta/\lambda$ and $\beta, \lambda, \delta > 0$

Special cases of the IDB distribution (relation to other distribution):

- $\theta = 0$; the Rayleigh distribution
- $\delta = \gamma = 0$; the Exponential distribution
- $\delta = 0$; decreasing failure rate
- $\delta \geq \theta\gamma$; increasing failure rate
- $0<\delta< \theta\gamma$; bathtub curve

Methodology and materials

Primary aim of this research is to estimate the parameters of the desire distribution. The genesis and the method of estimation involved in the research is the subject matter of methodology, while a discussion about the data and the computing aids constitutes the materials. The product limit estimate (Kaplan and Meier, 1958) of the survival function S(t) is used to estimate the survival probabilities for grouped data. For models in which some transformation of the survivor function is linear in the parameters, least squares estimation can be used to estimate the parameters. But the survival function of IDB distribution

$$S(t) = \left(1+\frac{t}{\lambda}\right)^{-\beta} e^{-\frac{\delta t^2}{2}} \qquad (3)$$

which is non linear and least squares method may not be used to estimate the parameters. Since there exists no censoring expect the last interval, so the likelihood function can be written as (Lawless, 1982);

$$L(\theta|\underline{t}) = \prod_{j=1}^{k+1} \left[S(t_{j-1}) - S(t_j)\right]^{d_j} \qquad (4)$$

Putting the value of S(t) and taking log on both sides, we have

$$\text{Log}\, L(\theta|\underline{t}) = \sum_{j=1}^{k+1} d_j \log\left[\left(1+\frac{t_{j-1}}{\lambda}\right)^{-\beta} e^{-\frac{\delta t_{j-1}^2}{2}} - \left(1+\frac{t_j}{\lambda}\right)^{-\beta} e^{-\frac{\delta t_j^2}{2}}\right] \qquad (5)$$

Differentiating eq.(5) with respect to λ, β and δ and setting to zero we get three non linear equations which are very difficult to solve analytically. The MLE equations are as follows

$$f = \sum_{j=1}^{k+1} \frac{d_j\left[\left(1+\frac{t_{j-1}}{\lambda}\right)^{-\beta-1} e^{-\frac{\delta t_{j-1}^2}{2}} t_{j-1} - \left(1+\frac{t_j}{\lambda}\right)^{-\beta-1} e^{-\frac{\delta t_j^2}{2}} t_j\right]}{\left[\left(1+\frac{t_{j-1}}{\lambda}\right)^{-\beta} e^{-\frac{\delta t_{j-1}^2}{2}} - \left(1+\frac{t_j}{\lambda}\right)^{-\beta} e^{-\frac{\delta t_j^2}{2}}\right]} = 0 \qquad (6)$$

$$g=\sum_{j=1}^{k+1}\frac{d_j\left[\left(1+\frac{t_{j-1}}{\lambda}\right)^{-\beta}e^{-\frac{\delta t_{j-1}^2}{2}}\log\left(1+\frac{t_{j-1}}{\lambda}\right)-\left(1+\frac{t_j}{\lambda}\right)^{-\beta}e^{-\frac{\delta t_j^2}{2}}\log\left(1+\frac{t_j}{\lambda}\right)\right]}{\left[\left(1+\frac{t_{j-1}}{\lambda}\right)^{-\beta}e^{-\frac{\delta t_{j-1}^2}{2}}-\left(1+\frac{t_j}{\lambda}\right)^{-\beta}e^{-\frac{\delta t_j^2}{2}}\right]}=0 \tag{7}$$

$$h=\sum_{j=1}^{k+1}\frac{d_j\left[\left(1+\frac{t_{j-1}}{\lambda}\right)^{-\beta}e^{-\frac{\delta t_{j-1}^2}{2}}t_{j-1}^2-\left(1+\frac{t_j}{\lambda}\right)^{-\beta}e^{-\frac{\delta t_j^2}{2}}t_j^2\right]}{\left[\left(1+\frac{t_{j-1}}{\lambda}\right)^{-\beta}e^{-\frac{\delta t_{j-1}^2}{2}}-\left(1+\frac{t_j}{\lambda}\right)^{-\beta}e^{-\frac{\delta t_j^2}{2}}\right]}=0 \tag{8}$$

