Recent Research in Science and Technology 2010, 2(3): 80-85
ISSN: 2076-5061
www.recent-science.com

STATISTICS

Left and Right Censoring for Expected Time to Seroconversion – A Shock Model Approach

R. Rajivgandhi, A. Ganesan, R. Vinoth and K. Kannadasan

[1]*Department of Statistics, Annamalai University, Tamil Nadu 608 002, India*
[2] *Department of Mathematics, Srinivas School of Engineering, Mangalore, Karnataka 575 021, India*

Abstract

The time to cross-antigenic threshold of the infected person's is a vital event. Once the accumulated number of events exists from the persons which as a certain threshold level, it could be viewed as a "Break down point". The important characteristic for the antigenic diversity threshold for a person is the time to attain the break down point. In this paper, a shock model approach is proposed to obtain the expectation time to attain the threshold level. Attempt to find patients censoring value is also seen. Graphical illustrations are provided for the use of the model.

Keywords: Antigenic diversity, Break down point, Censoring, Exponential distribution, shock model, Threshold

Introduction

Since the beginning of the HIV/AIDS, epidemic mathematicians and statisticians have developed models to describe and predict the course of the infection, both at the microbiological level and in a reference population, made up of one or more risk groups. The mathematical and statistical tools were derived from the firmly established theory of epidemic modeling, although some adjustments became necessary, because of specific characteristics of HIV infection. It has become clear that the extremely long asymptomatic period of the HIV infection was the main problem to be addressed in the model construction. Seroconversion studies have been conducted using various approaches.

The spreading of HIV has taken different directions since the virus was first discovered, or more correctly, our perception of how and where HIV is spreading has changed over the years. Initially, men were infected to a greater extent than women but this has evened out and now women are identified as the most affected group together with youth and poor.

In this paper a mathematical model is developed to obtain the expected time of breakdown point or the expected time to reach the threshold level, in the context of HIV/AIDS with the assumptions that the times between decision epochs are independent and identically distributed (i.i.d) random variable, the number of exits at each period time are i.i.d. random variables and that the threshold level is a random variable following a Exponential distribution. One can see for more detail in Esary *et al* (1973), Gurland (1955), Nowak and May (1991) and Stilianakis *et al* (1994). Needle sharing and Sexual contact for men and women is taken in two different modes as λ_1, λ_2 for men and λ_3, λ_4 for women. Right censoring means that we only know that their event time T_i lies in an interval (t, 1), sometimes some subjects have experienced the event before detailed observation commences. Thus their time to event lies in the interval [0, T). Assuming the starting point of a patient as unknown observation which is been recorded as left censoring, similarly if the end point is unknown it is observed as right censoring.

1. Assumption

These assumptions are somewhat artificial, but are made because of lack of detailed real-world information on the one hand and in order to illustrate the proceedings on the other hand.

- An uninfected partner has sexual contacts with an infected person and also shares unsterile needles for drug abuse.
- On every occasion of sexual contact and sharing of unsterile needle there is a random amount of transmission of HIV, which together contributes to the antigenic diversity.
- The damages due to the events namely sexual contacts and sharing of needles are statistically independent.
- If the cumulative damage due to successive events crosses the antigenic diversity threshold level the seroconversion takes place. The inter-arrival times between contacts and sharing needles are statistically independent.

2. Notations

X_i', X_i'' : a continuous random variable denoting the amount of contribution to the antigenic diversity due to the HIV transmitted in the i^{th} contact, in other words the damage caused to the immune system in the i^{th} contact, with p.d.f g (.) and c.d.f G (.).

Y_1, Y_2 : continuous random variable denoting the threshold levels for the two independent groups.

$g(.)$: The probability density functions of X_i

$g^*(.)$: Laplace transform of g (.)

$g_k(.)$: The k- fold convolution of g (.) i.e., p.d.f. of $\sum_{j=1}^{k} X_i$

$h(.)$: The probability density function of X_i''

$h^*(.)$: Laplace transform of h (.)

$h_k(.)$: The k- fold convolution of h (.), p.d.f. of $\sum_{i=1}^{k} X_i''$

$P_1(\cdot)$: Time to breakdown of the system due to depletion in grade one.

