On Modeling of Lifetime Data Using Two-Parameter Gamma and Weibull Distributions

Abstract

The analysis and modeling of lifetime data are crucial in almost all applied sciences including medicine, insurance, engineering, behavioral sciences and finance, amongst others. The main objective of this paper is to have a comparative study of two-parameter gamma and Weibull distributions for modeling lifetime data from various fields of knowledge. Since exponential distribution is a particular case of both gamma and Weibull distributions and the exponential distribution is a classical distribution for modeling lifetime data, the goodness of fit of both gamma and Weibull distributions are compared with exponential distribution.

Keywords: Gamma distribution; Weibull distribution; Exponential distribution; Lifetime data; Estimation of parameter; Goodness of fit

Research Article

Volume 4 Issue 5 - 2016

Rama Shanker[1]*, Kamlesh Kumar Shukla[1], Ravi Shanker[2] and Tekie Asehun Leonida[3]

[1]*Department of Statistics, Eritrea Institute of Technology, Eritrea*

[2]*Department of Mathematics, GLA College, NP University, India*

[3]*Department of Applied Mathematics, University of Twente, The Netherlands*

***Corresponding author:** Rama Shanker and Kamlesh Kumar Shukla, Department of Statistics, Eritrea Institute of Technology, Asmara, Eritrea, Email: shankerrama2009@gmail.com; kkshukla22@gmail.com

Received: September 19, 2016 | **Published:** October 07, 2016

Introduction

The lifetime or survival time or failure time in reliability analysis is the time to the occurrence of event of interest. The event may be failure of a piece of equipment, death of a person, development (or remission) of symptoms of disease, health code violation (or compliance). The modeling and statistical analysis of lifetime data are crucial for statisticians and research workers in almost all applied sciences including behavioral sciences, engineering, medical science/biological science, insurance and finance, amongst others.

The statistics literature is flooded with lifetime distributions including exponential distribution, gamma distribution, Lindley distribution, Weibull distribution and their generalizations, some amongst others.

Gamma Distribution

The probability density function (p.d.f.) and the cumulative distribution function (c.d.f.) of two-parameter gamma distribution (GD) having parameters θ and α are given by

$$f_1(x;\theta,\alpha)=\frac{\theta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\theta x};\ x>0,\theta>0,\alpha>0 \qquad (2.1)$$

$$F_2(x;\theta,\alpha)=1-\frac{\Gamma(\alpha,\theta x)}{\Gamma(\alpha)};\ x>0,\theta>0,\alpha>0 \qquad (2.2)$$

Where $\Gamma(\alpha,z)$ is the upper incomplete gamma function defined as

$$\Gamma(\alpha,z)=\int_z^{\infty}e^{-y}y^{\alpha-1}dy;\alpha>0,\ z\geq 0 \qquad (2.3)$$

It can be easily shown that the gamma distribution reduces to classical exponential distribution for $\alpha=1$ having p.d.f. and c.d.f.

$$f_2(x;\theta)=\theta e^{-\theta x};x>0,\theta>0 \qquad (2.4)$$

$$F_2(x;\theta)=1-e^{-\theta x};x>0,\theta>0 \qquad (2.5)$$

It should be noted that the gamma distribution is the weighted exponential distribution. Stacy [1] obtained the generalization of the gamma distribution. Stacy & Mihram [2] have detailed discussion about parametric estimation of generalized gamma distribution.

Weibull Distribution

The p.d.f. and the c.d.f. of two-parameter Weibull distribution having parameters θ and α are given by

$$f_3(x;\theta,\alpha)=\theta\alpha x^{\alpha-1}e^{-\theta x^{\alpha}}\ ;x>0,\theta>0,\alpha>0 \qquad (3.1)$$

$$F_3(x;\theta,\alpha)=1-e^{-\theta x^{\alpha}}\ ;x>0,\theta>0,\alpha>0 \qquad (3.2)$$

It can be easily shown that the Weibull distribution reduces to classical exponential distribution at $\alpha=1$. It should be noted that Weibull distribution is nothing but the power exponential distribution.

