A Note on Negation of a Probability Distribution

Manpreet Kaur
Jaypee Institute of Information Technology
AMIT SRIVASTAVA (**amit90554@gmail.com**)
Jaypee Institute of Information Technology https://orcid.org/0000-0002-4521-6732

Research Article

Keywords: Negation, Markov chain, Uncertainty, Probability distribution and Self information, Information generating function(IGF)

Posted Date: August 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1112844/v1

A Note on Negation of a Probability Distribution

Manpreet Kaur · **Amit Srivastava***

Received: date / Accepted: date

Abstract Evaluating the negation of an uncertain event is an open issue. Yager[14] suggested a transformation for evaluating the negation of a probability distribution. He used the idea that any event whose outcome is not certain can be negated by supporting the occurence of other events with no bias or prejudice for any particular outcome. Various authors have tried to generalize the negation transformation proposed by Yager[14]. However we need to focus on developing the basic structure of negation so that the behaviour of the process modelled by negation transformation can be understood in detail.Yager's negation is based on distribution of maximum entropy. If a probability distribution is uncertain(a state other than maximum entropy), the more iterations of negation, the more uncertain this probability event becomes eventually converging to a homogeneous state *i.e.* maximum entropy. In other words, it is the realization of the process. What is noted that during each negation, Yager's method ensures that the negation is intuitive, the next negation weakens the probability of the event occuring in the previous step. Since negation involves reallocation of probabilities at each step in such a way that the reallocation at each step can be determined from the reallocation at the previous step, therefore it is clear that Yager's negation has various attributes similar to that of a markov chain. In the present work, we have shown that Yager's definition of negation can be modelled as a markov chain which is irreducible, aperiodic with no absorbing states. Two examples has been discussed to strengthen and support the analytical results. Also we have defined an information generating function(IGF) whose derivative evaluated at specific points gives the moments of the self information of negation of a probability distribution. The properties of the generating function along with its relationship with the information generating

Manpreet Kaur
E-mail: manpreet1511@rediffmail.com
Department of Mathematics, Jaypee Institute of Information Technology, Noida-201304, India.

Amit Srivastava
E-mail: amit.srivastava@jiit.ac.in *Corresponding author
Department of Mathematics, Jaypee Institute of Information Technology, Noida-201304, India.

function proposed by S. Golomb[2] has been explored. A closer look at the properties of IGF confirms the existence of markovian structure of Yager's Negation.

Keywords Negation · Markov chain · Uncertainty · Probability distribution and Self information · Information generating function(IGF)

Mathematics Subject Classification (2010) 90C70 · 90B06 · 90B50 · 90C05

1 Introduction

Affirmation and Negation are two key concepts in various forms of human communication. An affirmation form is generally used to express the validity or truth of a assertion whereas the negative form determines its falsehood. In classical logic, if a statement P is TRUE, then its negation $\sim P$ is FALSE and if a statement P is FALSE, then its negation $\sim P$ is TRUE. The negation gives a different perspective for any happening in either society or nature. Two persons can examine same situation from distinct (positive and negative) perspectives and both may have reasonable justification for it. For example, various psychological studies outlines four approaches for regulating behaviour based on the repercussions and desired objective - positive reinforcement, negative reinforcement, positive punishment, and negative punishment(See Table 1).

Table 1 Reinforcement vs Punishment

	Reinforcement	Punishment
Positive	A pleasant stimulus is added to strengthen the likelihood of a particular response.	An undesirable stimulus is added to discourage a behaviour.
Negative	An undesirable consequence is removed to strengthen the likelihood of a particular response.	A pleasant stimulus is removed to discourage a behaviour.

