



Variance Laplacian: Quadratic Forms in Statistics

Rama Murthy Garimella

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

August 9, 2019

VARIANCE LAPLACIAN: QUADRATIC FORMS in STATISTICS

Garimella Rama Murthy*

Department of Computer Science & Engineering

*Mahindra Ecole Centrale, Hyderabad, India

ABSTRACT

In this research paper, it is proved [RRN] that the variance of a discrete random variable, Z can be expressed as a quadratic form associated with a Laplacian matrix i.e.

$$\text{Variance}[Z] = X^T G X,$$

where X is the vector of values assumed by the discrete random variable and

G is the Laplacian matrix whose elements are expressed in terms of probabilities. We formally state and prove the properties of Variance Laplacian matrix, G . Some implications of the properties of such matrix to statistics are discussed. It is reasoned that several interesting quadratic forms can be naturally associated with statistical measures such as the covariance of two random variables. It is hoped that VARIANCE LAPLACIAN MATRIX G will be of significant interest in statistical applications. The results are generalized to continuous random variables also. It is reasoned that cross-fertilization of results from the theory of quadratic forms and probability theory/statistics will lead to new research directions.

1. Introduction:

Structured matrices such as Toeplitz matrix naturally arise in various application areas of Mathematics, Science and Engineering. Specifically, in Probability Theory as well as Statistics, the autocorrelation matrix of an Auto-Regressive (AR) random process is a Toeplitz matrix. Auto-Regressive stochastic processes find many applications in stochastic modeling. Motivated by practical considerations, detailed research efforts went into understanding the properties of Toeplitz matrices (such as connections to orthogonal polynomials). For instance, considerable research effort went into efficiently inverting a Toeplitz matrix (such as Levinson-Durbin algorithm).

In the research area of Graph theory, a structured matrix called Laplacian naturally arises. It is defined utilizing the adjacency matrix of a graph (which essentially summarizes the adjacency information associated with the vertices of graph). Thus, Graph Laplacian was subjected to detailed study and several new properties of it are discovered. Some of these properties have graph-theoretic significance.

Effectively, researchers are interested in discovering the connections between concepts in Probability/Statistics and Structured Matrices. Discrete random variables find many applications in Statistics. Thus, a curious natural question is to see if structured matrices are naturally associated with scalar measures of discrete random variables, such as the moments.

2. Review of Related Literature:

In the field of mathematics, research related to quadratic forms has long history dating back to the time of Fermat, Bhaskara and others. Several interesting results such as the Rayleigh's theorem were discovered and proved. Quadratic forms have connections to such diverse areas such as topology, differential geometry etc. To the best of our knowledge, the author discovered for the first time that the variance of a discrete random variable can be expressed as the quadratic form associated with a Laplacian matrix (of probabilities) [Rama]. This discovery motivated the author to express other statistical/probabilistic measures as quadratic forms. This line of research enables cross-fertiization of ideas between probability theory/ Statistics and the theory of quadratic forms.

3. Variance of a Discrete Random Variable: Laplacian Quadratic Form:

Consider a discrete random variable, Z with probability mass function $\{p_1, p_2, \dots, p_N\}$. The variance of Z is given by

$$\text{Variance}(Z) = \text{Var}(Z) = E(Z^2) - (E(Z))^2.$$

Let the values assumed by the random variable Z be given by $\{T_1, T_2, \dots, T_N\}$. Let the associated vector of values assumed by Z be denoted by \bar{T} . Hence, we have that

$$\begin{aligned} \text{Var}(Z) &= \sum_{i=1}^N T_i^2 p_i - \left(\sum_{i=1}^N T_i p_i \right)^2 \\ &= \sum_{i=1}^N T_i^2 p_i - \sum_{i=1}^N \sum_{j=1}^N T_i T_j p_i p_j \\ &= \bar{T}^T [\bar{D} - \bar{P}] \bar{T}, \end{aligned}$$

where \bar{D} is a diagonal matrix whose diagonal elements are $\{p_1, p_2, \dots, p_N\}$ and $\bar{P}_{ij} = p_i p_j$ for all $1 \leq i, j \leq N$.

Let $\bar{G} = \bar{D} - \bar{P}$. Hence, we have that

$$\text{Var}(Z) = \bar{T}^T \bar{G} \bar{T}.$$

Thus, we have shown that variance of discrete random variable Z constitutes a quadratic form associated with the matrix \bar{G} . We now introduce the following well known definition:

Definition 1: A square matrix \bar{G} is called a Laplacian matrix if and only if all diagonal elements of it are all positive, all non-diagonal elements are non-positive and all the row sums are all zero.

Now, we prove that the square matrix \bar{G} is a Laplacian matrix.

Lemma 1: The square matrix \bar{G} is a Laplacian matrix

Proof: From the definition of \bar{G} , we readily have that

$$G_{ii} = p_i - p_i^2 = p_i(1 - p_i).$$

Also, we have that

$$G_{ij} = -p_i p_j \text{ for } i \neq j.$$

Further

$$\sum_{j=1}^N G_{ij} = G_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^N G_{ij} = p_i(1-p_i) - \sum_{\substack{j=1 \\ j \neq i}}^N p_i p_j = p_i(1-p_i) - p_i(1-p_i) = 0.$$

Hence, the square matrix \bar{G} is a Laplacian matrix.

Q.E.D.

Note: In the case of a discrete random variable which assumes countably many values, if the variance is finite, then the associated quadratic form is based on an infinite dimensional Laplacian matrix.

Note: It readily follows that the sum of two variance Laplacian matrices is also a Laplacian matrix. Under some conditions on the elements of two matrices, the sum will also be a Variance Laplacian matrix

Note: In the case of specific discrete random variables (such as Bernoulli, Poisson, Binomial etc), the associated Laplacian matrix can easily be determined. Also, if the number of values assumed by the random variables is at most 5, the eigenvalues of Laplacian matrix (roots of the associated characteristic polynomial) can be determined by algebraic formulas (Galois Theory).

Example 1: Specifically when the dimension of \bar{G} is 2 (i.e. the random variable, Z is Bernoulli random variable), we determine its eigenvalues and eigenvectors explicitly. Let $\text{Probability}\{Z = 0\} = q$. Then we have that

$$\bar{G} = \begin{bmatrix} q(1-q) & -q(1-q) \\ -q(1-q) & q(1-q) \end{bmatrix} = q(1-q) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

The eigenvalues are $\{0, 2(q - q^2)\}$. The

orthonormal basis of eigenvectors are $\left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 1 \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -1 \end{bmatrix} \right\}$. When $q = \frac{1}{2}$, spectral radius is $\frac{1}{2}$.

Note: Suppose we consider a discrete random variable Z which assumes the values $\{+1, -1\}$. In such case, it is easy to show that

$$\text{Variance}(Z) = 4q(1-q).$$

Example 2: We now consider discrete uniform random variable whose probability mass function is given by $\left\{ \frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N} \right\}$. The Variance Laplacian associated with it is given by

$$\bar{G} = \begin{bmatrix} \frac{N-1}{N^2} & \frac{-1}{N^2} & \cdots & \frac{-1}{N^2} \\ \frac{-1}{N^2} & \frac{N-1}{N^2} & \cdots & \frac{-1}{N^2} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{-1}{N^2} & \frac{-1}{N^2} & \cdots & \frac{N-1}{N^2} \end{bmatrix}.$$

Since, the sum of absolute values of elements in every row is same, the spectral radius $Sp(\bar{G})$ is given by (using well known result in linear algebra)

$$Sp(\bar{G}) = \frac{2(N-1)}{N^2}, \quad Trace(\bar{G}) = \frac{N-1}{N}, \quad Determinant(\bar{G}) = 0.$$

Since \bar{G} is a right circulant matrix, from linear algebra, its eigenvalues as well as eigenvectors can be explicitly determined. From basic argument, we now determine the eigenvalues of \bar{G} .

\bar{G} can be rewritten in the following manner:

$$\bar{G} = \left(\frac{N-1}{N^2} \right) I - \frac{1}{N^2} \bar{B}, \quad \text{where } \bar{B} \text{ is a matrix with zero diagonal elements and all}$$

the off-diagonal elements are equal to '1'. Thus $\bar{B} = \bar{e} \bar{e}^T - I$, where I is the

identity matrix and \bar{e} is a column vector of '1's. It readily follows that the

eigenvalues of \bar{B} are $\{(N-1), -1, -1, \dots, -1\}$. From basic linear algebra,

if ' α ' is an eigenvalue of \bar{B} , then the corresponding eigenvalue μ of \bar{G} is

$$\mu = \left(\frac{N-1}{N^2} \right) - \frac{1}{N^2} \alpha.$$

Thus, we infer that the eigenvalues of \bar{G} are $\left\{ 0, \frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N} \right\}$.

