

Distribution Function and its Applications in Software Reliability

Hoang Pham^{*}

Department of Industrial and Systems Engineering, Rutgers University, Piscataway, 08854, USA

Abstract

In this paper, we present a new distribution function that characterizes using a Vtub-shaped failure rate function. We discuss the characteristics of the distribution and determine the confidence intervals of the Vtub-shaped failure rate. We also illustrate the proposed distribution function with an application in software reliability modeling by using the new Vtub-shaped failure rate as the time-dependent fault detection rate per fault in a recent generalized software reliability model with the uncertainty of operating environments.

Keywords: failure rate function; Vtub-shaped failure rate; bathtub shape; software reliability model; lifetime distribution; operating environments

(Submitted on November 23, 2017; Revised on February 23, 2018; Accepted on April 9, 2018)

© 2019 Totem Publisher, Inc. All rights reserved.

1. Introduction

The failure rate function is one of the few important quantified measures that can be used to characterize the phenomena of electronic system products and equipment, and it is related to the human mortality rate [1]. The bathtub shape is commonly considered in reliability and statistical data analysis [2]. It consists of three periods: the first period is infant mortality, where the failure rate decreases with time; the second period is known as useful life, where the failure rate is constant; and the third is the wear out period, where the failure rate increases with time. The characteristics of systems with bathtub shaped failure rates are likely difficult to find in practice due to the constant failure rate.

Even though the two-parameter Weibull distribution with shape parameter η and scale parameter φ is commonly used in statistical reliability modeling and survival data analysis, it is commonly assumed that the failure rate itself will always provide a bathtub shape [3]. Unfortunately, this is not the case. For example, when $\eta < 1$, the failure rate function is decreasing and will represent an infant mortality period; when $\eta = 1$, a constant failure rate accounts for a useful life period; and when $\eta > 1$, the failure rate is increasing and known as the wear out period.

It is worth noting that the failure rate functions of most of the common known distributions used in engineering and the applied sciences do not exhibit a bathtub shape nor a Vtub shape [4]. For the Vtub-shaped failure rate $h(x)$, there exists a change point x_0 such that the failure rate $h(x)$, is decreasing in $[0, x_0]$ and slowly increasing, as a Vtub shape, in (x_0, ∞) .

It is also worth noting that there is a slight difference between the Vtub-shaped and bathtub-shaped curves [4]. For the bathtub-shaped curve, for example, the failure rate is constant during its useful life period. For the Vtub-shaped curve, after the infant mortality period the system will begin to experience a rather low increasing failure rate but never be at a constant rate due to aging. The behavior of systems with Vtub shaped failure rates can be observed in many practical applications including human life expectancy, medical devices, and electronic system products.

In this paper, we discuss a new distribution function that exhibits a Vtub shape. We also discuss the characteristics and confidence intervals of the failure rate function. We then illustrate the proposed new density function with an application in

^{*} Corresponding author.

E-mail address: hopham@soe.rutgers.edu

software reliability engineering by considering the new Vtub-shaped failure rate as the time-dependent fault detection rate per fault in a generalized software reliability model with the uncertainty of operating environments.

2. A Distribution Function and Vtub-Shaped Failure Rate

Consider the time to failure random variable X has a probability density function (pdf) $f(x)$ and cumulative distribution function (cdf) $F(x)$. Then, the failure rate function is

$$h(x) = \frac{f(x)}{1 - F(x)}$$

Here, we present a new two-parameter lifetime distribution with parameters a and b . The proposed pdf is as follows:

$$f(x) = \left(ax \ln(bx) + \frac{a}{b} \right) e^{-\left\{ ax \left[\frac{x}{2} \ln(bx) - \frac{x}{4} + \frac{1}{b} \right] \right\}} \quad \text{for } x > 0, a > 0, b > 0 \quad (1)$$

The corresponding reliability function and failure rate of the new distribution are

$$R(x) = e^{-ax \left[\frac{x}{2} \ln(bx) - \frac{x}{4} + \frac{1}{b} \right]} \quad \text{for } x > 0, a > 0, b > 0 \quad (2)$$

and

$$h(x) = \frac{f(x)}{R(x)} = ax \ln(bx) + \frac{a}{b} \quad \text{for } x > 0, a > 0, b > 0 \quad (3)$$

respectively. The failure rate function $h(x)$ above has a Vtub shape.