Rewarding your child so that he/she performs an activity that is any case expected from him/her is an example of positive reinforcement, whereas the seat belt reminders installed in cars are examples of negative reinforcement(the irritating sound stops when you perform the desired behaviour). Scolding a student when he/she misbehaves in school is an example of positive punishment whereas taking away privileges by parents when children misbehaves proves that negative punishment can be an effective discipline strategy. Both Reinforcement and Punishment have notable drawbacks but it solely depends on the individual how he/she uses any of the above approaches to his/her benefit. Anything negative doesn't necessarily mean unpleasant or unacceptable, it represents the opposite side of different aspects of life. This opposite side of life may be fascinating, annoying and it will have lots of uncertainty inherent in it. Also the whole idea of any assertion being either true or false is very restrictive. In our daily lives, we experience situations which are neither true nor false, in fact we come across so many events which has uncertainty inherent in it. An event or a sequence of events whose occurrence is not guaranteed cannot be expressed via True false logic. The probability theory has been quite effective in handling such situations. How to express the negation(opposite side) of an uncertain event has been a matter of discussion for so many years. **If the happening of an event is uncertain, then we can oppose(negate) it by using its probability**. Keeping this in mind, Yager[14] gave the basic framework of negation of a probability distribution. The negation proposed by Yager is basically an unbiased reallocation of probabilities. Many studies focus on determining how much uncertainty/knowledge is embedded in negation of a probability distribution. Various authors [9], [10], [11] have shown that more uncertainty(information) is embedded in the negation of a probability distribution. However we need to focus more on the underlying mathematical structure of negation and its properties which can increase its applicability in various domains. We consider the probability distribution $P(3) = (p_1, p_2, p_3)$ such that $p_1 + p_2 + p_3 = 1$ and $0 \leq p_i \leq 1; i = 1,2,3$. For simplicity we consider the degenerate distribution $P(3) = (1,0,0)$. The negation of $P(3)$ is given as $\overline{P}(3) = (0,0.5,0.5)$(See Figure 1). While determining the negation, $p_1 = 1$ is equally distributed among the second and third components which signifies that by negating(opposing) the occurrence of first event(with probability 1), we are supporting the occurrence of second and third event without any bias. Therefore the probability $p_1 = 1$ is equally distributed among the probability of second and third event. The second and third probabilities $p_2 = 0$ and $p_3 = 0$, could not contribute anything to the negation. In other words, the support that the occurrence of first event had in the original distribution has been equally divided among the second and the third event. Applying the negation transformation again, we obtain $\overline{\overline{p}}(3) = (0.5,0.25,0.25)$. Here the first entry $\overline{\overline{p}}_1$ is a result of equal contributions of 0.25 and 0.25 from $\overline{p}_2$ and $\overline{p}_3$ respectively. Again the second entry $\overline{\overline{p}}_2$ is a result of equal contributions of 0 and 0.25 from $\overline{p}_1$ and $\overline{p}_3$ respectively. Similar is the case with the third entry. Here it is interesting to note that whatever support we have for $\overline{P}(3) = (\overline{p}_1, \overline{p}_2, \overline{p}_3)$ is a result of redistribution of entries of $P(3) = (p_1, p_2, p_3)$. Similarly whatever support we have for $\overline{\overline{p}}(3) = (\overline{\overline{p}}_1, \overline{\overline{p}}_2, \overline{\overline{p}}_3)$ is a result of redistribution of entries from $\overline{P}(3) = (\overline{p}_1, \overline{p}_2, \overline{p}_3)$. But the support for $\overline{\overline{P}}(3)$ can be easily determined from the support for $\overline{P}(3)$ ignoring the support for various events in $P(3)$. Therefore for determining the distribution at the second iteration, one needs knowledge of distribution at first iteration only(and not of the original distribution). Similar will be the case for further iterations of the negation transformation. This shows that the negation transformation proposed by Yager has various attributes identical to that of a markov chain. One more question that immediately arises is that whether there exists any generating function associated with the negation transformation. For a discrete finite complete probability distribution $P(n) = (p_1, p_2, \ldots, p_n)$, S. Golomb[2] introduced the information generating function(IGF) given as

$$I_t(P(n)) = \sum_{i=1}^{n} p_i^t \,, t \geq 1 \tag{1}$$

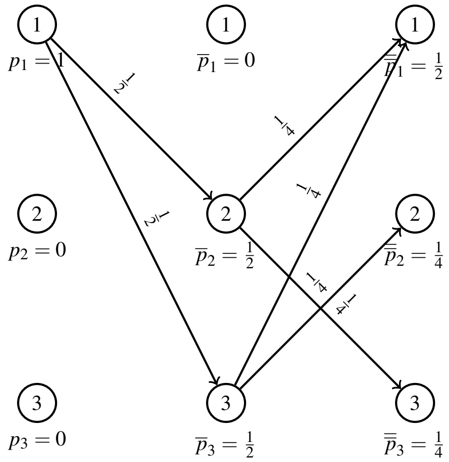

Fig. 1 Reallocation of probabilities in (1, 0, 0)

Differentiating (1.2) at point $t = 1$ gives

$$-\left(\frac{\partial}{\partial t} I_t(P(n))\right)_{t=1} = -\sum_{i=1}^{n} p_i \ln \ p_i = H\left(P(n)\right) \tag{2}$$

where $H\left(P(n)\right)$ is the well known shannon entropy function[7],[8]. On further differentiating $(k-1)$ times, we will get the k^{th} moment of the self information embedded in $P(n)$.
Various authors have tried to generalize the negation transformation proposed by Yager[14]. [16], proposed negation of probability distribution based on Tsallis entropy that degenerates into Yager's negation. The concept of negation is widely applied in various fields. In [3], negation of Z-numbers is proposed. Researchers applied concept of negation in D-S evidence theory also by defining negation of BPA based on pythagorean fuzzy numbers[5], maximum uncertainty allocation[1], reallocation[15], matrix method[4] and belief interval approach[6] and applied that in service supplier selection system[5], medical pattern recognition[6], decision making[13] and many more. In the present work, we have investigated the properties of markov chain that is embedded in the negation of a probability distribution. Some illustrative examples have been considered which shows obvious correlation between the markov chain and negation transformation. Also we have proposed an information generating function whose derivatives at specific points gives the moments of the self information(information content) embedded in the negation of a probability distribution. The discussed examples clearly indicate that the proposed generating function has evident connection with the information generating function proposed by S. Golomb [2].

