行政院國家科學委員會專題研究計畫　成果報告

有限個數的 i.i.d. Weibull 分布隨機變數之和的機率分布函數的近似值

計畫類別： 個別型計畫
計畫編號： NSC93-2118-M-009-018-
執行期間： 93 年 08 月 01 日至 94 年 07 月 31 日
執行單位： 國立交通大學統計學研究所

計畫主持人： 彭南夫

報告類型： 精簡報告

處理方式： 本計畫可公開查詢

中　華　民　國 94 年 10 月 31 日

有限個數的 i.i.d. 韋伯隨機變數之和的累積機率函數近似解

關鍵字 ： 韋伯隨機變數 ， 累積機率函數

An Approximation of the Distribution Function of the Sum of Finite i.i.d. Weibull Random Variables

Keywords : Weibull random variables, Cumulative Distribution function

1.INTRODUTION

The main purpose in this discussion is to find the distribution function of the sum of independent and identical distributed Weibull random variables. The Weibull distribution is a two-parameter distribution which by adjusting a scale parameter (denoted by c) and a shape parameter (denoted by β) we can obtain a variety of shapes to fit experimental data. Thus, this distribution is highly adaptable and widely used in practice. Without loss of generality, we can assume the scale parameter to be one. In addition, the Weibull distribution is widely used in reliability modeling since other distributions such as exponential (c=1), Rayleigh (c=2), and normal (if a suitable value for the shape parameter is chosen) are special cases of the Weibull distribution. In the version of hazard-rate function, when c>1, the hazard rate is a monotonically increasing function with no upper bound that describes the wear-out region, and when c=1, the hazard rate becomes constant (constant failure-rate region), and when c<1, the hazard rate is a monotonically decreasing with time (the early failure-rate region). This enables the Weibull distribution to describe the failure rate of many failure data in many kinds of application.

Unlike the renewal processes with the exponential inter arrivals, those with the Weibull interarrivals do not possess elegant properties. It is well known that by the nature of the renewal processes, the distribution of the process is related to the distribution of a finite sum of the independent and identically distributed inter arrivals, i.e. suppose that inter arrivals $X_1, X_2, ..., X_n$ are a random sample of $Weibull(c, \beta)$ distributed with probability density function

$$f(x) = \frac{c}{\beta} x^{c-1} e^{-\frac{x^c}{\beta}}. \qquad (1)$$

The probability of interest is

$$P(S_n \leq t) \text{ where } S_n = X_1 + ... + X_n, \qquad (2)$$

which has no simple formula to evaluate. The major result of this report is to present a simple, elegant and relatively accurate form of an approximation.

2. METHODOLOGY

In this section, we discuss the approximation of $F_n(t) = P(S_n \leq t)$, which could be seen as the convolution of *n* functions. When *n* is large, the evaluation of *n*-fold convolution includes *n*-dimensional integration, thus the evaluation becomes difficult. Even using the numerical analysis, the results of evaluation are obtained pointwisely. That is, when fixed one value of *t*, we can numerically evaluate $F_n(t)$, but just only one value in every evaluation. The numerical computation of the inverse Laplace transform [8] still directs to the answer pointwisely. Another approach is to use Edgeworth Expansion [3 and 7], but the shortage is that we have to includ numerical integration and when the shape parameter $c<1$, the accuracy of the approximation would not behave well relatively. Here we use three quantiles to obtain a simple approximation method, and the evaluation of approximation is obtained functionwisely, not pointwisely. Our approach describes as the following:

