Journal of Reliability and Statistical Studies (ISSN: 0974-8024)
Vol. 2, Issue 2(2009): 105-111

BAYESIAN INFORMATICS MODELNG OF SYNERGETIC DEVELOPMENT OF RURAL AGRO SYSTEMS

Rajeev Pandey*, Prasoon Verma* and Lalit M Joshi**
*Department of Statistics, Lucknow University, Lucknow (India)
**Department of Statistics, Kumaun University, Almora (India)
Email:*drpandey_r@yahoo.com and prasoon104@gmail.com

Abstract

Integrative thinking patterns have brought about a significant change in our approach towards analysis and applications of various statistical-informatics techniques in our surroundings. The present human society transforming into Information Society has undergone a tremendous metamorphosis from a mundane and monotonous survival to a more exotic and challenging existence. This has revolutionized the overall scenario to the extent that the whole living system can now be thought of as a large-scale complex synergetic system constituting of N dimensional vistas. Rural Agro System has always been the lifetime of the human civilization. Giving continuation to this trend, the present paper puts forward Bayesian-informatics approach dealing with socio-economic coordinates of Rural- Agro System.

The present study attempts in the direction of examining posterior risk and their analysis with the hope that the Bayesian aggregate model would simulate behaviour of the real system at the highest level of accuracy possible under the employed modeling strategy. It has been demonstrated that the Bayesian aggregate models are recursive in nature and they minimize the posterior risk with respect to the time of the system at present. It has also been depicted that the time of minimum posterior risk does not correspond to the time of minimum relative risk, as compared with General State Vector Linear Model (GSVLM) stated by Efraim Halfon [4]in Theoretical Systems Ecology: Advances and Case Studies.

Key words: General State Vector Linear Model, Aggregate Model, Bayesian Aggregation model, Posterior risk.

1. Introduction

The world of computing has been changing at a tremendous pace. The change is enormously rapid, the coordinates related to the various technologies along with the computing environments and other methodologies, and techniques are getting new dimensions at the core of all sorts of Bayesian computational activities. The explosion of data and prior information has obviously brought about significant changes in the overall scenario that deals with the retrieval, manipulation and usage of data. Data Warehouse is one such concept that has truly brought about metamorphosis and providing prior information that are so much required in today's changing world of Bayesian computing that is facing more and more complex and intricate problems with mega-dataset movement across various systems thus paving the way for paradigm shift from linear to non-linear environment, from static to dynamic programming.

Like any other system, Rural Agro System is no exception, unlike a Cartesian model, it is now projected as a living and well-connected network of such systems which are defined in terms of various parameters in the paper. Since these parameters are related in a linear fashion, therefore, a contextual Bayesian relation is established among these multiconnected parameters for a dynamic analytical interpretation.

In case of RAS, the Data Warehouse, as suggested by Neil Raden [7] works as a repository for data elements associated with the system. The central repository would help to increase large pool of data in terms of quality and consistency. Such sequential Data Based (DB) prior information provides a clear and unambiguous definition of every key entity, describing the way, each is used in RAS, as well as defining derivation, formulae, aggregation categories and refreshment time periods. This model linked with the system information architectures, as described by Alan Perkins [1] and Information Discovery [6].

2. General Formulation

In this section, we shall develop some formulae which are prerequisites for the development of a model. Let us denote a system composed of 'p' distinct systems:

$$S_i,\ i = 1, 2, \ldots .p \ \text{ i.e. } S = \bigcup_{i=1}^{p} S_i \tag{2.1}$$

A set **S =** {S} of systems will denote a collection of systems, each member **S** of the collection will necessary contain the same distinct subsystems in order to be in **S**.

For the real world problem being discussed in our work, the subsystems shall consists of (1) Spatial location; (2) Human recourse; (3) Male; (4) Female; (5) Household items; (6) Land preparation; (7) Energy consumption; (8) Farm machinery and equipments; (9) Settlement; (10) Live Stock; It is obvious that here p = 10.

Each subsystem S_i is also composed of set of distinct subsystems elaborated under set of systems mention above. Symbolically, we may represent them as S_{ij}, where j=1, 2,..., m i.e. $S_i = \bigcup_{i=1}^{m} S_{ij}; i = 1, 2, \ldots\ldots, 10$

For the sake of illustration, it may be cited that

S_{11} shall correspond to space.