Now, we arrive at the spectral representation of such a structured \bar{G} . Let '0' be the first eigenvalue and \bar{f}_1 be the corresponding right eigenvector.

$$\bar{G} = \frac{1}{N} \sum_{i=2}^N \bar{f}_i \bar{f}_i^T = \sum_{i=2}^N \bar{E}_i, \quad \text{where } \bar{E}_i \text{'s are residue matrices.}$$

It is well known that $\sum_{i=1}^N \bar{E}_i = \bar{I}$ i.e. identity matrix. Hence $\sum_{i=2}^N \bar{E}_i = \bar{I} - \frac{1}{N} \bar{e} \bar{e}^T$, where

\bar{e} is a column vector all of whose components are '1'.

Note: The matrix, $-\bar{G}$ constitutes a generator matrix of a finite state space Continuous Time Markov Chain (CTMC). Thus a discrete random variable can be associated with a CTMC.

In general, since, \bar{G} is a symmetric matrix, it is completely specified by the eigenvalues and eigenvectors.

Now, we briefly summarize few properties of \bar{G} matrix that readily follow.

- (i) Let \bar{e} be a column vector of 1's (ONES) i.e. $\bar{e} = [1 \ 1 \ \dots \ 1]^T$.
From Lemma 1, we have that $\bar{G}\bar{e} \equiv \bar{0}$. Hence '0' is an eigenvalue of \bar{G} and the corresponding eigenvector is \bar{e} .
- (ii) Since Variance [Z] is non-negative, we have that the quadratic form $\bar{T}^T \bar{G} \bar{T} \geq 0$ for all vectors \bar{T} . Hence the Laplacian matrix \bar{G} is a positive semi-definite matrix. Thus, all eigenvalues of \bar{G} are real and non-negative.

We now derive an important property of \bar{G} in the following Lemma.

- Property (iii):

Lemma 2: The spectral radius, μ_{max} (i.e. largest eigenvalue of \bar{G}) is less than or equal to $\frac{1}{2}$.

Proof: From linear algebra (particularly matrix norms), it is well known that the spectral radius of any square matrix \bar{A} i.e. $Sp(\bar{A})$ is bounded in the following manner:

$$\text{Minimum absolute row sum}(\bar{A}) \leq Sp(\bar{A}) \leq \text{Maximum absolute row sum}(\bar{A}).$$

But, in the case of Laplacian matrix \bar{G} , we have that

$$\sum_{j=1}^N |G_{ij}| = 2 p_i (1 - p_i) \text{ for all } i.$$

Hence, using the above fact from linear algebra, we have that

$$\text{Min}_i \{ 2 p_i (1 - p_i) \} \leq Sp(\bar{G}) \leq \text{Max}_i \{ 2 p_i (1 - p_i) \}.$$

Using the fact that, p_i 's are probabilities, we now bound $\text{Max}_i \{ 2 p_i (1 - p_i) \}$.

Let $f(p_i) = \{ 2 p_i (1 - p_i) \} = 2 p_i - 2 p_i^2$. We now calculate the critical points of $f(p_i)$

$$f'(p_i) = 2 - 4 p_i = 0. \text{ Hence } p_i = \frac{1}{2} \text{ is the unique critical point in feasible region.}$$

Also, we have that $f''(p_i) = -4$. Thus the critical point is maximum of $f(p_i)$.

Thus, $f\left(\frac{1}{2}\right) = \frac{1}{2}$. Hence we readily have that spectral radius of \bar{G} i.e. $Sp(\bar{G}) \leq \frac{1}{2}$. Q.E.D.

Note: The function $f(p_i)$ constitutes the well known logistic map whose properties were investigated by several researchers.

Goals:

- **Goal 1:** In view of the above discovery related to the variance of a discrete random variable (i.e. Laplacian quadratic form), we would like to discover other quadratic forms which naturally arise in probability/statistics.

- **Goal 2:** Once the interesting quadratic forms are identified, the results from the theory of quadratic forms (for example, Rayleigh’s Theorem) are applied to statistical/ probabilistic quadratic forms. On the other hand, results related to statistical / probabilistic quadratic forms are invoked to derive new results in the theory of quadratic forms (such as inequalities between quadratic forms).
- We now derive a specific inequality associated with quadratic forms based on statistical/probabilistic quadratic forms:

Consider a vector \bar{K} whose components are all positive real numbers. It readily follows that by means of the following normalization procedure, it can be converted into a probability vector \bar{p} (i.e vector whose components are probabilities and sum to one i.e. probability mass function of a random variable, say Z). Let the vector of values assumed by the random variable Z be \bar{T} .

$$\bar{p} = \frac{\bar{K}}{\sum_{i=1}^N K_i} = \frac{\bar{K}}{\alpha}.$$

But, we know that the variance of discrete random variable Z is non-negative. Hence $\bar{T}^T (diag(\bar{p}) - \bar{p} \bar{p}^T) \bar{T} \geq 0$, where $diag(\bar{p})$ is a diagonal matrix whose components are all the components of vector \bar{p} .

It readily follows that (on using the above normalization equation), we have the following inequality:

$$\alpha (\bar{T}^T (diag(\bar{K})) \bar{T}) \geq (\bar{T}^T (\bar{K} \bar{K}^T) \bar{T}) \text{ for all } \bar{T}, \bar{K}, \alpha.$$

We now state the following Theorem, useful in bounding the variance of Z.

Rayleigh’s Theorem: The local/global optimum values of a quadratic form evaluated on the unit Euclidean hypersphere (constraint set) are the eigenvalues and they are attained at the corresponding eigenvectors.

Using Rayleigh’s theorem, we arrive at the following result.

Lemma 3: $Variance(Z) \leq \frac{1}{2} (L^2 - norm(\bar{T}))^2$.

Proof: Formally, if the vector of values assumed by the random variable i.e. \bar{T} lies on the unit Euclidean hypersphere, then we have that

$$\mu_{min} \leq \bar{T}^T \bar{G} \bar{T} \leq \mu_{max} \leq \frac{1}{2}, \quad \text{if } L^2 - norm(\bar{T}) = 1.$$

Suppose $L^2 - norm(\bar{T}) \neq 1$. Then, we readily have that $\frac{\bar{T}}{L^2 - norm(\bar{T})}$ is a vector whose

$L^2 - norm$ is equal to one and the Rayleigh’s Theorem can be applied to the quadratic form based on it. Thus, it follows that

$$\mu_{min} (L^2 - norm(\bar{T}))^2 \leq Variance(Z) \leq \mu_{max} (L^2 - norm(\bar{T}))^2.$$

Hence, by applying the earlier upper bound on spectral radius, we have

$$\text{Variance}(Z) \leq \frac{1}{2} (L^2 - \text{norm}(\bar{T}))^2 . \quad Q.E.D.$$

Corollary: The non-zero lower bound on Variance(Z) is given by (using μ_{min})

$$\mu_{min} (L^2 - \text{norm}(\bar{T}))^2 \leq \text{Variance}(Z).$$

Note: We can consider a vector whose components are the finitely many eigenvalues of the Laplacian matrix. Let it be denoted by $\bar{\mu}$ i.e. $\bar{\mu} = (\mu_1 \mu_2 \dots \mu_N)^T$. As in the case of any vector, various L^p - norms of such vector can be defined.

- Property (iv) : Now, we consider sum of eigenvalues of \bar{G} i.e. $\text{Trace}(\bar{G})$.

It readily follows that

$$\text{Trace}(\bar{G}) = L^1 - \text{norm}(\bar{\mu}) = \sum_{i=1}^N p_i (1 - p_i) = \sum_{i=1}^N (p_i - p_i^2) = 1 - \sum_{i=1}^N p_i^2 = \sum_{i=1}^N \mu_i .$$

Since, $\text{Trace}(\bar{G})$ is the sum of eigenvalues, we have the following obvious bounds:

$$N \mu_{min} \leq \text{Trace}(\bar{G}) \leq N \mu_{max}$$

The following Lemma provides an interesting upper bound on $\text{Trace}(\bar{G})$.

Lemma 4: Let \bar{G} be an $N \times N$ matrix. Then $\text{Trace}(\bar{G})$ has the following upper bound.

$$\text{Trace}(\bar{G}) \leq \left(1 - \frac{1}{N}\right).$$

Proof: Let $\{p_1, p_2, \dots, p_N\}$ be the probability mass function of random variable Z.

