**International Journal of Mathematical Research**
*2016 Vol. 5, No. 2, 179-187.*
*ISSN(e): 2306-2223*
*ISSN(p): 2311-7427*
*DOI: 10.18488/journal.24/2016.5.2/24.2.179.187*


# CONSISTENCY OF A MIXTURE MODEL OF TWO DIFFERENT DISTRIBUTIONS APPROACH TO THE ANALYSIS OF BUYING BEHAVIOUR DATA

**Akomolafe Abayomi.A**[1]
[1]*Department of Statistics, Federal University of Technology, Akure, Nigeria*

## ABSTRACT
*A four-parameter probability distribution, which includes a wide variety of curve shapes, is presented. Because of the flexibility, generality, and simplicity of the distribution, it is useful in the representation of data when the underlying model is unknown. Further important applications of the distribution include the modeling and subsequent generation of random variates for simulation studies and Monte Carlo sampling studies of the robustness of statistical procedures. This research centered on combining these two distributions that will simultaneously capture the rate of occurrence of a phenomenon, especially buying behaviour and the actual performance of that phenomenon as well as tracking and forecasting future purchasing pattern based the data. Further important applications of the distribution include the modeling and subsequent generation of random variates for simulation studies of the robustness of statistical procedures. To do this, specification of the hybrid model named Exponential- Gamma mixture model is given and followed by its derivation. The concluding part of the paper depicts an example of the areas of its application.*

**Keywords:** Data fitting, Probability distribution, Timing process, Exponential – gamma model, Buying behavior, Nonstationarity.

**Received:** 19 December 2015/ **Revised:** 27 January 2016/ **Accepted:** 19 September 2016/ **Published:** 8 November 2016

## Contribution/ Originality

This study contributes to the existing literature by combining exponential-gamma mixture as a hybrid model to see its rate of tracking and forecasting purchasing pattern. This is carried out by using our newly arrived formula in analyzing both the real life and simulated data in other to ascertain the uniqueness of combined probability distribution in forecasting future customer buying behavior data for producers effective planning and administration.

## 1. INTRODUCTION

The stable distributions have a wide area of applications: probability theory, communications theory, physics, astronomy, economics, and sociology. Reasons for fitting a distribution to a set of data have been summarized by many researchers, the desire for objectivity, the need for auto-mating the data analysis, and interest in the values of the distribution parameters. Although various empir-ical distributions already exist, e.g., the Pearson sys-tem and the Johnson system (see Chapter 7 of Hahn and Shapiro) and the Burr distribution, we are presenting another distribution because of its simplicity, flexibility, and generality. The importance of exponential and gamma distributions in the field of mathematical statistics cannot be over emphasized. Both distributions are timing in nature; most importantly exponential distribution can also be termed as waiting time distribution which is analogous to the geometric and Poisson distributions. A mixture distribution describes a population which comprises two or more subpopulations combined in fixed proportions. A wide variety of curve shapes are possible with this distribution. Probability models of this type are utilized in situations where the sample observations can only be drawn from the whole population and not separately from the individual component populations. Such cases typically occur due to a difficulty in distinguishing the underlying subpopulations forming the mixed population. Consequently, mixture models are employed in fields which include pattern recognition, biology, chemistry, geology, medicine, and actuarial science. An example taken from the discipline of biology will serve to illustrate the type of situation in which mixture models arise. Biologists are often faced with the task of inferring population parameters on the basis of samples collected during the course of field experiments.

A particular characteristic of the customers' behaviour of interest is measured for each individual in the sample. An important consideration by Akomolafe [1] is that when taking these measurements incorporating age or sex of the individual, since the distribution of many characteristics is dependent on these. However, the age or sex of a customer is not always readily determined when measurements are performed in the natural environment. Hence, observations must be drawn from the mixed population, where the component populations are the different age groups or sexes of the customer. These two distributions had been found useful in any situation that is time variant. Given the robustness of these two distributions, it is then pertinent to simultaneously look at their behavior which is the subject matter of this paper.

An attempt has been made to specify the mixture of these two distributions which is tagged exponential-gamma model. Thereafter, a comprehensive derivation of the model is given for more clarification. The paper is concluded by an exposition of major area of its application which is capturing evolving customer visit buying behavior.

