International Journal of Mathematical Archive-8(1), 2017, 80-87
Available online through www.ijma.info ISSN 2229 – 5046

A STOCHASTIC APPROACH TO DETERMINE THE STATISTICAL MEASURE FOR TIME TO SEROCONVERSION OF HIV INFECTED USING LARGEST ORDER STATISTICS

R. KANNAN, R. KARTHI*

**Department of statistics,
Annamalai University, Annamalai Nagar -608002, Tamil Nadu, India.**

(Received On: 29-12-16; Revised & Accepted On: 25-01-17)

ABSTRACT

This paper focuses on the study of a stochastic model for predicting the seroconversion time of HIV transmission using order statistics. The estimation of statistical measure for time to seroconversion of HIV infected over the time interval (0, t] is an important aspect which help medical intervention. We propose a stochastic model under the assumption that, the interval time between the contacts forms an order statistics and threshold follows exponential- geometric distribution. A Statistical measures for seroconversion are obtained by taking time intervals between successive contacts as the largest order statistics. In developing such a stochastic model, the concept of shock model and cumulative damage process are used. Numerical illustration provided for different combination of parameter involved in the distribution of the random variable used in this model.

Key words: *Human Immuno Deficiency Virus (HIV), Acquired Immuno Deficiency Syndrome (AIDS), Largest order statistics, Antigenic Diversity Threshold, Seroconversion.*

INTRODUCTION

In the study of HIV infection and its progression, the antigenic diversity of the antigen namely HIV play an important role. The intensity of sexual contact is an important factor that adds to antigenic diversity, since more number of new antigens are acquire by the individual who is getting infected. The time to seroconversion from the point of infection depends upon what is known as antigenic diversity, which acts against immune ability of individual. Every individual has a threshold level of antigenic diversity. The successive contact of an individual with an infected partner a result acquired more and more of antigenic diversity which otherwise can be called as damage to the immune system. In the cumulative contribution on successive contact crosses this random level of threshold then the seroconversion takes place. Mathematical methods have been developed by Nowak and May (1991), Stiliankis *et al.* (1994) and Kirschner *et al.* (2000) developed suitable model to estimate the antigenic diversity threshold. Sathiyamoorthi and Kannan (2001) used the shock model and cumulative damage process evolved by Esary *et al.* (1973) to estimate the expected time to cross the antigenic diversity threshold.

In the estimation of statistical measure for time to seroconversion, there is an important role for the inter-arrival times between successive contacts; and it has a significant influence. Ratchagar *et al.* (2003) have derived a model for the estimation of expected time to seroconversion of HIV infected using order statistics. Kannan *et al.* (2008, 2009, 2011 and 2013) have obtained a stochastic model for estimation of expected time to seroconversion of HIV infected using order statistics and threshold follows Gamma, Erlang-2 and Exponentiated Exponential distribution. In this paper, it is assumed that the threshold follows Exponential-geometric distribution and inter-arrival times forms an order statistics; and so they are not independent. This is due to the fact that if the largest order statistics is taken, it implies that the inter-arrival times are becoming larger. Hence, frequent contacts would not be possible which will have its impact on the time to seroconversion. Numerical illustrations are provided using simulated data.

***Corresponding Author: R. Karthi*, Department of statistics,
Annamalai University, Annamalai Nagar - 608002, Tamil Nadu, India.***

ASSUMPTIONS OF THE MODEL

(i) The transmission of HIV is only through sexual contacts.
(ii) An uninfected individual has sexual contacts with HIV infected partner and a random number of HIV is getting transmitted, at each contact.
(iii) An individual is exposed to a damage process acting on the immune system and the damage is assumed to be linear and cumulative.
(iv) The inter–arrival times between successive contacts are taken to be identically and independently distributed random variables.
(v) The sequence of successive contacts and threshold level are independent.
(vi) From the collection of large number of inter-arrival times between successive contacts of a person, a random sample of 'k' observations are taken.

