

Application of Marshall - Olkin Discrete Uniform Distribution at Bayesian Approach of HIV-1 Analysis

Mary George K¹, Prasanth C B² and G Meenakshi³

¹Research Scholar, Department of Statistics, Annamalai University, Annamalai Nagar-608002
joicykunjumon@gmail.com

²Assistant Professor, Department of Statistics, Sree Keralavarma College, Thrissur, 680011
prasanthwarriercb@keralavarma.ac.in

³ Professor, Department of Statistics, Annamalai University, Annamalai Nagar-608002
meenakshichinna@rediffmail.com

Abstract

The advanced study about Human Immunodeficiency virus-1 (HIV-1) amounted to classify the stages of HIV patients into different categories according to the symptoms shown and the diseases appeared on one's body. Various interventions have been implemented for early detection and prevention of transmission of HIV infection. This has progressively changed the clinical profile of HIV infected individuals. By considering the distribution of CD_4^+ T count as a discrete distribution and adding stages of HIV patients as an additional parameter, Estimate posterior distribution for rate of CD_4^+ T count using Bayesian Approach. Here we calculate the CD_4^+ T count using discrete uniform distribution in any epoch of time. The time period follows an exponential distribution. The methodology used is Bayesian approach

Key words: exponential, time period, replication, Bayesian Approach.

Introduction

World health organization WHO classified the four clinical stages for HIV, ranging from Stage I asymptomatic to Stage IV AIDS. Patients remain at a higher stage after they recover from the clinical condition which placed them in that stage.

Stage I – Patients who are asymptomatic or have persistent generalized enlargement of lymph nodes of at least two sites for longer than 6 months. Stage II – Mildly symptomatic stage – range of dermatological conditions. Stage III- Additional clinical manifestations may appear moderately symptomatic stage, diarrhea, TB etc. Stage IV – Severely symptomatic stage pneumonia, extra pulmonary TB and other chronic diseases.

The treatment duration is associated with a reduction in HIV infected T cells or count of CD_4^+ T cells. The measurement of CD_4^+ T - cell (CD4) counts is a strong predictor of progression to AIDS and a means of monitoring Antiretroviral Therapy (ART). CD4 count is an important tool in determining treatment failure in HIV-positive patients. Possible increases or decreases in CD4 counts are directly related to HIV replication. The use of combinations of antiretroviral drugs generally results in the suppression of virus replication and hence increased levels of CD4. The success or failure in controlling levels in untreated patients or those on antiretroviral therapy may be associated with factors related to treatment adherence, habits, other correlated infections unrelated to HIV, cancer, immunosuppressive drugs (corticosteroids and chemotherapy),

duration between two consecutive ARTs, as well as socio-economic and psychosocial factors and access to healthcare. Hence the count of CD_4^+T is unpredictable without considering the influencing factors for decrease or increase in the count. The improvement in CD_4^+T count follows a discrete distribution. Here we calculate the $CD_4^+ T$ count using discrete uniform distribution in any epoch of time. The time period follows an exponential distribution. The methodology used is Bayesian approach.

Review of Literature

Marshall and Olkin [1] established a new way of introducing a parameter to extend a family of Weibull and Exponential distributions. They discussed the density and hazard rate of new family of distributions, statistical properties of new family of distributions, estimated values for different values of parameters and geometric stability. The new method is extended to bivariate case also.

Alshangiti *et al.* [2], used the method of extending family of distribution by Marshall and Olkin. They studied as a special cases of new family. It gives more flexibility to the model along with the baseline distribution used. The paper discuss about new family of Marshall-Olkin distributions, the density and hazard rate of new family, statistical properties, MLE estimates and provide applications in the context of statistics and reliability.

Sandhya and Prasanth [3] applied Marshall-Olkin method to Discrete Uniform Distribution discussed its probability mass function, AR(1) model, MLE for its parameters and Application of Marshall-Olkin Discrete Uniform (MODU)distribution.

Meenakshi and Maheswari [4] derived a predictive distribution for HIV replication by using Multinomial Distribution by Bayesian Methodology.

Mary *et.al* [5], discussed about the estimators in HIV Replication by considering the series as An Integer Valued Auto Regressive Process of order one. In this paper, statistical properties and MLE and CLS estimators are derived and its comparative study is also done.