S_{12} shall correspond to Longitude.

S_{22} shall correspond to Female.

S_{34} shall correspond to migration.

S_{84} shall correspond to cultivated land holding.

$S_{(11)9}$ has been used for wood species for home construction under S_{11}(Settlement) and likewise for other.

The genesis of subsystems and their distinct subsystem can be understood on the basis of set theoretic properties, e.g. $S_2 = S_{21} \bigcup S_{22}$ Human recourse Inventory = *Male* $\bigcup$ *Female*

Similarly $S_9 = S_{91} \bigcup S_{92} \bigcup S_{93} \bigcup S_{94} \bigcup S_{95} \bigcup S_{96} \bigcup S_{97} \bigcup S_{98}$,

where the symbols have their usual meanings. Now let S' denote the Modeled system, then from (2.1)

$$S' = \bigcup_{i=1}^{p'} S'_i \tag{2.2}$$

It is obvious that p may not be equal to p' in the model, nor will the selection of subsystems be identical. But since $S' = \left[\bigcup_{i=1}^{p} S'_i\right] \bigcup \left[\underset{S_i \notin S'}{S'_i}\right] - \left[\bigcup_{S_i \notin S} S_i\right]$ (2.3) where $\underset{S_i \notin S'}{\mathbf{r}} S_i' \mathit{and} \underset{S_i \notin S}{\mathbf{r}} S_i$ are small and / or relatively unimportant in the analysis where S is used instead of S_i'.

The relationships have been marked as M:1 (Many to One) which symbolizes that to each system the left there is one system in the right and M:M (Many to Many) as a complex relationship glorifies the right that of the sub system on the left correspond to a number of elements of the subsystem on the right and vice versa. It suggest the within entities relationship as depicted is the Fig. I between the elements of these subsystems.