The survival or the cumulative hazard function of IDB distribution can't be expressed in a linear form and therefore, obtaining initial value of the parameters to solve the maximum likelihood equations (6-8) is difficult. A simple model is deduced from Kaplam-Meier non parametric method (Lawless, 1982) to search the initial value of the parameters of IDB distribution as

$$\left(1+\frac{t_2}{\lambda}\right)^{c_1-\frac{c_3t_3^{\,2}}{t_1^{\,2}}}-\left(1+\frac{t_3}{\lambda}\right)^{\frac{c_1t_2^{\,2}}{t_1^{\,2}}-c_2}\left(1+\frac{t_1}{\lambda}\right)^{\frac{t_3^{\,2}(1+c_1-\frac{c_3t_3^{\,2}}{t_1^{\,2}})}{t_2^{\,2}}}=0 \tag{9}$$

Where $c_i = S(t_i)$, i = 1, 2, 3. S(t) is computed according to Kaplan-Meier method. Using the positive root of λ, the later equations can be solved.

$$\beta=\frac{\frac{c_1t_2^{\,2}}{t_1^{\,2}}-c_2}{\log\left\{\frac{1+\frac{t_2}{\lambda}}{(1+\frac{t_1}{\lambda})^{\frac{t_2^{\,2}}{t_1^{\,2}}}}\right\}} \tag{10}$$

and

$$\delta=\frac{2}{t_1^{\,2}}\left[\frac{\log(1+\frac{t_1}{\lambda})^{c_2-\frac{c_1t_2^{\,2}}{t_1^{\,2}}}}{\log\left\{\frac{1+\frac{t_2}{\lambda}}{(1+\frac{t_1}{\lambda})^{\frac{t_2^{\,2}}{t_1^{\,2}}}}\right\}}-c_1\right] \tag{11}$$

Solution of the equations (9-11) for λ, β and δ are known as the initial value of the parameter of the MLE equations. Mortality data of Japan, Canada and Bangladesh have used for fitting the IDB model. In every case, data is taken from secondary sources (Statistical Year Book and Demographic Year book). It is to be noted that these data sets are not directly observed age-specific death figures rather, these are synthesized from abridged life tables (Table 1-2).

Table 1: Observed and expected mortality indices of Bangladesh

Age Group	Observed Death	Expected Death
0	117	177
1	20	17
2	13	11
3	10	8
4	6	7
5-9	12	34
10-14	6	37
15-19	8	43
20-24	11	49
25-29	12	54
30-34	12	56
35-39	14	58
40-44	21	57
45-49	24	55
50-54	41	52
55-59	63	47
60-64	78	42
65-69	123	37
70-74	114	32
75+	295	125

Table 2: Observed and expected mortality indices of Canada and Japan

Age Group	Canada		Japan	
	Observed Death	Expected Death	Observed Death	Expected Death
1	1410	1925	9969	11503
1-4	292	524	3637	4442
5-9	187	1336	2326	12000
10-14	200	2127	1717	19256
15-19	455	2856	4102	25949
20-24	476	3494	4627	31829
25-29	585	4022	5262	36727
30-34	633	4426	8863	40531
35-39	832	4701	10236	43187
40-44	1091	4848	15438	44695
45-49	1488	4874	25069	45106
50-54	2492	4789	34761	44511
55-59	3851	4609	40866	43037
60-64	5052	4350	46916	40833
65-69	7022	4033	68729	38062
70-74	8838	3675	97355	34887
75-79	10347	3293	114097	31467
80-84	11389	2905	113553	27946
85+	19719	13572	104360	135915

Result and discussion

The MLE equations of IDB distribution is very complex and difficult to solve analytically. To overcome the situation a simple technique is proposed to find out the initial value of the parameters of MLE. As a result, the positive real roots of λ can be found by solving eq.(9) and after that the value of β and δ is obtained by solving equations (10-11). These values are known as initial value of the parameters.