The failure rate distribution is

$$H(x) = \int_0^x h(s) ds = ax \left[\frac{x}{2} \ln(bx) - \frac{x}{4} + \frac{1}{b} \right] \quad \text{for } x > 0, a > 0, b > 0 \quad (4)$$

Figures 1-4 describe the pdf $f(x)$, reliability function $R(x)$, failure rate $h(x)$, and failure rate distribution function $H(x)$ for various values of b with $a = 0.55$, respectively.

Note that the distribution F is an increasing failure rate (IFR) if the failure rate $h(x)$ is an increasing function of x . Similarly, F is a decreasing failure rate (DFR) if $h(x)$ is a decreasing function of x .

Definition [4] The distribution function F has a Vtub-shaped failure rate if there exists a change point x_0 such that the distribution F is DFR for $x < x_0$ and IFR for $x > x_0$, $h'(x_0) = 0$, and the function $h(x)$ is convex.

Lemma Given $a > 0$ and $b > 0$, the function $h(x)$ in Equation (3) is a convex function.

It can be easily shown that $x \ln(x)$ is a convex function $x > 0$; therefore, the function $h(x) = ax \ln(bx) + \frac{a}{b}$ is convex for $x > 0$, $a > 0$, and $b > 0$.

Theorem Given the values a and b , the distribution F with pdf as given in Equation (1) has a Vtub-shaped failure rate $h(x)$. There exists a point x_0 such that $h'(x_0) = 0$ and $h(x)$ is convex, where

$$x_0 = \frac{1}{eb} = \frac{0.3679}{b} \quad (5)$$

Proof From Equation (3), upon differentiation, we obtain

$$h'(x) = a(\ln(bx) + 1)$$

Set $h'(x) = 0$, and after simplification, we obtain

$$x_0 = \frac{1}{eb} = \frac{0.3679}{b}$$

From the lemma above, since the function $h(x)$ is nonnegative and convex, $h(x)$ initially decreases in x for all $x \leq x_0$ and then increases in x for all $x \geq x_0$.