Suppose X_i is a random variable from $Weibull(c, 1)$ distribution. Via the transform of variable $Y_i = X_i^{\,c}$, we obtain that Y_i is a random variable from $Exponential(1)$ distribution with $i = 1,...,n$. Thus, we expect to use the distribution function of $Y_1 + ... + Y_n$ to approximate the distribution function of $X_1 + ... + X_n$. Since the two distribution functions are continuously differentiable functions, for every $t \ge 0$, we expect to find some continuously differentiable function $w(t)$, such that

$$P(X_1 + ... + X_n \le t) = P(Y_1 + ... + Y_n \le w(t)), \qquad (3)$$

with $w(0) = 0$. Thus, we consider the form of third degree polynomial $w(t) = \alpha t^{3c} + \gamma t^{2c} + \tau t^{c}$. (Note that we can consider the form $\alpha t^{2c} + \gamma t^{c}$, but the approximation isn't well relatively.) Three unknown parameters above can be determined by three quantiles of $X_1 + ... + X_n$ and $Y_1 + ... + Y_n$.

First, let t_p satisfy

$$P(X_1 + ... + X_n \le t_p) = p, \qquad (4)$$

and θ_p satisfy

$$P(Y_1 + ... + Y_n \le \theta_p) = p, \qquad (5)$$

Thus, t_p and θ_p are the p-quantile of $X_1 + ... + X_n$ and $Y_1 + ... + Y_n$, respectively. When $p = 0.1$, 0.5, and 0.9, we can apply $\alpha t_p^{3c} + \gamma t_p^{2c} + \tau t_p^{c} = \theta_p$ to construct three linear functions. Thus, three unknown parameters can be evaluated through the three linear functions.

Lemma 1:

Suppose $Y_1, Y_2, ..., Y_n$ are a random sample from $Exponential(1)$ distributed. Then

$$P(Y_1 + ... + Y_n \le \theta) = 1 - \sum_{i=0}^{n-1} e^{-\theta} \theta^i / i!. \quad (6)$$

Proof:

First let $S_n^* = \sum_{i=1}^{n} Y_i$, there is a counting process $\{N(\theta) = \sup(n : S_n^* \le \theta), \theta \ge 0\}$ which is a renewal process, conclude that

$$S_n^* \le \theta \Leftrightarrow N(\theta) \ge n. \quad (7)$$

As inter arrivals are a random sample from $Exponential(1)$ distributed, we know that $N(\theta)$ is a random variable from $Poisson(\theta)$ distributed. Thus

$$P(Y_1 + ... + Y_n \le \theta) = P(S_n^* \le \theta) = P(N(\theta) \ge n)$$

$$= 1 - P(N(\theta) < n) = 1 - \sum_{i=0}^{n-1} e^{-\theta} \theta^i / i! \quad (8)$$

■

Via lemma 1, the value of θ_p can be obtained by numerical method (We use the Newton-Raphson method). Different to θ_p which follows lemma1, we must simulate the value of t_p. We interpret the procedure of the approach as the follows:

(1) Through Newton-Raphson method we can obtain the solutions of the equation $1 - \sum_{i=0}^{n-1} e^{-\theta_p} \theta_p^i / i! = p$, where $p =$0.1, 0.5, 0.9, i.e. $\theta_{0.1}, \theta_{0.5}$, and $\theta_{0.9}$.

(2) Determine the numbers of simulate data $N = 10^7$, let $T_1, T_2, ..., T_N$ be a random sample from $F_n(x) = P(X_1 + ... + X_n \le x)$, and order them become to $T_{(1)}, T_{(2)}, ..., T_{(N)}$. Thus, let $\hat{t}_p = T_{(Np)}$ be the approximation of the true quantile t_p,

where $p = 0.1,\ 0.5,\ 0.9$.

(3) Via the three linear equations

$$\alpha\,\hat{t}_{0.1}^{3c} + \gamma\,\hat{t}_{0.1}^{2c} + \tau\,\hat{t}_{0.1}^{c} = \theta_{0.1},$$

$$\alpha\,\hat{t}_{0.5}^{3c} + \gamma\,\hat{t}_{0.5}^{2c} + \tau\,\hat{t}_{0.5}^{c} = \theta_{0.5}, \qquad (9)$$

$$\alpha\,\hat{t}_{0.9}^{3c} + \gamma\,\hat{t}_{0.9}^{2c} + \tau\,\hat{t}_{0.9}^{c} = \theta_{0.9}.$$

Three of unknown parameters $\alpha,\ \gamma,\ \tau$ can be determined.