$P_2(\cdot)$: Time to breakdown of the system due to depletion in grade two.

$f(.)$: p.d.f. of random variable denoting between successive policy announcement with the corresponding c.d.f. F (.)

$F_k(.)$: k-fold convolution of F (.)

$V_k(t)$: Probability of exactly k policy announcements.

$S(.)$: Survival function.

$L(t)$: 1 - S (t)

Results

Let Y be the random variable which follows exponential distribution, with memoryless property. Taking two modes of transmission in two different situations the observation is found with the threshold level. Censoring is when an observation is incomplete due to some random cause. The cause of the censoring must be independent of the event of interest if we are to use standard methods of analysis. Start/end points for measuring time are chosen according to the context so may not represent the entire life of an individual. It is possible for both right and left censoring to occur in the same dataset since censoring is independent of survival.

If the cumulative damage in any one of the damage system cross its threshold level, the transfer of HIV is not possible. The probability that the infected person survey even after the cumulative loss of damage in the respective grades after K decision is given by

$$Pr\left[\sum_{i=1}^{k} X_i' < Y_1 \cap \sum_{i=1}^{k} X_i'' < Y_2\right]$$

Because of independence of X_i' and X_i''.

$$P(T > t) = \sum_{k=0}^{\infty} Pr[\text{there are exactly } k \text{ instants of exit in } (0,t)] * Pr[\text{the system does not fail in } (0,t)]$$

$$P(T > t) = \sum_{k=0}^{\infty} V_k(t) Pr\left[\sum_{i=0}^{k} X_i < max(y_1, y_2) * (y_3, y_4)\right] \quad (4.1)$$

$$= \left[\int_0^{\infty} g_k(x)\left[\left(1-e^{-\lambda_1 x}\right)\left(1-e^{-\lambda_2 x}\right)\right]\int_0^{\infty} h_k(x)\left[\left(1-e^{-\lambda_3 x}\right)\left(1-e^{-\lambda_4 x}\right)\right]\right] dx$$

Therefore $S(t) = P[T > t]$ is the survival function which gives the probability that the system will fail only after time t. Survival analysis is a class of statistical methods for studying the occurrence and timing of events. These methods are most often applied to the study of deaths.

$$S(t) = \sum_{k=0}^{\infty} V_k(t)\left[Pr\left(\sum_{i=1}^{k} X_i' < Y_1\right) * Pr\left(\sum_{i=1}^{k} X_i'' < Y_2\right)\right] \quad (4.2)$$

$P(X_1 + X_2 + \cdots + X_k < Y) =$ P [the system does not fail, after k epochs of exits].

Let $P_1(\cdot) \sim$ Exponential with parameter λ_1, λ_2 and

$P_2(\cdot) \sim$ Exponential with parameter λ_3, λ_4

$$Pr\left(\sum_{i=1}^{k} X_i' < Y_1\right) = \int_0^{\infty} g_k(x)\, P_1(x)\, dx$$

Similarly for P_2

$$Pr\left(\sum_{i=1}^{k} X_i < Y\right) = \int_0^{\infty} g_k(x) P_1(x) dx \int_0^{\infty} g_k(x) P_2(x) dx \quad (4.3)$$

$$= \int_0^{\infty} g_k(x)\left(e^{-\lambda_1 x}\right)\left(e^{-\lambda_2 x}\right)\int_0^{\infty} h_k(x)\left(e^{-\lambda_3 x}\right)\left(e^{-\lambda_4 x}\right)$$

$$= \left[g^*{}_k(\lambda_1) - g^*{}_k(\lambda_2)\right]\left[h^*{}_k(\lambda_3) - h^*{}_k(\lambda_4)\right]$$

$$S(T) = \sum V_k(t)\left\{\left(e^{-\lambda_1 x} + e^{-\lambda_2 x} - e^{-(\lambda_1+\lambda_2)x}\right) * \left(e^{-\lambda_3 x} + e^{-\lambda_4 x} - e^{-(\lambda_3+\lambda_4)x}\right)\right\} \quad (4.4)$$

