Executive summary of the research work done by Rani Sebastian for the period of June 2015 to August 2017

UGC Reference No. : 1592-MRP/14-15/KLCA019/UGC-SWRO

Title of research project: Generalisation of Gompertz distribution and their Applications in Reliability.

The Gompertz distribution plays an important role in modeling survival times, human mortality and actuarial tables. According to the literature, the Gompertz distribution was formulated by Gompertz (1825) to fit mortality tables. Johnson et al. (1995) and Garg et al. (1970) studied the properties of the Gompertz distribution and obtained the maximum likelihood estimates for the parameters. Johnson et al (1994) note that the Gompertz distribution is a truncated extreme value distribution. Gompertz distributions can be viewed as extensions of the exponential distributions because exponential distributions are limits of sequences of Gompertz distributions. Makeham (1860) examined the fit to actuarial data provided by the Gompertz distribution and observed with specific examples that he could improve the fit with the modification now known as the Gompertz Makeham distribution. Burga et al. (2009) discussed the stress-strength reliability in Gompertz case.

A random variable X is said to have a Gompertz distribution if its pdf is

$$f_G(x) = \beta e^{\alpha x} e^{-\frac{\beta}{\alpha}(e^{\alpha x}-1)}; x \geq 0, \alpha, \beta > 0$$

Corresponding survival function is

$$\overline{F}_G(x) = e^{-\frac{\beta}{\alpha}(e^{\alpha x}-1)}$$

Hazard rate function is

$$h_G(x) = \beta e^{\alpha x}$$

Apply the Marshall Olkin technique to Gompertz distribution we get the distribution function of the Marshall Olkin Gompertz distribution as

$$G_G(x) = \frac{1 - e^{-\frac{\beta}{\alpha}(e^{\alpha x}-1)}}{1 + (p-1)e^{-\frac{\beta}{\alpha}(e^{\alpha x}-1)}}$$

Corresponding pdf is

$$g_G(x) = \frac{p\beta e^{-\frac{\beta}{\alpha}(e^{\alpha x}-1)+\alpha x}}{(1 + (p-1)e^{-\frac{\beta}{\alpha}(e^{\alpha x}-1)})^2}$$

hazard rate function is

$$h_G(x) = \frac{\beta e^{\alpha x}}{1 + (p-1)e^{-\frac{\beta}{\alpha}(e^{\alpha x}-1)}}$$

Characteristic Properties

Theorem 0.0.1. *Marshall-Olkin Gompertz distribution(MOG) is geometric extreme stable.*

Theorem 0.0.2. *Let* $\{X_i, i \geq 1\}$ *be a sequence of independent and identically distributed random variables with common survival function* $\overline{F}(x)$ *and* N *be a geometric random variable with parameter* p *and* $P(N = n) = pq^{n-1}; n = 1, 2, ...; 0 < p < 1, q = 1 - p$. *wich is independent of* $\{X_i\}$ *for all* $i \geq 1$. *Let* $U_N = \min_{1 \leq i \leq N} X_i$. *Then* $\{U_N\}$ *is distributed as MOG if and only if* $\{X_i\}$ *is distributed as Gompertz.*

Theorem 0.0.3. *Consider an* $AR(1)$ *structure given by*

$$X_n = \begin{cases} \varepsilon_n & \text{with probability } p \\ \min(X_{n-1}, \varepsilon_n) & \text{with probability } 1-p \end{cases} \tag{1}$$

where $0 \leq p \leq 1$, $\{\varepsilon_n\}$ *is a sequence of independent and identically distributed random variables independent of* $\{X_{n-1}, X_{n-2}, \dots\}$. *Then* $\{X_n\}$ *is a stationary Markovian* $AR(1)$ *process with MOG marginals if and only if* $\{\varepsilon_n\}$ *is distributed as Gompertz distribution.*

Theorem 0.0.4. *Consider an* AR (1) *structure given by*

$$X_n = \begin{cases} \epsilon_n & w.p. \quad q \\ X_{n-1} & w.p. \quad p(1-q) \\ \min(pX_{n-1}, \epsilon_n) & w.p. \quad (1-p)(1-q) \end{cases} \quad , \; n \geq 1. \tag{2}$$

where w.p. means with probability. If $q = 0$ *we get the ordinary process.*

where $0 \leq p \leq 1$, $\{\varepsilon_n\}$ *is a sequence of independent and identically distributed random variables independent of* $\{X_{n-1}, X_{n-2}, \dots\}$. *Then* $\{X_n\}$ *is a stationary Markovian* AR (1) *process with MOG marginals if and only if* $\{\varepsilon_n\}$ *is distributed as Gompertz distribution.*