We now apply the Lagrange-multipliers method to bound $\sum_{i=1}^N p_i^2$. The objective function for the optimization problem is given by

$J(p_1, p_2, \dots, p_N) = \sum_{i=1}^N p_i^2$ with the constraint that the probabilities sum to one. Hence the Lagrangian is given by

$$L(p_1, p_2, \dots, p_N) = \sum_{i=1}^N p_i^2 + \alpha (\sum_{i=1}^N p_i - 1).$$

Now, we compute the critical point and the components of the Hessian matrix:

$$\frac{\delta L}{\delta p_i} = 2 p_i + \alpha, \quad \frac{\delta^2 L}{\delta p_i^2} = 2 \text{ for all 'i'}, \quad \frac{\delta^2 L}{\delta p_i \delta p_j} = 0 \text{ for all } i \neq j.$$

Hence, there is a single critical point and the Hessian matrix is positive definite at the critical point. Thus, we conclude that the objective function has a unique minimum and occurs at

$\frac{\delta L}{\delta p_i} = 0$ i.e. $p_i = \frac{-\alpha}{2}$. Using the constraint that the probabilities sum to one, we have

$$\alpha = \frac{-2}{N}. \text{ Thus, the global minimum occurs at } p_i = \frac{1}{N} \text{ for all 'i'.$$

Equivalently, we have the following upper bound on $\text{Trace}(\bar{G})$.

$$\text{Trace}(\bar{G}) \leq \left(1 - \frac{1}{N}\right). \quad \text{Q.E.D.}$$

Corollary: We now bound the second smallest eigenvalue, μ_2 of \bar{G} . It is clear that

$\text{Trace}(\bar{G}) \leq \left(1 - \frac{1}{N}\right)$. Further $(N-1)\mu_2 \leq \text{Trace}(\bar{G})$. Hence $\mu_2 \leq \frac{1}{N}$. Thus, we have

$\mu_2 \in \left(0, \frac{1}{N}\right]$ and $\mu_i \in \left[\frac{1}{N}, \frac{1}{2}\right]$ for $i \geq 3$. It also readily follows that

$$\mu_2 + (N-2)\mu_3 \leq \text{Trace}(\bar{G}).$$

Since, μ_2 is positive, we have the following upper bound on μ_3 . Most generally, since the eigenvalues are non-negative, we have the following bounds

$$(N-j)\mu_{j+1} \leq \text{Trace}(\bar{G}) \leq \frac{N-1}{N} \quad \text{for } 1 \leq j \leq (N-1) \quad \text{Q.E.D.}$$

Note: The upper bound on $\text{Trace}(\bar{G})$ is attained for uniform probability mass function

$$\text{i.e. } p_i = \frac{1}{N} \text{ for all } i.$$

Note: Consider an arbitrary positive definite symmetric matrix, \bar{B} with positive eigenvalues $\{\delta_1, \delta_2, \dots, \delta_N\}$. The idea utilized in the above corollary can be used to bound the eigenvalues in terms of $\text{Trace}(\bar{B})$. We explicitly state the following bounds which follow from the argument used in the above corollary

$$0 < \delta_j \leq \frac{\text{Trace}(\bar{B})}{(N-j+1)} \leq \frac{N\delta_{\max}}{(N-j+1)} \quad \text{for } 1 \leq j \leq (N-1).$$

The bounding idea used in the above corollary applies to an arbitrary positive semi-definite matrix. Also, the bounding idea is easily utilized for bounding the eigenvalues of an arbitrary negative definite/negative semi-definite matrix. It should also be noted that the Gerschgorin Disc theorem can also be readily applied for bounding the eigenvalues.

Note: We now apply the well known inequality relating the arithmetic and geometric mean of finitely many non-negative real numbers i.e. $\{x_i\}_{i=1}^N$ to bound the trace of a positive definite (positive semi-definite) matrix. The inequality is given by

$$\frac{\sum_{i=1}^N x_i}{N} \geq \sqrt[N]{x_1 x_2 \dots x_N} \quad \text{with equality if and only } x_1 = x_2 = \dots = x_N.$$

Thus, in the case of a positive definite matrix, A , we have that

$$\frac{\text{Trace}(A)}{N} \geq \sqrt{a_{11} a_{22} \dots a_{NN}}, \quad \text{where } a_{ii}'\text{s are diagonal elements of } A.$$

Hence, in the case of variance Laplacian matrix \bar{G} , we have that

$$\frac{\text{Trace}(\bar{G})}{N} = \frac{\sum_{i=1}^N p_i(1-p_i)}{N} \geq \sqrt[N]{p_1 \dots p_N (1-p_1) \dots (1-p_N)} .$$

Note: The quantity in the brackets on the right hand side of the above inequality can be associated with any probability mass function and can be readily provided with probabilistic interpretation (using repeated trials). It can readily shown using Lagrange multiplier's method that the quantity is maximized for uniform probability mass function (i.e. the probability mass function with maximum Shannon entropy). Such a measure can be denoted by $M(p_1, p_2, \dots, p_N)$.

It can be readily seen that for $N=2$, $M(p_1, p_2) = 2(p_1 p_2)$ and

for $N=3$, $M(p_1, p_2, p_3) = p_1(p_2 + p_3) p_2(p_3 + p_1) p_3(p_1 + p_2)$ i.e. homogenous polynomial of degree '6' in the variables p_1, p_2, p_3 . For arbitrary N , it readily follows that

$M(p_1, p_2, \dots, p_N)$ is homogenous multi – variate polynomial of degree $2N$ in p_1, p_2, \dots, p_N . It is a scalar measure associated with the probability mass function.

Since μ_{max} is the spectral radius of \bar{G} , we have that

$$\frac{(N-1)\mu_{max}}{N} \geq \sqrt[N]{p_1 \dots p_N} \sqrt[N]{(1-p_1) \dots (1-p_N)} .$$

Thus, we have that $\mu_{max} \geq \frac{N}{N-1} (\sqrt[N]{p_1 \dots p_N} \sqrt[N]{(1-p_1) \dots (1-p_N)})$.

In the above equation, equality is attained for the uniform probability mass function i.e.

$$p_i = \frac{1}{N} \text{ for all } i.$$

Note: The finite condition number of Laplacian matrix \bar{G} is defined as $\frac{\mu_{max}}{\mu_{min}}$, where μ_{min} is the smallest non-zero eigenvalue of \bar{G} and μ_{max} is the spectral radius of \bar{G} . Using the content of Lemma 2 and above corollary, the following lower bound on finite condition number of \bar{G} follows:

$$\text{as } \frac{\mu_{max}}{\mu_{min}} \geq 2N p_{min}(1-p_{min}), \text{ where}$$

p_{min} is the minimum of all the probabilities in the PMF of random variable Z .

- **Connections to Statistical Mechanics:**

Note: The expression for $\text{Trace}(\bar{G})$ has familiar relationship to Tsallis Entropy concept from statistical mechanics. We have the following Definition:

Definition: Tsallis entropy of a probability mass function $\{p_1, p_2, \dots, p_N\}$ is defined as

$$S_q(\bar{p}) = \frac{k}{q-1} \left(1 - \sum_{i=1}^N p_i^q \right), \quad \text{where 'k' is Boltzmann constant and } q \text{ is real number.}$$

We can express $S_q(\bar{p})$ in the following form for integer values of 'q' :

$$S_q(\bar{p}) = \frac{k}{q-1} \left[(p_1 + p_2 + \dots + p_N)^q - \sum_{i=1}^N p_i^q \right].$$

Thus, using multinomial theorem, RHS can be expressed as a multivariate polynomial in p_i 's.

- We now explore the algebraic structure of Tsallis entropy, $S_q(\bar{p})$ for integer values of 'q'. Thus

$$S_q(\bar{p}) = \frac{k}{q-1} \left(1 - \sum_{i=1}^N p_i^q \right) = \frac{k}{q-1} \left(\sum_{i=1}^N p_i - \sum_{i=1}^N p_i^q \right) = \frac{k}{q-1} \left(\sum_{i=1}^N p_i (1 - p_i^{q-1}) \right).$$

Using the expression for sum of a geometric sequence, we have that

$$S_q(\bar{p}) = \frac{k}{q-1} \left(\sum_{i=1}^N p_i (1 - p_i^{q-1}) \right) = \frac{k}{q-1} \left(\sum_{i=1}^N p_i (1 - p_i) (1 + p_i + p_i^2 + \dots + p_i^{q-2}) \right).$$

We let $S_q(\bar{p}) = \sum_{i=1}^N h(p_i)$, where the polynomial $h(p_i)$ has zeroes at '0', '1',

$(q-2)^{th}$ roots of unity.

We, also readily have that

- $Trace(\bar{G}) = k S_2(\bar{p})$, where \bar{p} specifies the probability mass function.
- We now associate an interesting matrix denoted by \tilde{G} , whose Trace is

$$Trace(\tilde{G}) = \frac{k}{q-1} \left(1 - \sum_{i=1}^N p_i^q \right) = \frac{k}{q-1} \left(\sum_{i=1}^N p_i (1 - p_i) (1 + p_i + p_i^2 + \dots + p_i^{q-2}) \right).$$

We let $\theta_i = p_i (1 + p_i + p_i^2 + \dots + p_i^{q-2})$ and let $\tilde{G}_{ij} = -k p_j \theta_i$ for $i \neq j$ and $\tilde{G}_{ii} = k p_i (1 - p_i^{q-1})$. It can be readily verified that \tilde{G} has desired Trace value which equals the Tsallis entropy $S_q(\bar{p})$ for integer 'q'. Interestingly \tilde{G} can be provided with probabilistic interpretation like variance Laplacian matrix \bar{G} . As in the case of \bar{G} , the properties of eigenvalues, eigenvectors of \tilde{G} can be readily investigated along the same lines as in the case of \bar{G} .

