EasyChair Preprint

№ 821

Variance Laplacian: Quadratic Forms in Statistics

Rama Murthy Garimella

November 16, 2021

VARIANCE LAPLACIAN: QUADRATIC FORMS IN STATISTICS

Garimella Rama Murthy*

Department of Computer Science & Engineering

*Mahindra University , 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.

$$Variance\,[Z] = \quad X^T\, G\, X,$$
$$where\ X\ is\ the\ vector\ of\ values\ assumed\ by\ the\ discrete\ random\ variable\ and$$

$G\ is\ the\ Laplacian\ matrix\ whose\ elements\ are\ expressed\ in\ terms\ of\ probabilities.$ 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

$$Variance(Z) = Var(Z) = E(Z^2) - (E(Z))^2.$$

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

$$Var(Z) = \sum_{i=1}^{N} T_i^2 \, p_i - (\sum_{i=1}^{N} T_i \; p_i)^2$$

$$= \sum_{i=1}^{N} T_i^2 \, p_i - \sum_{i=1}^{N} \sum_{j=1}^{N} T_i T_j p_i \, p_j$$

$$= \bar{T}^T [\bar{D} - \tilde{P}] \; \bar{T},$$

where $\bar{D}$ is a diagonal matrix whose diagonal elements are $\{ p_1, p_2, \dots, p_N \}$ and

$$\tilde{P}_{ij} = p_i p_j \text{ for all } 1 \leq i, j \leq N.$$

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

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

Thus, we have shown that variance of discrete random variable Z constitutes a quadratic form associated with the matrix $\bar{G}$.

Note: Since $\bar{T}$ is the vector of values assumed by the random variable Z, the components of $\bar{T}$ are necessarily distinct real numbers.

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

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

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. The class of Variance Laplacian matrices of discrete random variables forms an interesting algebraic structure.

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 atmost 5, the eigenvalues of Laplacian matrix (roots of the associated characteristic polynomial) can be determined by algebraic formulas (Galois Theroy).

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 $Probability\{\; Z = \;\; 0\,\} = q\;.\; Then\; we\; have\; that$

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

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

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

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

$$Variance(Z) = 4\,q\;(\,1 - q\,).$$

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)

$$Trace(\bar{G}) = \frac{N-1}{N}, \qquad Determinant(\bar{G}) = 0.$$

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:

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

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

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

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

$$if\ '\alpha'\ is\ an\ eigenvalue\ of\ \bar{B}, \quad then\ the\ corresponding\ eigenvalue\ \mu\ of\ \bar{G}\ is$$

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

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

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.

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

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. Since in this case, the generator matrix of CTMC is symmetric, the equilibrium probability distribution is the uniform probability mass function.

Note: Now, we show that a Discrete Time Markov Chain (DTMC) can be naturally associated with a discrete random variable using the related Variance Laplacian $\bar{G}$. We have that

$\bar{G} = \bar{D} - \tilde{P} = \bar{D}\left(I - \bar{D}^{-1}\tilde{P}\right)$. $Since\ \bar{G}\ is\ Laplacian\ and\ \bar{D}\ is\ non-singular, it\ follows\ that$

$$\bar{D}^{-1}\tilde{P}\,\bar{e} = \overline{H}\,\bar{e} = \bar{e}.$$

Hence, $\overline{H}$ is a stochastic matrix (being non-negative and row sums are all one) which is naturally associated with the discrete random variable.

- There is another way to associate a stochastic matrix with $\bar{G}$ by letting

$$\overline{H} = I - G.$$

$It\ can\ be\ readily\ seen\ that$ $\overline{H}$ is a stochastic matrix since $\bar{G}$ is a Laplacian matrix. Based on Lemma 2 below, it readily follows that the smallest eigenvalue of $\overline{H}$ is greater than or equal to $\frac{1}{2}$.

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. $\bar{e} = [\,1\ 1\ \ldots 1\,]^{T.}$
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.

- Now, we consider a function of the random variable Z i.e. $Y = Z^2$.
We readily have that

$$Var(Y) = E(Z^4) - (\,E(Z^2)\,)^2.$$
$$Now\ substituting\ for\ E(Z^2)\ using\ Var(Z), \quad we\ have$$
$$Var(Y) = E(Z^4) + (\,Var(Z) + (E(Z)\,)^2\,)^2.$$
$$Hence, \quad we\ conclude\ that\ Var(Z^2) \geq (\,Var(Z)\,)^2.$$
$$More\ generally, \quad we\ have\ that$$
$$Var\left(Z^{2l}\right) \geq (\,Var(Z)\,)^{2l}\ for\ any\ integer\ l \geq 2.$$

$Thus, the\ above\ inequality\ can\ be\ restated\ in\ terms\ of\ the\ associated\ quadratic\ forms.$

More generally, Jensen inequality can be applied in the following manner:

Let $\varphi(.)\ be\ a\ convex\ function\ and\ X\ be\ a\ real\ valued\ random\ variable. Then,$ $we\ have\ that$ $\varphi[\,E(X)\,] \leq E[\,\varphi(X)\,]$. We readily have that $Var(Z) = E(\,(Z-\mu)^2)$.

$$\varphi[\,E(\,(Z-\mu)^2)\,] \leq E[\,\varphi(\,(Z-\mu)^2)\,]\,.$$

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:

$$Minimum\, absolute\ row\ sum\,(\bar{A}) \le\ Sp(\bar{A}\ ) \le Maximum\ absolute\ row\ sum\,(\bar{A}\ ).$$

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

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

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

$$\underset{i}{Min}\{\ 2\ p_i(\,1 - p_i)\,\} \le Sp(\,\bar{G}\ ) \le \underset{i}{Max}\{\ 2\ p_i(\,1 - p_i)\,\}\,.$$

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

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

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

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

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

Note: The function $f(p_i)$ consitutes 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

$\bar{T}^T \, (diag \, (\bar{p} \,) - \bar{p} \;\; \bar{p}^T \,) \, \bar{T} \geq 0, \quad where \; diag(\bar{p}\,) \; is \; a \; diagonal \; matrix \; whose \; components \; are \; all \; the \; components \; of \; vector \; \bar{p}.$

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

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

- We now derive an interesting equality between quadratic form based on symmetric matrices of different dimensions. This equality is derived based on the well known fact that the variance of the random variable which sum of two different independent random variables is the sum of their variances i.e. Suppose Z_1, Z_2 are two independent random variables, then

$$Variance(Z_3) = \; Variance(Z_1 + Z_2) = \; Variance \; (Z_1 \,) + Variance \; (Z_2 \,).$$

$Let \; the \; Variance \; Laplacian \; matrices \; associated \; with \; discrete \; variables$ $\{ Z_1, Z_2 \} \, be \; \overline{G_1} \;\; , \overline{G_2}$ respectively. Also, let the two discrete random variables assume the same set of values. It is well known that the probability mass function of $Z_3 (= \; Z_1 + Z_2)$ is the convolution of probability mass functions of $\{ Z_1, Z_2 \}$. Also, it readily follows that if the vector, $\bar{T}$ of set of values assumed by random variables $Z_1 \; and \; Z_2$ is an N x 1 vector, then the probability mass function of Z_3 is a vector $\tilde{T}$ of length 2N-1. Let the variance Laplacian of Z_3 be $\tilde{G}$. From the above well known fact from probability theory, It readily follows that

$$\tilde{T}^T \, \tilde{G} \, \tilde{T} \; = \; \bar{T}^T \, \bar{G}_1 \, \bar{T} + \; \bar{T}^T \, \bar{G}_2 \, \bar{T} \, .$$

The above is an equality between two quadratic forms based on matrices of different dimensions. Similar equality can be derived based on positive vectors.

- Consider two Random Variables which have the same associated probability mass function and hence the same variance Laplacian matrix $\bar{G}$. The random variables assume the associated value vectors $\overline{T_1}$ and $\overline{T_2}$. Let these two value vectors be orthogonal in the associated vector space i.e. their inner product is zero. Hence, it readily follows that the variance of random variable assuming the value vector ($\overline{T_1}$+ $\overline{T_2}$) with associated variance Laplacian $\bar{G}$ is

$$(\overline{T_1} + \; \overline{T_2})^T \, \bar{G} \, (\overline{T_1} + \; \overline{T_2} \,) \; = \; \overline{T_1}^T \, \bar{G} \, \overline{T_1} + \overline{T_2}^T \, \bar{G} \, \overline{T_2}$$

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

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

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

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

Corollary: Using the Rayleigh's Theorem and the reasoning in above Lemma, it follows that if the Variance of a random variable is non-zero, then it is lower bounded by

$$\mu_{min} \ (L^2 - norm\ (\bar{T}))^2 \leq Variance(Z), \qquad where \ \ \mu_{min} \ is \ the \ smallest \ non-zero \ eigenvalue \ of \ \bar{G} \ associated \ with \ andom \ variable \ Z.$$

Note: The above bound on variance of a discrete random variable can be used alongwith Chebyshev/Bienyme inequality in probability theory.\

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. $Trace(\bar{G})$.