Dravid Mrudula *et. al.* [6] studied prior CD4 counts and CD4 counts after ART intervention among patients and were analyzed and calculated standard error of mean, standard deviation and paired *t* tests and concluded that there is a difference in CD4 count and is depending on previous ART

Norbert Dojer, [7] the paper deals with the mixed data. While optimizing the distribution using the data, discretization is needed for the mixed data which consider continuous and discrete distribution. The paper detailed how to do the discretization while optimization.

Mohammadi *et al.*[8] in their paper highlights the potential role of treatment duration and timing as important factors for successful treatment. Time is an important factor during the ART. This is also considered in the current paper.

Prashant and Supriya [9] gave a detailed study about the importance of population and sample size calculation in clinical studies. The data size used in the current paper is based on this calculation.

Montarroyos *et. al.*[10] list out some of the factors that influence the changes in CD4+ T-Cell count is discussed in detail. The cohort is divided into different levels according to the severity of stage of disease.

Xiao Zhang *et.al*, [11] analyzed the Bayesian method of analyzing a combination of different types of data or mixed data of continuous, ordinal and categorical data. The similar method of application is used in the current paper also.

Model Description

Assumptions of the model

1. Future observations are dependent of current observations
2. Future observations have the same distribution $P(y/\theta)$ as current observations

In Bayesian approach the Prior distribution is based on the historical data.

1. Specifying prior distribution
2. Constructing posterior distribution
3. Characteristics of posterior distribution
4. Equivalence of prior information and extra data

The predictive distribution is the distribution of a future observation after having observed the sample (y_1, y_2, \dots, y_n)

A random sample of HIV infected persons of size 'n' is drawn from a hospital who have taken the ART treatment. After classification based on the immunological stages related to HIV, the Prior distribution used is uniform distribution and derived posterior distribution of HIV replication. Finally, obtain a prediction distribution by adding an additional parameter HIV stage to the existing distribution using Marshall-Olkin method.

The probability mass function is $f(y) = \frac{1}{N}$; $y = 1, 2, 3, \dots, N$

The c.m.f. is given by $F(y) = \frac{y}{N}$; $y = 1, 2, 3, \dots, N$

The CD₄⁺T count replication are equally likely to happen. Considering that it follows a discrete uniform distribution and also consider an additional information that the stages of the HIV. According to the change in the stages of HIV, the Count also changes. Let us consider an additional parameter of 'Y' as the stage changing in HIV.

Adding an additional parameter 'Y', stages of HIV, to the distribution, and applying Marshall Olkin extension to the distribution, we have

$$\bar{G}(y, \mathbf{Y}) = \frac{\bar{Y}\bar{F}(y)}{1 - \bar{Y}\bar{F}(y)}$$

$$G(y, \mathbf{Y}) = 1 - \bar{G}(y, \mathbf{Y}) = \frac{1 - \bar{F}(y)}{1 - \bar{Y}\bar{F}(y)} = \frac{F(y)}{1 - \bar{Y}\bar{F}(y)}$$

$$G(y - 1, \mathbf{Y}) = \frac{F(y - 1)}{1 - \bar{Y}\bar{F}(y - 1)}$$

$$F(y) = \frac{y}{N}; y = 1, 2, 3, \dots, N$$

$$F(y-1) = \frac{y-1}{N}; y = 1, 2, 3, \dots, N$$

$$G(y, \mathbf{Y}) = \frac{\frac{y}{N}}{1 - \bar{Y}(1 - \frac{y}{N})} = \frac{y}{N\bar{Y} + \bar{Y}y}$$

$$G(y - 1, \mathbf{Y}) = \frac{\frac{y - 1}{N}}{1 - \bar{Y}(1 - \frac{y - 1}{N})} = \frac{y - 1}{N\bar{Y} + \bar{Y}(y - 1)}$$

$$g(y) = G(y, \mathbf{Y}) - G(y-1, \mathbf{Y})$$

$$g(y) = \frac{N\bar{Y}}{[N\bar{Y} + \bar{Y}(y-1)][N\bar{Y} + \bar{Y}y]}$$

if y_i be CD4 count from i^{th} category,

Since the time period of replication of CD_4^+ T count follows a continuous distribution, let us assume that the variable 't' follows an exponential distribution. ' θ ' is the rate of change of viral replication.