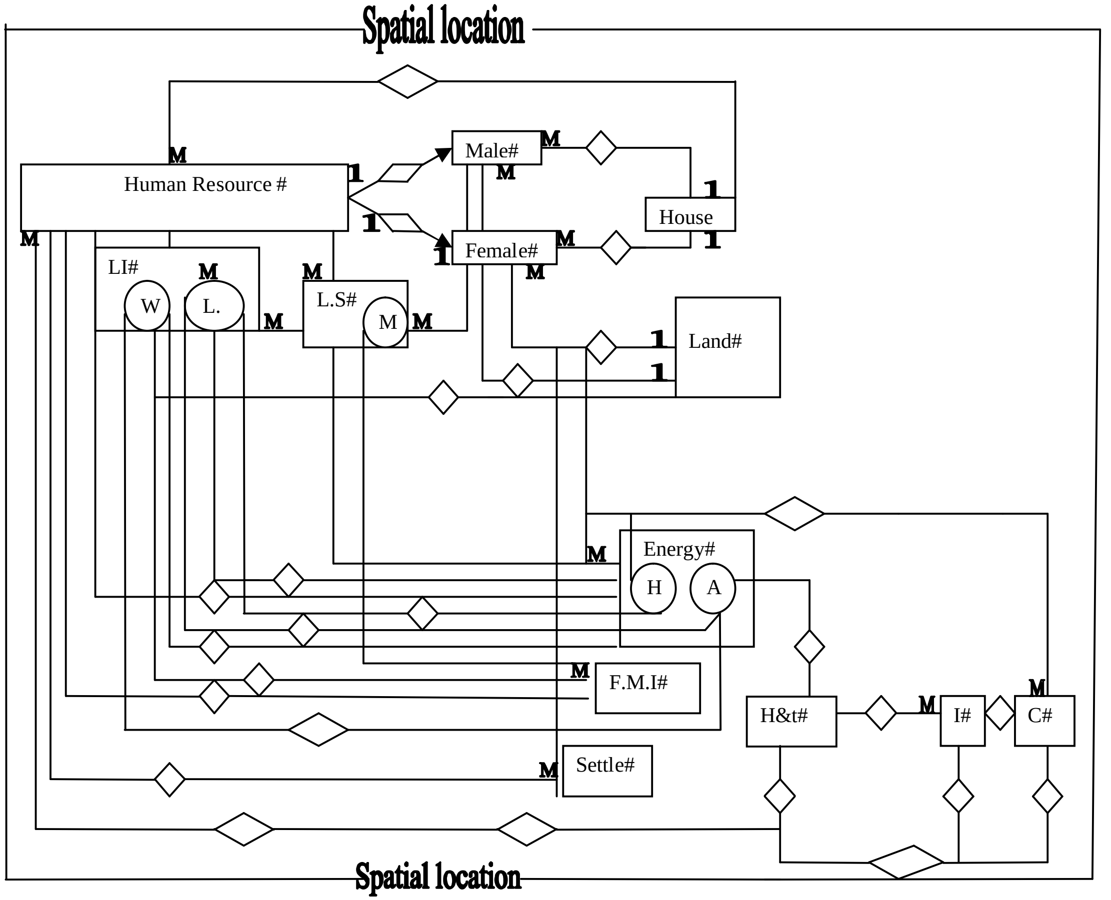

Fig. I

Spatial Location # (Space, longitude, latitude, additional variation, area, density, succession)

Human Recourse # Human Recourse Inventory

Male # (Male, age, education, marital status, migration, occupation, health and nutrition)

Female # (Female, age, education, marital status, migration, occupation, health and nutrition, fertility behavior)

House # (Household items, kitchen ware, entertainment items)
Land # (Land holding, own land, leased in, leased out, cultivated, Unirrigated, irrigated, abandon, fallow, protected etc.)
Energy # (Energy consumption, fuel wood, crop residues, electricity unit, kerosene, dung cake, L.P.G. cylinder, etc.)
H & t # Harvesting; I # Irrigation; C# (Cleaning of food grains)
F.M.I. # (Farm Machinery Implements Wooden plough, mud settler (Mai) spray pump, Storage bins); Settle# (Settlement); LI. # (Land Inventory); LP # (Land Preparation);
W # (Weeding); L.S. # (Live Stock) milking no., non-milking no., breading, initial investment, source of finance, fodder requirement, total; dung product, present value, Milk and milk products.

3. General State Vector Linear Model

Chipman [2] developed the aggregate model which group "true model" variables into sums or weighted averages. Zeigler [8] defined a base model to be a complete conceptualization of a system at a particular level of resolution. Halfon and Reggiani [5] discussed criteria for selecting an optimal model complexity which describes the system's behaviour. They also include an analysis of those errors introduced by using an aggregate model.

The General State Vector Linear Model (GSVLM) proposed by Halfon [4] in the following differential equation

$$x(t) = M(t)x(t) + u(t) \tag{3.1}$$

where $M(t)$ and $x(t)$ represents the optimized values generated by the matrix and values in the present case respectively. $u(t)$ represents a vector of input values as given by

$$u(t) = \sum_{i=1}^{k} m_i \times 1 \tag{3.2}$$

The additive model of the state of system $x(t)$, where $x(t)$ is given by solving (3.1) for $x_i(t)$; i = 1, 2, ...,p Halfon [4] suggested that the linear model has constant coefficients, then M is diagonal matrix defined by

$$M = \begin{bmatrix} -l_1 & 0 & 0 & 0 & 0 \\ 0 & l_2 & 0 & 0 & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & 0 & l_p \end{bmatrix} \tag{3.3}$$

where l_i, i = 1, 2,...,p denotes the turnover rate of one species. One can write

$$x_i(t) = \frac{u_i}{l_i} + \left[x_i(0) - \frac{u_i}{l_i} \right] \exp .[-l_i t] \tag{3.4}$$

for i = 1, 2,... ,p; state vector as well as the state of the system by substituting (3.3) in (3.1) and thus, the equilibrium turnover rate is defined by

$$l_i = \frac{u_i}{x_i^e(0)} \tag{3.5}$$

where $x_i^e(0) = x_i(\infty)$ for the equilibrium state for i = 1, 2,... ,p.