It readily follows that

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

Letting $\bar{p} = [\,p_1 \ p_2 \dots p_N\,]^T$ (*i.e. vector associated with probability mass function*), we have

$$L^1 - norm\ (\bar{\mu}) + L^2 - norm\ (\bar{p}) = 1.$$

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

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

Treating the matrix $\bar{G}$ as a vector, it readily follows that

$$L^1 - norm\,(\bar{G}) \;=\; 2\;Trace\,(\bar{G}\,) = 2\,(\,1 - \;L^2 - norm\,(\bar{p})\;\;).$$

Thus,

$$L^1 - norm\,(\bar{G}) + 2\,(\,L^2 - norm\,(\bar{p})\,)\;\;=\;\;2.$$

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

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

$$Trace(\,\bar{G}\,)\;\leq\left(1 - \tfrac{1}{N}\right).$$

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

We now apply the Lagrange-mulitpliers 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

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

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

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

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

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

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

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

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

Corollary 1: *We now bound the second smallest eigenvalue* , μ_2 *of* $\bar{G}$. *It is clear that*

$Trace(\,\bar{G}\,)\;\leq\left(1 - \tfrac{1}{N}\right)$. *Further* $(N-1)\mu_2 \leq Trace(\,\bar{G}\,)$. *Hence* $\mu_2 \leq \tfrac{1}{N}$. Thus, we have $\mu_2 \in \;\left(0, \dfrac{1}{N}\right]$ *and* $\mu_i \in \left[\dfrac{1}{N}\,,\quad\dfrac{1}{2}\right]$ *for* $i \geq 3$. *It also readily follows that*

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

Since, μ_2 *is positive, we have the following*

$upper\ bound\ on\ \mu_{3.}\ Most\ generally,\ since\ the\ eigenvalues\ are\ non-negative,$ $we\ have\ the\ following\ bounds$

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

Corollary 2: Using Lemma 2, *we* provide tighter bounds on the eigenvalues other than the spectral radius. Specifically,

$$Trace(\ \bar{G}\ ) = \sum_{j=2}^{N} \mu_j \leq \left(\frac{N-1}{N}\right).\ Since\ spectral\ radius\ of\ \bar{G}\ \ is\ upper\ bounded\ by\ \frac{1}{2},$$

$$we\ have\ that\quad (N-2)\ \mu_2\ \leq\ \frac{N-1}{N} - \frac{1}{2}\ .Thus\ \mu_2\ \leq\ \frac{1}{N-2}\frac{N-2}{2\,N}.$$

Hence $\mu_2\ \leq\ \frac{1}{2\,N}$. Using identical argument, it follows that

$$\mu_k \leq \left(\frac{N-2}{N-k}\right)\left(\frac{1}{2N}\right)\quad for\ \ 2 \leq k \leq (N-1).$$

Note: The upper bound on Trace($\bar{G}$) is attained for uniform probability mass function i.e. $p_i = \frac{1}{N}\ for\ all\ i.$

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

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

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

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

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

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

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

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

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

$$M(p_1 \,, \quad p_2 \,, ., \quad p_N)\ is\ homegenous\ multi-variate\ polynomial\ of\ degree\ 2N$$

$in\ p_1 \,, p_2 \,, \dots , p_N$. It is a scalar measure associated with the probability mass function.

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

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

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

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

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

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:

as $\frac{\mu_{max}}{\mu_{min}} \geq 2\, N\, p_{min}(1 - \ p_{min})$, where
$p_{min}\ is\ the\ minimum\ of\ all\ the\ probabilities\ in\ the\ PMF\ of\ random\ variable\ Z.$

- Property (v) : Now, we consider the sum of elements of leading diagonals above and below the main diagonal (**Auto-Correlation Sequence of PMF**)

It readily follows that since $\bar{G}$ is a symmetric matrix, the sum of elements in the leading diagonals above the main diagonal are same as the corresponding ones below the main diagonal. To specify the required sums, we need the following definition:

Definition: The auto-correlation sequence { R(j) : -(N-1)≤ $j \leq (N-1)$ } of the probability mass function ($p_1, p_2, \dots, p_N$) is the sequence of auto-correlations values of the sequence of probabilities for positive as well as negative values. Specifically we have that for lag 'j'

R(j) = R (-j) for -(N-1)≤ $j \leq (N-1)$.

It readily follows that the sum of elements of main diagonal is $Trace(\bar{G}) = 1 - R(0)$ and the sum of elements of leading diagonal (above/below main diagonal) adjacent to the main diagonal is

-R(1). The sum of other leading diagonal elements are { -R(2), -R(-3),…., -R(N-1) }.

Since, the sum of elements of the Laplacian matrix is zero, it readily follows that the autocorrelation sequence i.e.

> { R(j) : -(N-1)≤ $j \leq (N-1)$ } constitutes a Probability Mass Function. We denote it as the "Auto-correlation Probability Mass Function".

Note: Using the autocorrelation probability mass function, we can construct the variance Laplacian associated with such a PMF. Thus, given a random variable, a countable number of Variance Laplacian matrices are naturally associated with it (based on the countable sequence of autocorrelation probability mass functions derived from a given PMF).

Note: In a similar spirit, trailing diagonal sums can be provided with interesting interpretation. Details are avoided for brevity.