## 2. SPECIFICATION OF THE MODEL

Assume a timing process phenomenon;an appropriate robust starting point on repeated trials can be modeled as an exponential-gamma timing process. This is to say that each inter repeated trials time is assumed to be exponentially distributed governed by a rate,$\lambda_i$ . Furthermore, these individual trials rate of occurrence vary across the population. This heterogeneity can be captured by a gamma distribution with shape parameter r, and scale parameter α.

These distributions are given by the following

$$f\left(t_{ij} : \lambda_i\right) = \lambda_i e^{-\lambda_i\left(t_{ij} - t_i(j-1)\right)} \quad \ldots\ldots(1)$$

*And*

$$g\left(\lambda_i, r, \alpha\right) = \frac{\lambda_i^{r-1} \alpha^r e^{-\alpha\lambda i}}{\Gamma r} \quad \ldots\ldots(2)$$

Where $\lambda_i$ is individual trial,i's latent of occurrence,$t_{ij}$ is the time when the $j^{th}$ repeat of occurrence occurred, and $t_i$ is the time of first observed item occurred. For a single occurrence occasion, this leads to the following familiar Exponential – Gamma Mixture Model:

$$f\left(t_{ij} : r, \alpha\right) = \int_0^\infty f\left(t_{ij}, \lambda_i\right) \quad \ldots\ldots(3)$$

$$g(\lambda_i : r, \alpha) d\lambda = \frac{r}{\alpha}\left[\frac{\alpha}{\alpha + \left(t_{ij} - t_i(j-1)\right)}\right]^{r+1} \quad \ldots\ldots(4)$$

## 3. DERIVATION OF EXPONENTIAL – GAMMA MIXTURE MODEL

Given the Exponential model:

$$f(t_{ij} : \lambda_i) = \lambda_i e^{-\lambda_i\left(t_{ij} - t_i(j-1)\right)}$$

And the Gamma model:

$$g\,(\lambda_i\,, r, \alpha) = \frac{\lambda_i^{r-1}\alpha^r e^{-\alpha\lambda i}}{\Gamma \mathrm{r}}$$

Where

$$\lambda_i = individual\ i'sunpredicted\ rate\ of\ buying$$
$$t_{ij} = is\ the\ day\ when\ the\ jth\ repeat\ buying\ occurred$$
$$for\ j = 1$$
$$t_{i\,(j-1)} = the\ day\ of\ their\ first\ trial.$$

And

$r$ = shape parameter [is also a quantity of buying]

α = scale parameter [is also the time rate of buying]

for a single visit occasion, this leads to the following familiar exponential – gamma mixture model:

i.e.

$$f\left(t_{ij} : r, \alpha\right) = \int_0^\infty f\left(t_{ij}\,, \lambda_i\right)\ . g(\lambda_i : r, \alpha) d\lambda = \frac{r}{\alpha}\left[\frac{\alpha}{\alpha + \left(t_{ij} - t_i(j-1)\right)}\right]^{r+1} \ldots\ldots\ldots(5)$$

$$f\left(t_{ij}/\lambda_i\right). g\,(\lambda_i/r, \alpha) = \int \lambda_i e^{-\lambda_i\left(t_{ij} - t_i(j-1)\right)}. \frac{\lambda_i^{r-1}\alpha^r e^{-\alpha}}{\Gamma r}\ d\lambda \ldots\ldots\ldots(6)$$

$$f\left(t_{ij}/r, \alpha\right) = \int_0^\infty f\left(t_{ij}/\lambda_i\right). g\,(\lambda_i : r, \alpha) d\lambda \ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

$$= \int_0^\infty \alpha^r \lambda_i \lambda_i^{r-1} \frac{e^{-\alpha\lambda_i} e^{-\lambda_i\,(t_{ij} - t_i\,(j-1)}}{\Gamma r} d\lambda \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8)$$

$$= \frac{\alpha^r}{\Gamma r}\int_0^\infty \lambda_i^{r-r+r}\ e^{-\alpha\lambda_i - \lambda_i t_i + \lambda t_i\,(j-1)} d\lambda \ldots\ldots\ldots\ldots\ldots\ldots\ldots(9)$$