2 Preliminaries

2.1 Negation of a Probability distribution

Let $X = \{X_1, X_2, \dots, X_m\}$ be the frame of discernment(FOD), the set of all possible hypothesis under consideration and let, $P(n) = (p_1, p_2, \dots, p_n)$ be a discrete finite complete probability distribution defined on X with $p_i \in [0,1]$ for $i = 1, 2, \dots, n$ and $\sum\limits_{i=1}^{n} p_i = 1$. The negation of probability distribution proposed by Yager can be written as the set $\overline{P}(n) = (\overline{p}_1, \overline{p}_2, \dots, \overline{p}_n)$, where

$$\overline{p}_1 = \frac{1-p_1}{n-1} = 0.p_1 + \frac{1}{n-1}.p_2 + \frac{1}{n-1}.p_3 + ... + \frac{1}{n-1}.p_n$$

$$\overline{p}_2 = \frac{1-p_2}{n-1} = \frac{1}{n-1}.p_1 + 0.p_2 + \frac{1}{n-1}.p_3 + ... + \frac{1}{n-1}.p_n$$

$$\vdots$$

$$\overline{p}_n = \frac{1-p_n}{n-1} = \frac{1}{n-1}.p_1 + \frac{1}{n-1}.p_2 + \frac{1}{n-1}.p_3 + ... + 0.p_n$$

which can be further written in matrix form as

$$\overline{P}(n) = \begin{pmatrix} \overline{p}_1 \\ \overline{p}_2 \\ . \\ . \\ . \\ \overline{p}_n \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \\ \frac{1}{n-1} & 0 & \cdots & \frac{1}{n-1} \\ . & . & .... & . \\ . & . & .... & . \\ . & . & .... & . \\ \frac{1}{n-1} & \frac{1}{n-1} & \cdots & 0 \end{pmatrix} . \begin{pmatrix} p_1 \\ p_2 \\ . \\ . \\ . \\ p_n \end{pmatrix}$$

Here the probabilities in the set $\overline{P}(n)$ satisfies $0 \le \overline{p}_i \le 1$ and $\sum_{i=1}^{n} \overline{p}_i = 1$. Yager[14] specified that there can be many distinct negations embedded in a probability distribution and the above is the one that has the maximum entropy allocation among all the possible negations. In particular let $Q(n) = (N(p_1), N(p_2), \dots, N(p_n))$ denote the unbiased rearrangement of probabilities $P(n) = (p_1, p_2, \dots, p_n)$ with $N(p_i) \in [0,1]$ for $i = 1, 2, \dots, n$ and $\sum\limits_{i=1}^{n} N(p_i) = 1$. Then $Q(n)$ is called negation of $P(n)$ if

1. $p_i \le p_j$ gives $N(p_i) \ge N(p_j)$ for all $i, j = 1, 2, \dots, n$
2. If the probabilities in $P(n)$ are all equal, then all entries of $Q(n)$ are also equal. Further the pooled opinion of $P(n)$ and $Q(n)$(Convex combination of $P(n)$ and $Q(n)$) should not reflect any additional knowledge about the occurrence of events.
3. The set $Q(n) = \{N(p_1), N(p_2), \dots, N(p_n)\}$ should preserve the underlying mathematical structure of the probability distribution P.

For better understanding of third condition,
consider a random variable $X = (x_1, x_2, x_3, x_4, x_5) = (10, 11, 12, 13, 14)$ with probabilities $P(5) = (p_1, p_2, p_3, p_4, p_5) = (0.1, 0.2, 0.4, 0.2, 0.1)$.
The expectation of X is given as
$E_{P(5)}(X) = \sum_{i=1}^{5} p_i x_i = 10(0.1) + 11(0.2) + 12(0.4) + 13(0.2) + 14(0.1) = 12 = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$.
The Yager's negation of $P(5)$ is
$\overline{P}(5) = (\overline{p}_1, \overline{p}_2, \overline{p}_3, \overline{p}_4, \overline{p}_5) = (0.225, 0.2, 0.15, 0.2, 0.225)$ and
$E_{\overline{P}(5)}(X) = \sum_{i=1}^{n} \overline{p}_i x_i = 10(0.225) + 11(0.2) + 12(0.15) + 13(0.2) + 14(0.225) = 12 = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$.
Here the negation transformation has preserved the symmetry of $P(5)$ since the probabilities are redistributed among all the alternatives equally(basic structure of $P(5)$ remains unaltered). If the rearrangement is biased *i.e.* we distribute more to some and less to others, then the symmetry may get disturbed. Repeatedly applying the negation transformation on $P(n)$ will yield the probability distribution $(0.2, 0.2, 0.2, 0.2, 0.2)$.

2.2 Markov Chain

Markov chain gives a mathematical framework which characterizes transitions from one state to another using some probabilistic rules. Markov chains are the stochastic processes for which the description of the present state fully captures all the information that could influence the subsequent developments of the process. Mathematically a stochastic process $X = \{X_n : n \geq 0\}$ on a countable set S is a Markov Chain if, for any $i, j, k_1, k_2, \ldots k_{n-1} \in S$ and $n \geq 0$,

$$P(X_{n+1} = j | X_n = i, X_{n-1} = k_1, X_{n-2} = k_2, \ldots, X_0 = k_{n-1}) = P(X_{n+1} = j | X_n = i) = p_{ij}$$

Here p_{ij} is the probability representing transitions from state i to state j. These transition probabilities sum up to 1 *i.e.*

$$\sum_{j=1}^{n} p_{ij} = 1.$$

for each $i \in S$. In general, we list all the transition probabilities in a matrix. This matrix is called the stochastic matrix or state transition matrix or transition probability matrix. For n states, the transition probability matrix is a $n \times n$ square matrix given as

