

1 Introduction

Application of stochastic network to contact tracing as measured by electronic communication devices

2 Extended probability space

2.1 The ProbAdd Monoid

The probability of an event is a real number in the closed interval $\mathbb{P} = [0, 1]$. We seek to equip this interval with the OR operation.

Consider two elements of this interval, $p, q \in \mathbb{P}$. Define the probabilistic sum of p and q as:

$$p \vee q = p + q - pq \quad (1)$$

where pq is the standard multiplication of p and q .

The terms probadd, probaddition and probsum will indicate the use of the probabilistic sum.

It is the t-conorm of the product t-norm, referred to as the AND operation here. The product t-norm is a strict Archimedean t-norm.

Lemma 1. (\mathbb{P}, \vee) is a commutative monoid.

2.2 The ProbAdd Group

Consider $p \in \mathbb{P}$. To construct a group out of the ProbAdd monoid, one must find a probadd inverse of p . That is, we search for q such that

$$p \vee q = 0.$$

It is straightforward to show that $q = -p/(1 - p)$.

To simplify notation, we represent the probadd inverse of p as $\dagger p$. For $p \in \mathbb{P}$ it is trivial to show that $\dagger p \in \mathbb{R}_- \cup \{-\infty\}$. Define $\dagger \mathbb{P} = \mathbb{P} \cup \mathbb{R}_- \cup \{-\infty\}$.

Lemma 2. $(\dagger \mathbb{P}, \vee)$ is a commutative group.

2.3 The Probabilistic Semiring

In addition to the OR operation, one can equip \mathbb{P} with the AND operation. This operation is equivalent to standard multiplication and shall be referred to as such.

Lemma 3. $(\mathbb{P}, \vee, \cdot)$ is a commutative semiring.

2.4 The Probabilistic Field

The probadd inverses have already been included in $\dagger\mathbb{P}$. The multiplicative inverses that remain lie in $(1, \infty)$. Taken together, this forms the extended real numbers, \mathbb{R}^* .

Lemma 4. $(\mathbb{R}^*, \vee, \cdot)$ is a field.

Proof. Not true, since the operations do not allow for distributivity. \square

2.5 The Probabilistic Inner Product Space

Since $(\mathbb{R}^*, \vee, \cdot)$ is a field, the space $(\mathbb{R}^{*n}, \mathbb{R}^*, \vee, \cdot)$ is a vector space. -not true, see proof of prev. lemma

Define the inner product over this vector space as

$$\langle \mathbf{p}, \mathbf{q} \rangle = p_1 q_1 \vee p_2 q_2 \vee \cdots \vee p_n q_n. \quad (2)$$

Lemma 5. $(\mathbb{R}^{*n}, \mathbb{R}^*, \vee, \cdot, \langle \cdot, \cdot \rangle)$ is an inner product space.

2.6 Probabilistic Matrices

To represent the linear transformations of probabilistic vectors, one can construct matrices containing elements of \mathbb{R}^* . One must then define a matrix-vector product. This is made trivial by using the probabilistic inner product:

$$A\mathbf{p} = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_n^\top \end{bmatrix} \mathbf{p} = \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{p} \rangle \\ \vdots \\ \langle \mathbf{a}_n, \mathbf{p} \rangle \end{bmatrix}.$$

3 Stochastic Networks

We define a stochastic network as a graph with vertices taking on a discrete set of k values. The values of each vertex is unknown and has an associated probability vector $\mathbf{v} \in \mathbb{P}^k$ giving the probability distribution of the vertex. Naturally, $\|\mathbf{v}\|_1 = 1$.

As we are considering as an application the transmission of a virus, we restrict ourselves to stochastic networks with the possibility of neighbouring vertices affecting each other's values. The edges of the graph then have associated probabilities, corresponding to the likelihood of transforming one value into another.

3.1 Static Networks

Suppose we examine the spread of an epidemic in a certain time frame. We have measured exactly the contact structure of the population for this time frame, giving a graph with known probabilities of transmission for each edge. For

simplicity, we suppose vertices are binary: 0 indicates no infection, 1 indicates infection. We also suppose that a person infected during this time frame cannot infect another in the same time frame.