0.0.1 Estimation of reliability

Let X and Y be two independent random variables following Marshall Olkin Gompertz distribution with parameters α_1, β, p and α_2, β, p respectively. Then according to Gupta et al (2009) the reliability of the system given by $P(X > Y)$ where X is the strength and Y is the stress is given by

$$
\begin{aligned}
R = P(X > Y) &= \int_{-\infty}^{\infty} P(X > Y/Y = y) g_Y(y) dy \\
&= \int_{0}^{\infty} \frac{\alpha_1 \beta e^{-\frac{\beta}{\alpha}(e^{\alpha y}-1)+\alpha y}}{(1+(\alpha_1 - 1)e^{-\frac{\beta}{\alpha}(e^{\alpha y}-1)})^2} \frac{\alpha_2 e^{-\frac{\beta}{\alpha}(e^{\alpha y}-1)}}{1-(1-\alpha_2)e^{-\frac{\beta}{\alpha}(e^{\alpha y}-1)}} dy \\
&= \frac{\frac{\alpha_1}{\alpha_2}}{(\frac{\alpha_1}{\alpha_2} - 1)^2} \left[-\ln \frac{\alpha_1}{\alpha_2} + \frac{\alpha_1}{\alpha_2} - 1 \right]
\end{aligned}
$$

Let $(x_1, \ldots, x_m)$ and $(y_1, \ldots, y_n)$ be two independent random samples of sizes m and n from Marshall-Olkin Gompertz distribution with Marshall-Olkin parameters α_1 and α_2, respectively, and common unknown parameters β and p. L is the log likelihood function, then maximum likelihood estimates of the unknown parameters α_1, α_2 are the solutions of the non-linear equations $\frac{\partial L}{\partial \alpha_1} = 0$ and $\frac{\partial L}{\partial \alpha_2} = 0$ respectively. The elements of information matrix are

$$
\begin{aligned}
I_{11} &= -E\left(\frac{\partial^2 L}{\partial \alpha_1^2}\right) \\
&= \frac{m}{3\alpha_1^2}
\end{aligned}
$$

Similarly,

$$
I_{22} = -E\left(\frac{\partial^2 L}{\partial \alpha_2^2}\right) = \frac{n}{3\alpha_2^2}
$$

$$
I_{12} = I_{21} = -E\left(\frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_2}\right) = 0.
$$

By the property of m.l.e for $m \to \infty, n \to \infty$, we obtain that

$$(\sqrt{m}(\hat{\alpha}_1 - \alpha_1), \sqrt{n}(\hat{\alpha}_2 - \alpha_2))^T \xrightarrow{d} N_2\left(\mathbf{0}, \operatorname{diag}\{a_{11}^{-1}, a_{22}^{-1}\}\right),$$

where $a_{11} = \lim_{m,n\to\infty} \frac{1}{m} I_{11} = \frac{1}{3\alpha_1^2}$ and $a_{22} = \lim_{m,n\to\infty} \frac{1}{n} I_{22} = \frac{1}{3\alpha_2^2}$. The 95% confidence interval for R is given by

$$\hat{R} \mp 1.96\ \hat{\alpha}_1 b_1(\hat{\alpha}_1, \hat{\alpha}_2)\sqrt{\frac{3}{m} + \frac{3}{n}},$$

where $\hat{R} = R(\hat{\alpha}_1, \hat{\alpha}_2)$ is the estimator of R and

$$b_1(\alpha_1, \alpha_2) = \frac{\partial R}{\partial \alpha_1} = \frac{\alpha_2}{(\alpha_1 - \alpha_2)^3}\left[2(\alpha_1 - \alpha_2) + (\alpha_1 + \alpha_2)\log\frac{\alpha_2}{\alpha_1}\right].$$

0.0.2 Data Analysis and Modeling

In this section we analyze a real data set and compare Marshall Olkin Gompertz distribution with Gompertz distribution. The P-P plots for the two distributions are given. Estimated values are given. From that we can conclude that the Marshall Olkin Gompertz distribution is a better fit.