NOTATIONS

X_i : A random variable denoting the increase in the antigenic diversity arising due to the HIV transmitted during the i^{th} contact $X_1, X_2,\ldots, X_k$ are continuous i.i.d. random variables, with p.d.f. g(.) and c.d.f. G(.).

Y : A random variable representing antigenic diversity threshold and follows exponential-geometric distribution with parameters β and p, the p.d.f. being h(.) and c.d.f. H(.).

$U_{(k)}$: A continuous random variable denoting the inter-arrival times between the contacts follows largest order statistics with p.d.f. $f_{u(k)}(t)$ and c.d.f. $F_{u(k)}(t)$

$g_k(.)$: The p.d.f of the random variable $\sum_{i=1}^{k} X_i$.

$F_k(.)$: The k^{th} convolution of F(.).

T : A continuous random variable denoting the time to seroconversion with p.d.f. $l(.)$ and c.d.f. L(.).

$V_k(t)$: Probability of exactly k contacts in (0, t].

$l^*(s)$: The Laplace Stieltjes transform of $l(t)$.

$f^*(s)$: The Laplace Stieltjes transform of $f(t)$.

RESULTS

It can be shown that,

$$P\left[\sum_{i=1}^{k} X_i < Y\right] = \int_0^{\infty} g_k(x)\bar{H}(x)dx \tag{1}$$

Where, $\bar{H}(x) = 1 - H(x)$

The probability density function of exponential-geometric distribution is,

$$h(y) = \beta(1-p)\, e^{-\beta y}(1 - pe^{-\beta y})^2$$

and its distribution function is

$$H(y) = \left(1 - e^{-\beta y}\right)(1 - pe^{-\beta y})^{-1}$$

And

$$\bar{H}(y) = \frac{e^{-\beta y}(1-p)}{(1 - pe^{-\beta y})}$$

Since Y is taken to be exponential-geometric distribution (β, p).
Hence

$$P\left[\sum_{i=1}^{k} X_i < y\right] = \theta(1-p)\sum_{i=1}^{k}\left[\frac{p^{i-1}}{\theta + i\beta}\right] \tag{2}$$

We define the survival function as S(t)

$$S(t) = P(T > t)$$
$$= \sum_{k=0}^{\infty} Pr\{there\ are\ exactly\ k\ contacts\ in\ (0.t]\}$$
$$\times Pr\{the\ cumulative\ total\ of\ antigenic\ diversity < Y\}$$

$$S(t) = \sum_{k=0}^{\infty} V_k(t)\, P\left[\sum_{i=1}^{k} X_i < y\right]$$

$$S(t) = \sum_{k=0}^{\infty} [F_k(t) - F_{k+1}(t)] \tag{3}$$

L(t) = 1 – S(t)

$$= 1 - \left\{\sum_{k=0}^{\infty} [F_k(t) - F_{k+1}(t)] \left[\theta(1-p) \sum_{i=1}^{k} \left[\frac{p^{i-1}}{\theta + i\beta}\right]\right]\right\}$$

$$l(t) = -\sum_{k=0}^{\infty} [f_k^*(t) - f_{k+1}^*(t)] \left[\theta(1-p) \sum_{i=1}^{k} \left[\frac{p^{i-1}}{\theta + i\beta}\right]\right]$$

Taking Laplace transform of l(t) is,

$$l^*(s) = -\sum_{k=0}^{\infty} [f_k^*(s) - f_{k+1}^*(s)] \left[\theta(1-p) \sum_{i=1}^{k} \left[\frac{p^{i-1}}{\theta + i\beta}\right]\right] \tag{4}$$

Where $f^*(s)$ is the Laplace transform of f(.).

The inter–arrival times $U_1, U_2, U_3,\ldots,U_k$ are i.i.d random variables. Now arranging $U_1, U_2, U_3,\ldots,U_k$ in the increasing order of magnitude we get,

$U_{(1)} \leq U_{(2)} \leq \ldots \leq U_{(k)}$.