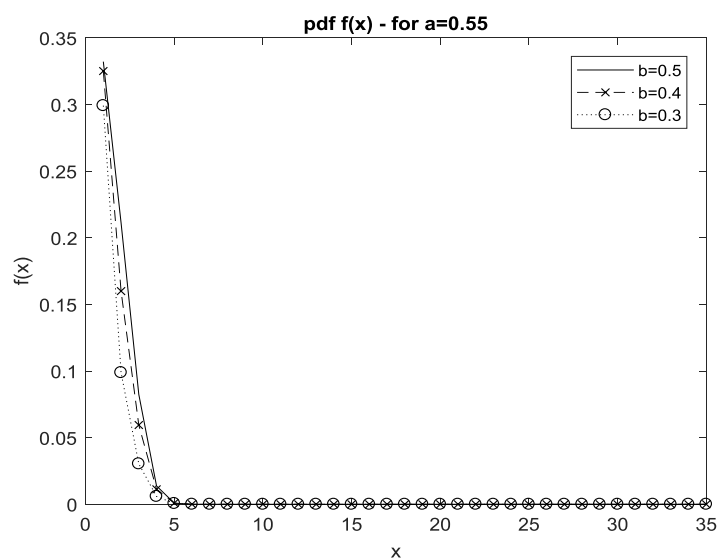


Figure 1. Probability density function with $a = 0.55$

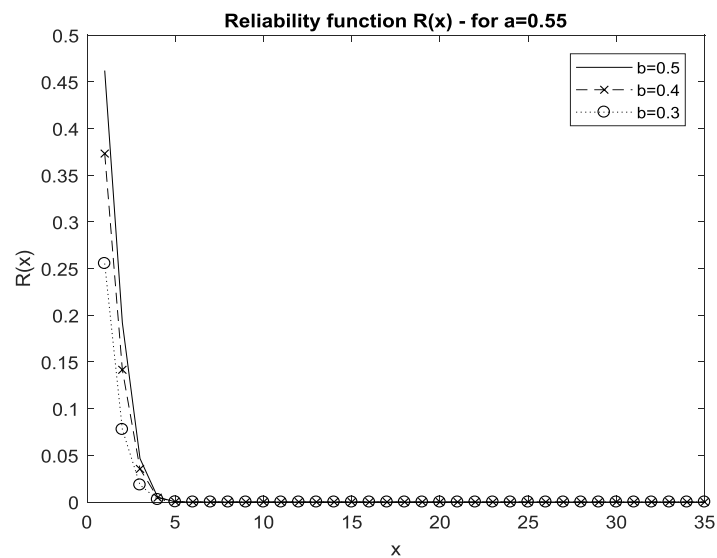
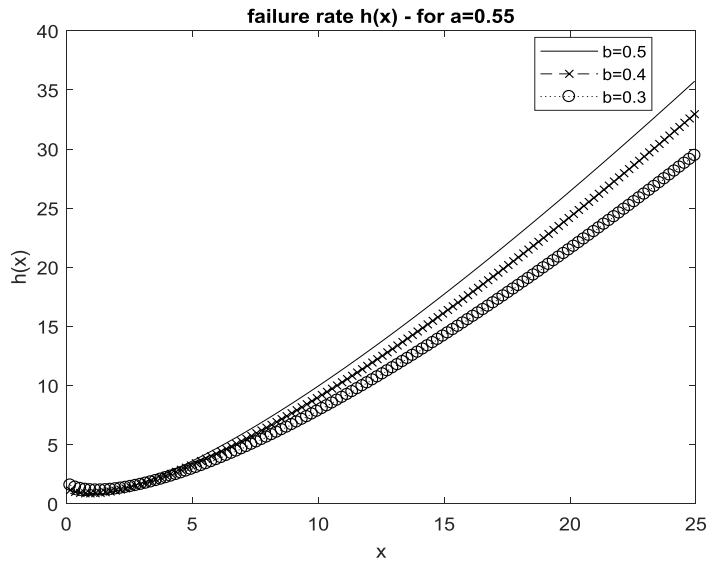
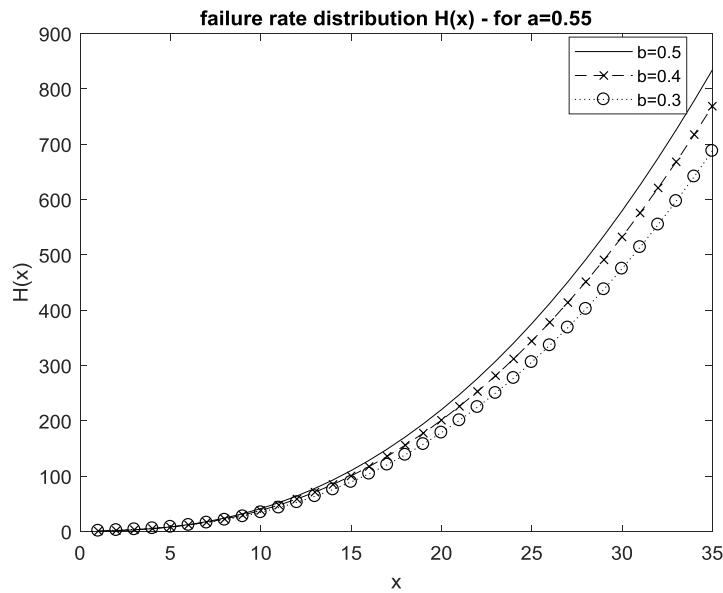