At the start of this time frame each vertex has an associated probability of being infected, $a_i \in \mathbb{P}$. Each vertex with a non-zero a_i can pass the virus onto its neighbour indexed by j with probability $b_{i,j}$. The probability that vertex j has the virus at the end of the time frame is the probabilistic sum over all possible paths: $b_{i,j}a_i$ for all i and the probability that the vertex already has the virus, a_j . That is

$$a_j \rightarrow a_j \vee b_{1,j}a_1 \vee \cdots \vee b_{N,j}a_N = \langle \mathbf{b}_j, \mathbf{a} \rangle. \quad (3)$$

Based on the definition of probabilistic matrix-vector multiplication, the transformation of the probabilities over this time frame may then be represented by

$$\mathbf{a} \rightarrow B\mathbf{a} \quad (4)$$

where the elements of the matrix B are the transmission probabilities $b_{i,j}$.

The properties of B depend on the assumptions of the model of the epidemic. We suppose an overly simple SI model: An individual is susceptible until infected, at which point they remain infected indefinitely. In this case the diagonal elements of B are 1. We also suppose that the probability of transmission is symmetric, then B is symmetric. Finally, we suppose that the number of edges on this graph is small in comparison to the number of vertices. The corresponding B is sparse.

All of the mentioned probabilities exist solely in \mathbb{P} . The propagation of probability is then modeled by matrices $B \in \mathbb{P}^{N \times N}$. As \mathbb{P} is not a fullfledged field when equipped with the appropriate operations, the vectors \mathbf{a} exist in a module.

3.1.1 Certainty and the Probabilistic Space

In addition to the propagation of the virus one must also model the effects of accurate testing on the probabilities. Suppose that the i -th vertex, with probability a_i of being infected, tests positive for the virus. Assuming the test is 100% accurate then the transformation of a_i over the time frame in which the test occurred is represented by

$$a_i \rightarrow 1$$

which may be represented by the multiplication of a_i by its multiplicative inverse, $1/a_i$. If the i -th vertex tests negative, this becomes multiplication by zero. Testing on the set of vertices may then be represented by multiplication by a diagonal matrix D whose elements are either 0, 1 or multiplicative inverses.

(nb: need tikz picture of example graphs)

Let the i -th and j -th vertices be neighbours. During a given time frame the probability of these vertices being infected are a_i and a_j , respectively, and there is a probability of $b_{i,j}$ of transmitting the virus between them. Suppose that after this time frame the j -th vertex tests positive for the virus. There are

then two questions to answer: what is the probability that the i -th vertex was infected, and; what is the probability that the i -th vertex is now infected?

Let us begin by answering the first of these questions. We apply Bayes' formula to the problem.

$$\begin{aligned} P(i \text{ was infected} | j \text{ is infected}) &= \frac{P(j \text{ is infected} | i \text{ was infected})P(i \text{ was infected})}{P(j \text{ is infected})} \\ &= a_i \frac{a_j \vee b_{i,j} \vee x_j}{a_j \vee a_i b_{i,j} \vee x_j} \end{aligned}$$

where x_j is the probability of the j -th vertex being infected through an alternative path. This probability is greater than a_i , and so certainty propagates as multiplication by superlinear numbers.

We can also answer this question given that j is not infected in the same way.

$$\begin{aligned} P(i \text{ was infected} | j \text{ is not infected}) &= \frac{P(j \text{ is not infected} | i \text{ was infected})P(i \text{ was infected})}{P(j \text{ is not infected})} \\ &= a_i \frac{1 - a_j \vee b_{i,j} \vee x_j}{1 - a_j \vee a_i b_{i,j} \vee x_j}. \end{aligned}$$

This creates problems for the definitions of our spaces, as subtraction is not properly defined yet. However, one can represent the negation of a probability $p \in \mathbb{P}$ by

$$\neg p = \left(\dagger \frac{1}{2} \right) (p) (\dagger p)^{-1} \quad (5)$$

where $(\dagger p)^{-1}$ is the multiplicative inverse of $\dagger p$ and $\dagger 1/2 = -1$.

4 Retry

We are concerned with the propagation of probability over time through a stochastic network. First we must define the stochastic network we are using. The network changes over time. While the set of nodes $\mathcal{N} = \{1, \dots, N\}$ remains constant, the set of weighted edges $\mathcal{E}(t) = \{(i, j, w) \mid i, j \in \mathcal{N}\}$ is measured during each time frame.

Each edge may be represented by a single number $b_{i,j}(t) \in \mathbb{P}$, indicating the probability of passing the infected condition from node i to node j , or vice versa. A value of zero indicates no edge between the two nodes. The network at time t is then defined by the symmetric matrix

$$B(t) = \begin{pmatrix} 1 & b_{1,2}(t) & \dots & b_{1,N}(t) \\ b_{1,2}(t) & 1 & & \vdots \\ \vdots & & \ddots & b_{N-1,N}(t) \\ b_{1,N}(t) & \dots & b_{N-1,N}(t) & 1 \end{pmatrix}. \quad (6)$$

Note that the diagonal elements are specified by the model. In our case, we suppose that an infected node remains infected.

