Journal of Data Science **7**(2009), 487-495

A Fractional Survival Model

Cheng K. Lee[1] and Jenq-Daw Lee[2]
[1]*Bank of America and* [2]*National Cheng Kung University*

Abstract: A survival model is derived from the exponential function using the concept of fractional differentiation. The hazard function of the proposed model generates various shapes of curves including increasing, increasing-constant-increasing, increasing-decreasing-increasing, and so-called bathtub hazard curve. The model also contains a parameter that is the maximum of the survival time.

Key words: Bathtub hazard, confluent hypergeometric function, survival function.

1. Introduction

In survival analysis, some well known functions are applied to explain the hazard curve of increasing (Weibull with shape parameter > 1), decreasing (Weibull with shape parameter < 1) and increasing-decreasing (log-logistic and lognormal) properties. Numerous articles have also proposed models with decreasing-constant-increasing or so-called bathtub hazard curves. To name some them on bathtub hazard curve, Xie and Lai (1995) proposed an additive Weibull model. Chen (2000) proposed a two-parameter lifetime model. Wang (2000) proposed an additive Burr XII distribution. A function with various shapes of hazard curves is always desirable in survival analysis. In this paper, we proposed a four-parameter survival function derived from the exponential function. The proposed model has increasing, increasing-constant-increasing, increasing-decreasing-increasing, and bathtub hazard curve. Moreover, the function includes a parameter that the value of the random variable often known as the survival time cannot exceed. In some special cases, the model can be reduced to a two-parameter function. For applications, two real world data sets are fitted wit the proposed model.

2. The Model

The one-parameter exponential distribution has been applied on many fields. The common form of the probability density function (PDF) is $f(t) = \mu e^{-\mu t}$, where $\mu > 0$ and $t \geq$. The cumulative density function (CDF) is $Z(t) = 1 - e^{-\mu t}$. The mean and the variance of this distribution are, respectively, μ^{-1} and μ^{-2}. The hazard function is equal to the first derivative of the CDF that is PDF divided by the survival function which is equal to 1 minus the CDF also known as the mortality function in survival analysis. Therefore, the hazard function of the one-parameter exponential distribution is μ, a constant. A constant hazard rate does not describe all the observed phenomena in many fields. It is desirable to have a model with a non-constant hazard function that has flexible shapes of the hazard curve.

To begin with, let $DZ(t)$ denote the first derivative of the CDF of the one-parameter exponential function, then the hazard function is $DZ(t)/(1 - Z(t)) = \mu$. Many works have been done to derive models with various hazard rates. Instead of varying the hazard rates, we here employ the idea of Stiassnie (1979) who used a model with an arbitrary order of differentiation to explain the dynamics of viscoelastic materials. Moreover, the arbitrary order is not necessarily an integer. Therefore, we may choose to take the derivative to some arbitrary order of the CDF to be given by $D_t^\lambda Z(t) = \mu e^{-\mu t}$, where $D_t^\lambda Z(t)$ denotes the differentiation of the arbitrary order λ with respect to t. Using Cauchy formula for repeated integration to solve for $Z(t)$ (see Appendix A1), we have

$$Z(t) = \frac{\mu t^\lambda}{\Gamma(\lambda + 1)} {}_1F_1[1; \lambda + 1; -\mu t],$$

Where $\Gamma(\cdot)$ is the Gamma function and ${}_1F_1[a; b; c]$ is the confluent hypergeom4etric function (Gurland, 1958; Muller, 2001) with 3 arguments. However, $Z(t)$ is just the incomplete Gamma distribution, a special case of the confluent hypergeometric function (Luke, 1959). To make the model more general, the first argument of $Z(t)$ is substituted by a. Then,

$$Z(t) = \frac{\mu t^\lambda}{\Gamma(\lambda + 1)} {}_1F_1[a; \lambda + 1; -\mu t].$$

In order for $Z(t)$ to be interpreted as a PDF, it is necessary that, for $t \leq T$, $Z(T) = 1$. After the normalization (see Appendix A2), we obtain the fractional mortality function

$$F(t) = \left(\frac{t}{T}\right) \frac{{}_1F_1[a; \lambda + 1; -\mu t]}{{}_1F_1[a; \lambda + 1; -\mu T]} \tag{2.1}$$

The fractional survival function $S(t)$ is, therefore, $1 - F(t)$. The negative sign in the third argument in equation (2.2) can be eliminated using Kummer's formula (Gurland, 1958) such that

$$F(t) = e^{\mu(T-t)} \left(\frac{t}{T}\right)^{\lambda} \frac{{}_1F_1[a; \lambda+1; -\mu t]}{{}_1F_1[a; \lambda+1; -\mu T]} \tag{2.2}$$