Now, we Consider the k^{th} order statistics, $U_{(k)}$

The p.d.f of the largest order statistics is

$$f_{u(k)}^*(t) = K[F(t)]^{k-1} f(t)$$

The Laplace Stieltjes transform the same is given by

$$f_{u(k)}^*(t) = \int_0^{\infty} e^{-st}\, K[F(t)]^{k-1} f(t)dt$$

Assuming that *f*(t) follows exp(c), it can be shown that

$$f_{u(k)}^*(s) = \frac{k!\, c^k}{(kc+s)(2c+s)(3c+s)\ldots\ldots\ldots(kc+s)} \tag{5}$$

Consider equation (4),

$$l^*(s) = -\sum_{k=0}^{\infty} [[f^*(s)]^k - [f^*(s)]^{k+1}] \left[\theta(1-p) \sum_{i=1}^{k} \left[\frac{p^{i-1}}{\theta + i\beta}\right]\right]$$

$$l^*(s) = -\{[f^*(s)]^k [1 - f^*(s)]\} \left[\theta(1-p) \sum_{i=1}^{k} \left[\frac{p^{i-1}}{\theta + i\beta}\right]\right] \tag{6}$$

Substitute eqn (5) in eqn (6), we get,

$$= -\left\{\left[\frac{k!\, c^k}{(kc+s)(2c+s)(3c+s)\ldots\ldots(kc+s)}\right]^k \left[1 - \frac{k!\, c^k}{(kc+s)(2c+s)(3c+s)\ldots\ldots(kc+s)}\right]\right\} \times \left[\theta(1-p) \sum_{i=1}^{k} \left[\frac{p^{i-1}}{\theta + i\beta}\right]\right]$$

$$l^*(s) = -\left[\frac{(k!\, c^k)^k s}{(k!\, c^k + s)^{k+1}}\right] \left[\theta(1-p) \sum_{i=1}^{k} \left[\frac{p^{i-1}}{\theta + i\beta}\right]\right]$$

$$E(T) = -\left.\frac{dl^*(s)}{ds}\right|_{s=0}$$

$$= (k!\, c^k)^k \frac{d}{ds}\left[\frac{s}{(k!\, c^k + s)^{k+1}}\right] \left[\theta(1-p) \sum_{i=1}^{k} \left[\frac{p^{i-1}}{\theta + i\beta}\right]\right]$$

$$= (k!\,c^k)^k \left[\frac{(k!\,c^k+s)^{k+1}(1) - \mathrm{s}\,(\mathrm{k}+1)(k!\,c^k+s)^k}{(k!\,c^k+s)^{2k+2}}\right] \times \left[\theta(1-p)\sum_{i=1}^{k}\left[\frac{p^{i-1}}{\theta+i\beta}\right]\right]$$

Put s = 0,

$$= (k!\,c^k)^k \left[\frac{(k!\,c^k)^{k+1}}{(k!\,c^k)^{2k+2}}\right] \times \left[\theta(1-p)\sum_{i=1}^{k}\left[\frac{p^{i-1}}{\theta+i\beta}\right]\right]$$

$$E(T) = -\frac{dl^*(s)}{ds}\bigg|_{s=0} = \frac{1}{k!\,c^k}\left[\theta(1-p)\sum_{i=1}^{k}\left[\frac{p^{i-1}}{\theta+i\beta}\right]\right] \quad (7)$$

(On simplification)

$$E(T^2) = \frac{d^2l^*(s)}{ds^2}\bigg|_{s=0}$$

$$\frac{d^2l^*(s)}{ds^2} = (k!\,c^k)^k \left[\frac{(k!\,c^k+s)^{k+2}(k) - (\,\mathrm{s\,k} - k!\,c^k)(k+2)(k!\,c^k+s)^{k+1}}{(k!\,c^k+s)^{2k+2}}\right] \times \left[\theta(1-p)\sum_{i=1}^{k}\left[\frac{p^{i-1}}{\theta+i\beta}\right]\right]^2$$