The time frames are discrete units, and so the variable t is a natural number. The network is then a matrix-valued single variable function, $B : \mathbb{N} \rightarrow \mathbb{P}^{N \times N}$.

The probability that node i is infected at time t is represented by $p_i(t)$, with $p_i : \mathbb{N} \rightarrow \mathbb{P}^N$. The probability vector for the network is then the vector of these numbers:

$$\mathbf{P}(t) = \begin{pmatrix} p_1(t) \\ \vdots \\ p_N(t) \end{pmatrix}. \quad (7)$$

Consider the probability of node i being infected at time $t + 1$. There are a number of possible ways for the node to have the infection. Chiefly, it can have been infected at a previous time ($p_i(t)$). Secondly, it can be infected during time frame t by node j ($b_{i,j}(t)p_j(t)$).

As there are N nodes in the network, there are N possible ways of being infected. Naturally, $p_i(t)$ is the negation of the probability that none of the ways have occurred:

$$p_i(t) = 1 - \prod_{j=1}^N (1 - b_{i,j}(t)p_j(t)).$$

This may also be represented by a probsum, and defines an inner product:

$$p_i(t) = b_{i,1}(t)p_1(t) \vee \cdots \vee b_{i,N}(t)p_N(t) = \mathbf{b}_i(t)^\top \mathbf{P}(t). \quad (8)$$

The values of $\mathbf{P}(t + 1)$ can be written as

$$\mathbf{P}(t + 1) = \begin{pmatrix} \mathbf{b}_1(t)^\top \mathbf{P}(t) \\ \vdots \\ \mathbf{b}_N(t)^\top \mathbf{P}(t) \end{pmatrix}. \quad (9)$$

This helps define a new matrix-vector product, represented by $\otimes : \mathbb{P}^{N \times N} \times \mathbb{P}^N \rightarrow \mathbb{P}^N$. This product simplifies the notation on the propagation of probability through the network:

$$B(t) \otimes \mathbf{P}(t) = \begin{pmatrix} \mathbf{b}_1(t)^\top \\ \vdots \\ \mathbf{b}_N(t)^\top \end{pmatrix} \otimes \mathbf{P}(t) = \begin{pmatrix} \mathbf{b}_1(t)^\top \mathbf{P}(t) \\ \vdots \\ \mathbf{b}_N(t)^\top \mathbf{P}(t) \end{pmatrix} = \mathbf{P}(t + 1). \quad (10)$$

4.1 Results of testing

Suppose at time t that node i is tested for infection. We suppose that this test is perfect. The question then is what effect does this test have on the probabilities of the network. For this, we use Bayes' formula. Consider an event A and a

second event B that has some dependence on event A . The probability that event A occurred given that event B occurs is given by

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (11)$$

Let $p_j^*(t_0)$ represent the probability that node j is infected at time t_0 after it becomes known that node i is infected at time t . By Bayes' formula, this number is equal to

$$p_j^*(t_0) = \frac{p_j(t_0)}{p_i(t)} P(i(t)|j(t_0)). \quad (12)$$

The last term, the probability that node i is infected at time t given that j is infected at time t_0 , is found by supposing that $p_j(t_0) = 1$ and propagating this information through the network to time t . Since our model assumes persistence of the virus, $P(i(t)|i(t_0)) = 1$ and $p_i^*(t_0) = p_i(t_0)/p_i(t)$.

In theory, only those paths dictated by $B(t)$ that connect nodes i and j need be considered, but the selection of these paths may prove more computationally expensive than the use of the \otimes operator. However, this must be done for all time frames and all nodes not equal to i . In applications, it will likely be necessary to consider only recent time frames and short paths.

In the case of a single time frame (ignoring propagating the new probabilities forward, which I don't think apply) the value of $p_j^*(t-1)$ may be expressed as

$$\begin{aligned} p_j^*(t-1) &= \frac{p_j(t-1)}{p_i(t)} \mathbf{b}_i(t-1)^\top \begin{pmatrix} p_1(t-1) \\ \vdots \\ 1 \\ \vdots \\ p_N(t-1) \end{pmatrix} \\ &= \frac{p_j(t-1)}{p_i(t)} \mathbf{b}_i^\top(t-1) \mathbf{P}_j^*(t-1) \\ &= \frac{p_j(t-1)}{p_i(t)} (p_i(t) \vee \dagger(b_{i,j}(t-1)p_j(t-1)) \vee b_{i,j}(t-1)). \end{aligned}$$

The vector on the second line, $\mathbf{P}_j^*(t-1)$, is equal to the vector $\mathbf{P}(t-1)$ with the j -th element replaced by 1. The third line indicates that $p_j^*(t-1)$ may be found by removing the probability created by $p_j(t-1)$ on $p_i(t)$ and replacing it with that created by an infected node j at time $t-1$.