- **Connections to Statistical Mechanics**:

Note: The expression for $Trace(\bar{G})$ 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

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

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

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

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

- 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

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

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. In view of this result, other coefficients of the characteristic polynomial can be interpreted as entropy type functions.*

- *We now associate an interesting matrix denoted by* $\tilde{G}$, *whose Trace is*
$$Trace(\tilde{G}) = \frac{k}{q-1}\left(1-\sum_{i=1}^{N} p_i^q\right) = \frac{k}{q-1}\left(\sum_{i=1}^{N} p_i(1-p_i)(1+p_i+p_i^2+\cdots.+p_i^{q-2})\right).$$
We let $\theta_i = p_i(1+p_i+p_i^2+\cdots.+p_i^{q-2})$ *and let* $\widetilde{G_{ij}} = -k\, p_j\theta_i$ *for* $i \neq j$ *and* $\widetilde{G_{ii}} = k\ p_i(1-p_i^{q-1})$. 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* $k\ S_2(\bar{p}) \geq N\ M(p_1, p_2, \ldots, p_N)$, *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*
$$i.e\ \ p_i = \frac{1}{N}\ \ for\ all\ \ 1 \leq i \leq N.\ Details\ are\ avoided\ for\ brevity.$$

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

Lemma 5: $\lim_{N\to\infty} (S_2(\bar{p}))^N = \lim_{N\to\infty} (Trace(\bar{G}))^N \leq \frac{1}{e}$. 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 $Trace(\bar{G}) \leq \left(1-\frac{1}{N}\right)$ *with equality if and only if*

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

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

Thus, we readily infer that

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

Taking the limit on both sides , we have

$$\underset{N \to \infty}{Lim} (\; Trace(\bar{G}) \;)^N \leq \underset{N \to \infty}{Lim} \left(1 - \frac{1}{N} \right)^N .$$

But, from basic calculus, we know that

$$\underset{N \to \infty}{Lim} \left(1 - \frac{1}{N} \right)^N = \frac{1}{e} .$$

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

We also readily have that

$$\underset{N \to \infty}{Lim} \left(\frac{L^1 - norm\ (\bar{G})}{2} \right)^N \leq \frac{1}{e} \;\; with\ equality\ if\ and\ only\ if$$

the probability mass function is a uniform PMF. Q.E.D.

The following corollary is a generalization and readily follows

Corollary: $\underset{N \to \infty}{Lim} \left(S_q(\bar{p}) \right)^N \leq \frac{1}{e^{(q-1)}}$.

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

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

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

Thus,

$$\underset{M \to \infty}{Lim} (S_2(\bar{p}))^M = \underset{M \to \infty}{Lim} (Prob \{ X \neq Y \})^M .$$

X, Y can correspond two independent trials (assuming values { 1, 2, 3,…, 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

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

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

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,..., -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 symmetrix 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 symmetrix 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 $\begin{bmatrix} 0 & \alpha \\ \alpha & 0 \end{bmatrix}$ $with\ \alpha\ being\ positive$. 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

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

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

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

$Case\ (ii)$: 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 $'q'$ $independent\ identically\ distributed\ random\ variables\ i.e. X_1, X_2, \ldots, X_q$ $and\ \ N = q$. Consider the following quantity:

$$\sum_{i=1}^{q} Prob(\, X_1 = i\,)\, (\, 1 - Prob(\, X_2 = i, X_3 = i, \ldots, X_q = i\,))\; = \; \sum_{i=1}^{q} p_i\, (1 - p_i^{q-1}) = \left(1 - \sum_{i=1}^{N} p_i^q\right)$$

$$= \frac{q-1}{k}\, S_q(\bar{p}\,).$$

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

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

Note: It readily follows that $Trace(\,\bar{G}\,)$ 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

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

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 Renyii entropy is approximated by Tsallis Entropy under some conditions.

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

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

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

$$\log(1-r) \approx -r \qquad for\ |r| < 1.$$

.

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

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

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

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

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

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

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

Note: The above approach of approximating entropy (such as Shannon Entropy) was first proposed in [Rama2], [Rama4]. Specfically Shannon Entropy is approximated by Tsallis Entropy for a linear approximation i.e. $\log(1-r) \approx -r \quad for\ |r| < 1$. It is shown that higher order approximations are different from Tsallis entropy except in the case of approximation.

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

The higher order polynomial approximation of $\log(1-r)$ leads to interesting entropy functions. We are currently deriving those polynomials approximating Renyi entropy.

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

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

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

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

Using the definition of Tsallis entropy, we have

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

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

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

- 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

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

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

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

Note: Entropic functions are associated with a probability mass function. We are naturally led to the idea of associating a single probability value with a probability mass function. The following discussion proposes one such approach.

