A Software Reliability Model Using Quantile Function

Bijamma Thomas, 1 Midhu Narayanan Nellikkattu, 2 and Sankaran Godan Paduthol 2

1 Department of Mathematics, University College of Applied Sciences, Mahatma Gandhi University, Kottayam 686041, India
2 Department of Statistics, Cochin University of Science and Technology, Cochin 686022, India

Correspondence should be addressed to Sankaran Godan Paduthol; sankaranpg@yahoo.com

Received 17 October 2013; Revised 16 January 2014; Accepted 28 January 2014; Published 11 March 2014

1. Introduction

Software reliability models play an important role in developing software systems and enhancing the performance of computer software. In general, software reliability model can be classified into two types, depending on the operating domain. The most popular category of models depends on time, which uses the concepts such as the mean time between failures and the failure intensity function. The second category of software models measures reliability as the ratio of successful runs to the total number of runs. The intensity (or failure rate) function plays a pivotal role for modelling software failure time data. Throughout the literature on failure time of software systems, certain parametric models have been used repeatedly such as the Rayleigh model by Schick and Wolverton [1], the exponential model by Moranda [2] and Musa [3], and the power model by Crow [4]. These distributions have closed form exponents for tail area probability and simple formula for intensity functions. For more details, see Lyu [5]. There are other approaches based on Markov process, nonhomogeneous Poisson process, Bayesian techniques, and software metric approach for modelling and analysis of software failure time data. For various models using the above approaches, one could refer to Xie [6].

The models described above are based on distribution function of failure time and reliability measures derived from it. An alternative and equivalent approach for modelling statistical data is to use quantile function. Even though both functions convey the same information about the distribution, the methodologies and concepts based on distribution function are more popular in practice. However, quantile functions have several properties that are not shared by distributions, which makes it more convenient for analysis. There are explicit general distribution forms for the quantile function of order statistics. Also random numbers from any distribution can be generated using appropriate quantile functions, a purpose for which lambda distributions were originally conceived. There are many simple quantile functions which are very good in empirical model building where distribution functions are not effective. In such situations, conventional methods of analysis using distribution functions are not appropriate.

For various properties and applications of quantile functions, we refer to Parzen [7], Gilchrist [8], Sarabia [9], and Sarabia et al. [10]. Recently, Nair and Sankaran [11] introduced the basic concepts in reliability theory in terms of quantile functions. In reliability, a single long term survivor can have a marked effect on mean life, especially in the case of heavy tailed models which are very common. In such cases, quantile based estimates are generally found to be more precise and robust against outliers. Another advantage in choosing quantiles is that, in life testing experiments, one does not need to wait till the failure of all the items on the test but only a percentage of them for proposing useful estimates. For more properties and applications of quantile functions in reliability analysis, one could refer to Nair et al. [12], Nair and
Vinesh kumar [13], Unnikrishnan Nair and Vinesh kumar [14], Midhu et al. [15], Midhu et al. [16], and Nair et al. [17]. Thus, there is a case for adopting quantile functions as models of lifetime and basing their analysis with the aid of functions derived from them. Motivated by this fact, in the present work, we introduce a class of software reliability models using quantile function.

The rest of the paper is organized as follows. In Section 2, we present model and the properties of the model. The proposed class of quantile functions has several desirable distributional properties. The existing well-known lifetime models are members of the class as special cases or through approximations. Various reliability characteristics of the model are discussed in Section 3. The proposed class is a family of flexible lifetime models as it can be used for modelling and analysis of lifetime data having different ageing criteria by choosing different combinations of parametric values. Approximation to some well-known distributions is carried out in Section 4. The inference on parameters of the model is discussed in Section 5. We also apply the model to a software failure time data. Section 6 provides brief conclusion of the study.