$$\pi(\theta) = \theta e^{-\theta t}; t > 0, \theta > 0$$

The posterior density function is given by

$$P(\theta/y_1, y_2, \dots, y_n, \mathbf{Y}) = \frac{P(y_1, y_2, y_3, \dots, y_n, \mathbf{Y}/\theta) \cdot \pi(\theta)}{P(y_1, y_2, y_3, \dots, y_n)} = \frac{L(y, \mathbf{Y}, \theta) \cdot \pi(\theta)}{\sum_{i=1}^n L_n(y_i)}$$

The probability distribution of replication of CD_4^+ T having considering the HIV stage of patient and the previous CD_4^+ T count of patient, the proposed distribution proportionality is given by

$$P(\theta/y_i, \mathbf{Y}) = C \cdot P(y_i, \mathbf{Y}/\theta) \cdot \pi(\theta)$$

ie, $P(\theta/y_i, \mathbf{Y}) \propto P(y_i, \mathbf{Y}/\theta) \cdot \pi(\theta)$

The probability density function $\pi(\theta)$ is called prior distribution that express the beliefs about a parameter θ before we see any data.

Select the statistical model $p(y, \mathbf{Y}/\theta)$ that reflects our beliefs about y given θ -- the likelihood distribution. After observing data $\mathbf{Y} = y_1, y_2, y_3, \dots, y_n$, we update our beliefs and calculate posterior distribution.

This distribution considers not only the pattern of CD_{4+}^T cells count but also the stages of HIV patient. Posterior distribution gives the count of CD_{4+}^T cells, by considering all the available information about the patients.

The Likelihood function on partial differentiation is given by

$$\frac{\partial \log L}{\partial \mathbf{Y}} = \frac{n}{\mathbf{Y}} - \sum \frac{n(N - y + 1)}{N\mathbf{Y} + (1 - \mathbf{Y})(y - 1)} - \sum \frac{\mathbf{Y} - y}{N\mathbf{Y} + (1 - \mathbf{Y})y}$$

$$\frac{\partial \text{Log} L}{\partial N} = \frac{n}{N} - \sum \frac{1}{N\mathbf{Y} + (1 - \mathbf{Y})y} - \sum \frac{1}{N\mathbf{Y} + (1 - \mathbf{Y})(y - 1)}$$

$$\frac{\partial \text{Log} L}{\partial \theta} = n \text{Log} \theta - n\theta t$$

$$\theta = 1/t$$

Data Analysis

The data supporting this article are from previously reported studies and datasets were collected and used for deriving MLE estimate of ' \mathbf{Y} '. For uniform distribution the MLE for N is the maximum value of y_i . The MLE for θ is $1/t$.

The visual structure of secondary data of CD_4 count is given below (Figure 1)

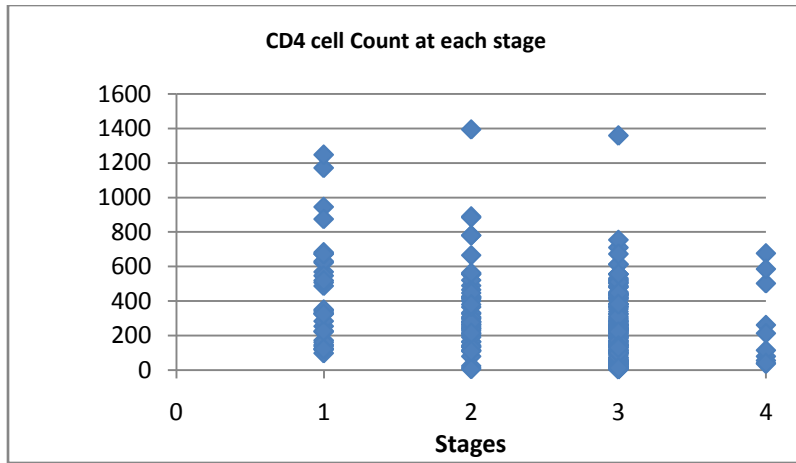


Figure 1. The visual structure of secondary data of CD4 count

Following figures represents the data in uniform and modified uniform distribution for stages 3 and 4