- From the earlier discussion, we have that $k S_2(\bar{p}) \geq N M(p_1, p_2, \dots, p_N)$, where $M(\bar{p})$ is an entropy-like measure associated with the PMF.
- In the spirit of Lemma 4, it can easily be proved that Tsallis entropy for arbitrary real parameter 'q' assumes maximum possible value for uniform PMF

$$i.e. p_i = \frac{1}{N} \text{ for all } 1 \leq i \leq N. \text{ Details are avoided for brevity.}$$

Now, we derive the following interesting result where we set $k = 1$. :

Lemma 5: $\lim_{N \rightarrow \infty} (S_2(\bar{p}))^N = \lim_{N \rightarrow \infty} (\text{Trace}(\bar{G}))^N \leq \frac{1}{e}$. It will be an equality if and only if the probability mass function corresponds to a uniform PMF i.e. $p_i = \frac{1}{N}$ for all i .

Proof: From Lemma (4), we have that $\text{Trace}(\bar{G}) \leq \left(1 - \frac{1}{N}\right)$ with equality if and only if

$$p_i = \frac{1}{N} \text{ for all } i \text{ i.e. uniform Probability Mass Function (PMF).}$$

From the above discussion, we have that with $k = 1$, $\text{Trace}(\bar{G}) = S_2(\bar{p})$.

Thus, we readily infer that

$$(\text{Trace}(\bar{G}))^N \leq \left(1 - \frac{1}{N}\right)^N.$$

Taking the limit on both sides, we have

$$\lim_{N \rightarrow \infty} (\text{Trace}(\bar{G}))^N \leq \lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right)^N.$$

But, from basic calculus, we know that

$$\lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right)^N = \frac{1}{e}.$$

Thus, claim in the Lemma follows with equality for uniform Probability Mass Function (PMF).

The following corollary is a generalization and readily follows

Corollary: $\lim_{N \rightarrow \infty} (S_q(\bar{p}))^N \leq \frac{1}{e^{(q-1)}}$.

Note: We now provide the probabilistic interpretation of claim in Lemma 5. Let X, Y be independent random variables assuming the values $\{1, 2, \dots, N\}$. We have that

$$\text{Trace}(\bar{G}) = \sum_{i=1}^N p_i (1 - p_i) = \sum_{i=1}^N \text{Prob}\{X = i, Y \neq i\} = \text{Prob}\{X \neq Y\} = S_2(\bar{p}).$$

Now, M such independent “pairs of trials” are performed and the following probability is computed

Thus,

$$\lim_{M \rightarrow \infty} (S_2(\bar{p}))^M = \lim_{M \rightarrow \infty} (\text{Prob}\{X \neq Y\})^M.$$

X, Y can correspond two independent trials (assuming values $\{1, 2, 3, \dots, N\}$).

- **Entropic Quadratic and Higher Degree Forms:**

We now express $S_2(\bar{p})$ in an equivalent form and show the relationship to a quadratic form. We have that

$$S_2(\bar{p}) = \left[(p_1 + p_2 + \dots + p_N)^2 - \sum_{i=1}^N p_i^2 \right] = \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N p_i p_j.$$

$$= \bar{p}^T \bar{R} \bar{p},$$

where \bar{R} is a symmetric matrix all of whose diagonal elements are zero and the off-diagonal elements are all equal to ONE. Also, it readily follows that

- (i) $\bar{R} = \bar{e}\bar{e}^T - I$, where \bar{e} is a column vector all of whose elements are 1 and I is the identity matrix. Hence, the eigenvalues of \bar{R} are $\{(N-1), -1, -1, \dots, -1\}$ i.e. \bar{R} is an indefinite matrix with -1 is an eigenvalue of multiplicity (N-1).

- Similar interpretation of $S_q(\bar{p})$ as a higher degree form based on a tensor can be readily given. It is avoided for brevity.

- We are naturally led to the following discussion:

Entropic measures such as Shannon entropy, Tsallis entropy satisfy the following conditions (axioms):

AXIOM (i): For all probability vectors, one of whose elements is one and all other elements are zero (i.e there are N such vectors), the Shannon and Tsallis entropies are equal to zero i.e. For such probability vectors i.e. $\bar{p}'s$, $S_2(\bar{p}) = 0$.

AXIOM (ii): For probability vector, all of whose elements are equal to $\frac{1}{N}$, Shannon entropy as well as Tsallis entropy assume the maximum value i.e. Those entropies attain maximum value for the uniform probability mass function. Specifically Shannon entropy attains the value $\log N$ and Tsallis entropy attains the value $1 - \frac{1}{N}$ for uniform PMF.

AXIOM (iii): For all probability vectors, the entropic quadratic form is non-negative.

Note: By multiplying the matrix \bar{R} by the scaling factor $\frac{N \log N}{N-1}$, it can be ensured that the dynamic range of Shannon entropy as well as Tsallis entropy $S_2(\bar{p})$ are same.

It is highly reasonable that the above three conditions/axioms must be satisfied by any reasonable entropy measure. Thus, we are led to the following problem.

- **PROBLEM:** Provide complete characterization of entropic quadratic forms i.e. specify the class of matrices \bar{R} 's for which the quadratic forms $\bar{p}^T \bar{R} \bar{p}$'s satisfy the above three conditions/axioms.

We now provide a partial solution to the above problem. Let us consider a non-negative matrix, A all of whose diagonal elements are zero. It readily follows that for the axiom (i) to be satisfied, it is necessary and sufficient the diagonal elements of matrix defining the quadratic form must all be zero. This result follows from the fact that for any probability vector one of whose elements is one and all the others are zero, the associated quadratic form becomes equal to the diagonal element of the defining symmetric matrix. Also, from simple arithmetic argument, the above axiom (ii) is satisfied at the probability vector $\left(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}\right)$. Thus, any such non-negative matrix leads to an Entropic Quadratic Form.

- Hence, it now remains to see if there are any matrices with positive as well as negative elements (with zero diagonal elements) and the quadratic form associated with such matrices satisfies the axiom (ii) also.

In view of the above point, we are naturally led to the following Lemma.

Lemma 6: There is no symmetric matrix, other than \bar{R} or scaled versions of it (i.e. with all diagonal elements being zero and all the off-diagonal elements being equal to a constant value) with \bar{e} (i.e. a column vector all of whose elements are '1'), as an eigenvector

Proof: We first consider the case where N, the dimension of matrix defining the quadratic form is 2. In this, in view of axiom (i), axiom (ii), axiom (iii), the defining matrix is of the form

$\begin{bmatrix} 0 & \alpha \\ \alpha & 0 \end{bmatrix}$ with α being positive. Thus, the claim is true. In the case of $S_2(\bar{p})$, $\alpha = 1$. Any real number α will also lead to entropic quadratic form satisfying the three axioms.

Now, we consider $N = 3$. Let the defining symmetric matrix be

$$\begin{bmatrix} 0 & a & b \\ a & 0 & c \\ b & c & 0 \end{bmatrix} \text{ for real numbers, } a, b, c.$$

Case(i): Suppose $a + b = a + c$ i.e. first two row sums are equal. Hence $b = c$.

Now further suppose $b + c = a + b$. In this $a = c$. Thus claim is true.

Case (ii): Suppose $a + b \neq a + c$. In this case also, the claim is true.

For $N > 3$, the proof follows by constructing the symmetric matrix incrementally ensuring that the row sum is constant. Also, another proof by mathematical induction can be readily provided. Q.E.D.

Note:

The above arguments can be generalized for tensor based higher degree forms that satisfy the above three axioms expected of an entropy measure.

- **Probabilistic Interpretation of Tsallis Entropy for Integer Valued Parameter 'q':**

The following probabilistic interpretation follows from a generalization of the probabilistic interpretation of diagonal elements of Variance Laplacian matrix \bar{G} . Specifically consider ' q ' independent identically distributed random variables i.e. X_1, X_2, \dots, X_q

and $N = q$. Consider the following quantity:

$$\begin{aligned} \sum_{i=1}^q \text{Prob}(X_1 = i) (1 - \text{Prob}(X_2 = i, X_3 = i, \dots, X_q = i)) &= \sum_{i=1}^q p_i (1 - p_i^{q-1}) = \left(1 - \sum_{i=1}^N p_i^q\right) \\ &= \frac{q-1}{k} S_q(\bar{p}). \end{aligned}$$

The above interpretation can also be given in terms of arbitrary independent trials.

Note: As in the case of $q = 2$, for arbitrary integer q , we can associate a matrix \bar{G} such that $\text{Trace}(\bar{G}) = S_q(\bar{G})$ with $k = 1$. Details are avoided for brevity.

Note: It readily follows that $\text{Trace}(\bar{G})$ is the DC/constant contribution to the variance Laplacian based quadratic form evaluated on the unit hypercube (i.e. set of all vectors whose components are +1 or -1). We readily have that

$$\bar{T}^T \bar{G} \bar{T} = \text{Trace}(\bar{G}) + \text{terms dependent on } \bar{T}.$$

It is exactly equal to the scaled Tsallis entropy, $k S_2(\bar{p})$ associated with the probability mass function of the discrete random variable.