Figure 2. Reliability function with $a = 0.55$

Figure 3. Failure rate function with $a = 0.55$ Figure 4. Failure rate distribution function with $a = 0.55$

3. Parameter Estimation

We now wish to estimate the two unknown parameters a and b in the pdf from Equation (1). Using the maximum likelihood estimation (MLE) method, the likelihood function is

$$L(a, b) = \prod_{i=1}^n \left(ax_i \ln(bx_i) + \frac{a}{b} \right) e^{-a \sum_{i=1}^n x_i \left[\frac{x_i}{2} \ln(bx_i) - \frac{x_i}{4} + \frac{1}{b} \right]}$$

The log likelihood function is

$$\ln L(a, b) = \sum_{i=1}^n \ln \left(ax_i \ln(bx_i) + \frac{a}{b} \right) - a \sum_{i=1}^n x_i \left[\frac{x_i}{2} \ln(bx_i) - \frac{x_i}{4} + \frac{1}{b} \right] \quad (6)$$

The first derivatives of the log likelihood function with respect to a and b are

$$\frac{\partial}{\partial a} \ln L(a, b) = \sum_{i=1}^n \frac{bx_i \ln(bx_i) + 1}{(abx_i \ln(bx_i) + a)} - \sum_{i=1}^n x_i \left[\frac{x_i}{2} \ln(bx_i) - \frac{x_i}{4} + \frac{1}{b} \right] \quad (7)$$

and

$$\frac{\partial}{\partial b} \ln L(a, b) = \frac{1}{b} \sum_{i=1}^n \frac{bx_i - 1}{(bx_i \ln(bx_i) + 1)} - \frac{a}{2b^2} \sum_{i=1}^n x_i (bx_i - 2) \quad (8)$$

respectively. Set Equations (7) and (8) equate to zero, and after rearrangements, the MLE of a and b can be obtained by solving the following equations:

$$\begin{cases} a = \frac{2b \sum_{i=1}^n \frac{bx_i - 1}{(bx_i \ln(bx_i) + 1)}}{\sum_{i=1}^n x_i (bx_i - 2)} \\ \sum_{i=1}^n \frac{bx_i \ln(bx_i) + 1}{abx_i \ln(bx_i) + a} = \sum_{i=1}^n x_i \left[\frac{x_i}{2} \ln(bx_i) - \frac{x_i}{4} + \frac{1}{b} \right] \end{cases} \quad (9)$$

We can easily obtain the confidence intervals for a and b . For the log-likelihood function in Equation (6), the Fisher information matrix H is given by

$$H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$$

Where

$$h_{11} = E \left[-\frac{\partial^2 \ln L}{\partial a^2} \right], h_{12} = h_{21} = E \left[-\frac{\partial^2 \ln L}{\partial a \partial b} \right], h_{22} = E \left[-\frac{\partial^2 \ln L}{\partial b^2} \right]$$

The variance matrix V is given by

$$V = [H]^{-1} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \quad (10)$$

The variances of estimated parameters a and b are

$$\text{Var}(a) = v_{11}, \text{Var}(b) = v_{22}$$

Based on the normal distribution approximation, we can obtain the $(1-\gamma)100\%$ confidence intervals for parameters a and b as follows:

$$[\hat{a} - z_{\gamma/2} \sqrt{v_{11}}, \hat{a} + z_{\gamma/2} \sqrt{v_{11}}]$$

and

$$[\hat{b} - z_{\gamma/2} \sqrt{v_{22}}, \hat{b} + z_{\gamma/2} \sqrt{v_{22}}]$$

respectively, where v_{ij} is described in Equation (10) and $z_{\gamma/2}$ is the $(1-\gamma/2)100\%$ percentile of the standard normal distribution.