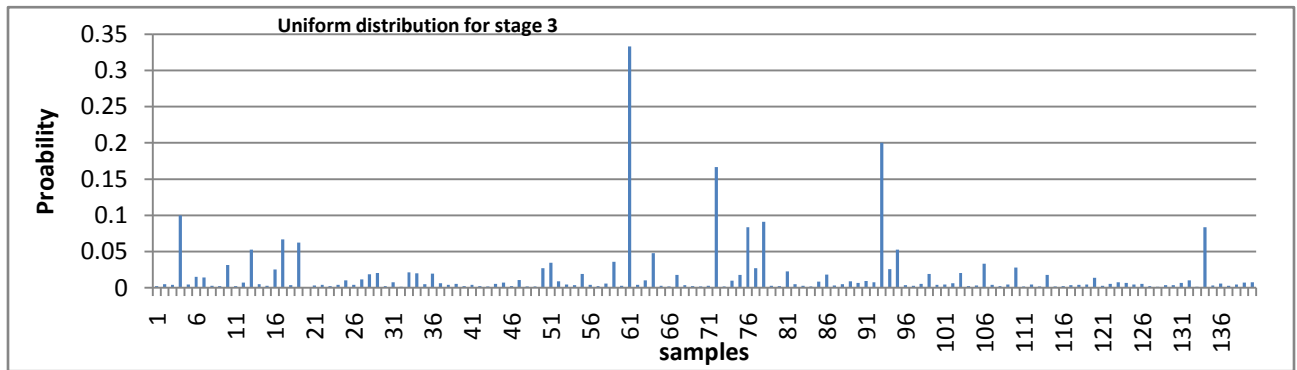


Figure 2. Uniform distribution for stage 3

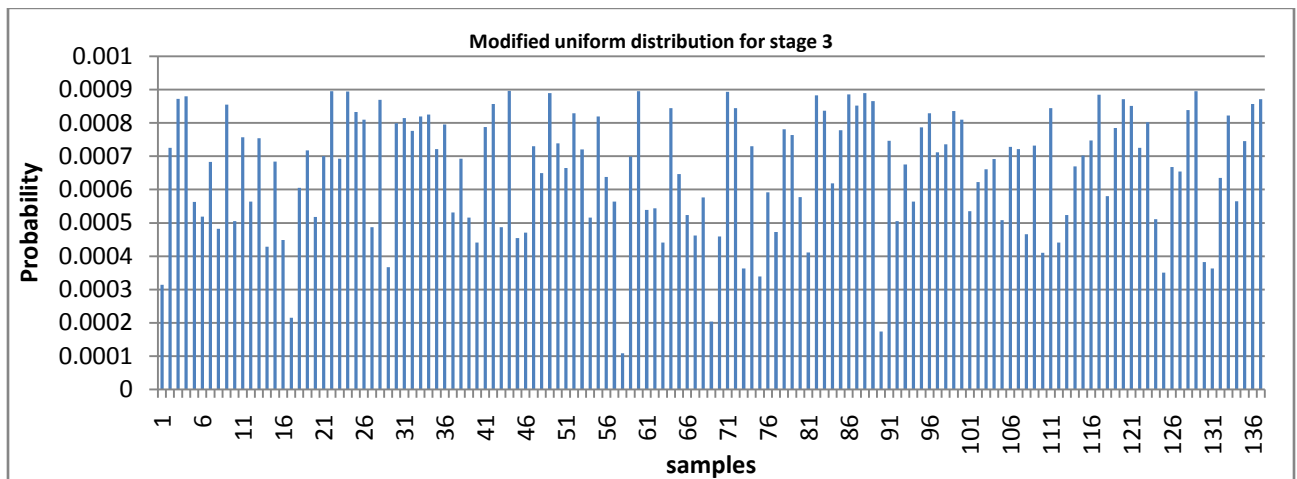


Figure 3. Modified Uniform distribution for stage 3

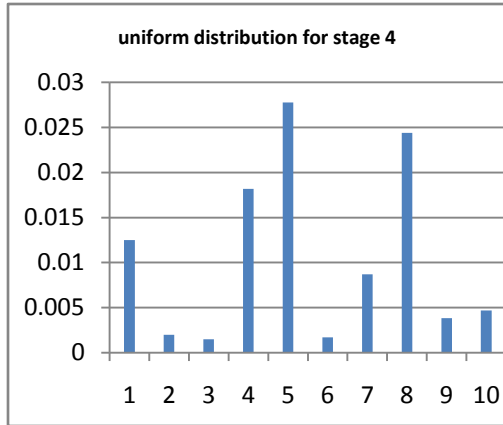


Figure 4. Uniform distribution for stage 4

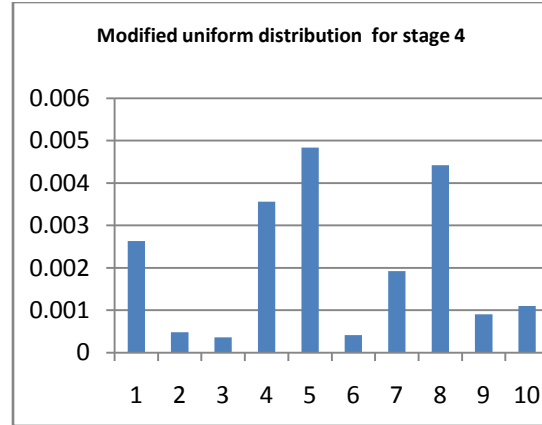


Figure 5. Modified Uniform distribution for stage 4

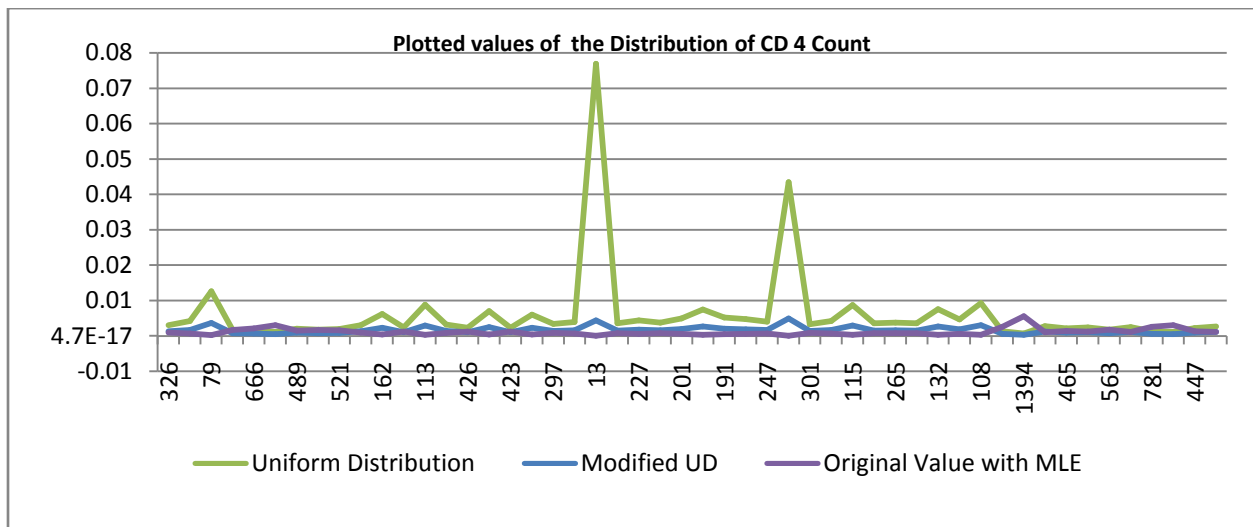


Figure 6. Plotted values of the Distribution of CD 4 Count

Generally, the sample size for any study depends on the acceptable level of significance, power of the study, expected effect size, underlying event rate in the population and standard deviation in the population. According to the sample size deciding strategy, MLE is derived using programming done in Python. The value of 'Y' is converging to 0.00667. The predictive probability distribution of viral replication for different time slot is given below.