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

The above matrix can be right stochastic(each row summing to one) or doubly stochastic(each row and each column summing to 1). Moreover the above matrix represent one step transitions only. The probability of transition from state i to state j in m steps is represented as $p_{ij}^{m} = P(X_{n+m} = j/X_n = i)$. The m step transition probabilities can be evaluated by multiplying P with itself m times.Further a markov chain is said to be irreducible if transitions are possible between every pair of states(in a finite number of steps) with positive probability. Also if the return to a particular state occurs at equal intervals of time, then that state is said to be periodic otherwise aperiodic. A markov chain is said to be aperiodic if all its states are aperiodic. Here it is worth mentioning that if one of the states in an irreducible markov chain is aperiodic, then all the other states are also aperiodic. Irreducibility and aperiodicity properties are important for characterizing the ergodicity of a Markov chain. There is possibility that once we reach any particular state in a markov process,then it is impossible to leave that state *i.e.* $p_{ii} = P(X_{n+i} = i/X_n = i) = 1$. Such states are called absorbing states. A markov chain is an absorbing markov chain if(a)there is atleast one absorbing state among all the states characterizing the markov chain; and(b)it is possible to go from any state to at least one absorbing state in a finite number of steps. A state which is not absorbing is called a transient state. Further as the time index approaches infinity, some markov chain may exhibit steady state behaviour. The steady state distribution of a markov chain is generally represented as a row vector π whose entries sum to one and satisfies $\pi P = \pi$, P being the transition probability matrix. The stability of a random process can be determined using steady state distribution and in some cases, it describes the limiting behaviour of the markov chain. An example of doubly stochastic matrix is permutation matrix which is very important from combinatorial point of view. Given a permutation f of k elements, $f : \{1,2,\ldots,k\} \to \{1,2,\ldots,k\}$ which can be represented as

$$\begin{pmatrix} 1 & 2 & \cdots & k \\ f(1) & f(2) & \cdots & f(k) \end{pmatrix},$$

a permutation matrix is a $k \times k$ matrix $P_f = (p_{ij}); i,j = 1,2,\ldots,n$ obtained by setting $p_{ij} = 1$ if $j = f(i)$ and $p_{ij} = 0$ otherwise for all $i = 1,2,\ldots,n$; or alternatively $p_{ij} = 1$ if $i = f(j)$ and $p_{ij} = 0$ otherwise for all $j = 1,2,\ldots,n$. It is possible to define the negation of a probability distribution via the entries of a doubly stochastic matrix. Consider a $n \times n$ doubly stochastic matrix $A = (a_{ij})$ *i.e.*

$$\sum_{i=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ij} = 1.$$

In particular if $a_{ii} = 0$ for all $i = 1,2,\ldots,n$ and if $a_{ij} = \frac{1}{n-1}$ for all $i \neq j$; $i,j = 1,2,\ldots,n$, then given a probability distribution $P(n) = (p_1, p_2, \ldots, p_n)$, we can define its negation $\overline{P}(n) = (\overline{p}_1, \overline{p}_2, \ldots, \overline{p}_n)$ as

$$\overline{p}_i = \sum_{j=1}^{n} a_{ij} p_i \tag{3}$$

for all $i = 1, 2, \ldots, n$. Since the elements in negation of a probability distribution can be represented in terms of entries of a doubly stochastic matrix and there are many markov chains where the transition probability matrix is doubly stochastic, it is clear that the underlying mathematical structure of negation can be modelled in a markov chain. In the next section, we will discuss some examples which will validate the above discussion.

3 Negation and Markov chain

Let Y_n represents the sum of n independent rolls of a fair die and O_i be the outcome on i^{th} die, $i = 1, \ldots, n$. Further let X_n denote the remainder when Y_n is divided by 7. Then $Y_n = O_1 + O_2 + \ldots + O_n = Y_{n-1} + O_n$ and $X_n = (X_{n-1} + O_n \text{mod} 7)\text{mod} 7$ will represent a Markov chain with states 0,1,2,3,4,5,6. As O_n can take values $1, 2, \ldots, 6$; $O_n\text{mod}7$ can't be zero. So transition from one state to itself is not possible and thus probability of transition from a state to itself is always zero. The probability of transition from state any state i to state j, $j = 0, 1, \ldots, 6$, $j \neq i$ is same as probability of getting outcome O_n as $1, 2, \ldots,$ *or* 6 i.e., $\frac{1}{6}$. Thus transition probability matrix is given by

$$P = \begin{pmatrix} 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \end{pmatrix}$$

Now suppose we obtain a 3 in first roll of the die. Then the initial distribution is given by

$$\begin{aligned} P(1) &= \begin{pmatrix} P(X_1 = 0) & P(X_1 = 1) & P(X_1 = 2) & P(X_1 = 3) & P(X_1 = 4) & P(X_1 = 5) & P(X_1 = 6) \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Here $P(1)$ represents the initial probability vector. The probability distribution after one transition is given as

$$\begin{aligned} P(2) &= \begin{pmatrix} P(X_2 = 0) & P(X_2 = 1) & P(X_2 = 2) & P(X_2 = 3) & P(X_2 = 4) & P(X_2 = 5) & P(X_2 = 6) \end{pmatrix} \\ &= P(1).P \\ &= \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \end{aligned}$$

It is clear that $P(2)$ is the negation of $P(1)$. Similarly $P(3)$ will be the negation of $P(2)$ and so on. It is interesting to note that the transition probability matrix P in the above example basically represents the negation transformation proposed by Yager. The matrix P is a doubly stochastic transition probability matrix on seven states 0,1,2,3,4,5,6 and since it is regular(P^2 has only strictly positive entries), the limiting distribution is given as

$$\left(\tfrac{1}{7} \; \tfrac{1}{7} \; \tfrac{1}{7} \; \tfrac{1}{7} \; \tfrac{1}{7} \; \tfrac{1}{7} \; \tfrac{1}{7} \right)$$