- **Relationship between Renyi Entropy and Tsallis Entropy:**

We now reason that Renyi entropy is approximated by Tsallis Entropy under some conditions.

Definition: Renyi entropy (of a discrete random variable X) of order ' α ', where $\alpha \geq 0$ and $\alpha \neq 1$ is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^N p_i^\alpha \right) = \frac{1}{1-\alpha} \log \left(1 - \left(1 - \sum_{i=1}^N p_i^\alpha \right) \right).$$

Letting $\sum_{i=1}^N p_i^\alpha = r$, we have that $H_\alpha(X) = \frac{1}{1-\alpha} \log(1-r)$.

But from the basic theory of infinite series [Kno], we have that

$$\log(1-r) = -r + \frac{r^2}{2} - \frac{r^3}{3} + \dots \text{ for } |r| < 1.$$

We consider non-degenerate probability mass functions. For such PMF's it readily follows that for $\alpha \geq 1$, $0 < r < 1$ (We expect that even for $0 < \alpha < 1$, $|r| < 1$). Thus, if we truncate the infinite series for $\log(1-r)$, we have that

$$\log(1-r) \approx -r \quad \text{for } |r| < 1.$$

Hence, it readily follows that with such approximation, we have

$$H_\alpha(X) \approx \frac{1}{1-\alpha} \left(- \left(1 - \sum_{i=1}^N p_i^\alpha \right) \right) = \frac{1}{\alpha-1} \left(1 - \sum_{i=1}^N p_i^\alpha \right) = S_\alpha(\bar{p}), \quad \text{where}$$

$S_\alpha(\bar{p})$ is Tsallis entropy with ' α ' is the real parameter.

Now, we bound the error term in approximating $\log(1-r)$ by $-r$. The error term is

$$\frac{r^2}{2} - \frac{r^3}{3} + \frac{r^4}{4} - \frac{r^5}{5} \dots = r^2 \left(\frac{1}{2} - \frac{r}{3} \right) + r^4 \left(\frac{1}{4} - \frac{r}{5} \right) + \dots \text{ with } |r| < 1.$$

Thus, the error term can be bounded by the following geometric series i.e.

$$r^2 + r^4 + r^6 + \dots = \frac{r^2}{1-r^2}$$

Note: The above approach of approximating entropy (such as Shannon Entropy) was first proposed in [Rama2]. Specifically Shannon Entropy is approximated by Tsallis Entropy.

Note: We can consider higher order approximations in association with $\log(1-r)$ and arrive at better approximations of Renyi entropy of order ' α '. We now consider second order approximation:

$$\log(1-r) \approx -r + \frac{r^2}{2} \text{ for } |r| < 1.$$

Using this approximation, we have that

$$H_\alpha(X) \approx \frac{1}{1-\alpha} \left(\left(-1 + \sum_{i=1}^N p_i^\alpha \right) + \frac{1}{2} \left(1 - \sum_{i=1}^N p_i^\alpha \right)^2 \right).$$

Expanding $\left(1 - \sum_{i=1}^N p_i^\alpha \right)^2$ and simplifying, we have that

$$H_\alpha(X) \approx \frac{1}{2(\alpha-1)} \left(1 - \sum_{i=1}^N \sum_{j=1}^N p_i^\alpha p_j^\alpha \right).$$

It can also be rewritten as

$$H_\alpha(X) \approx \frac{1}{2(\alpha-1)} \left(\left(\sum_{i=1}^N p_i \right)^\alpha - \sum_{i=1}^N \sum_{j=1}^N p_i^\alpha p_j^\alpha \right).$$

In the above expression, multinomial theorem can be used for further simplification.

Simplifying the above, we have that

$$H_\alpha(X) \approx \frac{1}{2(\alpha-1)} \left(1 - \sum_{k=1}^N (p_k)^{2\alpha} - \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N p_i^\alpha p_j^\alpha \right).$$

Using the definition of Tsallis entropy, we have

$$H_\alpha(X) \approx \frac{(2\alpha-1)}{(2\alpha-2)} S_{2\alpha}(\bar{p}) - \frac{1}{2(\alpha-1)} \left(\sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N p_i^\alpha p_j^\alpha \right).$$

We now obtain an equivalent expression for Renyi Entropy. Letting $t_i = p_i^\alpha$, we arrive at the vector $\bar{t} = (t_1 \ t_2 \ \dots \ t_N)^T$. In terms of that vector, the following approximation based on quadratic form is readily obtained

$$H_\alpha(X) \approx \frac{1}{2(\alpha-1)} (1 - \bar{t}^T \bar{B} \bar{t}), \quad \text{where } \bar{B} = \bar{e} \bar{e}^T \text{ with } \bar{e}, \quad \text{a column vector of 1's.}$$

- Now, we consider a specific value of α i.e. $\alpha = 2$ and arrive at an expression for approximating the Renyi entropy:

$$H_2(X) \approx \frac{1}{2} \left(\left(\sum_{i=1}^N p_i \right)^2 - \sum_{i=1}^N \sum_{j=1}^N p_i^2 p_j^2 \right).$$

Initial simplification of the above expression leads to

$$H_2(X) \approx \frac{1}{2} \left(S_2(\bar{p}) (1 - S_2(\bar{p})) + \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N p_i p_j \right).$$

On further simplification, we arrive at the following approximation for $H_2(X)$ in terms of Tsallis entropy $S_2(\bar{p})$.

$$H_2(X) \approx \left(S_2(\bar{p}) - \frac{1}{2} (S_2(\bar{p}))^2 \right).$$

In the following lemma, we derive interesting results related to $\sum_{i=1}^N p_i^q$. Specifically, the set of inequalities can have interesting consequences for Tsallis entropy,

Lemma 7: Consider probability mass function $\{p_1, p_2, \dots, p_N\}$. The following inequalities hold true:

$$\sum_{i=1}^N p_i^{2m+1} \leq \sum_{i=1}^N p_i^{m+1} \text{ for all integer 'm'. But}$$

$$\sum_{i=1}^N p_i^{2m+1} \geq \left(\sum_{i=1}^N p_i^{m+1} \right)^2 \text{ for all 'm'. Hence } S_{2m+1}(\bar{p}) \leq \frac{k}{2m} \left(1 - \left(\sum_{i=1}^N p_i^{m+1} \right)^2 \right)$$

Proof: Since p_i 's are probabilities, we readily have that $p_i^{2m+1} \leq p_i^{m+1}$ for any integer 'm'. Thus,

$$\sum_{i=1}^N p_i^{2m+1} \leq \sum_{i=1}^N p_i^{m+1} \text{ for all integer 'm'.$$

Now, consider a random variable Z which assume the values $\{p_1^m, p_2^m, \dots, p_N^m\}$ i.e. values assumed are higher integer powers of the probabilities in the associated PMF. We know that the variance of Z is non-negative.

$$\text{Variance}(Z) = \text{Var}(Z) = E(Z^2) - (E(Z))^2 \geq 0.$$

Thus, it readily follows that $E(Z^2) \geq (E(Z))^2$ and hence

$$\sum_{i=1}^N p_i^{2m+1} \geq \left(\sum_{i=1}^N p_i^{m+1} \right)^2 \text{ for all 'm'.$$

Thus, effectively we have that

$$\sum_{i=1}^N p_i^{m+1} \geq \sum_{i=1}^N p_i^{2m+1} \geq \left(\sum_{i=1}^N p_i^{m+1} \right)^2 \text{ for all m.}$$

Thus effectively $S_{2m+1}(\bar{p}) \leq \frac{k}{2m} \left(1 - \left(\sum_{i=1}^N p_i^{m+1} \right)^2 \right)$ or equivalently

$$S_{2m+1}(\bar{p}) \leq \left(S_{m+1}(\bar{p}) - \frac{m}{2k} (S_{m+1}(\bar{p}))^2 \right).$$

Corollary 1: Suppose the random variable Z assumes probability values q_i 's different from p_i 's.

Then, using the fact that Variance of Z is non-negative, we have the following inequality

$$\sum_{i=1}^N q_i^2 p_i \geq \left(\sum_{i=1}^N q_i p_i \right)^2.$$

It should be noted that both sides of inequality are convex combinations of real numbers Q.E.D.

Note: In terms of the Laplacian matrix \bar{G} , the above Lemma based inequality can be restated.

Let $\bar{T} = [p_1^m \ p_2^m \ \dots \ p_N^m]$ for a fixed integer 'm'. We readily have that $\bar{G} = \bar{D} - \bar{P}$ and

$$\text{Variance}(Z) \geq 0. \text{ Hence we have that } \bar{T}^T \bar{D} \bar{T} \geq \bar{T}^T \bar{P} \bar{T}.$$

Such a type of inequality can also be associated with positive real numbers which can be normalized into probabilities (using their sum). Details are avoided for brevity.