- **Single Probability Representing a Probability Mass Function: Renyi Entropy:**

From the above discussion, for $\alpha \neq 1$, we have that

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

Denoting the $L^\alpha - norm$ of the probability vector $\bar{p} = (p_1\ p_2 \ldots . p_N)^T$ by $||\bar{p}||_\alpha$, we have that (We consider integer values of α>1, where ever required in the following)

$$H_\alpha(X) = \frac{\alpha}{1-\alpha} \log(||\bar{p}||_\alpha). \ Equivalently, \quad we\ have\ that$$

$$H_\alpha(X) = \log\left(\frac{1}{(||\bar{p}||_\alpha)^{\frac{\alpha}{\alpha-1}}} \right). \ Hence$$

$$e^{-H_\alpha(X)} = (||\bar{p}||_\alpha)^{\frac{\alpha}{\alpha-1}}.$$

Now, we consider the sigmoidal function i.e. $Sigmoid(y) = \frac{1}{1+e^{-y}}$. It readily follows that

$$2\, Sigmoid(y) - 1\ represents\ a\ probability\ i.e.\ value\ lies\ between\ '0'and\ '1'.$$

$Thus, given\ any\ probability\ mass\ function\ i.e.$ vector $\bar{p} = (p_1\ p_2 \ldots . p_N)^T$,

$$Sigmoid(H_\alpha(X)) = \frac{1}{1 + (||\bar{p}||_\alpha)^{\frac{\alpha}{\alpha-1}}}.$$

$$Hence, \quad in\ view\ of\ above\ discussion, \quad the\ following\ quantity\ correspond\ to\ a\ probability\ value\ associated\ with\ the\ PMF.$$

$$2\, Sigmoid\big(H_\alpha(X)\big) - 1 = \frac{1 - (||\bar{p}||_\alpha)^{\frac{\alpha}{\alpha-1}}}{1 + (||\bar{p}||_\alpha)^{\frac{\alpha}{\alpha-1}}}.$$

From the above expression, the probability value representing a degenerate PMF as well as Uniform PMF can be readily calculated.

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 as wells as $L^p - norms\ of\ vectors\ whose\ elements\ are\ non-negative\ real\ numbers$.

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

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

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

Furthermore,

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

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

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

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.

$$Variance(Z) = Var(Z) = E(Z^2) - (E(Z))^2 \ge 0.$$

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

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

Thus, effectively we have that

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

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

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

Using Lemma 3, we have that

$$Variance(Z) \le \frac{1}{2}\left(L^2 - norm(\bar{T})\right)^2 = \frac{1}{2}\left(\sum_{i=1}^{N} p_i^{2m}\right).\ Hence$$

$$\sum_{i=1}^{N} p_i^{2m+1} \le \left(\sum_{i=1}^{N} p_i^{m+1}\right)^2 + \frac{1}{2}\left(\sum_{i=1}^{N} p_i^{2m+1}\right). \qquad Q.E.D.$$

Corollary 1: Suppose the random variable Z assumes probability values $q_i^{\prime}s$ different from $p_i^{\prime}s$.

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

$$\sum_{i=1}^{N} q_i^2\, p_i \;\geq\; \frac{1}{2}(\sum_{i=1}^{N} q_i p_i\;)^2 \,.$$

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

$$Also, \quad using\;\; Lemma\;\, 3, \quad we\;\; have\;\; that$$

$$\sum_{i=1}^{N} q_i^2\, p_i \;\leq\; (\sum_{i=1}^{N} q_i p_i\;)^2 + \frac{1}{2}\left(\sum_{i=1}^{N} q_i^2\right) \qquad Q.E.D.$$

Note: The result in Lemma 7 can be restated in terms of $L^p - norms\;\, of\;\, probability\;\, vectors$ $i.e.\;\, vectors\;(\,\bar{p}\,)\;whose\;\, elements\;\, are\;\, probabilities\;\, that\;\, add\;\, upto\;\, one$. Specifically, we have that

$$\left(\,L^{2m+1} - norm(\bar{p})\right)^{2m+1} \geq \left(\,L^{m+1} - norm(\bar{p})\right)^{2m+2}.$$

We also, have that

$$\left(\,L^{2m+1} - norm(\bar{p})\right)^{2m+1} \leq \left(\,L^{m+1} - norm(\bar{p})\right)^{2m+2} + \frac{1}{2}\;\left(\,L^{2m} - norm(\bar{p})\right)^{2m}.$$

Since vectors whose elements are non-negative real numbers can always be normalized (by the sum of their elements) to arrive at probability vectors, the above lemma leads to interesting inequaity between $L^p - norms\;\, of\, such\;\, vectors$.