The value of λ can be found by solving eq.(6) using the initial values of β and δ. After fixing the value of λ (new value) and δ (initial value), the value of β can be found by solving eq.(7). Lastly, fixing the values of β (new value) and λ (new value) then the value of δ can be obtained by solving eq.(8). Continue the process until the convergence (less than 10^{-07}) is obtained. Sometimes if the equation for λ is chosen first, after certain number of iterations, the value of t/λ may becomes <-1 and then the MLE equations can't be solved. To avoid this situation better to choose either β or δ first and then solve for λ. Final estimation of the parameters is shown in table 3.

Table 3: MLE estimation of the parameters and its significance

Source	**Parameter**			**Chi-square**
	λ	**β**	**δ**	
Bangladesh	0.0014183789	0.029576646	0.00062557	936
Japan	6.975279E-13	0.00057389	0.00045322	796920
Canada	3.753157E-13	0.000884407	0.0004701	75713

This process is easy to find out the initial value of the parameters for MLE if we don't have any idea about the approximate value of the parameters. As for example, if we choose the initial value for λ= 100, β= 3.0 and δ= 1.5 to solve equations (6-8), it takes a lot of time to reach the real value than that of the initial value chosen under equations (10-12) as λ= 40.017, β= 0.7178 and δ= 0.0005 for Bangladeshi data. Comparing these two initial value data sets, the iteration time for second set takes less time at least 30% than first set to solve the maximum likelihood equations until the final approximated value reach to λ= 0.0014183, β= 0.02957665 and δ = 0.0006255701. Five sets of mortality data from Japan, Canada and Bangladesh are used to verify the acceptability of the proposed technique and found almost same results (results of only three data sets are shown in this study). This technique is also applied in Jaisingh's (1987) data set and found similar results.

A careful examination of the results shown in these tables and graphs, it is clear that IDB distribution graduate mortality data of Bangladesh up to age 8 well (Fig. 1) while Gompertz model graduates mortality data beyond the age 15 (Alam, 1996). In case of Japan and Canada, the models exhibit the nature of Gompertz model. In case of Bangladesh, IBD model over estimates mortality in between the ages 10-50 while for Canada and Japan the over estimation is in between the age 08-55 (Fig. 2). In case of Canada, this over estimation is parallel to the observed mortality pattern while for Bangladesh, it is not parallel to the observed mortality pattern and the extent of over estimation is higher than that of Canada and Japan. It may be due to the fact that in case of Bangladesh the fluctuation in the mortality pattern is much more prominent than that of Japan and Canada. The extent of under estimation due to the model beyond the age range of over estimation is much more prominent than the over estimation range.

Two sets of data is used in the study of Dimitrakopoulou et al., (2007). Depending on the nature of the data sets, they have shown that the estimated hazard function may not always produce bathtub shape curve. Second data set of their study follows bathtub shaped hazard pattern like this work for Bangladeshi mortality data (Fig. 3) and first set produces unimodal pattern like Japanese and Canadian mortality data (Fig. 4).

Conclusion

To reach the initial solution of MLE equations for IDB distribution, our proposed technique may take less time to solve the MLE equations. The model used here for graduating mortality data is not found suitable, although the model fit the data of Bangladesh in the age range of 0-8 well. Development of a new model with the combination of increasing failure rate models and this model, may graduate the mortality data of Bangladesh to a better extent.

We are unable to make any comments regarding the efficiency of the estimation process in estimating parameters of the models studied. Because, the information used for estimating parameters are not from the respective distributions. A study of the efficiency of the estimation models as well as the initial solution may be taken as a problem for the future studies.