Note: Suppose the values assumed by the random variable are $\left\{ \frac{1}{p_1^m}, \frac{1}{p_2^m}, \dots, \frac{1}{p_N^m} \right\}$, then using the idea in the above proof, we have that

$$\sum_{i=1}^N \frac{1}{p_i^{2m-1}} \geq \left(\sum_{i=1}^N \frac{1}{p_i^{m-1}} \right)^2.$$

In the above inequalities, the probabilities can be rational numbers less than one. Hence the above inequalities hold true between rational numbers.

Now, we compute the $\text{Trace}(\bar{G}^2)$ (in the same spirit of $\text{Trace}(\bar{G})$) and briefly study its properties. It readily follows that, treating \bar{G} as a vector, we have that

$$\text{Trace}(\bar{G}^2) = (L^2 - \text{norm}(\bar{G}))^2 = \sum_{i=1}^N \mu_i^2 = (L^2 - \text{norm}(\bar{\mu}))^2.$$

i.e. treating the set of eigenvalues leading to eigenvalue vector, $\text{Trace}(\bar{G}^2)$ is the square of $L^2 - \text{norm}$ of such vector (of eigenvalues, the smallest of which is zero). Also, from the theory of matrix norms, the $L^2 - \text{norm}$ of a matrix is related to the spectral radius. We have

$$\begin{aligned} \text{Trace}(\bar{G}^2) &= \sum_{i=1}^N p_i^2 (1 - p_i)^2 + \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N p_i^2 p_j^2 \\ &= \sum_{i=1}^N p_i^2 \sum_{\substack{j=1 \\ j \neq i}}^N p_j^2 + \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N p_i^2 p_j^2 \\ &= 2 \left[\sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N p_i^2 p_j^2 \right] = \sum_{i=1}^N \mu_i^2 \text{ (with } \mu_{\min} = 0 \text{)}. \end{aligned}$$

Hence, $\text{Trace}(\bar{G}^2)$ is divisible by 2. Using the definition of Tsallis entropy $S_q(\bar{p})$, it can be readily seen that

$$\text{Trace}(\bar{G}^2) = 2 \left[\frac{1}{k^2} (S_2(\bar{p}))^2 - \frac{2}{k} (S_2(\bar{p})) + \frac{3}{k} (S_4(\bar{p})) \right].$$

We now determine an upper bound on $\text{Trace}(\bar{G}^2)$.

- Now, we derive interesting property related to the eigenvectors of \bar{G} .

Lemma 8: Let \bar{G} be an $N \times N$ matrix. Then $\text{Trace}(\bar{G}^2)$ has the following lower bound.

$$\text{Trace}(\bar{G}^2) \geq \frac{2(N-1)}{N^3}.$$

Proof: We apply Lagrange Multiplier's method to bound

$$\text{Trace}(\bar{G}^2) = 2 \left[\sum_{i=1}^N \sum_{j=1}^N p_i^2 p_j^2 - \sum_{k=1}^N p_k^4 \right].$$

Thus the objective function for the optimization problem is given by

$$J(p_1, p_2, \dots, p_N) = 2 \left[\sum_{i=1}^N \sum_{j=1}^N p_i^2 p_j^2 - \sum_{k=1}^N p_k^4 \right].$$

Using the constraint that the probabilities sum upto one, we have that the Lagrangian is given by

$$L(p_1, p_2, \dots, p_N) = 2 \left[\sum_{i=1}^N \sum_{j=1}^N p_i^2 p_j^2 - \sum_{k=1}^N p_k^4 \right] + \beta (\sum_{i=1}^N p_i - 1).$$

The critical point and the components of Hessian matrix are given by:

$$\frac{\delta L}{\delta p_i} = \left[4 \left(\sum_{\substack{k=1 \\ k \neq i}}^N p_k^2 \right) (p_i) + \beta \right], \quad \frac{\delta^2 L}{\delta p_i^2} = 4 \left(\sum_{\substack{k=1 \\ k \neq i}}^N p_k^2 \right) \text{ for all 'i',}$$

$$\frac{\delta^2 L}{\delta p_i \delta p_j} = 0 \text{ for all } i \neq j.$$

Thus, there is a single critical point and the Hessian matrix is positive definite at the critical point. Hence, we infer that the objective function has a unique minimum and occurs at

$$\frac{\delta L}{\delta p_i} = 0 \text{ i.e. } p_i = \frac{-\beta}{4 \left(\sum_{\substack{k=1 \\ k \neq i}}^N p_k^2 \right)}. \text{ Using the constraint that the probabilities sum to one,}$$

we have

$$\beta = \frac{-4 \left(\sum_{\substack{k=1 \\ k \neq i}}^N p_k^2 \right)}{N}. \text{ Thus, the global minimum occurs at } p_i = \frac{1}{N} \text{ for all 'i'.$$

$\text{Trace}(\bar{G}^2)$ at the unique minimum point is given by

$$\text{Trace}(\bar{G}^2) = \frac{2(N-1)}{N^3} \quad \text{Q.E.D.}$$

Corollary: Using the above lower bound on $\text{Trace}(\bar{G}^2)$, we lower bound the

spectral radius of \bar{G} i.e. $(N-1) \mu_{\max}^2 \geq \frac{2(N-1)}{N^3}$. Thus $\mu_{\max} \geq \frac{1}{N} \sqrt{\frac{2}{N}}$. Q.E.D.

Lemma 9: The right eigenvectors $\overline{g}'s$ (whose transpose are the left eigenvectors) of the variance Laplacian \overline{G} that are different from the all-ones vector (i.e. \overline{e} which lies in the right null space of \overline{G}) are such that they lie in the null space of matrix of all ones, \overline{S} i.e. $S_{ij} = 1$ for all i, j .

Proof: Since \overline{G} is a symmetric matrix, the set of eigenvectors forms an orthonormal basis. Also, the eigenvector corresponding to the ZERO eigenvalue of \overline{G} is the column vector of all ONES. Hence, we readily have the following fact:

$\overline{g}_i^T \overline{e} = 0$ for all i . Thus, the components of all other eigenvectors sum to zero.

Also, it readily follows that $\overline{g}^T \overline{S} \overline{g} = 0$. Since \overline{S} is a rank one matrix with the only non-zero eigenvalue being 'N' (with \overline{e} being the associated eigenvector), all the vectors $\overline{g}'s$ lie in the null space of \overline{S} (in fact they form the basis of the null space of \overline{S}).

Hence, L^1 - norm (\overline{g}_i) is divisible by 2 for all eigenvectors $\overline{g}'s$.

Also, let \overline{g} be an eigenvector of \overline{G} , other than all ones vector i.e. \overline{e} . We have that

$$\left(\sum_{i=1}^N g_i \right)^2 = \sum_{i=1}^N g_i^2 + 2 \left(\text{pairwise product of distinct components of } \overline{g} \right) = 0.$$

Hence, it follows that $\overline{g}^T \overline{S} \overline{g} = -1$, where \overline{S} is a matrix all of whose diagonal elements are zero and all the non-diagonal elements are 1.

Since L^2 - norm of \overline{g} is ONE, it readily follows that

$$\text{pairwise product of distinct components of } \overline{g} = -\frac{1}{2}. \text{ Q.E.D.}$$

Similar result can be derived based on the L^p - norm of \overline{g} . Details are avoided for brevity.

We now propose an interesting orthonormal basis which satisfies all the properties required of the set of eigenvectors of an arbitrary Laplacian matrix.

Definition: Hadamard basis (orthonormal) is the normalized set of rows/columns of a symmetric

Hadamard matrix, H_m . For instance, it is well known that $H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Hence the Hadamard

basis is given by $\left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \right\}$.

Note: Two $\{+1, -1\}$ vectors are orthogonal if and only if the number of +1's is equal to the number of -1's. Such vectors exist if and only if the dimension of vectors is an even number. Further the sum of elements in such vectors is zero (as required by the eigen vectors of an arbitrary Laplacian matrix which is not necessarily a variance Laplacian matrix).

Note: In view of Rayleigh's Theorem, if the orthonormal basis of eigenvectors of a Variance Laplacian \bar{G} is the Hadamard basis, then the global maximum value of associated quadratic form evaluated on the unit hypercube is attained at the eigenvector corresponding to its spectral radius.

- **Spectral Representation of Symmetric Laplacian Matrix \bar{G} :**

We now arrive at the spectral representation of variance Laplacian matrix \bar{G} i.e.

$$\bar{G} = \bar{P} D \bar{P}^T = \sum_{i=2}^N \mu_i \bar{f}_i \bar{f}_i^T \text{ where } \mu_i \text{'s are eigenvalues with } \mu_1 = 0 \text{ and } \bar{f}_i \text{ are}$$

Normalized eigenvectors of \bar{G} . It should be noted that the column vector of ALL ONES i.e. $\bar{e} = (1 \ 1 \dots 1)^T$ is an eigenvector corresponding to the zero eigenvalue and $\frac{1}{\sqrt{N}} \bar{e}$ is the associated normalized eigenvector.