Taking $x=y^{\frac{1}{\alpha}}$ and thus $y=x^{\alpha}$ in (2.4), we have

$$g(y;\theta,\alpha)=f\left(y^{\alpha}\right)f'(y)=\theta e^{-\theta y^{\alpha}}\alpha y^{\alpha-1}=\theta\alpha y^{\alpha-1}e^{-\theta y^{\alpha}}$$

Which is the p.d.f. of Weibull distribution defined in (3.1)

Maximum Likelihood Estimation

Maximum likelihood estimates of the parameters of gamma distribution (GD): Assuming $(x_1, x_2, x_3, \ldots, x_n)$ be a random sample of size n from Gamma distribution (2.1), the likelihood function is given by

$$L = \left(\frac{\theta^\alpha}{\Gamma(\alpha)}\right)^n \left(\prod_{i=1}^n x_i\right)^{\alpha-1} e^{-n\theta\bar{x}}, \ \bar{x} \text{ being the sample mean}$$

The natural log likelihood function, ln L of Gamma distribution is thus given by

$$\ln L = n\left[\alpha \ln\theta - \ln(\Gamma(\alpha))\right] + (\alpha-1)\sum_{i=1}^n \ln x_i - n\theta\bar{x}$$

The maximum likelihood estimate (MLE) $\hat{\theta}$ and $\hat{\alpha}$ of parameters θ and α of gamma distribution can be obtained by solving the natural log likelihood equation using R software (Package Stat 4).

Maximum likelihood estimates of the parameters of weibull distribution

Assuming $(x_1, x_2, x_3, \ldots, x_n)$ be a random sample of size n from GD (3.1), the natural log likelihood function, ln L of Weibull distribution is given by

$$\ln L = \sum_{i=1}^n \ln f(x_i;\theta,\alpha) \ = n(\ln\theta + \ln\alpha) + (\alpha-1)\sum_{i=1}^n \ln x_i - \theta\sum_{i=1}^n x_i^\alpha$$

The maximum likelihood estimate (MLE) $\hat{\theta}$ and $\hat{\alpha}$ of parameters θ and α of Weibull distribution can be obtained by solving the natural log likelihood equation using R software (Package Stat 4).

Goodness of Fit and Applications

In this section, the goodness of fit and applications of gamma and Weibull distributions discussed for several lifetime data and fit is compared with exponential distribution. In order to compare gamma, Weibull, and exponential distributions, $-2\ln L$ and K-S Statistics (Kolmogorov-Smirnov Statistics) for fifteen data sets have been computed and presented in Table 1. The formula for K-S Statistics is defined as follow:

$K\text{-}S = \text{Sup}\left|F_n(x) - F_0(x)\right|$, where $F_n(x)$ is the empirical distribution function. The best distribution corresponds to lower values of $-2\ln L$ and K-S statistics.

From the table 1 it is clear that gamma distribution gives better fit in data sets 2,3,4,6,8,10,11,12,13, and 15 while Weibull distribution gives better fit in data sets 1,5,7,9, and 14 Data sets (1-15).

Data Set 1: The data set represents the strength of 1.5cm glass fibers measured at the National Physical Laboratory, England. Unfortunately, the units of measurements are not given in the paper, and they are taken from Smith & Naylor [3].

0.55	0.93	1.25	1.36	1.49	1.52	1.58	1.61	1.64	1.68	1.73	1.81
2.00	0.74	1.04	1.27	1.39	1.49	1.53	1.59	1.61	1.66	1.68	1.76
1.82	2.01	0.77	1.11	1.28	1.42	1.5	1.54	1.6	1.62	1.66	1.69
1.76	1.84	2.24	0.81	1.13	1.29	1.48	1.5	1.55	1.61	1.62	1.66
1.7	1.77	1.84	0.84	1.24	1.3	1.48	1.51	1.55	1.61	1.63	1.67
1.7	1.78	1.89									

Data Set 2: The data set is from Lawless [4]. The data given arose in tests on endurance of deep groove ball bearings. The data are the number of million revolutions before failure for each of the 23 ball bearings in the life tests and they are:

17.88	28.92	33.00	41.52	42.12	45.60	48.8	51.84	51.96	54.12	55.56	67.80
68.44	68.64	68.88	84.12	93.12	98.64	105.12	105.84	127.92	128.04	173.40	

Data Set 3: This data represents the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli, observed and reported by Bjerkedal [5].