Put s = 0,

$$= (k!\,c^k)^k \left[\frac{(k!\,c^k)^{k+2}(k) + (k!\,c^k)(k+2)(k!\,c^k)^{k+1}}{(k!\,c^k)^{2k+2}}\right] \times \left[\theta(1-p)\sum_{i=1}^{k}\left[\frac{p^{i-1}}{\theta+i\beta}\right]\right]^2$$

$$E(T^2) = \frac{d^2l^*(s)}{ds^2}\bigg|_{s=0} = \frac{2k+2}{(k!\,c^k)^2}\left[\theta(1-p)\sum_{i=1}^{k}\left[\frac{p^{i-1}}{\theta+i\beta}\right]\right]^2 \quad (8)$$

(On simplification)

Hence,

$$V(T) = E(T^2) - [E(T)]^2 \quad (9)$$

Substitute equation (7) and (8) in equation (9), we get,

$$V(T) = \frac{2k+2}{(k!\,c^k)^2}\left[\theta(1-p)\sum_{i=1}^{k}\left[\frac{p^{i-1}}{\theta+i\beta}\right]\right]^2 - \left[\frac{1}{k!\,c^k}\left[\theta(1-p)\sum_{i=1}^{k}\left[\frac{p^{i-1}}{\theta+i\beta}\right]\right]\right]^2$$

$$V(T) = \frac{2k+1}{(k!\,c^k)^2}\left[\theta(1-p)\sum_{i=1}^{k}\left[\frac{p^{i-1}}{\theta+i\beta}\right]\right]^2 \quad (10)$$

(On simplification)

NUMERICAL ILLUSTRATIONS

Table - 1

	$\theta = 0.3, p = 0.5, \beta = 0.2, c = 0.1$	
k	E(T)	V(T)
1	1.8750	10.54688
2	9.3750	439.4531
3	31.2500	6835.938
4	78.1250	54931.64
5	156.2500	268554.7
6	260.4167	881618.9
7	372.0238	2076026
8	465.0298	3676296
9	516.6997	5072594

Figure -1

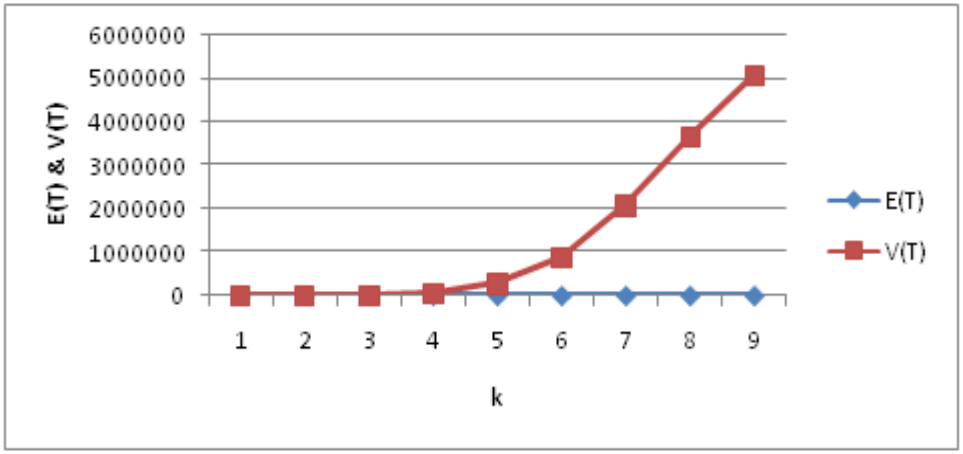


Table - 2

	$\theta = 0.3, p = 0.6, \beta = 0.2, k = 1$	
c	E(T)	V(T)
1	0.133333	0.053333
2	0.066667	0.013333
3	0.044444	0.005926
4	0.033333	0.003333
5	0.026667	0.002133
6	0.022222	0.001481
7	0.019048	0.001088
8	0.016667	0.000833
9	0.014815	0.000658