Note: In terms of the Laplacian matrix $\bar{G}$, the above Lemma based inequality can be restated. Let $\bar{T} = [\,p_1^m\;\; p_2^m \ldots p_N^m\,]\;\, for\;\, a\;\, fixed\;\; integer\;\, 'm.\; We\;\, readily\;\, have\;\, that\;\, \bar{G} = \bar{D} - \tilde{P}\;\;\, and$

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

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 $\left\{\frac{1}{p_1^m}\;\;,\;\;\; \frac{1}{p_2^m}, \ldots .\frac{1}{,p_N^m}\right\}$, then using the idea in the above proof, we have that

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

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

- **Comparision with Cauchy-Schwarz (CS) Inequality: Tighter bound**

Now, we consider a function $g(.)$ of the discrete random variable $X\; i.e.\;\, Y = g(X),\;\, where\;\, Y\;\, is$ also a discrete random variable. We readily have that

$$Var(Y) \;=\; E(Y^2) - \;(\,E(Y)\,)^2 \geq 0.$$

Thus, from basic probability theory, we have that

$$E(Y^2) = \sum_{i=1}^{N} g_i^2 p_i \quad and \;\; E(Y) = \sum_{i=1}^{N} g_i\, p_i \;\; where \;\; g_i' s \;\; are \; the \; values \; assumed \; by \; Y.$$

From the above, we have that

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

But from Cauchy-Schwarz inequality (applied to the vectors $\overline{g}$, $\overline{p}$), we have that

$$\left(\sum_{i=1}^{N} g_i\, p_i \;\right)^2 \leq \left(\sum_{i=1}^{N} g_i^2 \right) \left(\sum_{i=1}^{N} p_i^2 \right).$$

We now reason that our inequality (based on the fact that Var(Z) ≥ 0) is tighter than the above inequality provided by the Cauchy-Schwarz inequality. From Lemma 4, we readily have

$$\sum_{i=1}^{N} p_i^2 \geq \frac{1}{N} \;\; with \; equality \; for \; uniform \; PMF \; i.e. \; p_i = \frac{1}{N} \;\; for \; all \; 'i'. \;\; Thus$$

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

A more direct way to reason the above inequality is by the following fact:

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

Note: In view of the above result, a tighter inequality than Cauchy-Schwarz is obtained when vector $\overline{g}$ is arbitrary, but vector $\overline{p}$ is a probability vector (It should be noted that vectors with positive components can be normalized to arrive at probability vectors. For such positive vectors, the above inequality can be suitably modified. Also, the absolute value of components of a vector leads to a non-negative/positive vector)

Also, using the Cauchy Schwarz inequality, the following inequality readily follows:

$$\left(\sum_{i=1}^{N} g_i^2 p_i \right)^2 \leq \left(\sum_{i=1}^{N} g_i^4 \right) \left(\sum_{i=1}^{N} p_i^2 \right).$$

More generally,

$$\left(\sum_{i=1}^{N} g_i^q p_i \right)^2 \leq \left(\sum_{i=1}^{N} g_i^{2q} \right) \left(\sum_{i=1}^{N} p_i^2 \right) \;\; for \; any \; integer \; 'q' \,.$$

Note: In the spirit of above lemma, inequalities in probability theory (such as Var (Z) ≥ 0)

lead to new inequalities associated with real numbers. For instance, the following is the statement of Jensen inequality: Let
$\varphi(.)$ *be a convex function and* X *be a real valued random variable. Then, we have that* $\varphi[\,E(X)\,] \leq E[\,\varphi(X)\,]$.

It is well known that $\varphi(x) = x^{2m}$ *is a convex function for any integer 'm.*

$$\left(\sum_{i=1}^{N} g_i\, p_i \right)^{2m} \leq \sum_{i=1}^{N} g_i^{2m}\, p_i.$$

Now, suppose we let $g_i = p_i^{l}$ *for any integer* $l \geq 1$.

$$\left(\sum_{i=1}^{N} p_i^{l+1} \right)^{2m} \leq \left(\sum_{i=1}^{N} p_i^{2\,l\,m+1} \right) \text{ for integers 'l' and 'm'}.$$

- Now, we consider continuous random variable and derive the continuous version of Lemma 7. As in the discrete case, the following inequality can readily be used to derive interesting inequalities associated with Tsallis entropy of continuous random variable. Details are avoided for brevity.

Lemma 8 : Consider a continuous random variable *X* with probability density function $f_X(x)$. The following inequality holds true:

$$\int_{-\infty}^{+\infty} f_X^{2m+1}(x)\; dx \;\geq \left(\int_{-\infty}^{+\infty} f_X^{m+1}(x)\; dx \right)^2 \text{for an integer } m.$$

Proof: Consider a new random variable *Y* which is a function *g(.)* of the random variable X i.e.

$$Y = g(X).$$

We readily have that $\quad Variance(Y) = E(Y^2) - (\,E(Y)\,)^2 = Var(Y) \;\geq 0.$

From basic probability tbeory, we have that

$$E\big(g(X)\big) = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\; dx\,.$$

Hence

$$\int_{-\infty}^{+\infty} (\,g(x)\,)^2\, f_X(x)\; dx \geq \left(\int_{-\infty}^{+\infty} g(x)\, f_X(x)\; dx \right)^2.$$