Taking the first derivative of $F(t)$ in equation (2.1) and using the differential formula of the confluent hypergeometric function (Abramowitz & Stegun, 1972), the probability density function is

$$f(t) = \frac{\lambda}{T} \left(\frac{t}{T}\right)^{\lambda-1} \frac{{}_1F_1[a; \lambda; -\mu t]}{{}_1F_1[a; \lambda; -\mu T]}. \tag{2.3}$$

Or, using Kummer's formula, the PDF is

$$f(t) = e^{\mu(T-t)} \left(\frac{\lambda}{T}\right)^{\lambda-1} \frac{{}_1F_1[\lambda - a; \lambda; \mu t]}{{}_1F_1[\lambda - a + 1; \lambda+1; \mu T]} \tag{2.4}$$

Due to some properties of the confluent hypergeometric function (Muller 2001), when $a = 1$, Equation (2.3) becomes $\lambda t^{\lambda-1}/T^{\lambda}$ and when $a = \lambda$, Equation (2.3) becomes $\mu^{\lambda} t^{\lambda-1} \exp(-\mu t)/\gamma(\mu T; \lambda)$ where γ is the incomplete Gamma function.

The confluent hypergeometric function can be represented by a series or an integral expression (Gurlan, 1958; Muller 2001). Either expression has its own restrictions on the values of the first and the second arguments. In this article, the confluent hypergeometric function is numerically evaluated by Muller's (20001) algorithm based on the series expression in which the second argument cannot be zero or a negative integer. Applying the restrictions to our fractional survival model, λ cannot be zero or a negative integer, and $0 \le t \le T$.

The hazard function for this fractional survival model provides various shapes of curves (Figure 1) including increasing-constant-increasing (hazard 1 with $\alpha = 3, \lambda = 3, \mu = 1.5, T = 20$), decreasing-constant-increasing or so-called bathtub hazard curve (hazard 2 with $\alpha = 0.01, \lambda = 0.01, \mu = 0.7, T = 20$), increasing-decreasing-increasing (hazard 3 with $\alpha = 3.4, \lambda = 3.5, \mu = 2, T = 20$) and increasing (hazard 4 with $\alpha = -1, \lambda = 11, \mu = 2, T = 20$) hazard rates. However, the mean and the variance of the model do not exist (see Appendix A3).

3. Application

Case 1

We use the data of Marriage History File 1985 – 2003 of the Panel Study of Income Dynamics (PSID) to fit the model on time to the first marriage. The sample was censored in 2003 with 17338 married individuals and 6029 unmarried. The average year to the first marriage was 24.01. The data is available at http://simba.isr.umich.edu. The maximum likelihood estimates and the standard errors based on the second derivatives valued at the maximized log-likelihood function are in Table 1.

Table 1: Parameter Estimates of Time to First Marriage

Parameter	Estimate	Standard Error
α	34.025	0.551
λ	34.094	0.026
μ	1.483	0.552
T	94.512	0.487

The increasing-decreasing-increasing hazard curve of the fitted model is in figure 2 in which the peak of the hazard rates occurs around age 28.

Case 2

The data in this application contains survival times in month from 5880 patients after they received coronary artery bypass grafting (CABG). Among these patients, 545 died during the study. The average survival time to death after receiving the procedure was 47 months. The data is available at www.clevelandclinic.org/heartcenter/hazard/default.htm. The maximum likelihood estimates and the standard errors based on the second derivatives valued at the maximized log-likelihood function are in Table 2.

Table 2: Parameter Estimates of Time to Death after Receiving CABG

Parameter	Estimate	Standard Error
Parameter	Estimate	Standard Error
α	0.396	0.129
λ	0.302	0.023
μ	-0.021	0.003
T	225.265	8.639

The decreasing-constant-increasing hazard curve of the fitted model is in figure 3. The hazard is known as the bathtub curve which is also the three phase hazard function described by Sergeant, Blackstone and Meyns (1997) who analyzed CABG data as well.