Figure -2

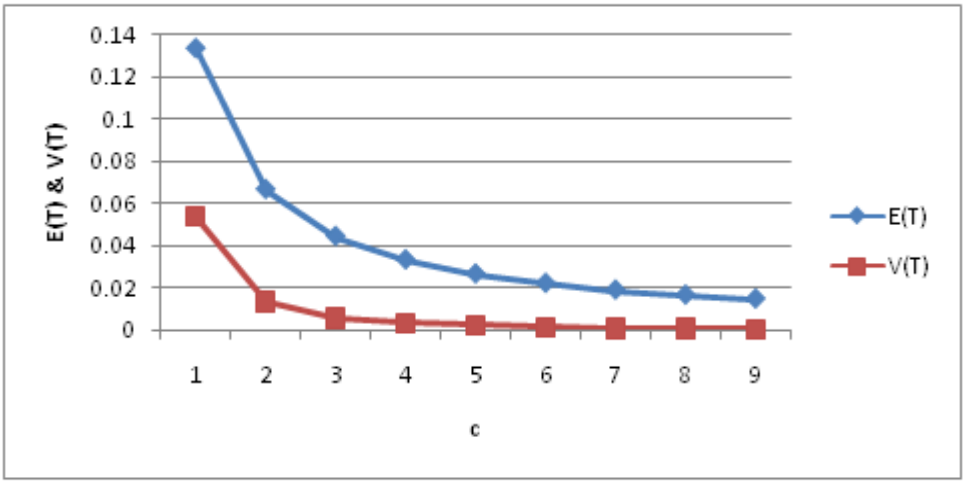


Table -3

	$p = 0.5, \beta = 0.3, k = 1, c = 2$	
θ	E(T)	V(T)
0.1	0.041667	0.005208
0.2	0.071429	0.015306
0.3	0.093750	0.026367
0.4	0.111111	0.037037
0.5	0.125000	0.046875
0.6	0.136364	0.055785
0.7	0.145833	0.063802
0.8	0.153846	0.071006
0.9	0.160714	0.077487

Figure -3

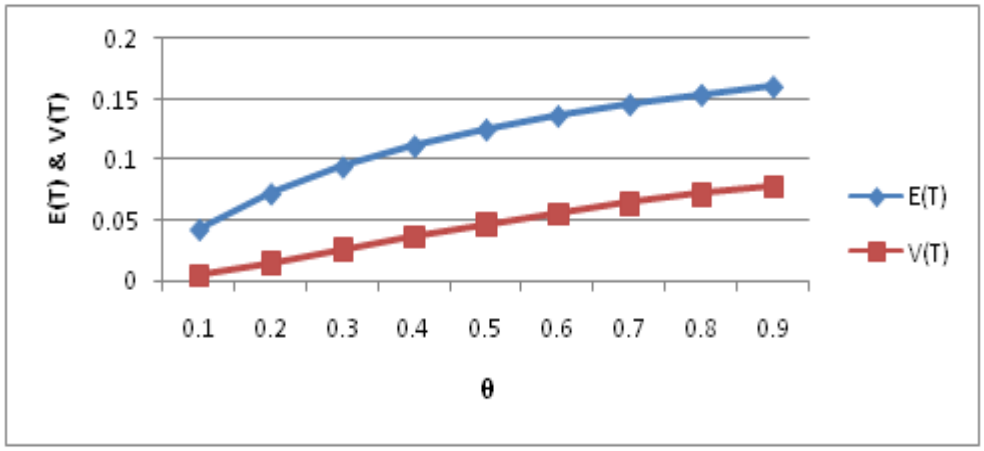


Table - 4

	$\theta = 0.3, \beta = 0.4, k = 1, c = 2$	
p	E(T)	V(T)
0.1	0.337500	0.341719
0.2	0.240000	0.172800
0.3	0.175000	0.091875
0.4	0.128571	0.049592
0.5	0.093750	0.026367
0.6	0.066667	0.013333
0.7	0.045000	0.006075
0.8	0.027273	0.002231
0.9	0.012500	0.000469