Reference

1. C.D. Lai, M. Xie and D.N.P. Murthy, Bathtub-shaped failure rate life distributions. In: *Handbook of statisticsAdvances in reliability,* 20, 69–104, Elsevier, London, 2001.
2. D.P. Gaver and M. Acar, Analytical hazard representations for use in reliability, mortality and simulation studies. *Com.in statist., B(Sim.& Com.)*, 8, 91-111, 1979.
3. E.L. Kaplan and P. Meier, Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53 (282), 457-481, 1958.
4. F.K. Wang, A new model with bathtub-shaped failure rate using an additive Burr XII distribution. *Reliab Engng Syst Safety,* 70, 305–312, 2000.
5. G.S. Mudholkar and D.K. Srivastava, Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Trans Reliab,* 42, 299–302, 1993.
6. J.F. Lawless, Statistical models and methods for lifetime data. John Wiley & Sons, New York, 1982.
7. L.F. Zhang, M. Xie, and L.C. Tang, Bias correction for the least squares estimator of Weibull shape parameter with complete and censored data. *Reliability Engineering &System Safety,* 91, 930–939, 2006.

8. L.R. Jaisingh, W.J. Kolarik and D.K. Dey, A flexible bathtub hazard model for non-repairable systems with uncensored data. *Microelectron. Reliab*, 27(1), 87-103, 1987.

9. M. Xie and C.D. Lai, Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. *Reliab Engng Syst Safety,* 52, 87–93, 1996.

10. M.M. Alam, A study of some mixed failure rate models in graduating mortality data. Unpublished M.Sc thesis, 1996.

11. R. Jiang and D.N.P. Murthy, Exponentiated Weibull family: a graphical approach. *IEEE Trans Reliab* 47, 68–72, 1999.

12. R. Jiang and D.N.P. Murthy, Modeling failure-data by mixture of 2 Weibull distributions: a graphical approach. *IEEE Trans Reliab,* 44, 477–488, 1995.

13. S. Jiang and D. Kececioglu, Graphical representation of two mixed Weibull distributions. *IEEE Trans Reliab* 41, 241–247, 1992.

14. S. Rajarshi and M.B. Rajarshi, Bathtub distributions—a review. *Commun Stat—Theory Meth,* 17, 2597–2621, 1988.

15. T. Dimitrakopoulou, K. Adamidis, and S. Loukas, A lifetime distribution with an upside-down bathtub-shaped hazard function. *IEEE Transactions on Reliability*, 56(2), 308-311, 2007.

16. T. Zhang and M. Xie, Failure data analysis with extended Weibull distribution. *Communications in Statistics—Simulation and Computation,* 36, 579–592, 2007.

17. U. Hjorth, A reliability distribution with increasing, decreasing, constant and bathtub-shaped failure rates. *Technometrics* 22, 99–107, 1980.

18. Z. Chen, A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Stat Probab Lett*, 49, 155–161, (2000).

19. Z.J. Gong, Estimation of mixed Weibull distribution parameters using the SCEMUA algorithm: application and comparison with MLE in automotive reliability analysis. *Reliability Engineering & System Safety,* 91, 915–922, 2006.