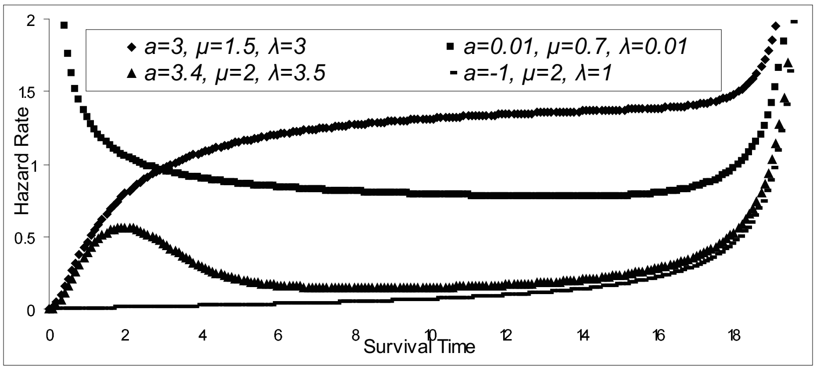

Figure 1: Hazard curves with various sets of parameters

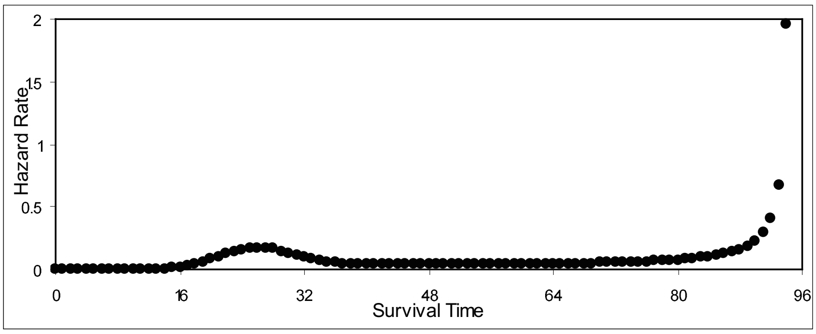

Figure 2: Hazard rate of the proposed model on time to first marriage

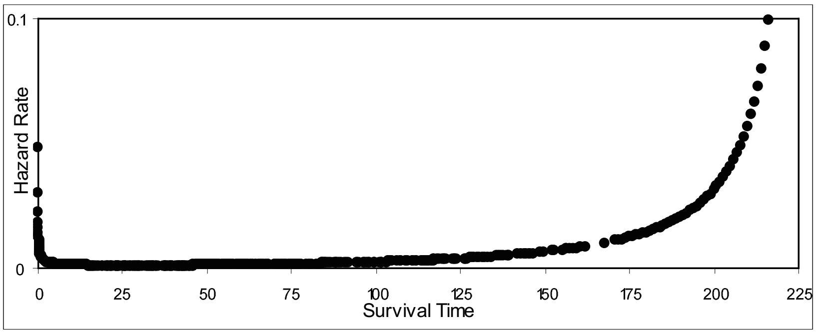

Figure 3: Hazard rate of the proposed model on survival time after CABG

Discussion

In this study, we proposed a survival function with flexible hazard curves with the parameter T that can be regarded as the maximum survival time. In case

1, the estimated T is 94.512 which suggests that the maximum age to the first marriage be 94.512. In case 2, the estimated T suggests that the maximum survival month after CABG be 225.265. In both cases, no covariates are considered. For future studies, we suggest that T can be a function of covariates so that each individual would have its own maximum survival time.

Acknowledgement

This work is part of the first author's master thesis finished at the University of Alabama at Birmingham in 1995 under the supervision of Professor Emeritus Malcolm E. Turner.

Appendix A1

Solving for $Z(t)$ in Equation (2.1)

$$\begin{aligned}
&Z(t)\\
&= Dt^{-\lambda}\{D_t^{\lambda}Z(t)\} = D_t^{-\lambda}\mu e^{-\mu t} = \mu D_t^{-\lambda}e^{-\mu t}\\
&= \frac{\mu}{\Gamma(\lambda)}\int_0^1 (t-y)^{\lambda-1}E^{\mu y}dy, \qquad \text{using Cauchy formula for repeated integration}\\
&= \frac{\mu}{\Gamma(\lambda})\int_0^1 (t-ts)^{\lambda-1}e^{-\mu ts}(tds), \quad \text{let } s = y/t\\
&= \frac{\mu t^{\lambda}}{\Gamma(\lambda)}\int_0^1 (1-s)^{\lambda-1}e^{-\mu s}ds
\end{aligned}$$

The confluent hypergeometric (Gurland, 1958) is defined by series as

$${}_1F_1(a;c,z) = \sum_{k=0}^{\infty}\frac{(a)_k z^k}{(c)_k k!}, \quad \text{where } (a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}.$$

It converges for all real values of a, c and z, and c cannot be a negative integer or zero. The confluent hypergeoemtric function can also be presented as an integral form (Gurland, 1958)

$${}_1F_1(a;c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 e^{zt}t^{a-1}(1-t)^{c-a-a}dt, \quad (c > a > 0).$$

Then,

$$\int_0^1 (1-s)^{\lambda-1}e^{-\mu s}ds = \frac{{}_1F_1[1;\lambda+1;-\mu t]}{\frac{\Gamma(\lambda+1)}{\Gamma(1)\Gamma(\lambda)}} = \frac{{}_1F_1[1;\lambda+1;-\mu t]}{\lambda}.$$

Thus,

$$D_t^{-\lambda}\mu e^{-\mu t} = \frac{\mu t^\lambda}{\Gamma(\lambda)}{}_1F_1[1;\lambda+1;-\mu t]$$

Therefore, $Z(t) = \mu t^\lambda {}_1F_1[1;\lambda+1;-\mu t]/\Gamma(\lambda+1)$.

