Stochastic Modeling and Applications
Vol. 25 No. 2 (July-December, 2021)
ISSN: 0972-3641

Received: 30th July 2021 *Revised: 01st November 2021* *Selected: 03rd December 2021*

A STUDY OF PROBLEMS IN GRAPH THEORY AND PROBABILITY WITH APPLICATION

SWATI & DR. CHINTA MANI TIWARI

Abstract: Graphical model for probabilistic relationships among set of variables, has become a popular representation for encoding uncertain expert knowledge in expert systems. The techniques, that have been developed, are new and still evolving, but they have been shown to be remarkably effective for some data analysis problems. More specifically, these techniques can be conveniently employed in the study of data in Medical Sciences, Environmental Science, Business and Management Studies, Economics and various branches of Social Science. Formation of networks makes the problems easier to be solved and hence exact and advance inference can be derived from the analysis. The research work features a worked example of Bayesian graphical modelling. Our data for this exercise are (i) the antibody measurements after oral polio vaccination (OPV) obtained from the publication of W.H.O. (ii) The statistics of Pulse Polio immunization held in Jhunjhunu district of Rajasthan State. We begin our analysis by finding a relationship between antibody and time after Trivalent Oral Polio Vaccination (TOPV). Also, we shall formulate graphical models of the variables involved using relevant Probability distribution. Moreover, the success and efficacy of pulse polio immunization programme in Jhunjhunu district also be analysed.

Keywords: graphical model, Graph theory, Probability theory, Bayesian networks

Introduction

A "graphical model" is a type of probabilistic network that has roots in several different research communities, including artificial intelligence (Pearl, 1988), statistics (Lauritzen, 1996), error-control coding (Gallager, 1963), and neural networks. The graphical models framework provides a clean mathematical formalism that has made it possible to understand the relationships among a wide variety of network-based approaches to computation, and in particular to understand many neural network algorithms and architectures as instances of a broader probabilistic methodology.

A graph of order p and size q is called a (p,q)-graph.

It has become a tradition to describe graphs by means of diagrams in which each element of the vertex set of the graph is represented by a dot and an edge $e = uv$ is represented by a curve joining the dots that represent the vertices u and v.

For example, consider the graph G with $V(G) = \{v1, v2, v3, v4, v5\}$ and $E(G) = \{v1v2, v1v3, v2v3, v4v5\}$. Then a possible diagram for this graphs is shown in Figure 1.1

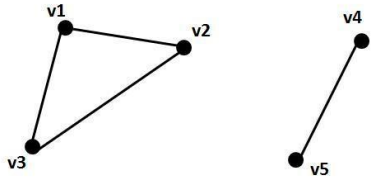

Figure 1.1: A disconnected graph on five vertices

Many situations in real life can be represented by graphs. Graphs has many features like data flow diagram, decision making ability, displays relationships among objects, easy alterations and modifications in existing system etc.

AREAS OF GRAPH THEORY

A concept of graph theory is widely growing and moving into the mainstream of mathematics as it is playing a vital role in following areas:

- In Pure Mathematics
- In Computer Science
- In Operation Research
- In Sociology
- In Science

Graph theory is a delightful playground for the exploration of proof techniques in discrete mathematics, and its results have applications in many areas of computing, social and natural sciences. One can model a road network by a graph having edges that correspond two road segments between intersections. One can assign edge weights to measure distance or travel time. One may want to know the shortest route from x to y.

Inference for Bayesian networks

Probability theory, also known as inductive logic, is a system of reasoning under uncertainty, that is under the absence of certainty. Within the Bayesian framework, probability is interpreted as a numerical measure of the degree of consistent belief in a proposition, consistency being with the data at hand.

The probability of an event A, denoted by P *(A)*, is a number in the interval [0,1], which obeys the following axioms:

P *(A)* = 1 if and only if A is certain,

If A and B are mutually exclusive, then *P (A or B) = P (A) + P (B)*

We will be dealing exclusively with discrete random variables and their probability distributions. Capital letters will denote a variable, or perhaps a set of variables, lower case letters will denote values of variables. Thus, suppose *A* is a random variable having a finite number of mutually exclusive states $(a_1, \ldots, a_n)$.

Then *P*(*A*) will be represented by a vector of nonnegative real numbers *P*(*A*) = $(x_1, \ldots, x_n)$ where $P(A = a_i) = x_i$ is a scalar and $\sum_i x_i = 1$.

A basic concept is that of conditional probability, a statement of which takes the form: Given the event $B = b$, the probability of the event $A = a$ is x, written $P(A = a \mid B = b) = x$. It is important to understand that this is not same as saying: "If $B = b$ is true then the probability of $A = a$ is x". Instead it says: "If $B = b$ is true, and any other information to hand is irrelevant to *A*, then $P(A = a) = x$". To be clear, consider what the probabilities would be if the state of *A* was part of the extra information.