(4) Under fixed n and shape parameter c, the estimated form of $w(t)$ is

$$\hat{w}(t) = \alpha\, t^{3c} + \gamma\, t^{2c} + \tau\, t^{c}. \qquad (10)$$

(5) Thus the approximation of the distribution function of interest is

$$P(X_1 + \ldots + X_n \le t) \cong P(Y_1 + \ldots + Y_n \le \hat{w}(t)) = 1 - \sum_{i=0}^{n-1} e^{-\hat{w}(t)}\hat{w}(t)^i / i!. \qquad (11)$$

We choose the number of simulate data is $N = 10^7$, and $T_{(Np)}$ is the p-th sample quantile. From [6] we can know that $T_{(Np)}$ is asymptotically following the $N(t_p, \frac{p(1-p)}{f_n^2(t_p)N})$ distributed, the error is proportional to $\frac{1}{\sqrt{N}} = 10^{-\frac{7}{2}}$, thus $T_{(Np)}$ would close to the true p-th quantile. Equation (11) is the most important result of this dissertation and offers a simple approximation via the correlation between the Exponential and the Weibull distributed. Moreover, the approximation approach will reduce a lot of time relative to simulation. In the next section we will discuss the accuracy of the approach.

3.APPROXIMATION RESULTS

In order to appraise the accuracy of our approximation, the criterion that be used is the relative error. We denote the relative error is $\frac{P-\hat{P}}{P}$, where p is the true probability and $\hat{P} = P(Y_1 + ... + Y_n \leq \hat{w}(T_{(Np)}))$ is the approximation of p. Total cases that we have done are as $n = 2,\ 3, ...,\ 10$, $c = 0.1,\ 0.2, ...,\ 0.9$ and $c = 2,\ 3, ...,\ 9$. In table 3.1, we list the relative error table includes only cases of $n = 2,\ 5$ and 9, c = 0.5 and 5, the others have the analogous result and tendency. Fixed n and c, in every block of table 3.1 includes the relative errors of probabilities 0.05, 0.1, …0.95. Further, we list the same cases of three of parameters in table A.1 of Appendix A.

We can discover that when c=0.5, the maximum relative error is 0.14%; and c=5, the maximum relative error is 1.4%, noted that happened as p=0.05, concludes that the absolute error isn't very large. Hence, we believe that Equation (11) brings a nice accuracy of approximation.

[Table 3.1]

Fixed as n=2

p c	0.5	5
0.05	0.000726445242646	-0.010161527424567
0.1	-0.000000002616051	-0.000000000000001
0.15	0.000119820278476	0.003828824693817
0.2	-0.000194908815901	0.005143137693705
0.25	0.000138893117596	0.005324085016400
0.3	-0.000011260430451	0.005005107033746
0.35	0.000043543016760	0.004334074875617
0.4	0.000186524455173	0.002847272084006
0.45	0.000024819173472	0.001416814541063
0.5	-0.000000000567722	0.000000000000000
0.55	-0.000148473532176	-0.001696971491006
0.6	-0.000059519398334	-0.003003892633153

0.65	-0.000040240442348	-0.004332428288305
0.7	-0.000105114490566	-0.005246866270862
0.75	-0.000172923165788	-0.005520852102506
0.8	-0.000064298286200	-0.005102997297659
0.85	0.000007023458372	-0.003342411510635
0.9	-0.000000001380996	0.000000000000000
0.95	0.000165231376151	0.004920445151018