MLE of Failure Rate Function

After we obtain \hat{a} and \hat{b} , the MLE of the failure rate, $h(x)$, can be computed as follows:

$$\hat{h}(x) = \hat{a}x \ln(\hat{b}x) + \frac{\hat{a}}{\hat{b}}$$

Let us define a partial derivative vector for the failure rate $h(x)$ as

$$v[h(x)] = \left[\frac{\partial h(x)}{\partial a}, \frac{\partial h(x)}{\partial b} \right]$$

Then, the variance of $h(x)$ can be obtained as follows:

$$Var[h(x)] = v[h(x)] \cdot V \cdot (v[h(x)])^T \quad (11)$$

Where V is given in Equation (10).

An approximate $(1-\gamma)100\%$ confidence interval for $h(x)$ can be computed as follows:

$$[\hat{h}(x) - z_{\gamma/2} \sqrt{Var[h(x)]}, \hat{h}(x) + z_{\gamma/2} \sqrt{Var[h(x)]}]$$

4. Applications in Software Reliability Engineering

Pham [3] developed a software reliability model considering the fault-detection rate with a Vtub-shaped [4] subject to the uncertainty of operating environments. Pham [5] recently proposed a generalized software reliability model where the uncertainty of operating environments is a random variable with a generalized probability density function of having two parameters $\alpha > 0$ and $\beta > 0$. The mean value function $m(x)$ defined as the expected number of software failures detected by time x is given by [5]:

$$m(x) = N \left(1 - \frac{\beta}{\beta + \int_0^x b(s) ds} \right)^\alpha \quad (12)$$

Where $b(x)$ is the fault detection rate per fault of a software system and N is the expected number of faults that exist in the software testing. Some recent studies [3, 5-6] have been carried out through various fault detection rates for single and multi-release of software systems. To illustrate the proposed new probability density function, we consider the Vtub-shaped failure rate $h(x)$ (Equation (3)) for the fault detection rate per fault, $b(x)$:

$$b(x) = ax \ln(bx) + \frac{a}{b} \quad \text{for } x > 0, a > 0, b > 0 \quad (13)$$

From Equation (12), we can easily obtain the mean value function $m(x)$ that represents the expected number of software failures to be detected by the time x , considering the uncertainty of operating environments as follows:

$$m(x) = N \left(1 - \frac{\beta}{\beta + ax \left(\frac{x}{2} \ln(bx) - \frac{x}{4} + \frac{1}{b} \right)} \right)^\alpha \quad (14)$$

There are some other software reliability models that have been developed based on the mean value functions $m(x)$, and these can be found in [5-6].

To obtain the numerical calculation, let us assume that $N = 25$, $a = 0.55$, $\alpha = 1.5$, and $\beta = 2.3$. From Equation (14), we obtain the expected number of failures to be detected by time x (i.e., m_i function, $i = 1, 2, 3$) for various values of b , as shown in Table 1 and Figures 5 and 6.

Further research work in software reliability modeling can be done by collecting real data from software system applications to assess the parameter estimates of the model function in Equation (14) and compare it to other existing models with respect to various performance criteria.

Table 1. Expected number of software failures for $N = 25$, $a = 0.55$, $\alpha = 1.5$, $\beta = 2.3$ for various b

b	0.5	0.4	0.3
Time x	m_1 function	m_2 function	m_3 function
1	3.15	4.11	5.68
2	6.75	7.78	9.54
3	10.80	11.39	12.65
4	14.50	14.64	15.30
5	17.36	17.26	17.47
6	19.41	19.21	19.17
7	20.84	20.61	20.46
8	21.84	21.63	21.43
9	22.55	22.36	22.16
10	23.06	22.90	22.71