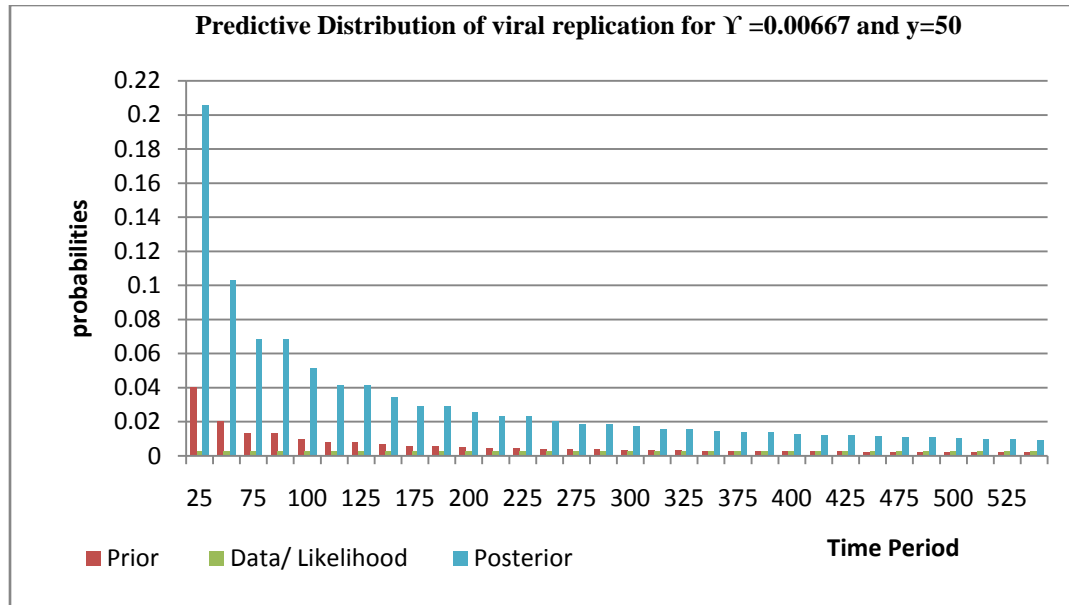


Figure 7. The predictive probability distribution of viral replication for different time slot

It is clear from the predictive distribution; the viral replication is decreasing from the initial stage

Table 1 Illustration of the posterior function for different values of ‘ Υ ’ and ‘t’

t	25	50	75	100	125	150	175	200	225	250	275
$\Upsilon=3.6667$	0.015295079	0.00764754	0.00509836	0.0050984	0.003824	0.003059	0.00305902	0.002549	0.00218501	0.002185011	0.0019119
$\Upsilon=2.6667$	0.020876385	0.010438192	0.00695879	0.0069588	0.005219	0.004175	0.00417528	0.003479	0.00298234	0.002982341	0.0026095
$\Upsilon=0.006667$	0.205267641	0.10263382	0.06842255	0.0684225	0.051317	0.041054	0.04105353	0.034211	0.02932395	0.029323949	0.0256585

Contd...

t	300	325	350	375	400	425	450	475	500	525	550
$\Upsilon=3.6667$	0.001699453	0.001699453	0.00152951	0.0013905	0.00139	0.001275	0.00117654	0.001177	0.00109251	0.001019672	0.0010197
$\Upsilon=2.6667$	0.002319598	0.002319598	0.00208764	0.0018979	0.001898	0.00174	0.00160588	0.001606	0.00149117	0.001391759	0.0013918
$\Upsilon=0.006667$	0.022807516	0.022807516	0.02052676	0.0186607	0.018661	0.017106	0.01578982	0.01579	0.01466197	0.013684509	0.0136845

Contd...

t	575	600	625	650	675	700	725	750	775	800
$\Upsilon=3.6667$	0.000955942	0.000899711	0.00089971	0.0008497	0.000805	0.000805	0.00076475	0.000728	0.00072834	0.000695231
$\Upsilon=2.6667$	0.001304774	0.001228023	0.00122802	0.0011598	0.001099	0.001099	0.00104382	0.000994	0.00099411	0.000948927
$\Upsilon=0.006667$	0.012829228	0.012074567	0.01207457	0.0114038	0.010804	0.010804	0.01026338	0.009775	0.00977465	0.009330347

Table 1 reveals the comparative study of posterior function at different values of ‘ Υ ’