2. The Model and Properties

Let $X$ be a nonnegative continuous random variable representing the failure time of a software with right continuous distribution function $F(x)$. Then, the quantile function of $X$ is defined as

$$Q(u) = F^{-1}(u) = \inf \{ x : F(x) \geq u \}, \quad 0 \leq u \leq 1. \quad (1)$$

For every $0 < x < \infty$ and $0 \leq u \leq 1$, we have $F(x) \geq u$ if and only if $Q(u) \leq x$. Thus, if there exists an $x$ such that $F(x) = u$, then $F(Q(u)) = u$ and we have that $Q(u)$ is the smallest value of $x$ for which $F(x) = u$. Further, if $F(x) = u$ is continuous and strictly increasing, $Q(u)$ is the unique value of $x$ such that $F(x) = u$. Then, we can find $x$ in terms of $u$ which is the quantile function of $X$.

If $f(x)$ is the probability density function of $X$, then $f(Q(u))$ is called the density quantile function. The derivative of $Q(u)$ is expressed as $Q'(u) = q(u)$ which is known as the quantile density function of $X$. Differentiating $F(Q(u)) = u$, we have

$$f(Q(u))q(u) = 1, \quad (2)$$

where $f(x)$ is the density of $X$. Now, we consider a class of distributions defined by the quantile density function

$$q(u) = ku^a(1-u)^b. \quad (3)$$

This distribution satisfied the general properties like symmetry, modality, tail behaviour, order statistics, shape properties based on the mode, L-moments, and transformations between members of the family. This class of distributions in (3) is the same as that in Jones [18]. Quantile function $Q(u)$ is defined in terms of quantile density function as

$$Q(u) = \int_0^u q(u) \, du. \quad (4)$$

Quantile function for the class of distributions in (3) can be obtained as

$$Q(u) = k\beta(u, a+1, b+1) \quad (5)$$

where $k > 0$, $a$, $b$ real, $0 \leq u \leq 1$, where $\beta(u, a+1, b+1)$ is the incomplete beta function with parameters $a$ and $b$. The support of the distribution (5) is $(Q(0), Q(1)) = (0, \beta(1, a+1, b+1))$.

The derivative of (3) gives

$$q'(u) = k[a^b(1-u)^{b-1} - 1 + (1-u)^b]u^{a-1}. \quad (6)$$

Equating $q'(u) = 0$ and solving for $u$, we get $u = a/(a+b)$. Thus, the members of the family have either unimodal density with mode (antimode) at $u = a/(a+b)$ or monotone density when $u \neq a/(a+b)$. Figure 1 gives the quantile density function for different values of parameters.

The quantile function defined in (5) becomes exponential when $a = 0$ and $b = -1$ and linear hazard quantile distribution Midhu et al. [15] when $a = -1$ and $b = -1$ or $a = 0$ and $b = -2$.

The distributional characteristics such as location, dispersion, skewness, and kurtosis can be expressed through quantile terms. For the class of distributions in (5), the quantile based measure of location is the median defined by

$$M = Q\left(\frac{1}{2}\right) = \beta\left(\frac{1}{2}, a+1, b+1\right). \quad (7)$$

Dispersion is measured by the interquartile range which is expressed as

$$\text{IQR} = Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right) = \beta\left(\frac{3}{4}, a+1, b+1\right) - \beta\left(\frac{1}{4}, a+1, b+1\right). \quad (8)$$

Galton’s coefficient of skewness is obtained from (5) as

$$S = \frac{[Q(3/4) + Q(1/4) - 2M]}{\text{IQR}}, \quad (9)$$

which can also be expressed as

$$S = \left(\beta\left(\frac{3}{4}, a+1, b+1\right) + \beta\left(\frac{1}{4}, a+1, b+1\right) - 2\beta\left(\frac{1}{2}, a+1, b+1\right)\right) \times \left(\beta\left(\frac{3}{4}, a+1, b+1\right) - \beta\left(\frac{1}{4}, a+1, b+1\right)\right)^{-1}. \quad (10)$$