$$= \frac{\alpha^r}{\Gamma r}\int_0^\infty \lambda_i^r\ e^{-\lambda_i(\alpha + t_{ij} + t_i\,(j-1)}\ d\lambda \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(10)$$

$$\text{Let} t_{ij} - t_i(j-1) = K$$

$$\therefore f\left(t_{ij}/r, \alpha\right) = \frac{\alpha^r}{\Gamma r}\int_0^\infty \lambda_i^r\ e^{-\lambda_i(\alpha + k)}\ d\lambda \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(11)$$

Put

$$\lambda_i(\alpha + k) = m \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(12)$$

$$\lambda = \frac{m}{(\alpha + k)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(13)$$

$$m = \lambda_i\,(\alpha + k) \ldots\ldots\ldots\ldots\ldots\ldots(14)$$

Differentiate equation 10 with respect to λ:

$$\frac{dm}{d\lambda} = (\alpha + k) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(15)$$

$$d\lambda = \frac{dm}{(\alpha + k)\ldots..} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(16)$$

Substitute for λ and dλ in equation (7)

i.e.

$$f\left(t_{ij}/r,\alpha\right) = \frac{\alpha^r}{\Gamma r}\int_0^\infty \left(\frac{m}{\alpha+k}\right)^r e^{-m}\,\frac{dm}{(\alpha+k)^1} \ldots\ldots\ldots\ldots\ldots\ldots(17)$$

$$= \frac{\alpha^r}{\Gamma r}\int_0^\infty \frac{m^r}{(\alpha+k)^r} e^{-m}\,.\frac{1}{(\alpha+k)^1}\,dm \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(18)$$

$$= \frac{\alpha^r}{\Gamma r}\int_0^\infty \frac{m^r}{(\alpha+k)^{r+1}} e^{-m}\,dm \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(19)$$

$$= \frac{\alpha^r}{\Gamma r}\,.\frac{1}{(\alpha+k)^{r+1}}\int_0^\infty m^r e^{-m}\,dm \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(20)$$

By Gamma distribution or Gamma function

Where

$$\int_0^\infty m^r e^{-m}\,dm = \Gamma r+1 = r\Gamma r = \frac{\alpha^r}{\Gamma r}\,.\frac{1}{(\alpha+k)^{r+1}}\,.r\Gamma r \ldots\ldots\ldots\ldots\ldots\ldots(21)$$

$$= r\,\frac{\alpha^r}{(\alpha+k)^{r+1}} \equiv r\,.\frac{\alpha^{r+1}}{\alpha}\,.\left(\frac{1}{(\alpha+k)^{r+1}}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(22)$$

$$\equiv \frac{r}{\alpha}\,.\frac{\alpha^{r+1}}{(\alpha+k)^{r+1}} \equiv \frac{r}{\alpha}\,.\left(\frac{\alpha}{\alpha+k}\right)^{r+1}$$

## 4. LIKELIHOOD SPECIFICATION

When estimating the ordinary (stationary) exponential-gamma model, there are two ways of obtaining the likelihood function for a given individual. The usual approach is to specify the individual-level likelihood function, conditional on that person's (unobserved) value of $\lambda_i$. This likelihood is the product of $J_i$ exponential timing terms, where $J_i$ is the number of repeat buying made by panelist $i$, plus an additional term to account for the right-censoring that occurs between that customer's last arrival and the end of the observed calibration period (at time $T$):

$$L_i/\lambda_i = \lambda_i e^{-\lambda_i\left(t_{i1}-t_{10}\right)}.\lambda_i e^{-\lambda_i\left(t_{12}-t_{i1}\right)}\ldots.\lambda_i e^{-\lambda_i\left(t_i J_i - t_{i(J_i-1)}\right)}.e^{-\lambda_i\left(T-t_i J_i\right)} \tag{23}$$

To get the unconditional likelihood we then integrate across all possible values of $\lambda$, using the gamma distribution as a weighting function:

$$L_i\left|r,\alpha = \int_0^\infty L_i\right|\lambda_i.gamma\left(\lambda_i;r,\alpha\right)d\lambda_i \tag{24}$$