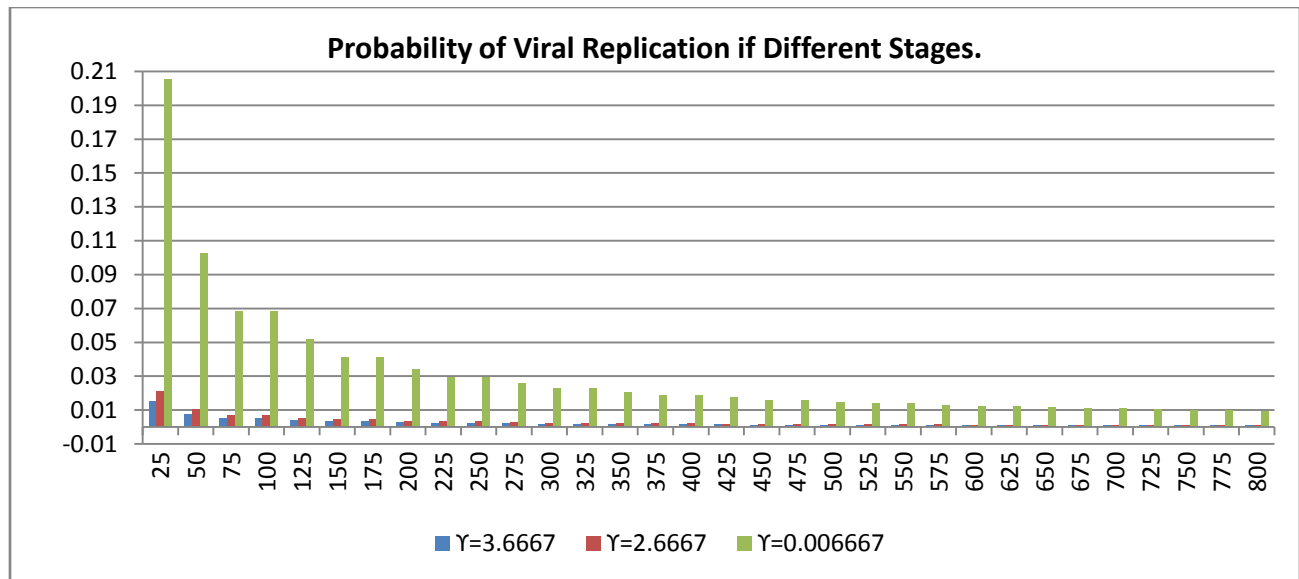


Figure 8. Probability of Viral Replication in Different stages.

The posterior distribution shows that the HIV replication in the preliminary stage is more than that of other stages. The HIV replication is increasing according to the decrease in CD4⁺ Cells

To check whether the data follow the distribution suggested, we use Chi- Squared Goodness of fit Test. This has done by comparing probability mass functions MODU probability mass function and Estimated probability mass function. The test result shows that the data follows suggested probability mass function.

. The mean square Error obtained for the Loss function calculated for the stage 3, as the stage is more consistent than compared to other stages, the value obtained is 0.005177. Hence, we can conclude that the estimator obtained is a reliable estimate.