In the case of extreme positive skewness, $Q(1/4) \to M$ and in the extreme negative skewness, $Q(3/4) \to M$, so that $S$ lies between $-1$ and $+1$. When the distribution is symmetric, $M = (Q(1/4) + Q(3/4))/2$ and hence $S = 0$. The measure of kurtosis can be defined as

$$T = \frac{[Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)]}{\text{IQR}}. \quad (11)$$
The $L$-moments are often found to be more desirable than the conventional moments in describing the characteristics of the distributions as well as for inference. The $L$-moments exist whenever $E(X)$ is finite, whereas for many distributions additional restrictions are required for the conventional moments to be finite. The $L$-moments have generally lower sampling variances and robust against outliers. See Hosking [19] and Sankarasubramanian and Srinivasan [20] for details. The first $L$-moment is the mean which is given by

$$L_1 = \int_0^1 (1 - u) q(u) du = k\beta (a + 1, b + 2).$$

The second $L$-moment can be expressed as

$$L_2 = \int_0^1 (u - u^2) q(u) du = k\beta (a + 2, b + 2).$$

The third $L$-moment is obtained as

$$L_3 = \int_0^1 (3u^2 - 2u^3 - u) q(u) du = 2k\beta (a + 3, b + 2) - k\beta (a + 2, b + 2)$$

and the fourth $L$-moment is given by

$$L_4 = \int_0^1 (u - 6u^2 + 10u^3 - 5u^4) q(u) du = k\beta (a + 2, b + 2) - 5k\beta (a + 3, b + 3).$$

$L$-coefficient of variation, analogous to the coefficient of variation based on ordinary moments, is obtained as

$$\tau_2 = \frac{L_2}{L_1} = \frac{k\beta (a + 2, b + 2)}{k\beta (a + 1, b + 2)} = \frac{\Gamma a + 2\Gamma b + 2/\Gamma a + b + 4}{\Gamma a + 1\Gamma b + 2/\Gamma a + b + 3} = \frac{a + 1}{a + b + 3}.$$
The $L$-coefficient of variation is 0 when $a = -1$ and it has the upper bound 1 when $b = -2$. The $L$-coefficient of skewness is measured as

\[
\tau_3 = \frac{L_3}{L_2} = \frac{2k\beta (a + 3, b + 2) - k\beta (a + 2, b + 2)}{k\beta (a + 2, b + 2)}
\]

\[
= \frac{2(\Gamma(a + 3)\Gamma(b + 2) + 2\Gamma(a + b + 5)}{\Gamma(a + 2)\Gamma(b + 2)\Gamma(a + b + 4)} - 1 \tag{18}
\]

\[
= 2\frac{(a + 2)}{a + b + 4} - 1 = \frac{a - b}{a + b + 4}.
\]

The $L$-coefficient of skewness lies between $(-1, 1)$, that is, $\tau_3 = -1$ when $a = -2$ and $\tau_3 = 1$ when $b = -2$. The $L$-coefficient of kurtosis is given by

\[
\tau_4 = \frac{L_4}{L_2} = \frac{k\beta (a + 2, b + 2) - 5k\beta (a + 3, b + 3)}{k\beta (a + 2, b + 2)}
\]

\[
= 1 - \frac{5\Gamma(a + 3)\Gamma(b + 2) + 2\Gamma(a + b + 5)}{\Gamma(a + 2)\Gamma(b + 2)\Gamma(a + b + 4)} \tag{19}
\]

\[
= 1 - \frac{5(a + 2)(b + 2)}{(a + b + 5)(a + b + 4)}.
\]

The $L$-coefficient of kurtosis attains the upper bound 1 when $a = -2$ or $b = -2$ and has the lower bound $-1/4$.