$$S(T) = \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)][g^*(\lambda_1)h^*(\lambda_3)] + \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)][g^*(\lambda_1)h^*(\lambda_4)] - \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)][g^*(\lambda_1)h^*(\lambda_3+\lambda_4)] + \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)][g^*(\lambda_2)h^*(\lambda_3)] + \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)][g^*(\lambda_2)h^*(\lambda_4)] - \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)][g^*(\lambda_2)h^*(\lambda_3+\lambda_4)] - \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)][g^*(\lambda_3)h^*(\lambda_1+\lambda_2)] - \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)][g^*(\lambda_4)h^*(\lambda_3+\lambda_4)] + \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)][g^*(\lambda_1+\lambda_2)h^*(\lambda_3+\lambda_4)] \quad (4.5)$$

$L(T) = 1 - S(T)$, then taking Laplace transform of L (T), we get

$$L^*(S) = \frac{[1-g^*(\lambda_1)h^*(\lambda_3)]f^*(S)}{[1-g^*(\lambda_1)h^*(\lambda_3)f^*(S)]} + \frac{[1-g^*(\lambda_1)h^*(\lambda_4)]f^*(S)}{[1-g^*(\lambda_1)h^*(\lambda_4)f^*(S)]} - \frac{[1-g^*(\lambda_1)h^*(\lambda_3+\lambda_4)]f^*(S)}{[1-g^*(\lambda_1)h^*(\lambda_3+\lambda_4)f^*(S)]} + \frac{[1-g^*(\lambda_2)h^*(\lambda_3)]f^*(S)}{[1-g^*(\lambda_2)h^*(\lambda_3)f^*(S)]} + \frac{[1-g^*(\lambda_2)h^*(\lambda_4)]f^*(S)}{[1-g^*(\lambda_2)h^*(\lambda_4)f^*(S)]} - \frac{[1-g^*(\lambda_2)h^*(\lambda_3+\lambda_4)]f^*(S)}{[1-g^*(\lambda_2)h^*(\lambda_3+\lambda_4)f^*(S)]} - \frac{[1-g^*(\lambda_3)h^*(\lambda_1+\lambda_2)]f^*(S)}{[1-g^*(\lambda_2)h^*(\lambda_1+\lambda_2)f^*(S)]} - \frac{[1-g^*(\lambda_4)h^*(\lambda_1+\lambda_2)]f^*(S)}{[1-g^*(\lambda_4)h^*(\lambda_1+\lambda_2)f^*(S)]} + \frac{[1-g^*(\lambda_1+\lambda_2)h^*(\lambda_3+\lambda_4)]f^*(S)}{[1-g^*(\lambda_1+\lambda_2)h^*(\lambda_3+\lambda_4)f^*(S)]} \quad (4.6)$$

Where $[f^*(S)]$ is Laplace transform of $F_k(t)$ since the inter arrival times are identically independently distributed. Equation (4.6) can be rewritten as

$$L^*(S) = \frac{c[1-g^*(\lambda_1)h^*(\lambda_3)]}{[c+s-g^*(\lambda_1)h^*(\lambda_3)c]^2} + \frac{c[1-g^*(\lambda_1)h^*(\lambda_4)]}{[c+s-g^*(\lambda_1)h^*(\lambda_4)c]^2} - \frac{c[1-g^*(\lambda_1)h^*(\lambda_3+\lambda_4)]}{[c+s-g^*(\lambda_1)h^*(\lambda_3+\lambda_4)c]^2} + \frac{c[1-g^*(\lambda_2)h^*(\lambda_3)]}{[c+s-g^*(\lambda_2)h^*(\lambda_3)c]^2} + \frac{c[1-g^*(\lambda_2)h^*(\lambda_4)]}{[c+s-g^*(\lambda_2)h^*(\lambda_4)c]^2} - \frac{c[1-g^*(\lambda_2)h^*(\lambda_3+\lambda_4)]}{[c+s-g^*(\lambda_2)h^*(\lambda_3+\lambda_4)c]^2} - \frac{c[1-g^*(\lambda_3)h^*(\lambda_1+\lambda_2)]}{[c+s-g^*(\lambda_3)h^*(\lambda_1+\lambda_2)c]^2} - \frac{c[1-g^*(\lambda_4)h^*(\lambda_1+\lambda_2)]}{[c+s-g^*(\lambda_4)h^*(\lambda_1+\lambda_2)c]^2} + \frac{c[1-g^*(\lambda_1+\lambda_2)h^*(\lambda_3+\lambda_4)]}{[c+s-g^*(\lambda_1+\lambda_2)h^*(\lambda_3+\lambda_4)c]} \quad (4.7)$$