Where gamma( $\lambda_i$; $r$, $\alpha$ ) denotes the gamma distribution as shown in (1). This yields the usual exponential-gamma likelihood, which can be multiplied across the $N$ panelists to get the overall likelihood for parameter estimation purposes:

$$L = \prod_{i=1}^{N}\frac{\Gamma\left(r+J_i\right)}{\Gamma(r)}\left(\frac{\alpha}{\alpha+T-t_{i0}}\right)^r\left(\frac{1}{\alpha+T-t_{i0}}\right)^{J_i} \tag{25}$$

An alternative path that leads to the same result is to perform the gamma integration separately for each of the $J_i+1$ exponential terms, and then multiply them together at the end. This involves the use of Bayes Theorem to refine our "guess" about each individual's value of $\lambda_i$ after each arrival occurs. Specifically, it is easy to show that if someone's first repeat visit occurs at time $t_{ij}$, then:

$$g\left(\lambda_{i2}\middle|\text{arrival at } t_{i1}\right) = gamma\left(r+1, \alpha + t_{i1} - t_{i0}\right) \tag{26}$$

The gamma distribution governing the rate of buying for subsequent arrivals follows:

$$g\left(\lambda_{i(j+1)}\middle|\text{arrival at } t_{ij}\right) = gamma\left(r+j, \alpha + t_{ij} - t_{i0}\right) \tag{27}$$

Using this logic, we can re-express the likelihood as the product of separate EG terms

$$L = \prod_{i=1}^{N}\prod_{j=1}^{J}\left(\frac{r+j+1}{\alpha + t_{i(j-1)} - t_{i0}}\right)\left(\frac{\alpha + t_{i(j-1)} - t_{i0}}{\alpha + t_{ij} - t_{i0}}\right)^{r+j} .S\left(T - t_{iJ_i}\right) \tag{28}$$

Which collapses into the same expression as (24).

When we introduce the nonstationary updating distribution, the multipliers ($c_{ij}$) change the value of $\lambda_i$ from visit to visit, thereby requiring us to use the sequential approach given in (28) to derive the complete likelihood function. We need to capture two forms of updating after each visit: one due to the usual Bayesian updating process (which is associated with stationary behaviour given by (25)) and the other due to the effects of the stochastic evolution process. Therefore, the distribution of buying rates at each repeat visit level is the product of two gamma distributed random variables – one associated with the updating multiplier and one capturing the previous visiting rate. For the case of panelist *i* making her *j*th repeat visit at time $t_{ij}$:

$$G\left(\lambda_{i(j+1)}\middle|\text{arrival at } t_{ij}\right) = gamma\left(r+j, \alpha + t_{ij} - t_{i0}\right).\, gamma\left(s, \beta\right) \tag{29}$$

One issue with this approach is that the product of two gamma random variables does not lend itself to a tractable analytic solution. However, there is an established result (see, e.g., [2] suggesting that the product of two gamma distributed random variables can be approximated by yet another gamma distribution, obtained by multiplying the first two moments about the origins of the original distributions:

$$m_1^{(\lambda_{i(j+1)})} = m_1^{(\lambda_{ij})} \times m_1^{(c_{ij})}$$
$$\text{and}$$
$$m_2^{(\lambda_{i(j+1)})} = m_2^{(\lambda_{ij})} \times m_2^{(c_{ij})} \tag{30}$$

As shown in Appendix A, this moment-matching approximation, used in conjunction with Bayesian updating, allows us to recover the updated gamma parameters that determine the rate of buying, $\lambda_{ij}$, for panelist *i*'s *j*th repeat visit as follows:

$$r(i, j+1) = \frac{\left[r(i,j)+1\right].s}{\left[r(i,j)+2\right].(s+1) - \left[r(i,j)+1\right].s} \tag{31}$$

$$\alpha(i, j+1) = \frac{\lfloor \alpha(i, j) + t_{ij} - t_{i(j-1)} \rfloor . \beta}{[r(i, j)+2].(s+1) - [r(i, j)+1].s} \tag{32}$$

Where $r(i, 1)$ and j($i$, 1) are equal to the initial values of $r$ and s as estimated by maximizing the likelihood function specified in (28).