Figure -4

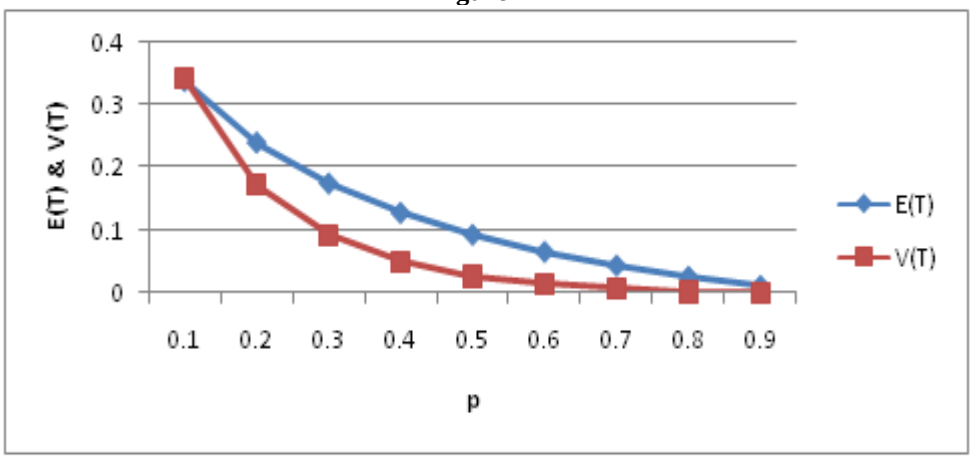
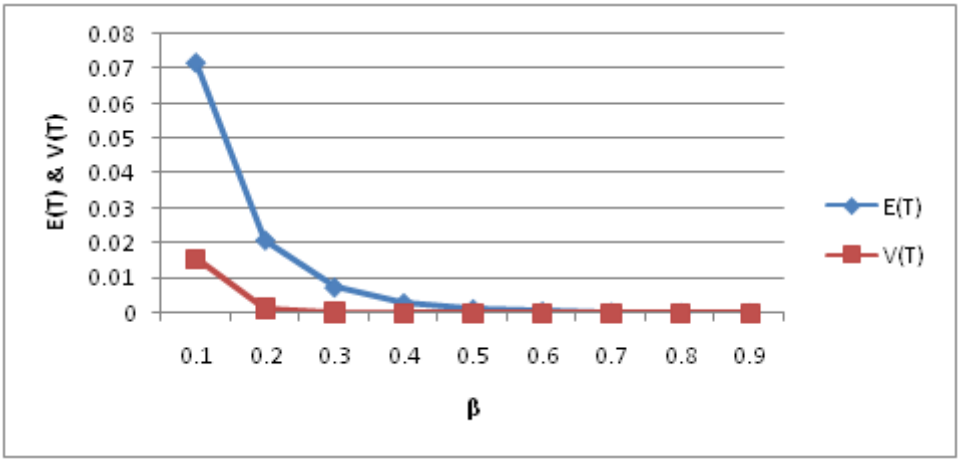


Table -5

	$\theta = 0.2, p = 0.5, k = 1, c = 2$	
β	E(T)	V(T)
0.1	0.071429	0.015306
0.2	0.020833	0.001302
0.3	0.007353	0.000162
0.4	0.002841	2.42E-05
0.5	0.001157	4.02E-06
0.6	0.000488	7.15E-07
0.7	0.000211	1.34E-07
0.8	9.30E-05	2.60E-08
0.9	4.16E-05	5.18E-09

Figure -5



CONCLUSIONS

The value of both mean and variance are increases with an increase in 'k' namely the number of contacts. If 'k' becomes larger then the corresponding $U_{(k)}$ the mean and variance of time to seroconversion also become larger thereby implying that it is the largest of the interarrival times in such a case the inter contact times are elongated thereby having a delayed time to seroconversion. It is easily seen from the table-1 and figure-1.