4. Bayesian Aggregation model

Dhami et.al [3] has suggested that the modeling parametric exponential function in (3.4) leads to a high degree of relative errors of approximated model and the summation of a finite number of terms is always preferable as it is most likely that summation of infinite number of terms of the exponential function may lead to wrong results. Dhami et.al [3], therefore, further introduced an Aggregate Model that is recursive in a more complicated form, as given by

$$x(t) = \sum_{i=1}^{p} \left\{ x_i(\infty) + \left(x_i(0) - x_i(\infty)\right)\left(\alpha + \frac{\beta}{l_i + 2t} + \gamma l_i\right)\right\} \tag{4.1}$$

where α, β and γ constraints can be worked out from Data Warehouse or the Pattern Warehouse of the present state of the system. This model when we will use in other location, will have no random effect of the present state of the system. We, therefore, introduce here Bayesian Aggregation model with random white noise.

$$x(t) = \sum_{i=1}^{p} \left\{ x_i(\infty) + \left(x_i(0) - x_i(\infty)\right)\left(\alpha + \frac{\beta}{l_i + 2t} + \gamma l_i\right)\right\} + u(t) \tag{4.2}$$

Where $u(t)$ white noise is normally distributed with zero mean and variance σ^2. Now the constrains α, β and γ can be worked out from the DB Prior, using Data warehouse or pattern warehouse of the present state of the system. The unknown σ^2 plays a dominant role, in turn depends on a number of members of complex elliptical family of shape distribution. Without loss of generality, we can have complex Watson Distribution as a special case to derive and develop the model of the system for the parameters involved including variance σ^2. Let $\pi(\underline{\theta})$ be the prior density for the model given in (4.2),$\underline{\theta} \in \Theta$ the parameter space in the Data Warehouse of the present state of the system, consisting of input variables, as stated in (2.1).

Now the state of system for different values of $i = 1, 2, ..., p$ is defined as the posterior state of the model, as given by

$$x(t) = \sum_{i=1}^{p} x_i(t)\pi(\underline{\theta}) \tag{4.3}$$

Here, $x_i(t)$ has been aggregated using the Data Warehouse entities of the state of the system in RAS. The relative posterior risk in using $x_A(t)$ to approximate $x(t)$ shall be defined by $q_{rpr}(t)$

$$q_{rpr}(t) = \frac{x(t) - x_A(t)}{x(t)} r\left[\underline{\theta}, \delta(\underline{\theta})\right] \tag{4.4}$$

where $r\left[\underline{\theta}, \delta(\underline{\theta})\right]$ is the Bayes risk i.e. the expected Loss in taking action $\delta(\underline{\theta})$ in the present state and the posterior risk given by numeration can be worked out as

$$b_{pr}(t) = [x(t) - x_A(t)]r[\underset{-}{\theta}, \delta(\underset{-}{\theta})]$$

$$= \sum_{i=1}^{p}\left[\{x_i(0) - x_i(\infty)\}\left\{\frac{(l_1 - l_A)\{\gamma(l_i + 2t)(l_A + 2t) - \beta\}}{(l_1 + 2t)(l_A + 2t)}\right\}\right]E\left[L\left\{\underset{-}{\theta}, \delta\left(\underset{-}{\theta}\right)\right\}\right] \tag{4.5}$$

where $L\left\{\underset{-}{\theta}, \delta\left(\underset{-}{\theta}\right)\right\}$ the loss function for the association of new status is affected variables in the present state of the system.

Hence if $l_1 = l_2 = ... = l_p$, then $b_{pr}(t)$ shall be zero for all t. The values of $b_{pr}(t)$ and $q_{rpr}(t)$ shall also be zero if $x_i(0) = x_i(\infty)$ for every value of $i = 1, 2, ..., p$. At $t = 0$, we shall have

$$x_A(t) = \sum_{i=1}^{p} x_i(0) = x(0) \qquad \text{and} \qquad q_{rpr}(0) = 0 \tag{4.6}$$

and as $t \to \infty$ so that $q(\infty) = 0$

Corresponding to value of't' given by (2.2), we shall have the expression for minimum posterior risk

$$[b_{pr}(t)]_{\min} = \sum_{i=0}^{p}[\{x_i(0) - x_i(\infty)\}(2\beta\gamma)]r\left[\underset{-}{\theta}, \delta\left(\underset{-}{\theta}\right)\right] \tag{4.7}$$

The time of minimum posterior risk does not correspond to the time of minimum relative posterior risk, as't' in this case is a solution of the quadratic equation.