## 5. APPLICATION

We performed 20 separate simulations to verify the accuracy of using such a moment-matching approximation. In each simulation, we first generated 1000 random draws from a gamma distribution with randomly determined shape and scale parameters to represent initial $\lambda$ values. Then, a matrix of updating multipliers was also simulated for a series of five updates (i.e., five future repeat visits). Each 1000x5 matrix was generated by taking draws from a gamma distribution, again with randomly determined shape and scale parameters, where columns, one through five represented the updates after one to five visits. The updated $\lambda$ series after five repeat visits was calculated using two methods

(1) direct (numerical) multiplication of the 1000 initial $\lambda$'s and the five updating series or

(2) Randomly drawing 1000 values from the distribution resulting from the moment-matching approximation across all five updates.

A Kolmogorov-Smirnov test of fit indicated that, for each of the 20 simulations, the distribution of values resulting from the moment-matching approximation is not significantly different from that resulting from the direct multiplication of these random variables. Therefore, we are confident that the moment-matching approximation accurately captures the gamma distributed updating process we wish to model.

After incorporating the evolution process into our model, the likelihood function to be maximized follows:

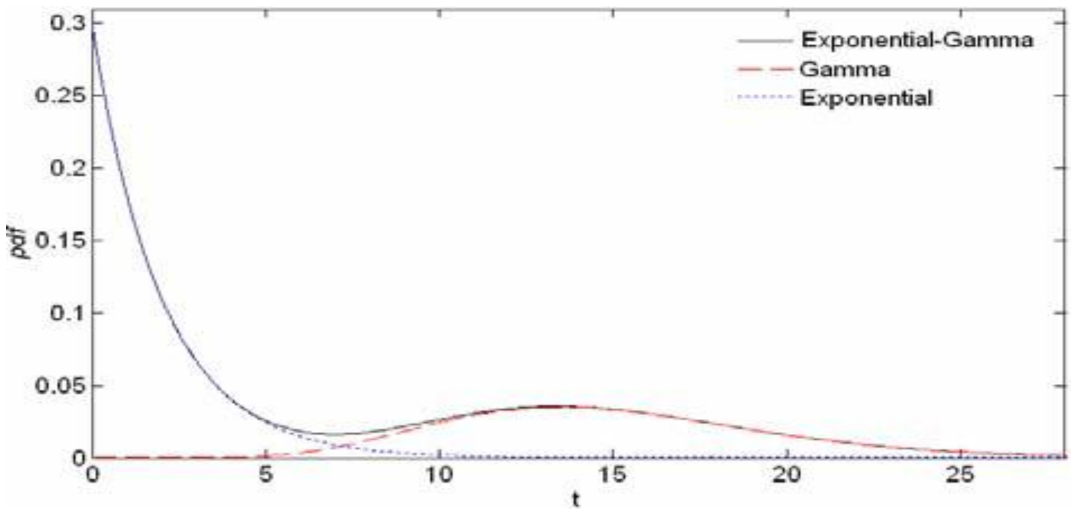
$$L = \prod_{i=1}^{N} \prod_{j=1}^{J} \left( \frac{r(i, j)}{\alpha(i, j)} \right) \left( \frac{\alpha(i, j)}{\alpha(i, j) + t_{ij} - t_{i(j-1)}} \right)^{r_{(i,j)}+1} . S\left(T - t_{iJ_i}\right) \tag{33}$$

Where $r(i,j)$ and $\alpha$ $(i,j)$ are defined in equations (31) and (32) while the survival function, $S(T\text{-}tij)$, is defined as:

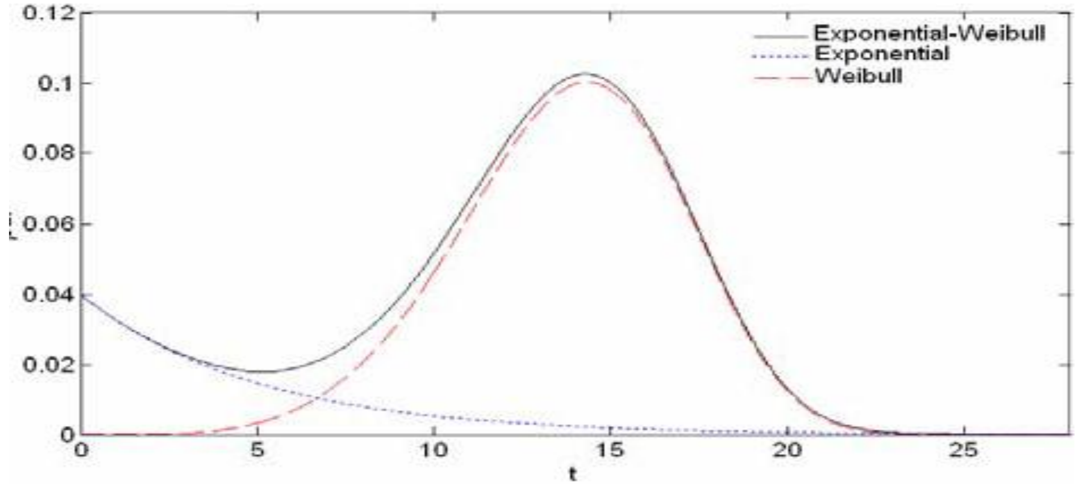
$$S\left(T - t_{ij}\right) = \left( \frac{\alpha(i, J_i + 1)}{\alpha(i, J_i + 1) + T - t_{iJ_i}} \right)^{r(i, J_i + 1)} \tag{34}$$

## 6. RESULTS

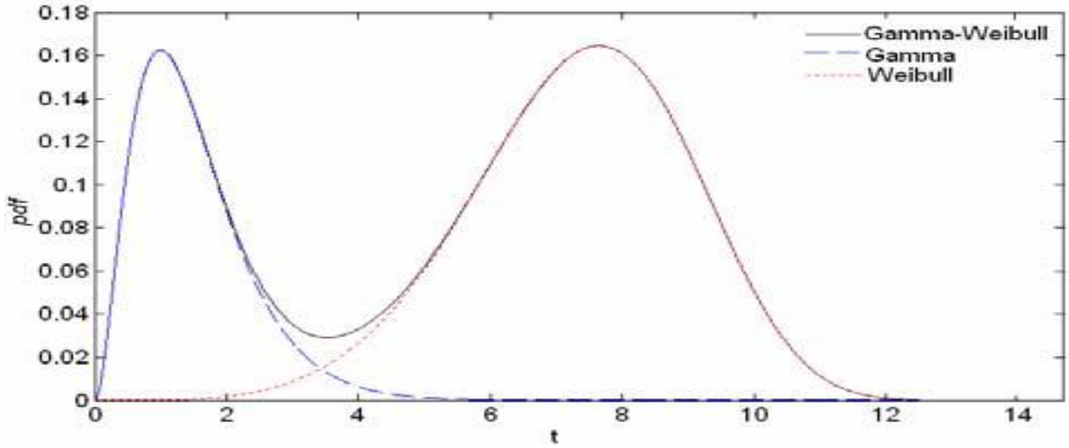
For the special case in which behaviour is not evolving and the nonstationary updating distribution degenerates to a spike at 1.0 (i.e., $s$ = $\beta$ = M, where M approaches infinity), then this equation collapses down exactly to the ordinary (stationary) exponential-gamma model. The graphs of the mixture of two different distributions as obtained from the data analysed for simulation parameters are shown in Figure a, b and c below.