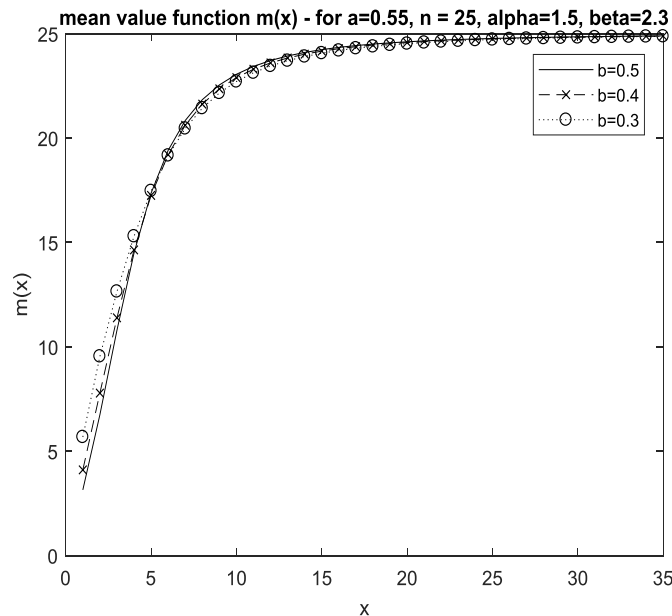


Figure 5. Mean value function with $a = 0.55$, $n = 25$, $\alpha = 1.5$, $\beta = 2.3$

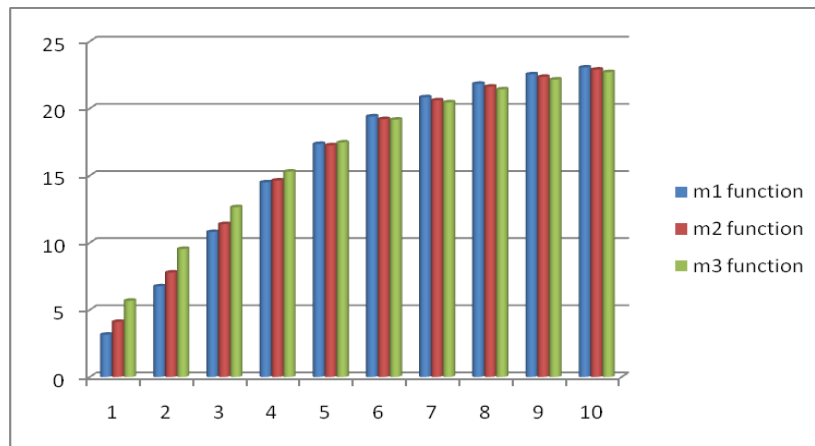


Figure 6. Expected number of failures for various values of b from Table 1

References

1. H. Pham, "Modeling U.S. Mortality and Risk-Cost Optimization on Life Expectancy," *IEEE Transactions on Reliability*, Vol. 60, No. 1, pp. 125-133, 2011
2. H. Pham and C. -D. Lai, "On Recent Generalization of the Weibull Distribution," *IEEE Transactions on Reliability*, Vol. 56, No. 3, pp. 454-458, 2007
3. H. Pham, "A New Software Reliability Model with Vtub-Shaped Fault-Detection Rate and the Uncertainty of Operating Environments," *Optimization-A Journal of Mathematical Programming and Operations Research*, Vol. 63, No. 10, pp. 1481-1490, 2014
4. H. Pham, "A Vtub-Shaped Hazard Rate Function with Applications to System Safety," *International Journal of Reliability and Applications*, Vol. 3, No. 1, pp. 1-16, 2002
5. H. Pham, "A Generalized Fault-Detection Software Reliability Model Subject to Random Operating Environments," *Vietnam Journal of Computer Science*, Vol. 3, No. 3, pp. 145-150, 2016
6. Q. Li and H. Pham, "NHPP Software Reliability Model Considering the Uncertainty of Operating Environments with Imperfect Debugging and Testing Coverage," *Applied Mathematical Modelling*, Vol. 41, No. 11, pp. 68-85, November 2017