10	33	44	56	59	72	74	77	92	93	96	100
100	102	105	107	107	108	108	108	109	112	113	115
116	120	121	122	122	124	130	134	136	139	144	146
153	159	160	163	163	168	171	172	176	183	195	196
197	202	213	215	216	222	230	231	240	245	251	253
254	254	278	293	327	342	347	361	402	432	458	555

Citation: Shanker R, Shukla KK, Shanker R, Leonida TS (2016) On Modeling of Lifetime Data Using Two-Parameter Gamma and Weibull Distributions. Biom Biostat Int J 4(5): 00107. DOI: 10.15406/bbij.2016.04.00107

Data Set 4: The data set reported by Efron [6] represent the survival times of a group of patients suffering from Head and Neck cancer disease and treated using radiotherapy (RT).

6.53	7	10.42	14.48	16.10	22.70	34	41.55	42	45.28	49.40	53.62
63	64	83	84	91	108	112	129	133	133	139	140
140	146	149	154	157	160	160	165	146	149	154	157
160	160	165	173	176	218	225	241	248	273	277	297
405	417	420	440	523	583	594	1101	1146	1417		

Data Set 5: The data set reported by Efron [6] represent the survival times of a group of patients suffering from Head and Neck cancer disease and treated using a combination of radiotherapy and chemotherapy (RT+CT).

12.20	23.56	23.74	25.87	31.98	37	41.35	47.38	55.46	58.36	63.47	68.46
78.26	74.47	81.43	84	92	94	110	112	119	127	130	133
140	146	155	159	173	179	194	195	209	249	281	319
339	432	469	519	633	725	817	1776				

Data set 6: This data set represents remission times (in months) of a random sample of 128 bladder cancer patients reported in Lee & Wang [7].

0.08	2.09	3.48	4.87	6.94	8.66	13.11	23.63	0.2	2.23	3.52	4.98
6.97	9.02	13.29	0.4	2.260	3.57	5.06	7.09	9.22	13.8	25.74	0.50
2.46	3.64	5.09	7.26	9.47	14.24	25.82	0.51	2.54	3.7	5.17	7.28
9.74	14.76	6.31	0.81	2.62	3.82	5.32	7.32	10.06	14.77	32.15	2.64
3.88	5.32	7.39	10.34	14.83	34.26	0.90	2.69	4.18	5.34	7.59	10.66
15.96	36.66	1.05	2.69	4.23	5.41	7.62	10.75	16.62	43.01	1.19	2.75
4.26	5.41	7.63	17.12	46.12	1.26	2.83	4.33	5.49	7.66	11.25	17.14
79.05	1.35	2.87	5.62	7.87	11.64	17.36	1.40	3.02	4.34	5.71	7.93
11.79	18.10	1.46	4.40	5.85	8.26	11.98	19.13	1.76	3.25	4.50	6.25
8.37	12.02	2.02	3.31	4.51	6.54	8.53	12.03	20.28	2.02	3.36	6.76
12.07	21.73	2.07	3.36	6.93	8.65	12.63	22.69				

Data Set 7: This data set is given by Linhart & Zucchini [8], which represents the failure times of the air conditioning system of an airplane:

23	261	87	7	120	14	62	47	225	71	246	21
42	20	5	12	120	11	3	14	71	11	14	11
16	90 1	16	52	95							

Data Set 8: This data set used by Bhaumik et al. [9], is vinyl chloride data obtained from clean upgradient monitoring wells in mg/l:

5.1	1.2	1.3	0.6	0.5	2.4	0.5	1.1	8	0.8	0.4	0.6
0.9	0.4	2	0.5	5.3	3.2	2.7	2.9	2.5	2.3	1	0.2
0.1	0.1	1.8	0.9	2	4	6.8	1.2	0.4	0.2		

Data set 9: This data set represents the waiting times (in minutes) before service of 100 Bank customers and examined and analyzed by Ghitany et al. [10] for fitting the Lindley [11] distribution.

0.8	0.8	1.3	1.5	1.8	1.9	1.9	2.1	2.6	2.7	2.9	3.1
3.2	3.3	3.5	3.6	4.0	4.1	4.2	4.2	4.3	4.3	4.4	4.4
4.6	4.7	4.7	4.8	4.9	4.9	5 .0	5.3	5.5	5.7	5.7	6.1
6.2	6.2	6.2	6.3	6.7	6.9	7.1	7.1	7.1	7.1	7.4	7.6
7.7	8.0	8.2	8.6	8.6	8.6	8.8	8.8	8.9	8.9	9.5	9.6
9.7	9.8	10.7	10.9	11 .0	11 .0	11.1	11.2	11.2	11.5	11.9	12.4
12.5	12.9	13 .0	13.1	13.3	13.6	13.7	13.9	14.1	15.4	15.4	17.3
17.3	18.1	18.2	18.4	18.9	19 .0	19.9	20.6	21.3	21.4	21.9	23 .0
27 .0	31.6	33.1	38.5								

Data Set 10: This data is for the times between successive failures of air conditioning equipment in a Boeing 720 airplane, Proschan [12].