$$E(T) = -\frac{d}{ds} L^*(S) \quad \text{given} \quad s = 0$$

Let the random variable U denoting inter arrival time which follows exponential with parameter c. Now $f^*(s) = \left(\frac{c}{c+s}\right)$, substituting in the above equation we get,

$$-\frac{dL^*(S)}{ds} = \frac{c[1-g^*(\lambda_1)h^*(\lambda_3)]}{c^2[1-g^*(\lambda_1)h^*(\lambda_3)]^2} + \frac{c[1-g^*(\lambda_1)h^*(\lambda_4)]}{c^2[1-g^*(\lambda_1)h^*(\lambda_4)]^2} - \frac{c[1-g^*(\lambda_1)h^*(\lambda_3+\lambda_4)]}{c^2[1-g^*(\lambda_1)h^*(\lambda_3+\lambda_4)]^2} + \frac{c[1-g^*(\lambda_2)h^*(\lambda_3)]}{c^2[1-g^*(\lambda_2)h^*(\lambda_3)]^2} + \frac{c[1-g^*(\lambda_2)h^*(\lambda_4)]}{c^2[1-g^*(\lambda_2)h^*(\lambda_4)]^2} - \frac{c[1-g^*(\lambda_2)h^*(\lambda_3+\lambda_4)]}{c^2[1-g^*(\lambda_2)h^*(\lambda_3+\lambda_4)]^2} - \frac{c[1-g^*(\lambda_3)h^*(\lambda_1+\lambda_2)]}{c^2[1-g^*(\lambda_2)h^*(\lambda_1+\lambda_2)]^2} - \frac{c[1-g^*(\lambda_4)h^*(\lambda_1+\lambda_2)]}{c^2[1-g^*(\lambda_4)h^*(\lambda_1+\lambda_2)]^2} + \frac{c[1-g^*(\lambda_1+\lambda_2)h^*(\lambda_3+\lambda_4)]}{c^2[1-g^*(\lambda_1+\lambda_2)h^*(\lambda_3+\lambda_4)]}$$

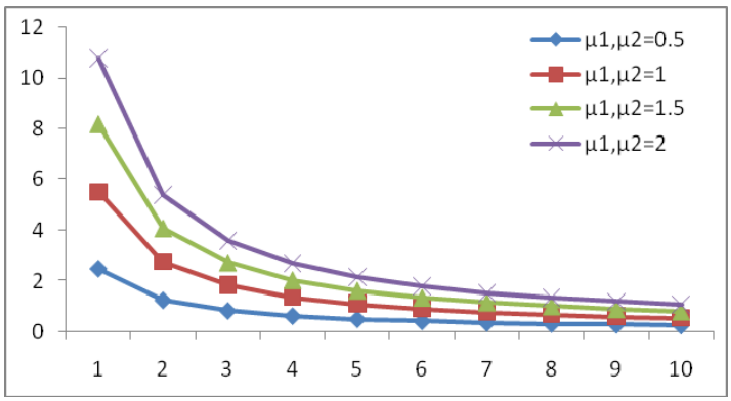
$$E(T) = \frac{1}{c[1-g^*(\lambda_1)h^*(\lambda_3)]} + \frac{1}{c[1-g^*(\lambda_1)h^*(\lambda_4)]} - \frac{1}{c[1-g^*(\lambda_1)h^*(\lambda_3+\lambda_4)]} + \frac{1}{c[1-g^*(\lambda_2)h^*(\lambda_3)]} + \frac{1}{c[1-g^*(\lambda_2)h^*(\lambda_4)]} - \frac{1}{c[1-g^*(\lambda_2)h^*(\lambda_3+\lambda_4)]} - \frac{1}{c[1-g^*(\lambda_3)h^*(\lambda_1+\lambda_2)]} - \frac{1}{c[1-g^*(\lambda_4)h^*(\lambda_1+\lambda_2)]} + \frac{1}{c[1-g^*(\lambda_1+\lambda_2)h^*(\lambda_3+\lambda_4)]} \quad (4.8)$$