5 Exhaustive spread with removal

As an infection spreads through a network it moves in a single direction. The infection cannot re-infect previous nodes. In a stochastic network probability can move in all directions. Probability can spread back to previous nodes.

To properly model this unidirectional spread, the probability needs to also move unidirectionally. The spread of probability can be written as paths from one infected node to other nodes. These are walks through a network that do not form loops (loop-erased random walks). The spread of probability is then the collection of all such walks.

Constructing these walks as they form can be done by spreading the infection one node at a time through networks with nodes already on the path removed. Let the contact structure be represented by the symmetric matrix $B(t)$ with elements $b_{i,j}(t)$. The probability of the network is represented by a list of probabilities indexed by node and path:

$$P(t) = \{p_{i,\mathcal{A}} \mid i \in \mathcal{N}, \mathcal{A} \subset \mathcal{N}\}. \quad (13)$$

For a given element of $P(t)$, $p_{i,\mathcal{A}}$, we take all elements $b_{i,j}(t)$ such that $j \notin \mathcal{A}$ and create new elements for each. Therefore, $\{b_{i,j}(t)p_{i,\mathcal{A}} \mid j \notin \mathcal{A}\} \subset P(t+1)$.

The total probability that the i -th node at time t is infected is the probabilistic sum of $p_{i,\mathcal{A}}$ for all \mathcal{A} . Let this be represented by $p_i = \vee_{\mathcal{A}} p_{i,\mathcal{A}}$.

5.1 Bayesian inference and testing

Suppose node i is tested for the virus. This necessarily affects the probabilities of all nodes on the path before and after it.

Suppose the i -th node tests positive for the virus. That is, the event predicted with the probability p_i occurs. This necessarily means that at least one of $p_{i,\mathcal{A}}$ has occurred. Using Bayes' formula, the new probability of $p_{i,\mathcal{A}}$ is

$$p_{i,\mathcal{A}} \rightarrow \frac{p_{i,\mathcal{A}}}{p_i}.$$

Naturally this change affects any probabilities further along the path. Consider the j -th node with a path \mathcal{B} that contains \mathcal{A} as a subset. While the i -th node is found to be infected, the particular path is unknown. Since $p_{i,\mathcal{A}}$ has been updated, $p_{j,\mathcal{B}}$ must be as well:

$$p_{j,\mathcal{B}} \rightarrow \frac{p_{j,\mathcal{B}}}{p_i}.$$

Bayes' formula is needed for probabilities appearing earlier on the path. Consider the k -th node with a path \mathcal{C} that is a subset of at least one path \mathcal{A} that includes the node i . The probability of the i -th node being infected along path \mathcal{A} given that the k -th node is infected along path \mathcal{C} is the probabilities along the path $\mathcal{A} \setminus \mathcal{C}$, which is simply $p_{i,\mathcal{A}}/p_{k,\mathcal{C}}$. For Bayes' formula, one must take the probabilistic sum over all such $\mathcal{A} \setminus \mathcal{C}$ and those that do not include \mathcal{C} . Bayes' formula then gives the following update to the probability $p_{k,\mathcal{C}}$:

$$p_{k,\mathcal{C}} \rightarrow \left(\left(\sum_{\mathcal{A} \supset \mathcal{C}} \frac{p_{i,\mathcal{A}}}{p_{k,\mathcal{C}}} \right) \vee \left(\sum_{\mathcal{A} \not\supset \mathcal{C}} p_{i,\mathcal{A}} \right) \right) \frac{p_{k,\mathcal{C}}}{p_i} = \frac{\gamma(i, k, \mathcal{C})}{p_i} p_{k,\mathcal{C}}.$$

Consider the l -th node along a path \mathcal{D} that contains the path \mathcal{C} as a subset. Since the probabilities along a path are linked only by multiplication, the update to $p_{l,\mathcal{D}}$ is the same multiplier applied to $p_{k,\mathcal{C}}$:

$$p_{l,\mathcal{D}} \rightarrow \frac{\gamma(i,k,\mathcal{C})}{p_i} p_{l,\mathcal{D}}.$$

Note that the path between nodes l and k does not contain any previously updated node-path combinations. This exhausts all types of updates to probabilities.