0.0.3 Simulation Study

We generate $N = 1000$ sets of X-samples and Y-samples from Marshall-Olkin Gompertz distribution with parameters α_1, β, p and α_2, β, p respectively. The combinations of samples of sizes $m = 20$, 30, 40 and

$n = 20, 30, 40$ are considered. The estimates of α_1 and α_2 are then obtained from each sample to obtain $\hat{R}$. The validity of the estimate of R is discussed by the measures:

1) Average bias of the simulated N estimates of R:

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{R}_i - R)$$

2) Average mean square error of the simulated N estimates of R:

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{R}_i - R)^2$$

3) Average length of the asymptotic 95% confidence intervals of R:

$$\frac{1}{N}\sum_{i=1}^{N}2(1.96)\hat{\alpha}_{1i}\ b_{1i}(\hat{\alpha}\alpha_{1i}, \hat{\alpha}\alpha_{2i})\sqrt{\frac{3}{m}+\frac{3}{n}}$$

4) The coverage pprobability of the N simulated confidence intervals given by the proportion of such interval that include the parameter R.

0.0.4 Conclusion

The Gompertz distribution plays an important role in modeling survival times, human mortality and actuarial tables. In this paper as a generalization of the Gompertz distribution, Marshall Olkin Gompertz distribution is considered. A three parameter $AR(1)$ process is also considered. Motivation behind this study is due to the following properties attained by the probability distribution obtained by applying Marshall Olkin technique. It gives added flexibility for the underlying model. It offer a wide range of behavior than the basic distribution from which they are derived. The property that extended distributions can have an interesting hazard function depending on the value of the added parameter. It can be used to model real situation in a better manner than the basic distribution. The geometric minimum stability property of newly formed distribution can be utilized to develop a stationary process. When X and Y are two independent random variables following Marshall Olkin Gompertz distribution, then average bias, average mean square error, average confidence length and coverage probability of the of the simulated estimates of reliability R is computed. We apply the models to a real data set and compare Marshall Olkin Gompertz distribution with Gompertz distribution and establishes that Marshall Olkin Gompertz distribution is a better fit.

Research papers presented/published/ Accepted for publication/ Communicated for publication

1. Presented a paper in the International Conference on Statistics for Twenty first Century-2015, December 17-19,2015;Organized by University of Kerala, Trivandrum.

2. Presented a paper in the National Seminar on Applied Statistical methodology with special emphasis on Time series Analysis, February12-13,2016; Organized by Nirmala College Muvattupuzha.

3. Presented a paper on Generalisations of Gompertz distribution and their applications in time series in the second international conference on Statistics for twenty first century during 21-23 December 2016 organized by Department of Statistics, University of Kerala.

4. Presented a paper "Gompertz distribution : Some Aspects" in the National Seminar on StatisticalTechniques in Applied Areas held during 27-28 February 2017 organised by Department of Statistics St.Thomas College Thrissur.

5. Generalisations of Gompertz distribution and their applications in reliability(2016)Journal of Kerala Statistical Association **Accepted for Publication.**

References

1. Burga, S., Mehmet, F. K., Abd-Elfattah, A.M. (2009) Comparison of estimators for stress-strength reliability in Gompertz case, *Hacettepe Journal of Mathematics and Statistics*, 38 (3), 339-349.

2. EL-Bassiouny, A. H., Abdo, N. F.(2009) Reliability Properties of Extended Makeham Distributions, *Computational methods in science and technology*,15(2), 143-149.

3. Garg, M.L., Rao, B.R. and Redmond, C.K. (1970). Maximum Likelihood Estimation of the Parameters of the Gompertz Survival Function, Journal of the Royal Statistical Society C, 19, 152-159.

4. Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality and on the New Mode of Determining the Value of Life Contingencies, Phil. Trans. R. Soc. A., 115, 513-580.

5. Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). Continuous Univariate Distributions, Vol. 1, 2nd ed., John Wiley and Sons, (pp. 25-26, 81-85).

6. Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distributions, Vol. 2, 2nd ed., John Wiley and Sons.

7. Jose, K.K., Rani Sebastian (2010)Marshall-Olkin Gumbel bivariate distributions and their applications in stress-strength reliability analysis and autoregressive minification processes, *Journal of Statistics and Applications*, 5 (203-223)

8. Makeham, W.M.(1860) On the Law of Mortality and the Construction of Annuity Tables,*The Assurance Magazine and Journal of the Institute of Actuaries VIII* 301-310.

9. Marshall, A. W., Olkin, I.(1997) A new method for adding a parameter to a family of distributions with application to the exponential and weibull families, *Biometrica.* 84(3), 641-652.

10. Tavares, V.L. (1980) An exponential Markovian stationary process, *J. Appl. Prob.*, 17, 1117-1120.