3. Reliability Characteristics

One of the important concepts in reliability analysis is the hazard function which is defined as

\[
h(x) = \lim_{\Delta x \to 0} \frac{P[x < X < x + \Delta x \mid X \geq x]}{\Delta x} \tag{20}
\]

Unnikrishnan Nair and Sankaran [11] studied the quantile form of the hazard function termed as hazard quantile function which is given by

\[
H(u) = \frac{1}{(1 - u)q(u)} \tag{21}
\]

$H(u)$ can be interpreted as the conditional probability of the failure of a unit in the next small interval of time given the survival of the unit at 100$(1 - u)$% point of the distribution. For the class of distributions (3), we have

\[
H(u) = \frac{1}{ku^a(1 - u)^{b+1}}. \tag{22}
\]

The shape of the hazard function is determined by the derivative of $H(u)$. For the hazard quantile function of (22), the derivative is given by

\[
H'(u) = \frac{(1 - u)^{-b-2}u^{-(a+1)}[(a + b + 1)u - a]}{k} \tag{23}
\]

Since $k > 0$, $H'(u)$ changes sign according to the term $g(u) = [(a + b + 1)u - a]$. The sign of $g(u)$ changes according to the values of $a$ and $b$. If $a < 0$ and $b > -(a + 1)$, $g(u)$ is positive and hence $H(u)$ has an increasing failure rate. If $a > 0$ and $b < -(a + 1)$, then $g(u)$ is negative and then $H(u)$ decreases. When $a = 0$ and $b < -1$, then $g(u)$ is negative which leads to the fact that $H(u)$ has a decreasing failure rate. When $a = 0$ and $b > -1$, then $g(u)$ is positive and hence $H(u)$ has an increasing failure rate. When $a < 0$ and $-1 \leq b < -(a + 1)$, $g(u)$ is positive and $H(u)$ has an increasing failure rate. If $a > 0$ and $-(a + 1) < b < -1$, $g(u)$ is negative. Thus, it follows that $H(u)$ has a decreasing failure rate. In the case when $a < 0$ and $b < -1$, $H(u)$ increases up to a maximum at $u = a/(a+b+1)$ and then decreases. So, $H(u)$ is upside bathtub here. And in the case when $a > 0$ and $b > -1$, $H(u)$ decreases to a minimum at $u = a/(a+b+1)$ and then increases. So, $H(u)$ is bathtub. Plots of the hazard quantile functions for different values of parameters are presented in Figure 2.

Mean residual function is a well-known measure, which has been widely used in various fields of reliability and survival analysis. In quantile setup, the mean residual quantile function is expressed as

\[
M(u) = (1 - u)^{-1} \int_u^1 [Q(p) - Q(u)] dp. \tag{24}
\]

The above identity can also be expressed as

\[
M(u) = \frac{1}{1-u} \int_u^1 (1 - p)q(p)dp. \tag{25}
\]

$M(u)$ is interpreted as the average remaining life beyond the 100$(1 - u)$% of the distribution. For our class of distributions, $M(u)$ is obtained as

\[
M(u) = \frac{1}{1-u} \int_u^1 k\beta p^a(1 - p)^{b+1} dp \tag{26}
\]

\[
= \frac{1}{1-u} \beta (p, a + 1, b + 2).
\]

4. Approximations

Since the class of distributions (5) cannot be converted to a tractable form for its distribution function, its relationship with other known standard distributions can be assessed only through approximations. When we have a quantile function that provides approximation to many types of distributions, only one functional form for $Q(u)$ and the related inferential aspects are sufficient for modelling and analysis, as the quantile function will adapt automatically to the suitable form. In this section, we attempt to fit our distribution to some well-known lifetime distributions like inverse Gaussian distribution and Weibull distribution. The same approach can be used to approximate other distributions. We use the method of $L$-moments for finding the values of the parameters.

4.1. Inverse Gaussian Distribution. Probability density function of inverse Gaussian is given by

\[
f(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \tag{27}
\]
The inverse Gaussian distribution with parameters $\mu$ and $\lambda$ has $L$-moments $L_1 = \mu$, $L_2 = \mu^2/\lambda$, and $L_3 = 3\sqrt{\mu/\lambda}$. We equate $L$-moments of inverse Gaussian distribution with the $L$-moments of the proposed model. There is no explicit closed form expression for $k$, $a$, and $b$ in terms of $\mu$ and $\lambda$. However, as an illustration, we assume that $\mu = 1$ and $\lambda = 3$ and the corresponding values of (5) are $k = 0.2960$, $a = -0.6683$, and $b = -1.22$. Figure 3 gives the probability density function (p.d.f) of the inverse Gaussian distribution and the dotted line represents that of the approximated quantile function. For measuring the closeness between the two models, we find $\sup_x |F_1(x) - F_3(x)|$, where $F_1(x)$ is the distribution function of the proposed model and $F_3(x)$ is the distribution function of the inverse Gaussian distribution. Now, we obtain $\sup_x |F_1(x) - F_3(x)| = 0.02203$ which is very small.