$$E(T) = \frac{1}{c\left[1-\left[\frac{\mu_1}{\mu_1+\lambda_1}\frac{\mu_2}{\mu_2+\lambda_3}\right]\right]} + \frac{1}{c\left[1-\left[\frac{\mu_1}{\mu_1+\lambda_1}\frac{\mu_2}{\mu_2+\lambda_4}\right]\right]} - \frac{1}{c\left[1-\frac{\mu_1}{\mu_1+\lambda_1}\frac{\mu_2}{\mu_2+(\lambda_3+\lambda_4)}\right]} + \frac{1}{c\left[1-\left[\frac{\mu_1}{\mu_1+\lambda_2}\frac{\mu_2}{\mu_2+\lambda_3}\right]\right]} + \frac{1}{c\left[1-\left[\frac{\mu_1}{\mu_1+\lambda_2}\frac{\mu_2}{\mu_2+\lambda_4}\right]\right]} - \frac{1}{c\left[1-\left[\frac{\mu_1}{\mu_1+\lambda_2}\frac{\mu_2}{\mu_2+(\lambda_3+\lambda_4)}\right]\right]} - \frac{1}{c\left[1-\left[\frac{\mu_1}{\mu_1+\lambda_3}\frac{\mu_2}{\mu_2+(\lambda_1+\lambda_2)}\right]\right]} - \frac{1}{c\left[1-\left[\frac{\mu_1}{\mu_1+\lambda_4}\frac{\mu_2}{\mu_2+(\lambda_1+\lambda_2)}\right]\right]} + \frac{1}{c\left[1-\left[\frac{\mu_2}{\mu_2+(\lambda_1+\lambda_2)}\frac{\mu_2}{\mu_2+(\lambda_3+\lambda_4)}\right]\right]}$$

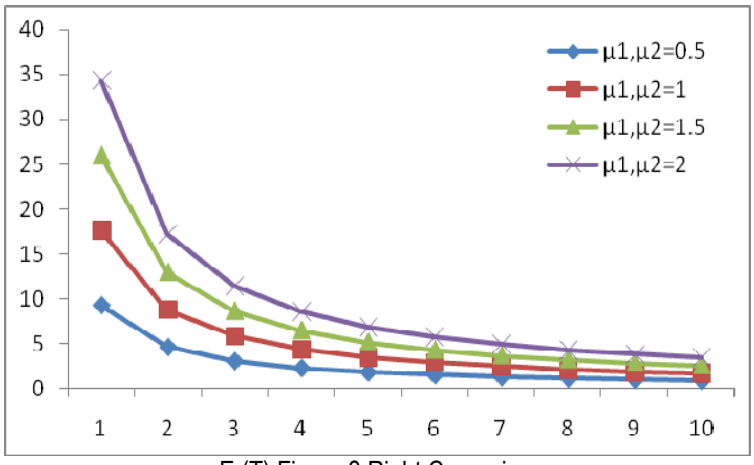
$$\therefore E(T) = \frac{1}{c}\left[\frac{(\mu_1+\lambda_1)(\mu_2+\lambda_3)}{(\mu_1+\lambda_1)(\mu_2+\lambda_3)-\mu_1\mu_2} \mid \frac{(\mu_1+\lambda_1)(\mu_2+\lambda_4)}{(\mu_1+\lambda_1)(\mu_2+\lambda_4)-\mu_1\mu_2} - \frac{(\mu_1+\lambda_1)(\mu_2+(\lambda_3+\lambda_4))}{(\mu_1+\lambda_1)(\mu_2+(\lambda_3+\lambda_4))-\mu_1\mu_2} + \frac{(\mu_1+\lambda_2)(\mu_2+\lambda_3)}{(\mu_1+\lambda_2)(\mu_2+\lambda_3)-\mu_1\mu_2} + \frac{(\mu_1+\lambda_2)(\mu_2+\lambda_4)}{(\mu_1+\lambda_2)(\mu_2+\lambda_4)-\mu_1\mu_2} - \frac{(\mu_1+\lambda_2)(\mu_2+(\lambda_3+\lambda_4))}{(\mu_1+\lambda_2)(\mu_2+(\lambda_3+\lambda_4))-\mu_1\mu_2} - \frac{(\mu_1+\lambda_3)(\mu_2+(\lambda_1+\lambda_2))}{(\mu_1+\lambda_3)(\mu_2+(\lambda_1+\lambda_2))-\mu_1\mu_2} - \frac{(\mu_1+\lambda_4)(\mu_2+(\lambda_1+\lambda_2))}{(\mu_1+\lambda_4)(\mu_2+(\lambda_1+\lambda_2))-\mu_1\mu_2} + \frac{\mu_1+(\lambda_1+\lambda_2)(\mu_2+(\lambda_3+\lambda_4))}{\mu_1+(\lambda_1+\lambda_2)(\mu_2+(\lambda_3+\lambda_4))-\mu_1\mu_2}\right] \quad (4.9)$$