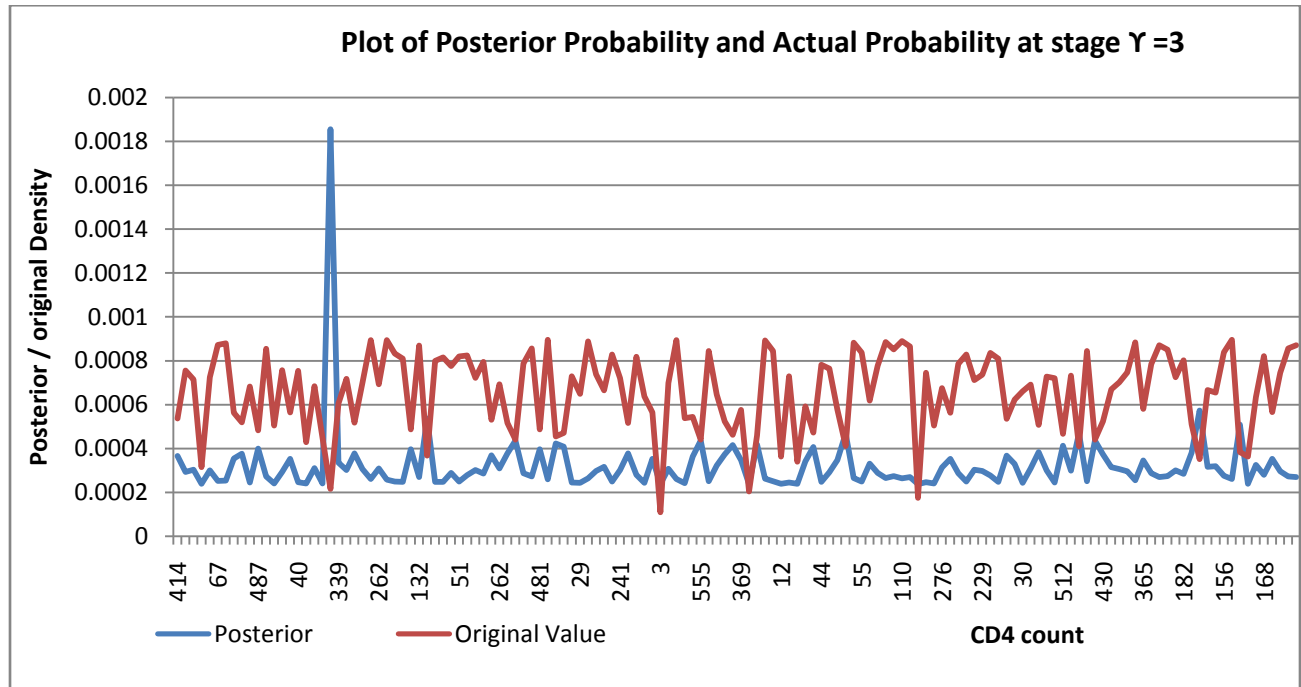


Figure 9 Plot of Posterior Probability and Probability for Actual values at stage 3

Conclusion

Here we detailed the Bayesian approach to viral replication by considering Uniform distribution using Marshall- Olkin approximation. The representation of Uniform distribution and Marshall –Olkin extension to Uniform distribution are depicted in fig 2, fig 3, fig 4 and fig 5. As we can see mode is same for both the distributions. The statistical properties of baseline distribution hold for the newly derived MODU Distribution in this particular study of viral replication. Once the likelihood values were known, the prior distribution has less influence on the posterior. This Bayesian property will lead to get a more reliable estimate. The reliability of fitted distribution is checked by doing chi - squared goodness of fit. For comparison and data analysis, we have used digital data collected from reliable sources. Posterior probability is derived for wide range of values to get a consistent result and for the comparison.

References

1. Albert W. Marshall and Ingram Olkin, "A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families," *Biometrika*, vol. 84, no. 3, pp.641–652,(1997)
2. Arwa M. Alshangiti, M. Kayid , B. Alarfajm A new family of Marshall–Olkin extended distributions, *Journal of Computational and Applied Mathematics* 271, pp. 369–379. (2014)
3. E. Sandhya and C. B. Prasanth, Marshall-Olkin Discrete Uniform Distribution. Hindawi Publishing Corporation, *Journal of Probability* Volume (2014), Article ID 979312, 10 pages
4. G Meenakshi, P R Maheswari, Prediction of HIV Repliation in the Human Immune System using Multinomial Distribution by Bayesian Methodology, *International Journal of Scientific Research in Mathematical and Statistical Sciences* Vol 6, Issue I,: pp 46-52, Feb (2019)
5. Mary George K, Prasanth C B, G Meenakshi, Estimators in HIV Replication which follows an Integer Valued Auto Regressive Process of order One, *Sambodhi*, Vol-44 No.-01 (I) January-March (2021)
6. Mrudula N D, Suwarna U P, Khadse R K, Minal P, Shubhangi D K. Statistical Analysis and Evaluation of CD4 Count after 6 Months on ART. *Indian J Community Med* 37:266-7(2012)
7. Norbert Dojer, Learning Bayesian networks from datasets joining continuous and discrete variables. *International Journal of Approximate Reasoning* 78, pp. 116–124. (2016)
8. Pejman Mohammadi, Angela Ciuffi, and Niko Beerenwinkel. Dynamic models of viral replication and latency. *Curr Opin HIV AIDS* 2015,. Volume 10 _ Number 2, 10:90–95, March (2015)
9. Prashant Kadam and Supriya Bhalerao, Sample size calculation, *Int J Ayurveda Res. ; 1(1): 55–57. PMID: 20532100, Jan-Mar (2010)*
10. Ulisses Ramos Montarroyos et. al.Factors Related to Changes in CD4+ T-Cell Counts over Time in Patients Living with HIV/AIDS: A Multilevel Analysis *Indian J Community Med. ; 37(4): pp. 266–267, Oct-Dec(2012)*
11. Xiao Zhang et.al A Bayesian method for analyzing combinations of continuous, ordinal, and nominal categorical data with missing values. *Journal of Multivariate Analysis* 135, pp. 43-58. (2015)