Conditional probabilities are important for building Bayesian networks. But Bayesian networks are also built to facilitate the calculation of conditional probabilities, namely the conditional probabilities for variables of interest, given the data (also called evidence) at hand.

The fundamental rule for probability calculus is the product rule

$P(A \text{ and } B) = P(A|B)P(B)$,

which gives how to combine conditional probabilities for individual variables to define joint probabilities for set of variables.

Review of Literature

There are three criteria for the eradication of an infectious disease: (1) biological and technical feasibility; (2) costs and benefits; and (3) societal and political considerations (Alyward et al. 2000). Current eradication programs include poliomyelitis (polio) (World Health Organization 2008), leprosy (Kealey and Smith.et.al 2010) and guinea worm disease (Smith? et al. 2012).

Pulse vaccination has been investigated in several mathematical models, often in disease models with seasonal transmission. Many diseases show seasonal patterns in circulation; thus inclusion of seasonality may be crucial. Agur et al. (1993) argued for pulse vaccination using a model of seasonal measles transmission, conjecturing that the pulses may antagonise the periodic disease dynamics and achieve control at a reduced cost of vaccination. Shulgin et al. (1998) investigated the local stability of the disease free periodic solution in a seasonally forced population model with three groups: susceptible (S), infected (I) and recovered (R). They considered pulse vaccination and explicitly found the threshold pulsing period (Shulgin et al. 1998). Recently, Onyango and Müller considered optimal periodic vaccination strategies in the seasonally forced SIR model and found that a well-timed pulse is optimal, but its effectiveness is often close to that of constant-rate vaccination (Onyango and Müller 2014). In addition to seasonality, spatial structure has been recognised as an important factor for disease dynamics and control (Xiao et al. 2013). Heterogeneity in the population movement, along with the patchy distribution of populations, suggests the use of metapopulation models describing disease transmission in patches or spatially structured populations or regions. Mobility can be incorporated and tracked in these models in various forms. Common models include linear constant fluxes representing long-term population motion [e.g., migration (Liebovitch and Schwartz 2004)] and nonlinear mass-action representing short-term mobility (Lloyd and Jansen 2004). Liu and Zhou (2009) and Burton et al. (2012) considered epidemic models with both types of movement. A possible inherent advantage of pulse vaccination in a spatially structured setting, discussed by Earn et al. (1998), is that the disease dynamics in coupled regions can become synchronised by pulse vaccination, thereby increasing the probability of global disease eradication. Earn et al. presented simulations of patch synchronisation after simultaneous pulse vaccinations in a seasonal SEIR metapopulation model in which the patch population dynamics were initially out of phase. Here an additional population class of Exposed (E) was considered.

Wagner and Neshat developed an approach based on graph theory to quantify and hence mitigate supply chain vulnerability.

Agrawal et al. explored the various disposition alternatives and developed an approach for the selection of best disposition alternative using Graph Theory and Matrix Approach (GTMA). In GTMA, nodes and edges are two basic elements, which represent the inter-relationship between nodes (different attributes) in the form of digraph and in the form of matrix. These nodes represent the attributes. Agrawal et al. assumed that the digraph consists of a set of nodes $N=\{n\_i\}$ where

i=1,2,3,⋯,Mand a set of directed edge$E$={a_ij}. They selected 10 attributes for the selection of disposition alternative.

Although above these works do not give direct information about the **Study of Problems in Graph Theory and Probability with Application** of pulse polio immunization programme on which this study is made. The works surveyed give vital knowledge about the several mathematical models, often in disease models with seasonal transmission. Most of the books surveyed are those available either in the libraries or in online archives such as JSTOR and Google Books.

Research Methodology and Plan of work

The research work features a worked example of Bayesian graphical modelling. Our data for this exercise are (i) the antibody measurements after oral polio vaccination (OPV) obtained from the publication of W.H.O. (ii) The statistics of Pulse Polio immunization held in Jhunjhunu district of Rajasthan State. We begin our analysis by finding a relationship between antibody and time after Trivalent Oral Polio Vaccination (TOPV). Also, we shall formulate graphical models of the variables involved using relevant Probability distribution. Moreover, the success and efficacy of pulse polio immunization programme in Jhunjhunu district also be analysed.

Results and Discussions

Poliomyelitis is an acute infectious disease mainly which affects the central nervous system. It is caused by polio virus with three sub types 1, 2 and 3. This disease is associated with poor environment hygiene especially the lack of safe water and poor sanitation. The clinical picture may manifest in the form of minor illness, non - paralytic polio myelitis or paralytic poliomyelitis.

A study on this reveal that around 2, 50000 new cases of poliomyelitis occurs in each year over the world. Of these fatality is 10% while 25% become severely impaired and another 50% have some residual weakness and 15% recovered by treatment.