4.2. Weibull Distribution. The probability density function of Weibull distribution is given by

$$f(x) = \frac{e^{-x/\beta^\alpha} \alpha(x/\beta)^{\alpha-1}}{\beta}. \quad (28)$$

The Weibull distribution with parameters $\alpha$ and $\beta$ has $L$-moments $L_1 = \beta\Gamma(1 + 1/\alpha)$, $L_2 = \beta(1 - 2^{-1/\alpha})\Gamma(1 + 1/\alpha)$, and $L_3 = (3 - 2(1 - 3^{-1/\alpha})L_2)/(1 - 2^{-1/\alpha})$. We equate $L$-moments of Weibull distribution with the $L$-moments of the proposed model. There is no explicit closed form expression for $k$, $a$, and $b$ in terms of $\alpha$ and $\beta$. However, as an illustration, we assume that $\alpha = 0.5$ and $\beta = 2$ and the corresponding values of (5) are $k = 5.1613$, $a = 1.142$, and $b = -1.2857$. Figure 4 gives p.d.f of the Weibull distribution and the dotted line represents the p.d.f of approximated quantile function. For measuring the closeness between the two models, we find $\sup_x |F_1(x) - F_3(x)|$, where $F_1(x)$ is the distribution function...
of the proposed model and \( F_3(x) \) is the distribution function of the Weibull distribution. Now, we obtain \( \sup_x |F_1(x) - F_3(x)| = 0.0072 \) which is also very small.

5. Estimation of Parameters and Data Analysis

For estimating the parameter of the distributions, which are expressed in terms of quantile function, there are different methods available (see Gilchrist [8]). Among these different methods, the commonly used techniques are method of minimum absolute deviation, method of least squares, method of maximum likelihood, and method of \( L \)-moments. Recently, method of \( L \)-moments is widely used as an alternative to the conventional methods in inference problems in view of the robustness in the estimates produce (see Hosking [19]). To estimate the parameters of the function given in (5), we use the method of \( L \)-moments. The simple algebraic expressions of the \( L \)-moments explained in Section 2 admit the applicability of the \( L \)-moments method for estimating the parameters of the model (5). Let \( x_1, x_2 \cdots x_n \) be a random sample of size \( n \) with quantile function (5). Since there are three parameters in the model, we take three sample \( L \)-moments \( l_r, r = 1, 2, 3 \), and those are given by

\[
\begin{align*}
l_1 &= \frac{1}{n} \sum_{k=1}^{n} x_{(k)} \\
l_2 &= \frac{1}{2} \left( \frac{n}{2} \right)^{-1} \sum_{k=1}^{n} \left\{ \left( \frac{n}{2} \right) - \left( \frac{n}{2} - k \right) \right\} x_{(k)} \\
l_3 &= \frac{1}{3} \left( \frac{n}{3} \right)^{-1} \sum_{k=1}^{n} \left\{ \left( \frac{n}{3} \right) - 2 \left( \frac{n}{3} - k \right) \left( \frac{n}{3} - k + 1 \right) \\
&\quad + \left( \frac{n}{3} - k \right) \right\} x_{(k)},
\end{align*}
\]

(29)

where \( x_k \) is the \( k \)th order statistic. We equate sample \( L \)-moments to population \( L \)-moments given by

\[
l_r = L_r : r = 1, 2, 3.
\]

(30)

Solutions of set of equations (30) give the estimates of \( k, a, \) and \( b \). The set of equations (30) are nonlinear in \( k, a, \) and \( b \). We use Newton Raphson method to find the values of \( k, a, \) and \( b \).

