International Journal of Applied
Mathematics & Statistical Sciences (IJAMSS)
ISSN: 2319-3972
Vol.1, Issue 2 Nov 2012 11-16

International Academy of Science,
Engineering and Technology
Connecting Researchers; Nurturing Innovations

LARGE DATA MATRICES; RANDOM WALK MODEL AND APPLICATION OF ENTROPY IN HIV- MOTHER TO CHILD TRANSMISSION (MTCT)

[1]BASAVARAJAIAH D M, [2]B NARASIMHAMURTHY, [3]B.LEELAVATHY & [4]MAHESHAPPA K

[1]National Institute of Epidemiology (NIE), ICMR, Affiliated Madras University, Chennai, India

[2]Scientist "G" National Institute of Epidemiology (NIE), ICMR, Chennai, India

[3]Professor, BMCRI, Bangalore, India

[4]Associate professor, UAS (B), G.K.V.K, Bangalore, India

ABSTRACT

Several factors influenced HIV –transmission from pregnant to Mother to child(MTCT), because of the not use of ARV-prophylaxis at the time of onset and before birth of neonatal baby, prolong breast feeding , placental absorption and high RNA Plasma viral load , drug induced toxicity; illness, lower CD4 count, WHO-Clinical stage IV, Opportunistic infection of pregnant women. Present study aims to fitting random walk on diseases free groups, free probability , and large deviation of entropy mathematical models of HIV, to reduce the large data matrices among HIV infected children and pregnant women with or without receiving ARV Prophylaxis single dose- NVP before and onset of birth of neonatal baby. The model was clearly demonstrated the new born child should underwent the HIV testing at different interval time, the probability were expected to meet the P (n, m) .entropy deviation model uses data coverage, density and Correlation to determine the reduced dimension.

KEY WORDS: HIV, MTCT, NVP, CD4, Free Probability, WHO-Stage, ARV

INTRODUCTION

Methods in HIV MTCT -Epidemiology typically involve a set of mathematical expressions or equations that aim to capture the essence of processes and phenomena for characterize the HIV –transmission. A mathematical model seeks to uncover and formalize the underlying mechanism or laws "that explain the observed phenomena of diseases transmission from mother to child (MTCT). The present study aims, how to reduce the large data matrices among HIV infected children's through random walk on free groups and free probability; when the matrices are asymptotically normal and free b) Application of entropy in large deviation among MTCT –HIV transmission.

METHODS

Model Formulation: In MTCT –HIV transmission ,large variables has generated on prospective study at fixed period of time ,we have correlate the variables between clinical ,laboratory and biological parameters without transformed scale, correlation was fluctuated ,in which we have to analyze high –dimensional data by looking at the Eigen value (or more precisely single value(or of random) .Let us we considered the correlation matrix 50X5 , $X(X_{ij}$ =ith children's were HIV tested at jth interval (the interval period j = tested @1weeks,6weeks,12weeks,24 weeks,36weeks,50weeks). 1≤i≤50, 1≤j≤5 and note that norm one vector α* ,which maximize the variance number $\{\alpha, X_I\}_{i=1}^{56\ weeks}$ is the first principle component and the vector α^{**} maximizing the variance subject to normalization and orthogonally to α* is the second and so on .The α* is the approximately the average of five times HIV test and α* is approximately the difference between the averages of the first two and last three test and so on.

The stability of the principle component under random fluctuation in the data X_{ij}. Wishart and R.A. Fisher studied the distribution of the principle component when the entries "$nxp\ data\ matrix\ X\ with\ i.i.d$".Gaussian considered the empirical distribution of the whole collection of singular values of such matrices, i.e. showed that,$P^{-1}\{\#\ eigen\ value\ \leq nt\}\ an\ absolutely\ continious\ functions$.G (t), when n and P↑∞ asymptotically increases with infinity.

Such that $\frac{n}{p} \to \gamma > 0$,let H be a $nxn\ real\ symmetrical\ matrix\ with\ i.i.d$ Gaussian $N\left(1,\frac{1}{n}\right)$, entries ,we consider the matrix family H_n as a non-commutative random variable with the expectation $\mathcal{T}_n(H_n) = \frac{1}{n}\sum_{i=1}^{50 weeks} E\,(H_{nii}),\frac{1}{n}E(t_r, H_n)$, where E is the classical expectations and also that,$\mathcal{T}_n = \left({H_n}^2\right)$ =1.