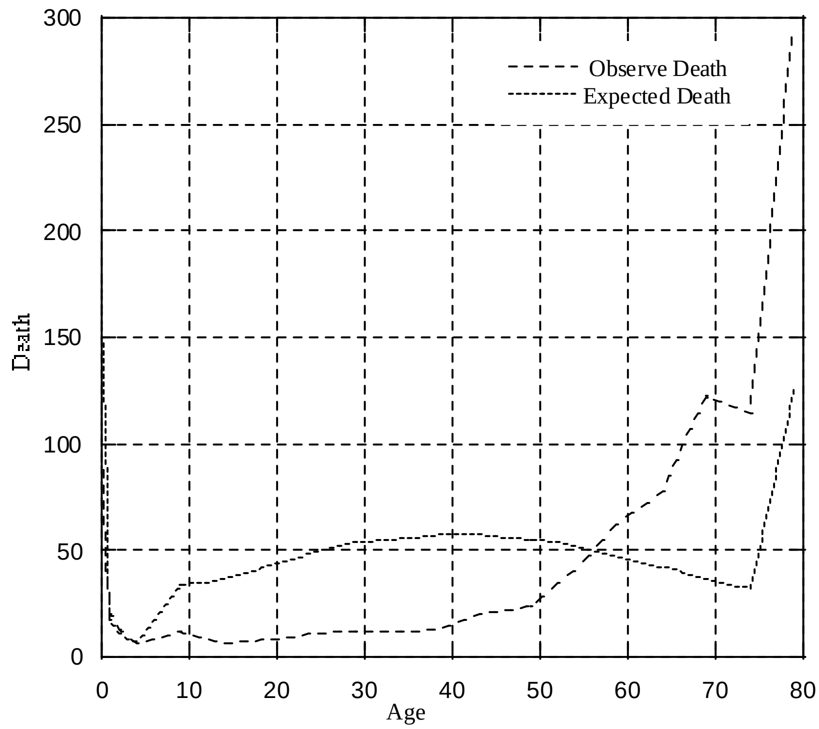

Figure 1: Representation of observe and expected death for Bangladesh

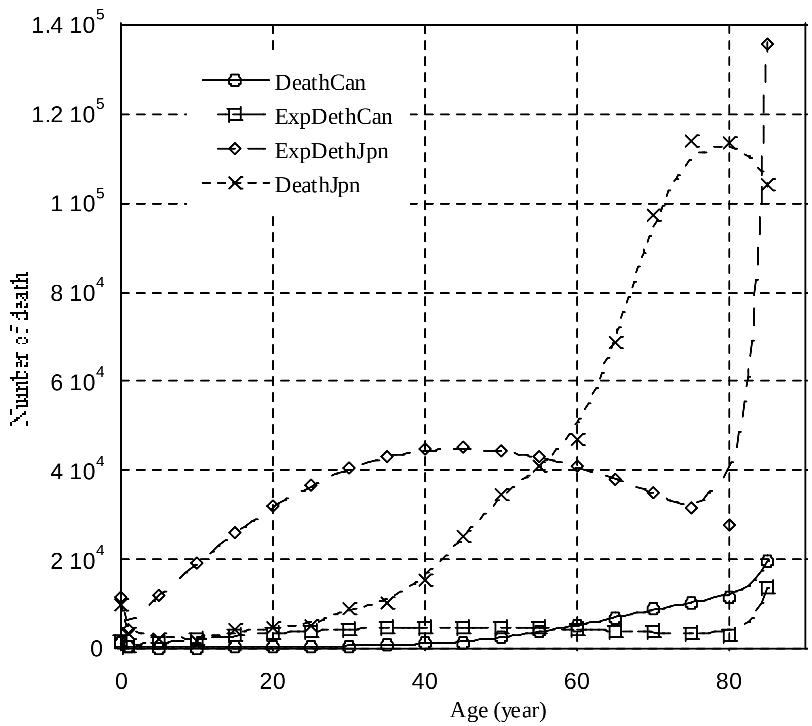

Figure 2: Observe Vs expected death for Canada and Japanese mortality data (DeathCan: Observed death; ExpDethCan: expected death for Canada and DeathJpn: Observed death; ExpDethJpn: expected death for Japan).