Appendix A2. Normalization of $Z(t)$

Let $Z(t) = g(a,\mu,\lambda,t)t^\lambda {}_1F_1[a;\lambda+1;-\mu T]$, where $g(a,\mu,\lambda,T)$ is a function of a,μ,λ, and t. The argument t is the normalizing constant. Then, for a constant T $(t \leq T)$,

$$Z(t) = g(a,\mu,\lambda,T)T^\lambda {}_1F_1[a;\lambda+1;-\mu T].$$

To normalize, for all T,

$$Z(T) = 1 = g(a,\mu,\lambda,T)T^\lambda {}_1F_1[a;\lambda+1;-\mu T].$$

Then,

$$g(a,\mu,\lambda,T) = \{T^\lambda {}_1F_1[a;\lambda+1;-\mu T]\}^{-1}.$$

Substituting into $Z(t)$, we obtain

$$Z(t) = \left(\frac{t}{T}\right)^\lambda \frac{{}_1F_1[a;\lambda+1;-\mu t]}{{}_1F_1[a;\lambda+1;-\mu T]}.$$

Appendix A3. Moment Generating Function

The moment generating function using the PDF in equation (2.3) is

$$M_t(s) = \int_0^\infty e^{st}\frac{\lambda}{T}\left(\frac{t}{T}\right)^{\lambda-1}\frac{{}_1F_1[a;\lambda;-\mu t]}{{}_1F_1[a;\lambda;-\mu T]} = \frac{\Gamma(\lambda+1)(-s)^{a-\lambda}(\mu-s)^{-a}}{T^\lambda {}_1F_1[a;\lambda;-\mu T]}$$

The first derivative of the moment generating function with respect to s is

$$M_t'(s) = \frac{\Gamma(\lambda+1)(-s)^{a-\lambda-1}(\mu-s)^{-a-1}(\mu\lambda-\mu a-\lambda s)}{T^\lambda {}_1F_1[a;\lambda;-\mu T]}.$$

The second derivative of the moment generating function with respect to s is

$$M_t''(s) = \frac{\Gamma(\lambda+1)(-s)^{a-\lambda-2}(\mu-s)^{-a-2}A}{T^\lambda {}_1F_1[a;\lambda+1;-\mu T]},$$

where $A = (s^2\lambda(\lambda+1) + 2s\mu(\lambda+1)(a-\lambda) + (a-\lambda)(a-\lambda-1)\mu^2)$.

When the confluent hypergeometric is evaluated under the integral expression, the value of the second argument subtracting the first argument must be positive

(Gurland, 1958). It implies that $a - \lambda$ must be less than 1 in our proposed model. Therefore, the first and the second derivative of the moment generating function are undefined due to raising zero to a negative power. The confluent hypergeometric can also be evaluated under the series expression and, in this case, both derivatives of the moment generating function are 0. It means that the expectation and the variance are both 0 which implies that t is not a random variable. Thus, the mean and the variance of the proposed distribution do not exit.

References

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions*, 10th Ed. Government Printing Office.

Chen, Z. (2000). A New two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters* **3**, 155-161.

Gurland, J. (1958). A Generalized Class of Contagious Distributions. *Biometrics* **14**, 229-249.

Luke, Y. L. (1959). Expansion of the Confluent Hypergeometric Function in Series of Bessel Functions. *Mathematical Tables and Other Aids to Computation* **13**, 261-271.

Muller, K. E. (2001). Computing the Confluent Hypergeometric Function, $M(a, b, x)$. *Numerical Mathematics* **90**, 179-196.

Sergeant, P., Blackstone, E. and Meyns, B. (1997). Validation and interdependence with patient-variables of the influence of procedural variables on early and late survival after CABG. *European Journal of Cardio-thoracic Surgery* **12**, 1-19.

Wang, F. K. (2000). A new model with bathtub-shaped failure rate using an additive Burr XII distribution. *Reliability Engineering and System Safety* **70**, 305-312.

Xie, M. and Lai, C. D. (1995). Reliability analysis using an additive Weibull model with bathtub-shaped failure function. *Reliability Engineering and System Safety* **52**, 87-93.

Received February 8, 2007; accepted May 10, 2008.

Cheng K. Lee
Home Equity Loss Forecasting
Home Loans & Insurance
Bank of America
Charlotte, North Carolina 28244, USA
lee.abcde@gmail.com

Jenq-Daw Lee
Graduate Institute of Political Economy
National Cheng Kung University
Tainan, Taiwan 70101
leewien@gmail.com