We know that \bar{G} is completely specified the probability mass function of the associated discrete random variable i.e. $\{p_1, p_2, \dots, p_N\}$. Hence we have that

$$\sum_{i=2}^N \mu_i f_{ij}^2 = p_j (1 - p_j) \text{ for } 1 \leq j \leq N \text{ (i.e. diagonal elements of } \bar{G} \text{)}.$$

Also, we have that

$$\sum_{i=2}^N \mu_i f_{il} f_{im} = -p_l p_m \text{ for } l \neq m \text{ and } 1 \leq l \leq N, 1 \leq m \leq N \text{ i.e.}$$

(off diagonal elements of \bar{G}).

The orthogonal matrix \bar{P} is of the following form:

$$\bar{P} = \begin{bmatrix} \frac{1}{\sqrt{N}} & f_{21} & f_{31} & \dots & f_{N1} \\ \frac{1}{\sqrt{N}} & f_{22} & f_{32} & \dots & f_{N2} \\ \frac{1}{\sqrt{N}} & f_{23} & f_{33} & \dots & f_{N3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{N}} & f_{2N} & f_{3N} & \dots & f_{NN} \end{bmatrix}.$$

Since, we have that $\bar{P} \bar{P}^T = \bar{P}^T \bar{P} = I$, the L^2 - norm of rows, column vectors of \bar{P} is one.

The residue matrices i.e. $\bar{E}_i = \bar{f}_i \bar{f}_i^T$ are such that

$$\sum_{i=1}^N \bar{E}_i = I. \text{ Hence } \sum_{i=2}^N \bar{E}_i = \bar{Q} \text{ with } Q_{ii} = \frac{N-1}{N} \text{ for all } i \text{ and } Q_{ij} = -\frac{1}{N} \text{ for } i \neq j.$$

It follows that $\bar{Q} \bar{e} = \bar{0}$, where \bar{e} is a column vector of all ones and \bar{Q} is Laplacian .

Also, we readily have that

$$\sum_{i=2}^N f_{ij}^2 = \frac{N-1}{N} \text{ and } \sum_{i=2}^N f_{il} f_{im} = -\frac{1}{N} \text{ for } l \neq m \text{ and } 1 \leq l \leq N, 1 \leq m \leq N.$$

Note: In the spirit of properties of Laplacian \bar{G} , we can derive new results related to Graph Laplacian. Thus new results in spectral graph theory can be readily derived.

- **Abstract Vector Space of Random Variables:**

Consider a collection of discrete random variables. All of them assume same values. Specifically consider two random variables X, Y . From research literature [PaP], $E(XY)$ (i.e. expected value of their product) can be regarded as an inner product between the random variables $\{X, Y\}$ (regarded as abstract vectors). Suppose \bar{T} be the set of values assumed by the random variables X, Y . It readily follows that

$$E(XY) = \bar{T}^T \tilde{P} \bar{T}, \text{ where } \tilde{P} \text{ can be considered as a symmetric matrix.}$$

$$\text{Using Dirac notion } E(XY) = \langle \bar{T}, \tilde{P} \bar{T} \rangle.$$

It readily follows that the inner product $E(XY)$ is zero i.e. the associated random variables are ORTHOGONAL if \bar{T} lies in the null space of the symmetric matrix \tilde{P} . Thus, the null space of matrix \tilde{P} determines the space of orthogonal random variables.

- **Connections to Stochastic Processes:**

Let us first consider a discrete time, discrete state space stochastic process i.e. a countable collection of discrete random variables. In view of the above results, the variance values of random variables constitute a sequence of quadratic forms. Thus, the sequence of scalar variance values constitute an infinite sequence of real/complex numbers. We consider the following special cases:

- (I) Consider the case where the random process is a strict sense stationary random process. Hence, the sequence of variance values (i.e. the associated quadratic forms) form a constant sequence (DC sequence).
- (II) Consider the case where the random process constitutes a homogeneous Discrete Time Markov Chain (DTMC). Since such a process exhibits an equilibrium behaviour, the sequence of variance values of the discrete random variables (i.e. associated quadratic forms) converges to an equilibrium variance value (based on the equilibrium probability mass function).

- **Unit Random Process: Connection to Verhulst Dynamical System:**

Now, we consider a UNIT Random Process (i.e. state, $X(n)$ of the random process assumes $\{+1, -1\}$ values only) whose marginal Probability Mass Function (PMF) is of the form $\{q(n), 1 - q(n)\}$ i.e. time-varying PMF depends on the time evolution of probability $q(n)$. Let the evolution of $q(n)$ be given by

$$q(n+1) = a q(n)(1 - q(n)), \quad \text{where } a \in [0, 4].$$

The above dynamics is same as that of a Verhulst dynamical system. Hence we

can utilize the known results for Verhulst dynamical system. Thus, it readily follows that for $a \in [0,3]$,

the marginal PMF converges to the equilibrium PMF of $\left\{ \frac{a-1}{a}, \frac{1}{a} \right\}$. Also, for $a \in (3,4]$, the PMF exhibits oscillatory behavior or chaotic behavior.

In the case of such UNIT RANDOM Process, the Variance Laplacian is given by

$$\bar{G}(n) = \begin{bmatrix} q(n)(1-q(n)) & -q(n)(1-q(n)) \\ -q(n)(1-q(n)) & q(n)(1-q(n)) \end{bmatrix}.$$

It also immediately follows that $\text{Trace}(\bar{G}(n)) = 2q(n)(1-q(n))$ and

$\text{Variance}(X(n)) = 4q(n)(1-q(n))$. Thus, for $a \in [0,3]$, the equilibrium value of Variance of the UNIT random process is given by $4 \frac{(a-1)}{a^2}$.

Note: The unit random process considered exhibits Markovian property. But, the time evolution of marginal PMF is non-linear. Under some conditions such a stochastic dynamical system exhibits steady state/equilibrium behavior.

Note: In view of algebraic interpretation of $S_q(\bar{p})$ for integer valued q , we define the

following **Generalized Verhulst type dynamical system: Generalized Logistic Map**

$$x(n+1) = a x(n)(1-x(n)) \left(1 + x(n) + x^2(n) + \dots + x^l(n) \right) \text{ for an integer } l \text{ with } x(n) > 0$$

We are currently investigating the dynamics of such a Generalized Verhulst dynamical system.

We now provide some related results. The generalized logistic map is given by

$f(y) = a y (1 - y^l)$ for ' l ' an integer. From basis calculus, it follows that the global maximum of $f(\cdot)$ occurs at $y = \sqrt[l]{\frac{1}{1+l}}$. Thus, for $f(y) < 1$, we require that $a \in \left(0, \frac{(l+1)\sqrt[l]{l+1}}{l} \right)$.

It can be readily verified that the above generalized logistic map reduces the ordinary logistic map for $l = 1$.

- **Functional Logistic Map: Sigmoid Function:**

We now introduce the following generalization of logistic map. The basis for the generalization is the following function of interest in artificial neural networks (called SIGMOID function).

$$\text{SIGMOID}(z) = \frac{1}{1 + e^{-z}} = f(z). \text{ It readily follows that}$$

$$\frac{df(z)}{dz} = f(z)(1-f(z)).$$

From the theory of ordinary non-linear differential equations, it could be reason that the above differential equation has a unique solution. We are naturally led to the following generalization:

$$\frac{df(z)}{dz} = a_0 + a_1 f(z) + a_2 (f(z))^2 + \dots + a_M (f(z))^M, \quad \text{where}$$

$g(y) = a_0 + a_1 y + a_2 y^2 + \dots + a_M y^M$ is a polynomial such that $g(1) = 0$.

Thus, we are led to the following definition:

Definition: A functional logistic map based on the function $f(z)$ is defined as

$g(f(z)) = b f(z) (1 - f(z))$ where b is a real number. More generally, in the spirit of above discussion, we have

$$h(f(z)) = b f(z) (1 - f^q(z)) \text{ for an integer 'q'.$$

It readily follows that if $f(z) = z$, then $g(z)$ is the well known logistic map.

We realize that Verhulst population dynamical system is defined based on the logistic map. In the same spirit, we can define interesting non-linear dynamical systems based on the above generalized logistic maps proposed above. Details are avoided for brevity. We expect detailed results to be derived in association with such non-linear differential equations and dynamical systems.

4. Other Interesting Quadratic Forms in Probability/Statistics:

In this section, we investigate several other quadratic forms which are naturally associated with measures such as covariance/Correlation of two random variables which assume same values.