For a probability distribution $P(n) = (p_1, p_2, \ldots, p_n)$, we have $0 \leq p_i \leq 1 \;\; \forall \;\; i = 1, 2, \ldots, n$

$\Rightarrow \quad 0 \leq 1 - p_i \leq 1$

$\Rightarrow \quad 0 \leq \frac{1-p_i}{n-1} \leq \frac{1}{n-1}$

$\Rightarrow \quad 0 \leq \overline{p}_i \leq \frac{1}{n-1} \;\; \forall \;\; i = 1, 2, \ldots, n$

Therefore

$$0 \leq P(X_2 = j) \leq \frac{1}{6}; \;\; j = 0, 1, 2, 3, 4, 5, 6$$

i.e. all the probabilities after the second roll of the die are bounded in the interval $[0, \frac{1}{6}]$. Similarly we have

$$\overline{\overline{p}}_i = \frac{p_i + n - 2}{(n-1)^2} \;\; \forall \;\; i = 1, 2, \ldots, n$$

and

$$\frac{1}{n-1} - \frac{1}{(n-1)^2} \leq \overline{\overline{p}}_i \leq \frac{1}{n-1} \;\; \forall \;\; i = 1, 2, \ldots, n$$

which gives

$$\frac{5}{36} \leq P(X_3 = j) \leq \frac{1}{6}; \;\; j = 0, 1, 2, 3, 4, 5, 6$$

i.e. all the probabilities after the third roll of the die are bounded in the interval $[\frac{5}{36}, \frac{1}{6}]$. Similarly we can obtain bounds for further probabilities. Uncertainty increases on repeatedly applying the negation *i.e.*, uncertainty embedded in $P(n) = (p_1, p_2, \ldots, p_n)$ is less than the uncertainty embedded in $\overline{P}(n) = (\overline{p}_1, \overline{p}_2, \ldots, \overline{p}_n)$ and so on. Therefore using the well known Shannon entropy function

$$H(P(X_n)] = -\sum_{j=0}^{6} P(X_n = j) \log[P(X_n = j)] \;\; \forall \;\; n$$

$$H[P(X_1)] \leq H[P(X_2)] \leq H[P(X_3)] \leq \ldots \leq H[P(X_n)] \ldots; \; \forall \;\; n$$

As the time index approaches infinity, uncertainty also approaches its maximum [9], [10], [11] and the probabilities gets constrained in intervals that keep on shrinking. Figure 2 shows after every transition, the process appears to move on the opposite side of the limiting distribution *i.e.* it flips back and forth in orientation. This is obvious because every transition is actually representing the negation of the probability vector at each step. For better understanding, we take another example. Consider an ant performing a random walk on vertices of a complete graph K_4(vertices of a tetrahedron). We assume that the ant begins at any of the four vertices taken at random (say A) and at each time step moves to another vertex. Also we assume that amount of time ant takes in turning is negligible as compared to the time ant takes travelling between the vertices. Considering the graph as undirected and unweighted,the vertex the ant moves to is chosen uniformly at random among the neighbours of the present vertex.This random walk can be modelled as a markov chain which is irreducible, aperiodic and whose transition probability matrix(TPM) can be written as

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Whatever vertex the ant starts from, it can move to any of the neighbouring vertices with probability $\frac{1}{3}$. The matrix P clearly represents the transitions associated with the negation transformation. If the ant starts at vertex A(say), then the initial probability vector is $P(1) = (1,0,0,0)$ and probability vector at the next step is the negation of $P(1)$ *i.e.* $P(2) = P(0).P = (0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Similarly we can obtain $P(2)$,$P(3)$... by applying the negation transformation again and again. Here $P(2)$ can be viewed as how one can opppose the event of ant being at vertex A(In the absence of any external information, we can assume ant being at vertices B, C and D with equal probabilities). Also the above random walk has a stationary distribution(in this case limiting also) given as $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Here stationary distribution indicate that as the number of transitions approach infinity, the probability of the ant being at vertices A,B,C and D becomes identical. Clearly the movements in the random walk are uncorrelated and unbiased. Unbiased means that the ant explores every possible direction with equal probability *i.e.* there is no preferred direction. Further uncorrelated means that the direction of movement at each step is independent of the previous directions moved; the location at each step is dependent only on the location in the previous step. We now summarize as follows.

The markov chain used to characterize negation of a probability distribution typically have the following attributes.

1. The transition probability matrix used to represent negation of $P(n) = (p_1, p_2, \ldots, p_n)$ is a $n \times n$ doubly stochastic matrix with all diagonal entries 0 and all off - diagonal entries equal to $\frac{1}{n-1}$.
2. All the states are transient(no absorbing state)so that $p_{ij} < 1$ for all i and j.
3. There exists a positive integer N such that P^N has no zero entries, which implies that each state may be reached from every other state in N transitions indicating that the markov chain is regular.
4. The stationary distribution is uniform, since the markov chain is irreducible and aperiodic.
5. If we alter any two(or more than two) entries of the initial probability vector, then the probability vector at the subsequent iterations will get altered at those two positions only, rest will remain unchanged. In the above example if

$$P(1) = \begin{pmatrix} 0-\delta & 0+\delta & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

then

$$P(2) = \begin{pmatrix} \frac{1}{6}-\delta & \frac{1}{6}+\delta & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