74	57	48	29	502	12	70	21	29	386	59	27
153	26	326									

Data set 11: This data set represents the lifetime's data relating to relief times (in minutes) of 20 patients receiving an analgesic and reported by Gross & Clark [13].

1.1	1.4	1.3	1.7	1.9	1.8	1.6	2.2	1.7	2.7	4.1	1.8
1.5	1.2	1.4	3	1.7	2.3	1.6	2				

Data Set 12: This data set is the strength data of glass of the aircraft window reported by Fuller et al. [14].

18.83	20.8	21.66	23.03	23.23	24.05	24.321	25.5	25.5	25.8	26.69	26.77
26.78	27.05	27.67	29.9	31.11	33.2	33.73	33.8	33.9	34.76	35.75	35.91
36.98	37.08	37.09	39.58	44.05	45.29	45.381					

Data Set 13: The following data represent the tensile strength, measured in GPa, of 69 carbon fibers tested under tension at gauge lengths of 20mm, Bader & Priest [15].

1.312	1.314	1.479	1.552	1.700	1.803	1.861	1.865	1.944	1.958	1.966	1.997
2.006	2.021	2.027	2.055	2.063	2.098	2.140	2.179	2.224	2.240	2.253	2.270
2.272	2.274	2.301	2.301	2.359	2.382	2.382	2.426	2.434	2.435	2.478	2.490
2.511	2.514	2.535	2.554	2.566	2.570	2.586	2.629	2.633	2.642	2.648	2.684
2.697	2.726	2.770	2.773	2.800	2.809	2.818	2.821	2.848	2.880	2.954	3.012
3.067	3.084	3.090	3.096	3.128	3.233	3.433	3.585	3.858			

Data Set 14: The following data set represents the failure times (in minutes) for a sample of 15 electronic components in an accelerated life test, Lawless [4].

1.4	5.1	6.3	10.8	12.1	18.5	19.7	22.2	23.0	30.6	37.3	46.3
53.9	59.8	66.2									

Data Set 15: The following data set represents the number of cycles to failure for 25 100-cm specimens of yarn, tested at a particular strain level, Lawless [4].

15	20	38	42	61	76	86	98	121	146	149	157
175	176	180	180	198	220	224	251	264	282	321	325
653											

Table 1: ML Estimates, -2ln L, K-S Statistics and p-values of the fitted distributions of data sets 1 to 15

	Distribution	ML Estimates		-2In L	K-S Statistics	P-value
		$\hat{\theta}$	$\hat{\alpha}$			
Data 1	Gamma	11.5711	17.4355	47.903	0.809	0.000
	Weibull	0.0598	5.7796	30.413	0.803	0.000
	Exponential	0.6636		177.660	0.564	0.000
Data 2	Gamma	0.0558	4.0280	226.045	0.123	0.838
	Weibull	0.0021	1.4377	232.269	0.229	0.152
	Exponential	0.0138		242.870	0.307	0.019
Data 3	Gamma	0.0209	2.0833	788.495	0.996	0.000
	Weibull	0.0029	1.2849	795.750	0.177	0.021
	Exponential	0.0057		889.220	0.297	0.000
Data 4	Gamma	0.0046	1.0320	744.834	0.166	0.079
	Weibull	0.0059	0.9521	744.845	0.151	0.139
	Exponential	0.0045		744.881	0.16	0.101
Data 5	Gamma	0.0047	1.0476	564.029	0.148	0.259
	Weibull	0.0064	0.9404	563.68	0.129	0.419
	Exponential	0.0045		564.03	0.139	0.33
Data 6	Gamma	0.1287	1.1851	822.169	0.878	0.000
	Weibull	0.0946	1.0514	823.785	0.873	0.000
	Exponential	0.1085		824.371	0.868	0.000
Data 7	Gamma	0.0136	0.8127	304.335	0.947	0.000
	Weibull	0.0329	0.853	303.874	0.944	0.000
	Exponential	0.0167		305.25	0.954	0.000
Data 8	Gamma	0.5654	1.0627	110.826	0.937	0.000
	Weibull	0.5263	1.0102	110.899	0.934	0.000
	Exponential	0.532		110.901	0.934	0.000
Data 9	Gamma	0.2034	2.0095	634.6	0.043	0.993
	Weibull	0.0306	1.4573	637.461	0.057	0.9
	Exponential	0.1012		658.041	0.173	0.005
Data 10	Gamma	0.0076	0.9157	173.852	0.719	0.000
	Weibull	0.0032	1.1731	175.978	0.797	0.000
	Exponential	0.0083		173.94	0.74	0.000