In the above inequality, the right hand side involves inner product of the arbitrary function $g(x)$ *and the probability density function* $f_X(x)$. Now let $g(x) = f_X^{m}(x)$ *for an arbitrary integer* m. Thus, substituting for *g(x)*, the above inequality leads to

$$\int_{-\infty}^{+\infty} f_X^{2m+1}(x)\; dx \;\geq \left(\int_{-\infty}^{+\infty} f_X^{m+1}(x)\; dx \right)^2.$$

Also, if the probability density $f_X(x)$ is bounded by one, we have the following sandwitch inequality

$$\int_{-\infty}^{+\infty} f_X^{m+1}(x)\ dx \ \geq \int_{-\infty}^{+\infty} f_X^{2m+1}(x)\, dx \ \geq \left(\int_{-\infty}^{+\infty} f_X^{m+1}(x)\ dx \right)^2 \ Q.E.D.$$

Note: It is clear that the above inequality can be generalized to non-negative functions which are integrable i.e. functions belong to the class $L^1(R)$, where R is the set of real numbers. For instance h(x) be such a function with $\int_{-\infty}^{+\infty} h(x)\ dx = S$. Then the inequality becomes

$$S \int_{-\infty}^{+\infty} h^{2m+1}(x)\ dx \ \geq \left(\int_{-\infty}^{+\infty} h^{m+1}(x)\ dx \right)^2.$$

Further, if $0 \leq h(x) < 1\ \ for\ all\ x$, then we have that

$$S^{m+1} \int_{-\infty}^{+\infty} h^{m+1}(x)\ dx \ \geq S \int_{-\infty}^{+\infty} h^{2m+1}\ (x)\, dx \ \geq \left(\int_{-\infty}^{+\infty} h^{m+1}(x)\ dx \right)^2.$$

The above sandwitch inequality can be easily restated in terms of $L^p - norms$.

Note: As in the case of discrete random variables, the inequality provided by above Lemma is tighter than the one provided by the Cauchy-Schwarz inequality for functions. Detailed derivation is avoided for brevity.

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

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

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

$$Trace(\bar{G}^2) \ = \ \sum_{i=1}^{N} p_i^2\ (1 - \ p_i\ )^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} p_i^2 p_j^2$$

$$= \sum_{i=1}^{N} p_i^2 \sum_{\substack{j=1 \\ j \neq i}}^{N} p_j^2 \ + \ \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} p_i^2 p_j^2$$ ....**FALSE**

$$= 2 \left[\sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} p_i^2 p_j^2 \right] = \ \sum_{i=1}^{N} \mu_i^2\ \ (\ with\ \ \mu_{min}\ = 0\ ).$$

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

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

Letting $[\,p_1^2\ p_2^2 \ldots .p_N^2] = \bar{V}^T,\ we\ have\ that$

$$Trace(\bar{G}^2) = \bar{V}^T\,\bar{R}\ \bar{V}\ ,$$

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

$$\bar{R} = \bar{e}\bar{e}^T - I, \qquad where\ \bar{e}\ is\ a\ column\ vector\ all\ of\ whose\ elements\ are\ 1$$

and $I\ is\ the\ identity\ matrix$. Hence, the eigenvalues of $\bar{R}$ are { (N-1), -1, -1,…, -1 } i.e. $\bar{R}$ is an indefinite matrix with -1 is an eigenvalue of multiplicity (N-1).

- In view of above result on $Trace(\bar{G}^2)$ and the earlier discussion on entropic quadratic and higher degree forms, we introduce the concept of GENERALIZED ENTROPIC FORMS. Let $\tilde{V}(m) = [\,p_1^{m+1}\ p_2^{m+1} \ldots .p_N^{m+1}\,]\ \ for\ m \geq 0.\ We\ realize\ that\ \tilde{V}(0) = \bar{p},$ vector of probabilities. Also, $\tilde{V}(1) = \bar{V},\ as\ defined\ above.\ Further, let\ us\ define\ generalized\ entropic\ forms\ \tilde{S}_m(\bar{p})\ by\ considering\ the\ quadratic\ forms\ in\ the\ vectors\ \tilde{V}(m)\ for\ m \geq 0.$ i.e.

$$\tilde{S}_m(\bar{p}) = \tilde{V}^T(m)\,\bar{R}\ \tilde{V}(m)\ \ for\ m\ \geq 0.$$

  $We\ investigate\ properties\ of\ such\ generalized\ entropic\ forms.$
  $It\ readily\ follows\ that$

$$\tilde{S}_0(\bar{p}) = S_2(p) = Trace(\bar{G}\ \ )\ i.e\ Tsallis\ \ entropy\ for\ q = 2.$$

  $Also$

$$2\,\tilde{S}_1(\bar{p}) = Trace(\bar{G}^2).$$

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

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

Lemma 9: Let $\bar{G}$ be an N x N matrix. Then Trace($\bar{G}^2$) has the following lower bound.