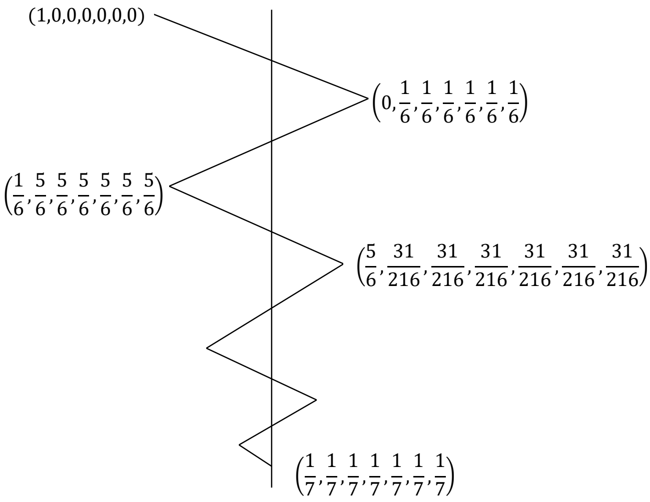

Fig. 2 Negation converging to uniform distribution

and so on. Mathematically we can represent negation by the recurrence relation

$$P(X_{k+1} = j) = \frac{1}{n-1} - \frac{P(X_k=j)}{n-1}; k = 1,2,\ldots.$$

4 IGF without any bias

We again consider a discrete finite complete probability distribution $P(n) = (p_1, p_2, \ldots, p_n)$ and its negation $\overline{P}(n) = (\overline{p}_1, \overline{p}_2, \ldots, \overline{p}_n)$. Then the information generating function(IGF) corresponding to $\overline{P}(n)$ can be defined as

$$I_t(\overline{P}(n)) \;=\; \sum_{i=1}^{n} \overline{p}_i^t \,; t \geq 1 \tag{4}$$

Here t is a real(or complex variable). Clearly $I_1(\overline{P}(n)) = 1$ and since $0 \leq p_i \leq 1$ for all $i \Rightarrow 0 \leq \frac{1-p_i}{n-1} \leq \frac{1}{n-1} \Rightarrow 0 \leq \overline{p}_i \leq \frac{1}{n-1}$ for all i therefore (4) is convergent for all $t \geq 1$.
Further we can write

$$I_t(\overline{P}(n)) \;=\; \sum_{i=1}^{n} (\tfrac{1-p_i}{n-1})^t = (\tfrac{1-p_1}{n-1})^t + (\tfrac{1-p_2}{n-1})^t + \ldots + (\tfrac{1-p_n}{n-1})^t \;; t \geq 1$$

Differentiating (4.1) at t = 1 gives

$$-\left(\frac{\partial}{\partial t} I_t(\overline{P}(n))\right)_{t\,=\,1} \;=\; -\sum_{i=1}^{n} \overline{p}_i \; log \;\; \overline{p}_i = H\left(\overline{P}(n)\right) \tag{5}$$

where $H\left(\overline{P}(n)\right)$ is the Shannon entropy of negation of a probability distribution. On further differentiating $(k-1)$ times, we obtain

$$(-1)^k \left(\frac{\partial^k}{\partial t^k} I_t(\overline{P}(n))\right)_{t=1} = (-1)^k \sum_{i=1}^{n} \overline{p}_i \, log^k \; \overline{p}_i$$

which represents the k^{th} moment of the self information embedded in $\overline{P}(n)$.

Using (3), we can write (4) in terms of entries of a doubly stochastic matrix as

$$I_t(\overline{P}(n)) = \sum_{i=1}^{n}(\sum_{j=1}^{n} a_{ij}p_i)^t \; ; t \geq 1 \tag{6}$$

which gives

$I_t(\overline{P}(n)) = \sum_{i=1}^{n}(a_{i1}p_i + a_{i2}p_i + ... + a_{in}p_i)^t = (a_{11}p_1 + a_{12}p_1 + ... + a_{1n}p_n)^t + (a_{21}p_1 + a_{22}p_2 + ... + a_{2n}p_n)^t + ... + (a_{n1}p_1 + a_{n2}p_1 + ... + a_{nn}p_n)^t \; ; t \geq 1.$

We now list some properties of IGF given by (4) as follows.

1. If all the entries of $P(n) = (p_1, p_2, \ldots, p_n)$ are equal then all the entries of $\overline{P}(n) = (\overline{p}_1, \overline{p}_2, \ldots, \overline{p}_n)$ are also equal *i.e.* if $p_i = \frac{1}{n} \forall i$ then $\overline{p}_i = \frac{1}{n} \forall i$. In this case

$$I_t(\overline{P}(n)) = n^{1-t} = I_t(P(n)) \; ; t \geq 1$$

which gives

$$-\left(\frac{\partial}{\partial t} I_t(\overline{P}(n))\right)_{t=1} = log \; n = -\left(\frac{\partial}{\partial t} I_t(P(n))\right)_{t=1}$$