Theorem (1)

$(\boldsymbol{H_n^{2m}}) \to \frac{\mathbf{1}}{\boldsymbol{m+1}}\binom{\boldsymbol{2m}}{\boldsymbol{m}}, \boldsymbol{Catalan\ numbers}, \boldsymbol{as\ n} \uparrow$
$\infty\ \boldsymbol{which\ are\ the\ 2mth\ moments\ of\ the\ semicircle\ distribution\ density}\ \boldsymbol{\omega}: \boldsymbol{\omega(x) = (2\pi)^{-1}(4 - x^2)^{-1})}\ \boldsymbol{for}\ |\boldsymbol{x}| \leq \mathbf{2}\ \boldsymbol{and} = \mathbf{0}\ \boldsymbol{other\ wise}$.

Proof: Real symmetrical matrices, dropping the Gaussian assumption .Let H be nxn random matrix with i.i.d N (0, 1) entries with finite moments of all orders, and let $\lambda_1(H), \lambda_2(H), \ldots \ldots \lambda_N$(H) be the Eigen values in increasing order and consider the random HIV infection measures.$\frac{1}{n}\left[\delta\left(\lambda_1(H)\right) + \delta\left(\lambda_2(H)\right) + \cdots .. + \delta(\lambda_n(H))\right]$ As the empirical Eigen value distribution of H .its expectation value $\mu_H = \frac{1}{n}\mathbb{E}\left(\sum_{j=1}^{n}\delta_1\,(\lambda_j\,(H))\right)$ called the mean Eigen value distribution of H and it is easy to find out $\int x^m \mu_H(x)dx = \frac{1}{n}\mathbb{E}tr(H^m) = \tau_n(H^m)$.

Theorem (2): Let $\{\boldsymbol{H_n}\}\ \boldsymbol{be}$ independent real symmetric matrix with finite moments such that $\mathbb{E}\left(\boldsymbol{H_{n,ij}}\right) = \mathbf{0}$ and $\mathbb{E}\left(\boldsymbol{H_{n,ij}^2}\right) = \frac{\mathbf{1}}{\boldsymbol{n}}$ for $\mathbf{1} \leq \boldsymbol{i} \leq \boldsymbol{j} \leq \boldsymbol{n}$.if furthermore,

$\boldsymbol{Sup_{1<i<j\leq n}}\mathbb{E}\left|\boldsymbol{H_{n,ij}}\right|\boldsymbol{k} = \mathbf{0}\left(\boldsymbol{n}^{-\frac{\boldsymbol{k}}{\mathbf{2}}}\right)\boldsymbol{for}$ Each $\boldsymbol{k} \in \mathbb{N}$ as $\boldsymbol{n} \to \infty$.

Proof: $\mu_{H,}$the mean Eigen value distribution of H tends to the semi –circle law .the way of interpreting result was $\{H_n\}$ is a family of non –co-commutative random variable with certain distribution family $\{\varphi_n\}$ (Positive linear functional on the $*-algebra$ generated by H and φ_n converges to the semicircle law and weak *-topology.

HIV Random Walk on Diseases free Groups and Free Probability Model

A diseases free groups F_n with n generator $\mathcal{g}_1,\mathcal{g}_{1,}\ldots.\mathcal{g}_n$ in the set of all individual children with these n population, Consider a random walk on Fn which starts from the unit and one step in the move from the group element $\mathcal{g}$ to $with$ probability $(2n^{-1})$ if $h\ \in \left\{\mathcal{g}_1,\mathcal{g}_{1,}\ldots.\mathcal{g}_1^{-1} \ldots \ldots_{\mathcal{g}_m^{-1}}\right\}$,then the probability of the return unit in m steps in the form of

P(n.m) = $\frac{1}{(2n)^m}\left\langle\left(L\mathcal{g}_{1+,}L\mathcal{g}_{1,}\ldots.L\mathcal{g}_1^{-1} \ldots + L\mathcal{g}_{n,}\ldots L\mathcal{g}_n^{-1}\right)m\ \delta_e, \delta_e\right\rangle$,

Where we have observed that in $L_2(G), g \to L_g$ is a unitary representation given by

$\left(L_g, \xi\right)(g^1) = \xi(g^{-1}, g^1)\ for\ \xi \in l^2(G) and\ \delta_g$ Stands for the characteristic functions of the element$g$.