- In general, quadratic form is of the form $\beta = \sum_{i=1}^N \sum_{j=1}^N T_i T_j B_{ij}$, where B_{ij} has statistical or probabilistic significance e.g. B could be Toeplitz auto – correlation matrix of an Auto – Regressive process. In fact B could be the state transition matrix of a Discrete – Time Markov Chain (DTMC). Further B could be $-Q$, where Q is the generator matrix of a CTMC.
- Variance Laplacian related investigation naturally leads to studying the following more general quadratic form associated with two jointly distributed random variables X, Y that are “symmetric” in the sense that their ‘marginal probability mass functions’ are exactly same and the values assumed by them are same. Let the common marginal probability mass function of the two random variables be $\{p_1, p_2, \dots, p_N\}$. In the spirit of Laplacian \bar{G} , we are motivated to introduce, a more general Laplacian matrix, \bar{H} i.e.

$$\bar{H} = \bar{D} - \bar{P}, \quad \text{where } \bar{D} = \text{diag}(p_1, p_2, \dots, p_N) \text{ i.e. a diagonal matrix and}$$

$$\bar{P}_{ij} = \text{Probability}\{X = i, Y = j\} \text{ i.e. matrix of joint probabilities.}$$

With such definition \bar{H} need not be symmetric but still is Laplacian. Suppose, \bar{P} is a symmetric matrix (a stronger condition which ensures that the random variables $\{X, Y\}$ are “symmetric”), \bar{H} will be a symmetric, Laplacian matrix.

Let the common vector of values assumed by the random variables, X, Y be \bar{T} . Hence, the quadratic form associated with \bar{H} is given by $\bar{T}^T \bar{H} \bar{T}$. Explicitly, we have the following novel measure associated with jointly distributed random variables $\{X, Y\}$.

$$\begin{aligned}\theta &= \bar{T}^T \bar{H} \bar{T} = \sum_{i=1}^N T_i^2 p_i - \sum_{i=1}^N \sum_{j=1}^N T_i T_j \text{Prob} \{X = i, Y = j\} \\ &= E(X^2) - E(XY) = E(Y^2) - E(XY)\end{aligned}$$

Note: If X, Y are independent and identically distributed random variables, then the above measure is the common variance of them. Also, if X, Y are same then θ is zero.

- We now introduce the concept of “symmetrization of Jointly Distributed Random variables” based on the following well known result associated with quadratic forms:

$$\bar{T}^T \tilde{P} \bar{T} = \frac{1}{2} \bar{T}^T (\tilde{P} + \tilde{P}^T) \bar{T} \text{ i.e. symmetric quadratic form.}$$

Definition: Two jointly distributed random variables with Joint PMF matrix \tilde{P} (not necessarily symmetric) are “symmetrized” when they are associated with the symmetric joint PMF matrix $\frac{1}{2} (\tilde{P} + \tilde{P}^T)$.

Lemma 10: Laplacian quadratic form $\bar{T}^T \bar{H} \bar{T}$ is always positive semi-definite.

Proof: It readily follows that if $E(XY)$ is non-positive, then ‘ θ ’ is non-negative. Thus, the more interesting case is when $E(XY)$ is non-negative. In this case, we invoke a well known result in the abstract vector space of random variables. From [PaP], the following definition is well known

Definition: The second moment of the random variables X, Y i.e. $E(XY)$ is defined as their inner product. Further, the ratio

$$\frac{E(XY)}{\sqrt{E(X^2) E(Y^2)}}$$

is the cosine of their angle, β i.e. say $\text{Cos}(\beta)$.

Hence, it is well known that $|\text{Cos}(\beta)| \leq 1$. Thus, in the case of random variables X, Y whose joint probability mass function matrix, \tilde{P} is symmetric, we have that

$$|E(XY)| \leq E(X^2). \text{ Thus, if } E(XY) \geq 0, E(XY) \leq E(X^2).$$

Thus, the Laplacian quadratic form $\bar{T}^T \bar{H} \bar{T}$ is always positive semi – definite.

Q.E.D.

Corollary: In this case, the covariance of random variables considered above can be bounded in the following manner:

$$C_{xy} = E(XY) - (E(X))^2.$$

Since Variance is non-negative, we have that $E(X^2) \geq (E(X))^2$ or $-E(X^2) \leq -(E(X))^2$. Hence, $C_{xy} \geq -\theta$. Q.E.D.

We now briefly consider familiar scalar measures routinely utilized in probabilistic/statistical investigations and provide them with quadratic form interpretation.

(I) Covariance: By definition, covariance of two random variables X, Y is given by

$$C_{xy} = E(XY) - E(X)E(Y).$$

Suppose the random variables X, Y assume the same vector of values \bar{T} .

Then, we have the following quadratic form interpretation of covariance of X, Y.

$$\begin{aligned} C_{xy} &= \sum_{i=1}^N \sum_{j=1}^N T_i T_j \text{Prob} \{ X = i, Y = j \} - \sum_{i=1}^N \sum_{j=1}^N T_i T_j p_i p_j \\ &= \bar{T}^T \tilde{P} \bar{T} - \bar{T}^T \tilde{J} \bar{T}, \quad \text{where } \tilde{J}_{ij} = p_i p_j. \\ &= \bar{T}^T (\tilde{P} - \tilde{J}) \bar{T}. \end{aligned}$$

Thus, we have a quadratic form that is not Laplacian.

Note: The correlation coefficient of two random variables X, Y is defined as

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}. \text{ It readily follows that } |\rho_{xy}| \leq 1 \text{ i.e. } |C_{xy}| \leq \sigma_x \sigma_y.$$

This well known result can be given a quadratic form based interpretation if the random variables X, Y assume the same values.

(II) From the above discussion, it readily follows that given a random variable, X ($E(X)$), $E(X^2)$ are arbitrary quadratic forms.

Note: Suppose, we consider two independent discrete random variables X, Y and their sum Z i.e. $Z = X + Y$. The following inferences are well known from probability theory:

- The Probability Mass Function (PMF) of Z is the convolution of the PMFs of the random variables X, Y
- Variance (Z) = Variance (X) + Variance (Y).

Suppose the discrete random variables X, Y assume the same set of finitely many values captured by the $N \times 1$ vector \bar{T} . Also, let \tilde{G}_x, \tilde{G}_y be the associated $N \times N$ variance Laplacian matrices. The set of values assumed by Z (i.e. convolution of the elements of \bar{T} with those of elements of \bar{T} itself) is an $2N - 1$ vector. Thus, the variance quadratic form associated with random variable Z is equal to the sum of variance quadratic forms associated with the random

variables X, Y. It should be noted that the Variance Laplacian matrix associated with the random variable Z i.e. \bar{G}_Z is a $(2N-1) \times (2N-1)$ matrix.

- Thus, the above result from probability theory shows the equivalence of two quadratic forms associated with matrices of different dimensions.

Note: Two vectors whose elements are positive real numbers can be normalized using their L^1 – norms to arrive at probability vectors. If the resulting probability vectors i.e. PMFs correspond to independent random variables, the above result on sum of variances can be invoked in association with their variance quadratic forms.

Correlation Matrix of Finitely Many Random Variables:

Let us consider finitely many real valued discrete random variables, all of which assume the same set of finitely many values. The correlation matrix of such random variables is given by

$$R_N = \begin{bmatrix} R_{11} & \cdots & R_{1N} \\ \vdots & \vdots & \vdots \\ R_{N1} & \cdots & R_{NN} \end{bmatrix}, \quad \text{where } R_{ij} = E(X_i X_j).$$

From the above discussion, it is clear that the elements of R_N are quadratic forms in the set of values assumed by the random variables \bar{T} (elements of \bar{T} are quadratic forms).

It is well known that R_N is non-negative definite. Using the above discussion, the correlation matrix R_N can be written as

$$R_N = \bar{T}^T \circ \bar{P} \circ \bar{T}$$

i.e. $R_{ij} = \bar{T}^T \bar{P}_{ij} \bar{T}$ where \bar{P}_{ij} is the sub – matrix of block symmetric matrix \bar{P} .

Note: 'o' is a suitable defined product like Kronecker or Schur product.

It should be noted that \bar{P} is the associated block symmetric matrix arising in association with quadratic forms.

Note: Covariance matrix is also well defined in statistics/probability theory. Suppose all the random variables (whose correlation matrix is considered) assume the same values. Then the diagonal elements of it are Laplacian quadratic forms and the off-diagonal elements are also quadratic forms which are not necessarily Laplacian.

5. Conclusions:

In this research paper, it is proved that the variance of a discrete random variable constitutes the quadratic form associated with a Laplacian matrix (whose elements are expressed in terms of probabilities). Various interesting properties of the associated Laplacian matrix are proved. Also, other quadratic forms which naturally arise in statistics are identified. It is shown that cross fertilization of results between the theory of quadratic forms and statistics/probability theory leads to new research directions.

References:

[Kno] K. Knopp, "Theory and Applications of Infinite Series," Dover Publications, Inc, New York

[PaP] A. Papoulis and S.U.Pillai, "Probability, Random Variables and Stochastic Processes," Tata-McGraw Hill, New Delhi, 2002.

[Rama1] G. Rama Murthy, "Time Optimal Spectrum Sensing," IIIT Technical Report No. 63, December 2015

[Rama2] G. Rama Murthy, "Gibbs, Shannon Entropy and related Measures: Tsallis Entropy," Easy Chair preprint # 866

[RRN] G. Rama Murthy, Rhishi Pratap Singh, Naveen Chilamkurti "Wide Band Time Optimal Spectrum Sensing" accepted for publication in the International Journal of Internet Technology and Secured Transactions, Inderscience Publishers