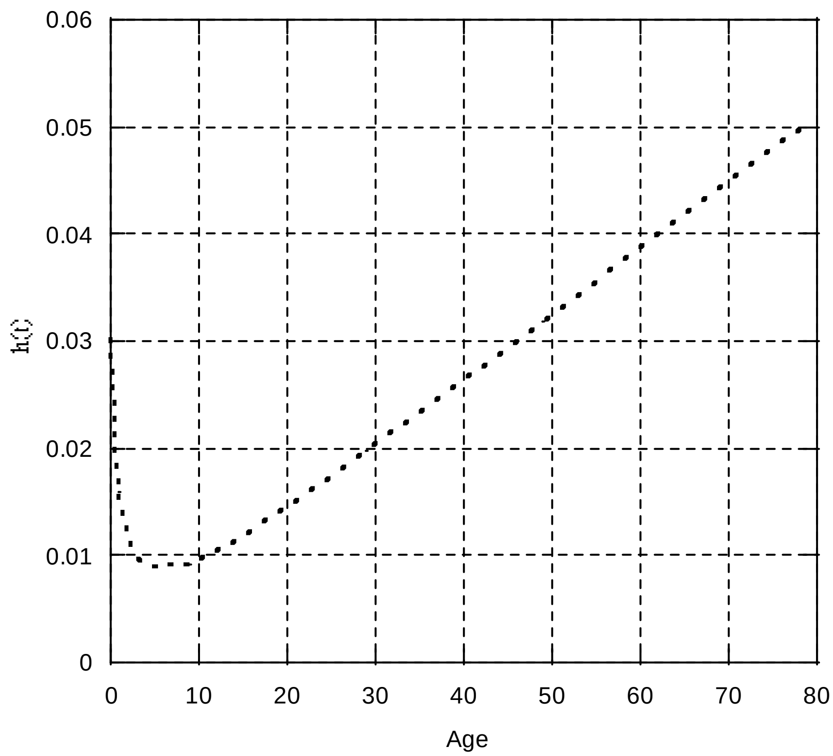

Figure 3: Hazard plot for Bangladesh.

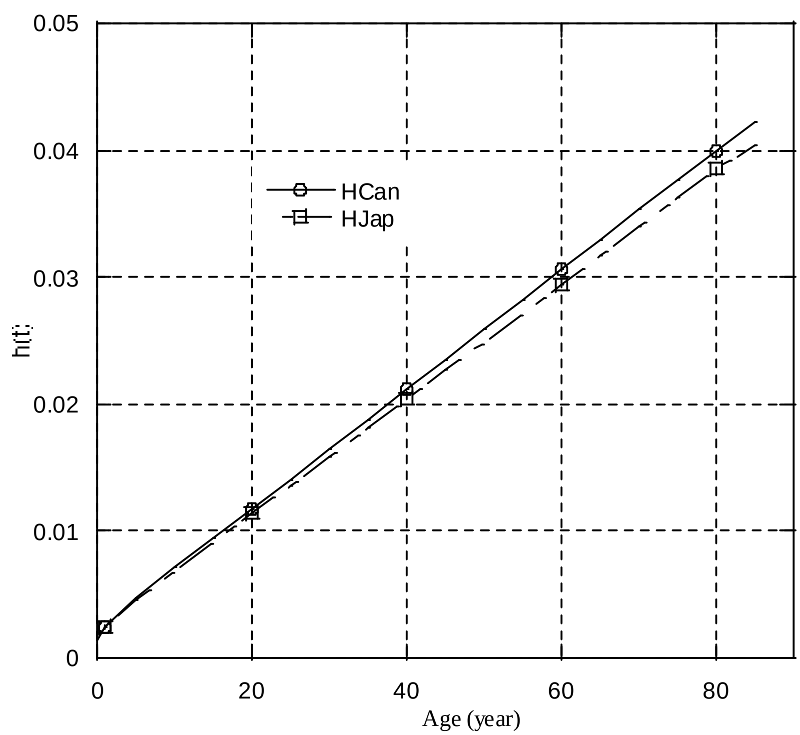

Figure 4: Hazard plot for Canada and Japan (HCan: hazard plot for Canadian data; HJap: hazard plot for Japan).