$p(n,m) = 0\ if\ m\ is\ odd\ and$

n= disease free, m=infected odd

$p(n,2m) = (2n)^{-2m}\ n^{-m}\langle(\sum_{j=iweek}^{50weeks} Xn,j)2m,\delta e,\delta e\rangle$, Where 2m= odd (HIV infected at different intervals.

$X_{n,j} = \frac{1}{\sqrt{2}}(Lg_i + Lg_i^{-1})$, asymptotic behavior as $n \to \infty$;

$p(n,m) \approx \frac{1}{(2n)^m}\left(\frac{1}{m+1}\right)\binom{2m}{m}$ Compared with Theorem (1), It was propounded by the central limit theorem for the array of HIV testing done at different intervals.

$\left(X_{n1(i\ I\ week)}, X_{n2(II\ Week)}, X_{n3(III\ Week\ test\ done)}, \ldots\ldots . X_{n(50weeks)}\right)$,of non computing variables of HIV .

Than the converges in distribution to the semicircle law

$\varphi(x) = \langle X\delta_e\delta_e\rangle, than\ X_{n,j}$, Satisfy the following property,

$\varphi(P1X_n, i(1))\ P2X_n, i(2) \ldots\ldots\ldots . Pk\left(X_n, i(k)\right) = 0$

For all polynomial $P_1, P_2 \ldots\ldots P_K$ such that $\varphi(P1X_n, i(j))$=0 and $i(1) \neq i(2) \ldots\ldots . . \neq i(k)$,non commuting random variables is called diseases free, Successive random variables was independent in nature. The model was clearly determined the new born child should underwent the HIV testing at different interval time, the probability were expected to meet the P(n,m) .

ENTROPY IN HIV

Modal Formulation

We considered the classical case. Let $\xi 1, \xi 2$ .... be independent standard real Gaussian random variable and let G be an open set in the space $\mathbb{M}(\mathbb{R})$ of probability measures on $\mathbb{R}$ (with weak *topology),standard Gaussian measures $\vartheta \notin \overline{G}$,then

$prob\left(the\ HIV\ testing\ measures\ at\ different\ interval \frac{1}{n}\sum_{j=1}^{n} \delta_J(\xi_J) \in G\right) \approx \exp(-n(v,G))$

Where G" is the subset prohibition measures, $\delta_J\ random\ variable$

Where $C(v,G) = inf\{I(\mu): \mu \in G\}$ and $I(\mu)$ is the rate of HIV testing done or relative entropy.

$$I(\mu) = -\int p(x)\log p(x)dx + \frac{1}{2}\int x^2\mu(dx) + \frac{1}{2}\log 2x$$

$-\int p(x)\log p(x)dx + \frac{1}{2}$ …….. Entropy function

$\frac{1}{2}\int x^2\mu(dx) + \frac{1}{2}\log 2x$----------Moments.

And $\mu(dx) = p(x)dx.The$ first part of the is the entropy or the Boltzmann-Gibbs entropy $S(\mu) = -\int p(x)logp(x)dx$. The above theorem says that the probability of the large deviation from Gaussianity decreases exponentially with n, and the co efficient multiplying n in the exponent is an infimum of an expression which is essentially the entropy.Large deviation for HIV testing among new born baby "statement that "

$Prob\left\{\frac{1}{n}\sum_j \delta\ \delta(\lambda_j\ (H_n) \in G\right\} \approx e^{-n^2 C(\omega,g),}$ Each entropy is i.i.d and $\frac{n}{n}$ real matrix, where G is an open set $\mathbb{M}(\mathbb{R})$ not containing the semi-circle law $\omega\ and\ C(\omega, G) = Inf\{I(\mu)/\mu\epsilon G\} with$

$$I(\mu) = -\frac{1}{2}\iint log|x-y|\mu(dx)\mu(dy) + \frac{1}{4}\int x^2\mu(dx) + constant$$

In analogy with the classical case, identify $S^f(\mu) = -\frac{1}{2}\iint log|x-$
$y|\mu(dx)\mu(dy) as\ the\ HIV\ non\ tested\ \ free\ entropy.$

The important point was noted that the probability that the empirical Eigen value distribution was different from the semicircle law and decreases sharply with increase of sample size ,i.e. as $\exp(-n^2 C(\omega, G)$ in contrast to the classical case in which the probability for" large deviation" goes like $\exp(-n^2 C(v, G)$