1. Numerical Illustration

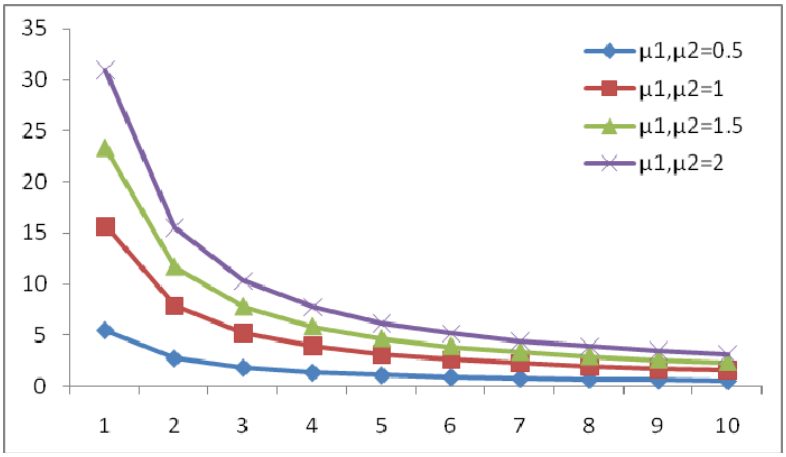
E (T) Figure 1 Threshold



E (T) Figure 2 Left Censoring



E (T) Figure 3 Right Censoring



Conclusion

Survival data is a term used for data measuring the time to some event. In the simplest case, the event is death, but the term also covers other events, like occurrence of a disease, a complication, or, e.g., the time to occurrence of an epileptic seizure. Censored data means that the observations are only partially known. A common reason is that the person studied is alive when the study is evaluated, and thus all that is known about his lifetime is that it exceeds the age he had at the time of evaluation. The presence of censored data is a major technical problem. The occurrence of censoring must be unrelated to the future lifetime,

To illustrate the method described in this paper, we give some limited simulation results. We consider three simulated data; in Figure 1 we observe the Threshold level of the inter-arrival time increases the expected time to cross the antigenic diversity decreases.

In Figure 2, Left censoring is much rarer. Event of interest already occurred at the observation time, but it is not known exactly when. Keeping this in practice the data were simulated and found that as inter-arrival time increases the Left censoring observation were decreasing.

In Figure 3, Subjects followed until some time, at which the event has yet to occur, but then takes no further part in the study. The most commonly used censoring, the Right censoring is found to decrease in the observed value as the inter-arrival between the contact increases.

In all the three figures it is found that the time to seroconversion decreases. The observed simulated data were found that it follows Chebychev inequality.

Reference

Esary. J. D., A.W. Marshall and F. Proschan (1973). Ann. Probability, 1(4), 627-649.

Gurland. J. (1955), Ann. Math. Statist. 26, 294-300.

Nowak, M. A. and R. M. May (1991). Mathematical Biosciences, 106, 1-21.

Stilianakis. N., D. Schenzle and K. Dietz (1994). Mathematical Biosciences. 121, 235-247.