As the value of the parameter 'c' is the distribution of times interval between successive contacts shown an increase it means that, the average time interval between successive contacts, which is given by $E(U) = \frac{1}{C}$, Since U ~ exp(c), therefore time interval between successive contacts becomes smaller and hence the statistical measure for time to seroconversion decreases, as indicated in table-2 and figure-2.

As the value of 'θ' which is namely, the parameter of the random variable X_i denoting contribution to the antigenic diversity increases, then it seen that, the mean time to seroconversion and variance time to seroconversion both increases, as indicated in table-3 and figure-3.

In the table-4, the variation in statistical measure that consequences to the change in parameter 'p' is noted as the parameter of the threshold parameter p increases, then the statistical measure for the time to seroconversion decreases, as shown in figure-4.

It is observed from the table-5, and also the figure-5, as the value of 'β' which is the parameter of the exponential-geometric distribution of the threshold increases, the expected time to seroconversion as well as variance time to seroconversion are decreases.

ACKNOWLEDGEMENT

We are immensely grateful to Dr.R.sathiyamoorthi, former Professor and Head, Department of Statistics, Annnamalai University, Chidambaram and Dr.G.S.Harisekharan, WIPRO, Chennai for their invaluable constructive suggestions, guidance and intellectual support.

REFERENCES

1. Esary, J. D., Marshall, A. W. and Proschan, F. Shock models and Wear processes, Ann.Probability, Vol.1 (4), (1973), 627-649.
2. Kannan, R., Ganesan, A., and Thirumurugan, A. Stochastic model for estimation of expected time to seroconversion of HIV infected using largest order statistics. Int. J. Agricult.Stat.Sci., 5(2), (2009), 397-404.
3. Kannan, R., Ganesan, A., Sathiyamoorthi, R., and Malarvizhi, G. A Stochastic Model for the Estimation of Time to seroconversion of HIV Infected using Order Statistics, Bio–Science Research Bulletin, Vol.24(1), (2008),1–7.
4. Kannan, R., Kavitha. S., and Sathiyamoorthi, R. A stochastic approach to determine the expected time to Seroconversion of HIV infected using order statistics. Journal of the Indian Academy of Mathematics, 35, (2013), 131-141.
5. Kannan, R., Thirumurugan, A., Sathiyamoorthi, R., and Malarvizhi, G. Estimation of expected time to seroconversion of HIV infected using order statistics. Int. J. Agricult. Stat.Sci., 7(1), (2011), 197-209.

6. Kirschner, D., Webb, G.F., and Cloyd, M. Model of HIV-1: Disease progression based on virus-induced lymphnode homing and homing induced apoptosis of CD4+ lymphocytes. Journal of AIDS, 24(4), (2000), 352-362.
7. Nowak, M.A. and May, R.M. Mathematical Biology of HIV Infections: Antigenic Variation and Diversity Threshold, Mathematical Biosciences, Vol.106, (1991), 1-21.
8. Ratchagar, N.P., Vijaya, S., Sathiyamoorthi, R., and Kannan, R. A stochastic model for estimation of expected time to seroconversion of HIV infected using order statistics, Proceedings of the National Conference on Mathematical and Computational Models, PSG College of Technology, Coimbatore, (2003),37-42.
9. Sathiyamoorthi, R. and Kannan, R. A stochastic model for time to seroconversion of HIV transmission. Journal of Kerala Statistical Association, 12, (2001), 23-39.
10. Stilianakis, N., Schenzle, D. and Dietz, K. On the Antigenic Diversity Threshold Model for AIDS, Mathematical Biosciences, Vol. 121, (1994), 235-247.

Source of support: Nil, Conflict of interest: None Declared.