2. Suppose we add k events with zero probability in the probability distribution $P(n) = (p_1, p_2, \ldots, p_n)$, then the revised probability distribution is $P(n+k) = (p_1, p_2, ..., p_n, \underbrace{0, 0, ..., 0}_{k-times})$. In this case, the IGF given by (1) remains unaltered since events with zero probability does not provide any information regarding the occurrence of events[7], [8]. However the revised negation in this case is

$$\overline{P}(n+k) = \left(\frac{1-p_1}{n+k-1}, \frac{1-p_2}{n+k-1}, \ldots, \frac{1-p_n}{n+k-1}, \underbrace{\frac{1}{n+k-1}, \ldots, \frac{1}{n+k-1}}_{k-times}\right)$$

and the corresponding IGF can be written as

$$I_t(\overline{P}(n+k)) = \sum_{i=1}^{n+k} \overline{p}_i^t =$$

$$\overline{p}_1^t + \overline{p}_2^t + ... + \overline{p}_n^t + \underbrace{(\frac{1}{n+k-1})^t + (\frac{1}{n+k-1})^t + ... + (\frac{1}{n+k-1})^t}_{k-times} \; ; t \geq 1; k \geq 1$$

$$= (\frac{1-p_1}{n+k-1})^t + (\frac{1-p_2}{n+k-1})^t + ... + (\frac{1-p_n}{n+k-1})^t + k(\frac{1}{n+k-1})^t \; ; t \geq 1; k \geq 1$$

The negation transformation defined by Yager[14] equally redistribute the probabilities among events with zero and non zero probabilities. Here $I_t(\overline{P}(n+k))$ has k additional terms(identical also) which signifies that $\overline{P}(n+k)$ has more uncertainty inherent in it than $P(n+k)$. In fact

$$-\left(\frac{\partial}{\partial t} I_t(P(n))\right)_{t=1} = -\sum_{i=1}^{n} p_i \; log \; p_i = -\left(\frac{\partial}{\partial t} I_t(P(n+k))\right)_{t=1}$$

and

$$-\left(\frac{\partial}{\partial t} I_t(\overline{P}(n+k))\right)_{t\,=\,1} = -\sum_{i=1}^{n+k} \overline{p}_i \; log \; \overline{p}_i =$$
$$-\sum_{i=1}^{n} \overline{p}_i \; log \; \overline{p}_i - \underbrace{(\frac{1}{n+k-1})log\frac{1}{n+k-1}}_{k-times}$$

3. Suppose we alter any two(or more) entries of $P(n) = (p_1, p_2, \ldots, p_n)$, then only the corresponding entries in $\overline{P}(n) = (\overline{p}_1, \overline{p}_2, \overline{p}_3, \ldots, \overline{p}_n)$ will change, rest will remain the same. For *e.g.* if we alter the first and last entries in $P(n)$ with resulting distribution as $P^*(n) = (p_1 - \vartheta, p_2, , p_3, \ldots, p_n + \vartheta)$, then the revised negation can be written as
$$\overline{P}^*(n) = \left(\frac{1-p_1+\vartheta}{n-1}, \frac{1-p_2}{n-1}, \frac{1-p_3}{n-1}, \ldots, \frac{1-p_n-\vartheta}{n-1}\right)$$
Similar will be the case with the IGF of $P(n)$ and its negation. In fact we can write
$$I_t(\overline{P}^*(n)) \; - I_t(P^*(n)) \; = (\frac{1-p_1+\vartheta}{n-1})^t + (\frac{1-p_2+\vartheta}{n-1})^t - p_1^t - p_2^t$$
4. The IGF corresponding to $\overline{P}(n)$ is given by (4). The IGF corresponding to $\overline{\overline{P}}(n)$(negation transformation applied two times on $P(n)$) can be defined as
$$I_t(\overline{\overline{P}}(n)) \; = \; \sum_{i=1}^{n} \overline{\overline{p}}_i \, ; t \geq 1 \tag{7}$$
Here the IGF given by (4) can be determined from the redistribution of entries in $P(n)$. Similarly IGF given by (7) can be determined from the redistribution of entries in $\overline{P}(n)$. Same applies for further iterations of negation transformation. Therefore the uncertainty(information) embedded in $\overline{P}(n)$ depends solely on the entries of $P(n)$, uncertainty(information) embedded in $\overline{\overline{P}}(n)$ depends solely on the entries of $\overline{P}(n)$(and not on the entries of $P(n)$) and so on. It is clear that the IGF given by (4) and (7) exhibits markovian behaviour. We take an example. Consider a random variable $X = (x_1, x_2, x_3, x_4, x_5)$ with corresponding probabilities $P(5) = (p(x_1), p(x_2), p(x_3), p(x_4), p(x_5)) = (0.1, 0.2, 0.4, 0.2, 0.1)$. The IGF corresponding to $P(5)$ can be written as
$$I_t(P(5)) \; = \; \sum_{i=1}^{5} p_i^t = (0.1)^t + (0.2)^t + (0.4)^t + (0.2)^t + (0.1)^t \; , t \geq 1.$$
Also
$$I_t(\overline{P}(5)) \; = \; \sum_{i=1}^{5} \overline{p}_i^t = (0.225)^t + (0.2)^t + (0.15)^t + (0.2)^t + (0.225)^t \; , t \geq 1.$$
and
$$I_t(\overline{\overline{P}}(5)) \; = \; \sum_{i=1}^{5} \overline{\overline{p}}_i^t = (0.19375)^t + (0.2)^t + (0.2125)^t + (0.2)^t + (0.19375)^t \; , t \geq 1.$$
In fact after applying 10 iterations of negation transformation on $P(5)$, we will obtain
$$I_t(\overline{\overline{P}}(5)) \; = \; \sum_{i=1}^{5} \overline{\overline{p}}_i^t = (0.2)^t + (0.2)^t + (0.2)^t + (0.2)^t + (0.2)^t \; , t \geq 1.$$
Clearly the IGF defined by (4) and (7) has preserved the symmetry of $P(5)$.