The calculation of $\gamma(i,k,\mathcal{C})$ can be done by starting from p_i rather than constructing it piecemeal.

$$\begin{aligned} \gamma(i,k,\mathcal{C}) &= \left(\sum_{\mathcal{A} \supset \mathcal{C}} \frac{p_{i,\mathcal{A}}}{p_{k,\mathcal{C}}} \right) \vee \left(\sum_{\mathcal{A} \not\supset \mathcal{C}} p_{i,\mathcal{A}} \right) \\ &= p_i \vee \left(\sum_{\mathcal{A} \supset \mathcal{C}} \dagger p_{i,\mathcal{A}} \right) \vee \left(\sum_{\mathcal{A} \supset \mathcal{C}} \frac{p_{i,\mathcal{A}}}{p_{k,\mathcal{C}}} \right) \\ &= p_i \vee \left(\sum_{\mathcal{A} \supset \mathcal{C}} \dagger p_{i,\mathcal{A}} \vee \frac{p_{i,\mathcal{A}}}{p_{k,\mathcal{C}}} \right) \\ &= p_i \vee \left(\sum_{\mathcal{A} \supset \mathcal{C}} \frac{p_{i,\mathcal{A}}(1 - p_{k,\mathcal{C}})}{p_{k,\mathcal{C}}(1 - p_{i,\mathcal{A}})} \right). \end{aligned}$$

Depending on the number of relevant paths \mathcal{A} this second method may prove faster.

Suppose instead that the i -th node tests negative for the virus. This necessarily means that none of $p_{i,\mathcal{A}}$ have occurred. Their probabilities are therefore zero. All paths $\mathcal{B} \supset \mathcal{A}$ now have probability zero.

$$p_{i,\mathcal{A}} \rightarrow 0, \quad p_{j,\mathcal{B}} \rightarrow 0.$$

To update the k -th and l -th node as we did before we again use Bayes' formula. The probability that the i -th node is not infected given that the k -th node along path \mathcal{C} is infected is the negation of $\gamma(i,k,\mathcal{C})$.

$$p_{k,\mathcal{C}} \rightarrow \frac{1 - \gamma(i,k,\mathcal{C})}{1 - p_i} p_{k,\mathcal{C}}, \quad p_{l,\mathcal{D}} \rightarrow \frac{1 - \gamma(i,k,\mathcal{C})}{1 - p_i} p_{l,\mathcal{D}}.$$

A Notes

This model includes the possibility of self-infection through loops in the network. If node i has a probability of infection then this probability may spread to node j . If node j comes into contact with node i , some of the probability it transfers will have originated in i .

Ideally, the propagation of probability would work only on loop-erased random walks or self-avoiding walks (not sure which).

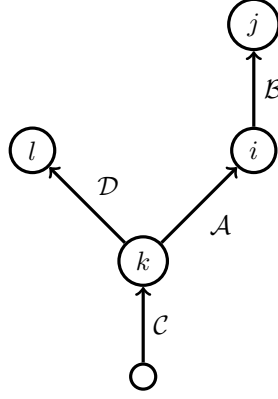


Figure 1: Example relation of nodes i , j , k and l , with paths \mathcal{A} , \mathcal{B} , \mathcal{C} and \mathcal{D} .

$p_i \rightarrow$	1	0
$p_{i,\mathcal{A}} \rightarrow$	$\frac{p_{i,\mathcal{A}}}{p_i}$	0
$p_{j,\mathcal{B}} \rightarrow$	$\frac{p_{j,\mathcal{B}}}{p_i}$	0
$p_{k,\mathcal{C}} \rightarrow$	$\frac{\gamma(i,k,\mathcal{C})}{p_i} p_{k,\mathcal{C}}$	$\frac{1-\gamma(i,k,\mathcal{C})}{1-p_i} p_{k,\mathcal{C}}$
$p_{l,\mathcal{D}} \rightarrow$	$\frac{\gamma(i,k,\mathcal{C})}{p_i} p_{l,\mathcal{D}}$	$\frac{1-\gamma(i,k,\mathcal{C})}{1-p_i} p_{l,\mathcal{D}}$

Table 1: Updates of probability based on the results of testing and Bayes' formula. The paths follow the relationship $\mathcal{D} \supset \mathcal{C} \subset \mathcal{A} \subset \mathcal{B}$.