Now, additive model can be generated as under

$$x^p(t) = \sum_{i=1}^{p}\left[\{x_i(\infty) + (x_i(0) - x_i(\infty))\}\left\{\alpha + \frac{\beta}{l_i + 2t} + \gamma l_i\right\}^p\right]E\left[L\left\{\underset{-}{\theta}, \delta\left(\underset{-}{\theta}\right)\right\}\right] \tag{4.8}$$

Which is of recursive nature as above expression can be expressed in the following form

$$[x^{p-1}(t) + x^p(t)][r[\underset{-}{\theta}, \delta(\underset{-}{\theta})]] \tag{4.9}$$

where $x^{p-1}(t)$ corresponds to

$$\sum_{i=1}^{p}\left[\{x_i(\infty) + (x_i(0) - x_i(\infty))\}p\gamma(\alpha + \gamma l_i)^{p-1}.\left(\frac{\beta}{l_i + 2t}\right)\right]E\left[\underset{-}{\theta}, \delta\left(\underset{-}{\theta}\right)\right] \tag{4.10}$$

and the remaining terms shall be covered in $x^p(t)$.

The Bayesian aggregate model is also recursive as is evident from following expression

$$x_A{}^p(t) =$$

$$\sum_{i=1}^{p} x_i(\infty)\pi\left(\underset{-}{\theta}\right) + \left[\left\{\sum_{i=1}^{p} x_i(0) - \sum_{i=1}^{p} x_i(\infty)\right\}\left\{\alpha + \frac{\beta}{l_A + 2t} + \gamma l_A{}^2\right\}^p\right]r\left(\underset{-}{\theta}, \delta\left(\underset{-}{\theta}\right)\right) \tag{4.11}$$

which by the explanation given by the expression (4.11), shall assume the form

$$x_A{}^p(t) = \sum_{i=1}^{p}\left[x_i(\infty) + x_p(\infty)\right] + \left[\sum_{i=1}^{p-1} x_i(0) - \sum_{i=1}^{p-1} x_i(\infty)\right]\left(\alpha + \frac{\beta}{l_A + 2t} + \gamma l_A{}^2\right)^p +$$
$$+\left[x_p(0) - x_p(\infty)\right]\left(\alpha + \frac{\beta}{l_A + 2t} + \gamma l_A{}^2\right)^p \tag{4.12}$$

Above equation can be written in recursive form as

$$x_A^p(t) = x_A^{p-1}(t) + \left\{\sum_{i=1}^{p} x_i(0) - \sum_{i=1}^{p-1} x_i(\infty)\right\}\left[\alpha + \frac{\beta}{l_A + 2t} + \gamma_A^2\right]$$
$$+\left\{x_p(0) - x_p(\infty)\right\}\left[\alpha + \frac{\beta}{l_A + 2t} + \gamma l_A^2\right]^p + x_p(\infty) \tag{4.13}$$

The Bayesian Aggregation model mentioned in (4.13) does provide a navigational angle to the complex work of optimizing resources which are multiple facets. The intent of finding posterior risk and its minimum value subject to value of 't' given by equations (4.7) and (4.8) is to ultimately posterior risk so that the output from this model may be more properly interpreted and from (4.8), it can be seen that the time of minimum posterior risk does not correspond to the time of minimum relative risk, as compared with General State Vector Linear Model (GSVLM) stated by Efraim Halfon [4] in Theoretical Systems Ecology: Advances and Case Studies, as given in (3.1) and (3.4).

References

1. Alan Perkins (1996). Data Warehouse Architecture - A Blueprint for Success, www.visible.com.
2. Chipman, J.S. (1975). Optimal Aggregation in Large Scale Econometric Models, Sankhya, Ser. C, 37, p. 121 - 159.
3. Dhami, H.S., Singh, A.K., Negi, G.S. and Shah, A. (2004). The General State vector linear model for sustainable Eco-Development applied on illustrative basis to a sample valley Village of Almora district, IMA Preprint Series # 1937, University of Minnesota, Minneapolis, USA.
4. Efraim Halfon (1979). Theoretical Systems Ecology: Advances and Case Studies, Academic Press, New York.
5. Halfon, E. and Reggiani, M.G. (1978). Adequacy of Ecosystem Models, Ecol. Modell. 4, p. 41 - 50.
6. Information Discovery Warehouse Pattern (2003). The Data Warehouse v/s White Paper, September 2003.
7. Neil Raden (1998). Modeling a Data Warehouse, www.archer-decision.com.
8. Zeigler, B.P. (1976). The Aggregation Problem in Systems Analysis and Simulation in Ecology, (B.C.ed), Vol. 4, p. 299 – 311, Academic Press, New York.