Hosking (1990) has studied asymptotic properties of \( L \)-moment estimates. The following theorem provides asymptotic normality of sample \( L \)-moments.

**Theorem 1** (Hosking [1990]). Let \( X \) be a real-valued random variable with quantile function \( Q(u, \theta) \), where \( \theta \) is a vector of \( m \) parameters. Assume that variance of \( X \) is finite. Let \( \ell_r, r = 1, 2, \ldots, m \), be sample \( L \)-moments calculated from a random sample of size \( n \) drawn from the distribution of \( X \). Then, \( \sqrt{n}(\ell_r - L_r), r = 1, 2, \ldots, m \), converge in distribution to the multivariate normal \( N(0, \Lambda) \), where the elements \( \Lambda_{rs} (r,s = 1, 2, \ldots, m) \) of \( \Lambda \) are given by

\[
\Lambda_{rs} = \int_{0 < u < v < 1} \{ P_{r-1}^*(u) P_{s-1}^*(v) + P_{s-1}^*(u) P_{r-1}^*(v) \}
\]

\[
\times u (1 - v) q(u) q(v) \, du \, dv,
\]

where \( P_r^*(x) \) is being the \( r \)th shifted Legendre polynomial defined by

\[
P_r^*(x) = \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} x^k.
\]

(32)

Since the set of equations (30) are nonlinear in \( k, a, \) and \( b \) asymptotic distributions of \( L \)-moment estimates of the parameters and are difficult to obtain, one can use bootstrap method to obtain the asymptotic variance of the estimates.

Now, we apply the model (5) to a real data set taken from Musa [21]. The data set represents failure time in seconds of a command and controlling system during in-house testing using a simulation of the real operational environment. The data set consists of failure time for a sample of size 136. Musa [21] fitted a software reliability growth model to the data. The proposed model (5) fitted to this data set. We estimated the parameters using the method of \( L \)-moments and the estimates are given as \( \hat{a} = 0.484752, \hat{b} = -1.21617, \) and \( \hat{k} = 712.679 \). Since the estimate \( a > 0 \) and \( b < 0 \), \( H(u) \) has a decreasing failure rate as shown in Figure 5. To check the goodness of fit, we use \( Q-Q \) plot which is given in Figure 6. Figure 6 shows that most of the data points are close to the straight line. This indicates that the quantile function given in (5) is a reasonable fit to the data. We also employed chi-square goodness of fit to check the adequacy of the model. The chi-square statistic value is 8.73 with \( P \) value 0.891. This indicates that the proposed model is a reasonable one for the given data set.

6. Conclusion

A probability distribution can be specified either in terms of the distribution function or by quantile function. Although
both convey the same information about the distribution, with different interpretations, the concepts and methodologies based on distribution functions are traditionally employed in most forms of statistical theory and practice. One reason for this is that quantile based studies were carried out mostly when the traditional approach fails to provide results of desired quality. Except in a few isolated areas, there have been no systematic parallel developments aimed at replacing distribution functions in modelling and analysis by quantile functions. However, the feeling that through an appropriate choice of the domain of observations a better understanding of a chance phenomenon can be achieved is fast gaining acceptance.

Motivated by this fact, in the present work, we have introduced a class of quantile function models, useful in software reliability analysis. The proposed class has several desirable properties and several existing well-known distributions that are members of the class of distributions as special cases or through approximations. Various reliability characteristics were discussed. The parameters of the model were estimated using $L$-moments and the model was applied to a real data set. The method of maximum likelihood can also be employed to find the estimates of the parameters (see Gilchrist [8]). The proposed class of quantile functions is a flexible model in the sense that it has the property of increasing hazard rate and decreasing hazard rate, bathtub hazard rate, and upside down hazard rate by changing the parametric values.

**Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**Acknowledgments**

The authors thank the referee and the editor for their helpful and constructive comments.

**References**