[a]




[b]




[c]

**Figure-(a),(b) and (c).** Graph of simulated data showing the consistency of the mix distributionS

The analysis of my result is as shown in the tables below

**Table-1.** Simulation Results

Exponential -Gamma

| | $\Pi_{1=0.6}$ | $\lambda=2$ | $\alpha_{1=10}$ | $\beta_{1=1.5}$ |
|---|---|---|---|---|
| | $\Pi_1$ | $\lambda$ | $\alpha_1$ | $\beta_1$ |
| Average (Π,θ) | 0.6041 | 2.0502 | 11.7699 | 1.4165 |
| Standard error (Π,θ) | 0.0081 | 0.0133 | 0,1434 | 0.0143 |

Exponential -Weibull

| | $\Pi_{1=}0.8$ | $\lambda=5$ | $\alpha_{2=}15$ | $B_{2=}5$ |
|---|---|---|---|---|
| | $\Pi_1$ | $\lambda$ | $\alpha_1$ | $\beta_1$ |
| Average (Π,θ) | 0.8110 | 4.4460 | 14.9555 | 5,0691 |
| Standard error (Π,θ) | 0.0019 | 0.0338 | 0.0128 | 0.0206 |

Convergence was achieved in all cases even when the starting value was poor and thus emphasizes the numerical stability of EM algorithm. The values of averages and standard error suggest that the EM estimates performed consistently. According to the simulation results, the EM approach work well with different mixture proportions. No restriction was imposed on the maximum number of iterations and convergence was assumed when the absolute differences between successive estimates were less. The results from the simulated data sets are reported in Table1, which gives the averages of the maximum likelihood estimators and standard errors.

## 7. CONCLUSION

This research has succeeded in deriving the Exponential – Gamma mixture Model. The model could be successively applied into capturing evolving customer visit buying behavior process that has three main components which has been addressed in the model viz.;

(a) A timing process governing an individual's rate of visiting

(b) A heterogeneity distribution that accommodates differences across people

(c) An evolutionary process that allows a given individual underlying visit rate to change from one visit to the next by providing a precise forecast of overall new product sales as well as tracking and forecasting future purchasing pattern based on consumer buying behaviour data.

The mixture models of two different distributions such as Exponential-Gamma and Exponential-Weibull to represent the heterogeneous survival data sets. The maximum likelihood estimations of parameters of the mixture models obtained with EM algorithm. Simulations were performed to investigate the convergence of the proposed EM algorithm. According to the simulation results, the EM algorithm was successful in estimation of parameters of the mixture models. The mixture models of two different distributions such as Exponential-Gamma and Exponential-Weibull successfully applied for modeling failure times however, while exponential – Gamma may be an excellent bench mark model, it fails to capture certain buying behaviors that are not stationary over time. Accounting for this condition will be considered in the subsequent work as well as introducing a multiplier effect into the equation that will capture different attrition pattern and some measurable demographic variables.

Funding: This study received no specific financial support.

Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.

## REFERENCES

[1] A. A. Akomolafe, "Analysis of consumer depth of repeat purchasing pattern: An exploratory study of beverages buying behaviour data," *Journal of Business and Organisational Development*, vol. 3, pp. 1-8, 2011.

[2] M. G. Kendall and S. Alan, *The advanced theory of statistics*, 3rd ed. New York: Hafner, 1977.

## BIBLIOGRAPHY

[1] K. Andreas, J. Streburger, and J. H. MandThomas, "Consumer and their brand: Developing relationship theory in consumer research," *Journal of Consumer Research*, vol. 24, pp. 343-373, 1995.

[2] D. Angelis, R., R. Capocaccia, T. Hakulinen, B. Soderman, and A. Verdecchia, "Mixture models for cancer survival analysis: Application to population-based data with covariates," *Statistics in Medicine*, vol. 18, pp. 441-454, 1999.

[3] R. E. Colvert and T. J. Boardman, "Estimation in the piece-wise constant hazard rate model," *Communication in Statistics-Theory Methods*, vol. 11, pp. 1013-1029, 1976.

[4] G. J. Eskin, "Dynamic forecasts of new product demand using a depth of repeat model," *Journal of Marketing Research*, vol. 10, pp. 115-129, 1973.

[5] D. Machin, C. Y.B., and M. K. Parmar, *Survival analysis: A practical approach*, 2nd ed. Chichester, West Sussex, England: John Wiley & Sons, 2006.

[6] J. M. Marin, M. T. Rodriguez-Bernal, and M. P. Wiper, "Using weibull mixture distributions to model heterogeneous survival data," *Communication in Statistics-Simulation and Computation*, vol. 34, pp. 673-684, 2005.