Fixed as n=5

p c	0.5	5
0.05	-0.000028135251537	-0.014064661989290
0.1	-0.000000007634631	0.000000000000002
0.15	0.000008069914397	0.003784904432964
0.2	0.000509988574720	0.005825262499093
0.25	0.000340433284307	0.006454284943475
0.3	0.000628625741176	0.005655125010712
0.35	0.000642236022542	0.004201874492865
0.4	0.000125130856831	0.002810794471418
0.45	0.000196852924741	0.001418848371266
0.5	-0.000000001023538	0.000000000000002
0.55	0.000325840009388	-0.001174124992608
0.6	0.000290632220714	-0.002266358442895
0.65	0.000252098046966	-0.002949416858121
0.7	0.000277480011670	-0.003413516974510
0.75	0.000254161667295	-0.003491304489550
0.8	0.000199371124861	-0.003091183240431
0.85	0.000174674060825	-0.001950997481086
0.9	-0.000000001927334	0.000000000000001
0.95	-0.000021837121247	0.002886342734995

Fixed as n=9

p c	0.5	5
0.05	0.001436234621018	-0.010512189365780

0.1	-0.000000012003200	0.000000000000005
0.15	0.000557973789439	0.004208106793534
0.2	0.000524827975903	0.004523974335599
0.25	0.000174871687116	0.004347404044753
0.3	0.000275117740350	0.003550649249438
0.35	0.000340236827250	0.002914461327970
0.4	0.000406740368423	0.001891905365702
0.45	0.000277355403994	0.000952014957686
0.5	-0.000000001422286	-0.000000000000003
0.55	0.000113761481000	-0.000865972087613
0.6	-0.000128186125497	-0.001469193796314
0.65	-0.000099682351945	-0.002062921597401
0.7	0.000002286246061	-0.002299894298618
0.75	0.000047694801632	-0.002261558612072
0.8	-0.000065649121020	-0.001909441671796
0.85	0.000010740192212	-0.001183948276717
0.9	-0.000000002419210	-0.000000000000003
0.95	0.000015724643715	0.001799676396769

APPENDIX A

[Table A.1]

Fixed n and c, there are three parameters in every block, where α, γ and τ are in the upper, middle and down side, respectively.

n \ c	0.5	5
2	0.000280751644006 -0.010114574919690 1.255212150831596	0.000004835157895 -0.000513436509664 0.087023676809991
3	-0.000395335224320 -0.012406564768126 1.464308967719061	0.000000025577837 -0.000019005934388 0.018978874641112
4	-0.000224864735582 -0.016947170359548 1.650435209948320	0.000000000555125 -0.000001658936240 0.006292385415443

5	-0.000493569188358 -0.018315452950023 1.816576635488150	0.000000000026227 -0.000000236200742 0.002643130317068
6	-0.000726355289725 -0.018704570924749 1.967622246201590	0.000000000002166 -0.000000047912910 0.001296831369222
7	-0.000497097072477 -0.022139525917010 2.112546353444382	0.000000000000270 -0.000000012598155 0.000709991929183
8	-0.000863384052326 -0.020185034372005 2.239924284448030	0.000000000000042 -0.000000003807886 0.000419152653162
9	-0.000835110482110 -0.021605902558135 2.367244177097372	0.000000000000008 -0.000000001360679 0.000263824260089
10	-0.000897769786661 -0.021741736990287 2.484964323969154	0.000000000000002 -0.000000000536868 0.000174175974127

REFERENCES

1. Bickel, P. J. & Doksum, K. A. (2001). *Mathematical Statistics, Basic Ideas and Selected Topics*. 2nd edition. Volume1. New Jersey: Prentice-Hall.
2. Elsayed, A. (1996). *Reliability Engineering*. MA: Addison-Wesley.
3. Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.
4. Ross, S. M. (1996). *Stochastic Processes*. 2nd edition. New York: Wiley.
5. Ross, S. M. (1997). *Simulation*. 2nd edition. New York: Academic Press.
6. Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: J. Wiley & sons.
7. Stuart, A. & Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics*. 5th ed., Volume 1, Distribution Theory. London: Charles Griffin & Co.
8. Trivedi, K. S. (2002). *Probability and Statistics with Reliability, Queuing and Computer Science Applications*. 2nd edition. New York: Wiley.