Oral polio vaccine and inactivated polio vaccine

Two types of polio vaccine are available. One is oral polio vaccine (OPV - Sabin) and the other is inactivated polio vaccine (IPV - Sark). Both are effective in controlling and eliminating poliomyelitis.

In India OPV is used in the national immunization programme. Hence, we shall concentrate our investigation only on OPV. Our investigation mainly contains two sections. In the first section, we study the persistence of antibody after booster vaccination on a data obtained from WHO with the help of graphical models. In the later section we shall investigate on the effectiveness of immunerisation programme launched in Jhunjhunu District of Rajasthan state on the basis of the data collected from this region.

Effect of vaccination and Analysis on antibody measurements using Beyesean networks

We have collected a raw data from the publication of WHO related to the persistence of antibody after vaccination and the response to booster vaccination with Trivalent OPV (T.O.P.V). A study was performed on this with graphical models and conclusions were drawn accordingly.

Methods

The study population consists of 175 children ranging from age 2 to 17 years. The age distribution of children receiving polio vaccination. The following table gives the persistence of Booster response in polio neutralizing antibody.

Table No. 1

Persistence of Booster response in polio neutralizing antibody

	No. change in titer at 3 weeks No. (%)	¿ for fold Booster at 3 weeks No. (%)	Sustained ¿ for fold Booster at 6 to 10 weeks No. (%)	Sustained ¿ for fold Booster at 6 months No. (%)
Polio 1	12/26 (46)	14/26 (54)	14/14 (100)	13/14 (93)
Polio 2	10/26 (36)	16/26 (62)	No sample available	13/16 (81)
Polio 3	14/26 (64)	12/26 (48)	12/12 (100)	11/12 (92)
Total	36/78 (46)	42/778 (54)		37/42 (88)

The data in Table No. 1 shows that at six months period in polio I, 93% of children were sustaining fourfold or greater antibody; in polio 2 it is 81% and in Polio 3 it is 92%. Hence a total of 88% of children were sustaining fourfold or greater antibody when it in measured at 6 months period.

A diagrammatic representation of the distribution is shown in Figure 1.2.

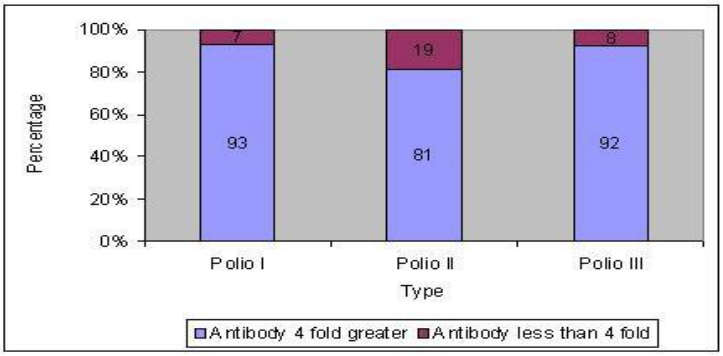

Figure 1.2

Results and findings

1. There is loss in antibody with respect to time since last vaccination and an exponential relation exist between antibody and time.
2. The distribution of antibody (y) can be approximated to normal probability distribution and the likelihood functions and models can be derived from this.
3. Antibody is not related to sex and age of children at the time of immunization
4. Four fold a greater antibody prevails at six months period after Booster vaccination.
5. From the analysis we suggest that for persons traveling to countries where poliomyelitis to prevalent is advisable to take booster TOPV as a preventive mea- sure irrespective of age.

The Polio immunization programme

The immunization programme (giving OPV) has commenced in India in 1998. As a part of the programme the same started in Jhunjhunu District in the same year. Since then the Programme is continuing every year. Under this programme all children below 5 years age will be given two doses OPV every year and as a result on completion of 5 years of age each children will be receiving 8-10 doses of OPV. The statistics collected from the medical authorities of Jhunjhunu shows that an average of 206011. Children were given 2 doses of OPV every

year during the period 2005 to 2009. (See table below), Also average of the target achieved is 99.16

Table No. 2

Sr. No.	Year	Target No. of children	No. of children Vaccinated	% of the target achieved (p^I)
1	2005	208554	206567	99.05
2	2006	208554	206936	99.22
3	2007	207226	205068	98.96
4	2008	207226	205794	99.31
5	2009	207226	20560	99.26
	Total		1030055	

Rajasthan Health Services

District: Jhunjhunu

Average of Children Vaccinated during the year 2005-09 = 206011

Average % of the target achieved (P^I) = 99.16

Moreover the data declare that since 2001 no polio affected case was reported in Jhunjhunu which shows the total success of the programme. Now we shall formulate a statistical procedure to test the truth of the claim made by the District medical authorities with the help of graphical Models.