5 Conclusion

In the present work, we have shown that the underlying mathematical structure of negation of a probability distribution has many(if not all) properties identical to that of a markov chain. In Yager's definition, we actually redistribute the number of cases favouring a particular outcome equally among all the other outcomes. One disadvantage that the yager model has that it does not allow any component of the probability distribution to retain something. Retainment is essential in many processes and it is possible to define an unbiased model which includes both retainment and redistribution. Suppose half of each probability is retained and the remaining is equally distributed among the other alternatives, then the revised probabilities are

$$\overline{p}_1 = \frac{p_1}{2} + \frac{p_2 + p_3 + \ldots + p_n}{2(n-1)} = \frac{p_1}{2} + \frac{\overline{p}_1}{2}$$

$$\overline{p}_2 = \frac{p_2}{2} + \frac{p_1 + p_3 + \ldots + p_n}{2(n-1)} = \frac{p_2}{2} + \frac{\overline{p}_2}{2}$$

$$\vdots$$

$$\overline{p}_n = \frac{p_n}{2} + \frac{p_1 + p_2 + \ldots + p_{n-1}}{2(n-1)} = \frac{p_n}{2} + \frac{\overline{p}_n}{2}$$

Here the set $(\overline{p}_1, \overline{p}_2, \ldots, \overline{p}_n)$ again satisfies $0 \leq \sum_{i=1}^{n} \overline{p}_i \leq 1$ and $\sum_{i=1}^{n} \overline{p}_i = 1$. The above model is totally unbiased since the amount retained and redistributed is exactly identical for all components. Also the above model can be characterized by a markov chain which is irreducible, aperiodic and has a unique stationary distribution. The only difference is that all outgoing probabilities have been dichotomized, while the probability of staying at the same state has been increased. The chain performs identical transitions as the original one but stays longer at each state. The above model can be interpreted as slowing down of the original one. Finally we can also investigate of behaviour of underlying markov chain when we negate any particular outcome by giving preference to some or a group from the remaining set of outcomes. Work on these extensions of negation is in progress and will be communicated elsewhere.

6 Declaration

1. The authors have no competing interests to declare that are relevant to the content of this article.
2. The manuscript has not been submitted to other journal for simultaneous consideration.
3. The submitted work is original and has not have been published elsewhere in any form or language (partially or in full).

4. All authors whose names appear on the submission
   1) made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work;
   2) drafted the work or revised it critically for important intellectual content;
   3) approved the version to be published; and
   4) agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
5. The authors did not receive support from any organization for the submitted work.
6. The authors have no relevant financial or non-financial interests to disclose.
7. Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Compliance with ethical standards

Conflict of interest On behalf of all the authors, the corresponding author declares that there is no conflict of interest.

References

1. Deng X, Jiang W (2020) On the negation of a dempster-shafer belief structure based on maximum uncertainty allocation. Inf Sci 516:346–352
2. Golomb SW (1966) The information generating function of a probability distribution. IEEE Transactions on Information Theory 12:75 - 77
3. Liu Q, Cui H, Tian Y, Kang B (2020) On the negation of discrete z-numbers. Inf Sci 537:1
4. Luo Z, Deng Y (2019) A matrix method of basic belief assignment's negation in dempster-shafer theory. IEEE Trans Fuzzy Syst
5. Mao H, Cai R (2020) Negation of pythagorean fuzzy number based on a new uncertainty measure applied in a service supplier selection system. Entropy 22(2):195
6. Mao H, Deng Y (2021) Negation of BPA: a belief interval approach and its application in medical pattern recognition. Appl Intell. https:// doi.org/10.1007/s10489-021-02641-7
7. Pal NR, Pal SK (1991) Entropy : A new definition and its applications. IEEE transactions on systems, man, and cybernetics 21(5): 1260-1270
8. Shannon CE (1948) A Mathematical theory of communication. Bell System Technical Journal 27(3): 379-423
9. Srivastava A, Maheshwari (2018) Some New Properties of Negation of a Probability Distribution. International Journal of Intelligent Systems 33(6): 1133-1145
10. Srivastava A, Kaur L (2019) Uncertainty and negation information theoretic applications. International Journal of Intelligent Systems 34(6) : 1248-1260
11. Srivastava A, Tanwar P(2021) Interplay between symmetry, convexity and negation of a probability distribution. International Journal of Intelligent Systems 36(4) :1876-1897
12. Wu Q, Deng Y, Xiong N (2021) Exponential negation of a probability distribution. Soft Computing https://doi.org/10.1007/s00500-021-06658-5
13. Xiao F (2021) CaFtR: A fuzzy complex event processing method. Int J Fuzzy Syst. https://doi.org/10.1007/s40815-021-01118-6
14. Yager RR (2014) On the maximum entropy negation of a probability distribution. IEEE Transactions on Fuzzy Systems 23(5): 1899- 1902
15. Yin L, Deng X, Deng Y (2018) The negation of a basic probability assignment. IEEE Trans Fuzzy Syst 27(1):135–143
16. Zhang J, Liu R, Zhang J, Kang B (2020) Extension of yager's negation of a probability distribution based on tsallis entropy. Int J Intell Syst 35(1):72–84