Data 11	Gamma	5.0874	9.6662	35.637	0.609	0.000
	Weibull	0.1215	2.7869	41.173	0.587	0.000
	Exponential	0.5263		65.67	0.471	0.000
Data 12	Gamma	0.6146	18.9374	208.231	0.135	0.577
	Weibull	0.0021	1.8108	241.63	0.368	0.000
	Exponential	0.0325		274.531	0.458	0.000
Data 13	Gamma	9.2878	22.8042	101.971	0.057	0.979
	Weibull	0.0065	5.1692	103.482	0.066	0.917
	Exponential	0.4079		261.701	0.448	0.000
Data 14	Gamma	0.0523	1.4412	128.372	0.102	0.992
	Weibull	0.0123	1.2978	128.041	0.099	0.995
	Exponential	0.0363		129.47	0.156	0.807
Data 15	Gamma	0.0101	1.8082	304.876	0.136	0.748
	Weibull	0.0027	1.1423	306.687	0.191	0.32
	Exponential	0.0056		309.181	0.202	0.257

Concluding Remarks

In this paper an attempt has been made to have the comparative and detailed study of two-parameter gamma and Weibull distributions for modeling lifetime data from various fields of knowledge. Since exponential distribution is a particular case of both gamma and Weibull distributions and the exponential distribution is a classical distribution for modeling lifetime data, the goodness of fit of both gamma and Weibull distributions are compared with exponential distribution. From the fitting of exponential, Weibull and gamma distributions it is obvious that in majority of data sets gamma distribution gives better fit than both Weibull and exponential distributions.

Acknowledgement

None.

Conflict of Interest

None.

References

1. Stacy EW (1962) A generalization of the gamma distribution. Annals of Mathematical Statistical 33(3): 1187-1192.
2. Stacy EW, Mihram GA (1965) Parametric estimation for a generalized gamma distribution. Technometrics 7(3): 349-358.
3. Smith RL, Naylor JC (1987) A comparison of Maximum likelihood and Bayesian estimators for the three parameter Weibull distribution. Applied Statistics 36(3): 358-369.
4. Lawless JF (2003) Statistical models and methods for lifetime data, John Wiley and Sons, New York, USA.
5. Bjerkedal T (1960) Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. Am J Hyg 72 (1): 130-148.
6. Efron B (1988) Logistic regression, survival analysis and the Kaplan-Meier curve. Journal of the American Statistical Association 83: 414-425.
7. Lee ET, Wang JW (2003) Statistical methods for survival data analysis, 3rd edition, John Wiley and Sons, New York, NY, USA.
8. Linhart H, Zucchini W (1986) Model Selection, John Wiley, New York, USA.
9. Bhaumik DK, Kapur K, Gibbons RD (2009) Testing Parameters of a Gamma Distribution for Small Samples. Technometrics 51: 326-334.
10. Ghitany ME, Atieh B, Nadarajah S (2008) Lindley distribution and its Application .Mathematics Computing and Simulation 78(4): 493-506.
11. Lindley DV (1958) Fiducial distributions and Bayes' Theorem. Journal of the Royal Statistical Society Series B 20: 102-107.
12. Proschan F (1963) Theoretical explanation of observed decreasing failure rate. Technometrics 5(3): 375-383.
13. Gross AJ, Clark VA (1975) Survival Distributions: Reliability Applications in the Biometrical Sciences, John Wiley, New York, USA.
14. Fuller EJ, Frieman S, Quinn J, Quinn G, Carter W (1994) Fracture mechanics approach to the design of glass aircraft windows: A case study SPIE Proc 2286, 419-430.
15. Bader MG, Priest AM (1982) Statistical aspects of fiber and bundle strength inhybrid composites. In: Hayashi T, Kawata K, Umekawa S (Eds), Progressin Science in Engineering Composites, ICCM-IV, Tokyo, 1129-1136.