From the table we see that an average of 206011children were vaccinated during the period 2005 to 2009. For a period of 5 years also the data claims that since 2001 no polio case was reported in the District of Jhunjhunu. Let P denote the proportion of success of the vaccination programme. From the data published by Dr. Sabin, he claims 98% efficiency of the programme.

So we shall choose the null hypothesis as $Ho : P = P_0 = 0.98$; Since in Jhunjhunu District it has produced in 100% Success we shall choose $H_1 : P = P^I = 1$ (where P^I is the value of P in Jhunjhunu District). We shall test H_0 against H_1.

According to statistical theory $P' \sim N(P_0, \sqrt{\frac{p_0 q_0}{n}})$. Consequently we have

$$Z = \frac{P' - P_0}{\sqrt{\frac{p_0 q_0}{n}}},$$

where Z is the standard normal variable and n is the average of the children vaccinated

From the sample study

$$Z = \frac{1 - 0.98}{\sqrt{\frac{0.98 \times 0.02}{206011}}},$$

where n denote the average number of children vaccinated during 2005 to 2009

Thus Calculated value of Z is 64.84

This Value of Z falls in the critical region at 05% level of significance.

Hence we reject the Hypothesis $P = 0.98$ which establishes the acceptance of the hypothesis $H1 : P = 1$. Therefore, the study establishes 100% success of the programme in Jhunjhunu District.

The likelihood function and the graphical model of P^I (the proportion of success)
are given as follows.

$$P' \sim N(\mu, \sigma^2) \text{ where } \mu = P_0 = 0.98 \text{ and } \sigma = \sqrt{\frac{0.98 \times 0.02}{n}},$$

where P^I denote the proportion of target achieved in the immunization programme.

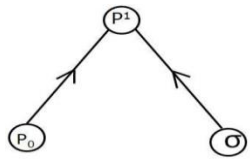

Figure 1.3

Conclusion.

In Summary our observation warrant the following conclusions.

1. The process of giving OPV to all children of age 0-5 years shall be continued without drop outs.
2. Providing 8-10 doses of OPV to children up to the age of 5 years results in 100% control of the disease.
3. Steps shall be taken to achieve 100% immunization coverage.

By implementing this kind of immunization programme all over India we can eradicate the acute infectious disease poliomyelitis from our country. In general, by forming graphical models and neural networks one can study various problems in real life situations and exact inference can be drawn from the investigation which are useful to mankind.

References:

[1] S Agrawal, RK Singh, Q Murtaza (2016) Disposition decisions in reverse logistics: Graph theory and matrix approach. Journal of Cleaner Production 137: 93-104.

[2] SM Wagner, N Neshat (2010) Assessing the vulnerability of supply chains using graph theory. International Journal of Production Economics 126: 121-129

[3] Alyward B, Hennessey KA, Zagario N, Olivé J-M, Cochi S (2000) When is a disease eradicable? 100 years of lessons learned. Am J Public Health 90(10):1515–1520

[4] Kealey AD, Smith? RJ (2010) Neglected tropical diseases: infection, modeling, and control. j health care poor underserved 21:53–69

[5] Agur Z, Cojocaru L, Mazor G, Anderson R, Danon Y (1993) Pulse mass measles vaccination across age cohorts. Proc. Natl. Acad. Sci. 90:11698–11702

[6] Shulgin B, Stone L, Agur Z (1998) Pulse vaccination strategy in the SIR epidemic model. Bull Math Biol 60:1123–1148

[7] Onyango NO, Müller J (2014) Determination of optimal vaccination strategies using an orbital stability threshold from periodically driven systems. J Math Biol 68(3):763–784

[8] Smith? RJ, Cloutier P, Harrison J, Desforges A (2012) A mathematical model for the eradication of Guinea worm disease. In: Mushayabasa S, Bhunu CP (eds) Understanding the dynamics of emerging and re-emerging infectious diseases using mathematical models, pp 133–156

[9] Liebovitch LS, Schwartz IB (2004) Migration induced epidemics: dynamics of flux based multipatch models. Phys Lett A 332:256–267

[10] Liu J, Zhou Y (2009) Global stability of an SIRS epidemic model with transport related infection. Chaos Solitons Fractals 40(1):145–158

[11] Earn DJD, Rohani P, Grenfell BT (1998) Persistence, chaos, and synchronicity in ecology and epidemiology. Proc R Soc Lond B 265:7–10

SWATI: PH.D. RESEARCH SCHOLAR, DEPT. OF MATHEMATICS, MUIT, LUCKNOW, U.P.

DR. CHINTA MANI TIWARI: PROFESSOR, RESEARCH GUIDE, DEPT. OF MATHEMATICS, MUIT, LUCKNOW, U.P.