RESULTS & DISCUSSIONS

HIV –MTCT has becomes increasingly NACO(2011), because of the not use ARV-prophylaxis at the time of onset and before birth of neonatal baby, prolong breast feeding , placental absorption and high viral load of pregnant women, drug induced toxicity; illness, lower CD4 count etc., is main charecristic- parameters has increase of MTCT rate. As a simple model, consider the linear model in the form of $y = H_x + n$, where x is K-input variable ,children undergo HIVtestingatdifferentintervals,

$$y\ is\ the\ N - dimentional\ output\ vector, n\ is\ the\ N - out\ put\ resut\ Reactive\ or\ Non\ reactivemodeling,$$

Orthogonally symmetric Gaussian HIV transmission and H is the complex valued random NXK matrix .We considered K is the number of transmitting children while N is the same for receiving infection from mother side.K=$\neq$ $of\ not\ HIV\ reactive\ and\ N = the\ HIV\ \ reactive/infection.$

In the first H is the propagation co efficient for each reactive and non reactive pair of result while second case , each entropy H depend on the transmission /not transmission co efficient ,and $K = n_T J, n = nR_G$ with n_T and n_R the test done and test not done algorithms received from children's and J and g the number of pregnant women,

Who are received not received single dose of NVP at time baby birth itself . Of course ,the simplest case is one where the entries H are i.i.d. was realistically true and correct ,they are not i.i.d.Set for NXN positive matrix A, the normalized distribution function as $F_A^N(x) = \frac{1}{n}\sum_{j=1}^N X(\lambda_j(A) \le x)$ so that
$F_{HH*}^N(x) - N\theta(X) = KF_{HH*}^N(x) - N\theta(X), where\ \theta$ is the Heavy side function .This is because the non-zero Eigen values of HH* and h* H are identical in nature .

CONCLUSIONS

Entropy and HIV random walk on diseases free groups and free probability model method that can be used for dimension Reduction of high dimensional data. This method uses data coverage, density and Correlation to determine the reduced dimension that has good model .While this method does not fluctuate the data, whose results are easy to interpret and more accuracy.

ACKNOWLEDGEMENTS

We thank National AIDS Control organization, and Karnataka State AIDS Prevention Society, Bangalore-560001, India

REFERENCES

1. J. Abello, P.M. Pardalos, and M.G.C. Resende (eds.), Handbook of Massive DataSets, Kluwer Academic Publishers, Norwell, 2002.
2. S.M. Andrijich and L. Caccetta, Solving the multisensor data association problem,Nonlinear Analysis, 47: 5525-5536, 2001.
3. R.K. Ahuja, T.L. Magnanti, and J.B. Orlin, Network Flows: Theory, Algorithm,and Applications, Prentice Hall, Englewood Cli®s, 1993.
4. R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghaven, Automatic subspaceclustering of high dimensional data for data mining. ACM SIGMOD Record, 27(2):94-105, 1998.
5. J. Bellingham, A. Richards, and J.P. How, Receding Horizon Control ofAutonomous Aerial Vehicles. In Proceedings of the American Control Conference,Anchorage, AK, 8-10, 2002.
6. M.J.A. Berry and G. Lino®, Data Mining Techniques for Marketing, Sales andCustomer Support, Wiley, New York, 1997.
7. C.L. Blake and C.J. Merz, (1998). UCI Repository of Machine Learning Databases http://www.ics.uci.edu/~mlearn/MLRepository.html. Oct. 24, 2004.
8. M. Brand, Pattern Discovery Via Entropy Minimization, Uncertainty 99: Inter-national Workshop on Arti¯cail Intelligence and Statistics, (AISTAT) TR98-21, 1998.
9. W. Chaovaliwongse, Optimization and Dynamical Approaches in Nonlinear TimeSeries Analysis with Applications in Bioengineering, Ph.D Thesis, University of Florida, 2003.
10. C. Cheng, A. W. Fu, and Y. Zhang. Entropy-based subspace clustering for miningnumerical data. In Proceedings of International Conference on Knowledge Discov-ery and Data Mining, 84-93, 1999
11. T. Cormen, C. Leiserson, and L. Rivest, Introduction to Algorithms, MIT Press,Cambridge, 2001.
12. R.O. Duba and P.E. Hart, Pattern Classi¯cation and Scene Analysis,Wiley-Interscience, New York, 1974
13. J.Wishart ,Biometricka ,(1928) Vol.20A,PP.32-52
14. R.A.Fisher, Annals of Eugenics (1939) Vol.9,, PP.238-249
15. V.Marcenko, L.Pastur, Distribution of some sets of random matrices, Mat.Sb.Vol.1,pp.507-536.