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

Proof: We apply Lagrange Multiplier's method to bound

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

$$Thus\ the\ objective\ function\ for\ the\ optimization\ problem\ is\ given\ by$$

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

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

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

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

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

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

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

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

we have

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

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

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

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

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

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

  Lemma 10 : The right eigenvectors $\overline{g's}$ (whose transpose are the left eigenvectors) of the variance Laplacian $\bar{G}$ that are different from the all-ones vector (i.e. $\bar{e}$ which lies in the right null space of $\bar{G}$) are such that they lie in the null space of matrix of all ones, $\bar{S}$ $i.e.$ $S_{ij} = 1\ for\ all\ i, j.$

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

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

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

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

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

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

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

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

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

Similar result can be derived based on the $L^p - norm\ of\ \bar{g}$. 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 $\bar{G}\ i.e.$

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

  Normalized eigenvectors of $\bar{G}$. It should be noted that the column vector of ALL ONEs i.e. $\bar{e} = (1\ 1 \ldots\ 1)^T\ is\ an\ eigenvector\ corresponding\ to\ the\ zero\ eigenvalue\ and$

$\frac{1}{\sqrt{N}}\ \bar{e}$ *is the associated normalized eigenvector.*

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

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

Also, we have that

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

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

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

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

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

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

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

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

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

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

$$\textit{Using Dirac notion } E(XY) = < \bar{T}, \tilde{P}\, \bar{T} >.$$

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 tha 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 bahviour, 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 evoluation of probability $q(n)$. Let the evolution of $q(n)$ be given by

$$q(n+1) = a\ q(n)(\,1 - q(n)\,), \qquad where\ a \in [0,4].$$

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

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

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

$Variance(\big(X(n)\big) = 4\, q(n)\big(1-q(n)\big)$. 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**

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

$0 < x(n) < 1$.

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

$f(y) = a\; y\left(1-y^l\right)$ for $'l'$ *an integer*. From basis calculus, it follows that the global maximum of $f(.)$ *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 $l = 1$.

Note: In [RaR], the authors consider a more generalized logistic map of the following form:

$$g\,(z) = r\, z^p\,(\,1 - z^q\,) \text{ with } z \in [0,1] \text{ and } p, q \text{ are positive valued parameters.}$$

Suppose, we constrain the parameters $\{\,p, q\,\}$ to be positive as well as negative integers. Then, it readily follows that (using sum of geometric progression)

$$g\,(z) = r(1-z)z^p\,(1 + z + z^2 + \cdots + z^{q-1}).$$

In the above expression, the geometric progression can be replaced by interesting (simple) arithmetic-geometric progression. Hence, we have an interesting generalized logistic map of the following form

$$g\,(z) = r(1-z)z^p\,(1 + 2\,z + 3\,z^2 + \cdots + q\,z^{q-1}).$$

It is expected that these generalizations of logistic map find interesting applications.

Note: We have shown that one type of generalized logistic map (based on geometric sum) has direct relationship to Tsallis entropy with integer parameter. It is very natural to associate a new entropy measure (like Tsallis entropy) using the above generalized logistic map based on the simple arithmetic-geometric progression.

- **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).

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

$$\frac{d\, f(z)}{d\, z} = f(z)\,(1 - f(z)\,).$$

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:

$$\frac{d\, f(z)}{d\, z} = a_0 + a_1\, f(z) + a_2 (\, f(z)\,)^2 + \cdots + a_M (\, f(z)\,)^M, \qquad where$$

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

Further, by constraining the coefficients i.e. a_i's to assume values { +1, -1 }, we get a group of 2^N differential equations whose solutions are of interest to us.

Thus, we are led to the following definition:

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

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

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

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.

- **Matrix Logistic Map: Matrix Verhulst Dynamical System:**

  We now introduce the concept of "Matrix Logistic Map".

  Definition: Given a constant matrix "C" and matrix Variable "A", matrix logistic map is defined as

  $$f(A) = C\, A\,(\, I - A\,),\ where\ I\ is\ the\ identity\ matrix.$$

  The associated matrix verhulst dynamical system is defined as

  $$A(n+1) = C\, A(n)\,(\, I - A(n)\,)\ for\ n \geq 0.$$

$The\ matrix\ fixed\ point, Z\ of\ above\ dynamical\ system\ is\ defined\ as$

$Z = C\, Z\, (\, I - Z\,)\ i.e.\ Z\ is\ a\ solution\ of\ the\ following\ matrix\ quadratic\ equation$

$$C\, Z^2 - C\, Z + Z\ =\ C\, Z^2 + (\, I - C\,)\, Z = \bar{0}.$$

- $Now,\ we\ briefly\ summarize\ the\ results\ associated\ with\ the\ solutions\ of\ an\ arbitrary\ matrix\ quadratic\ equation\ i.e.\ B_2 Z^2 + B_1\, Z + B_0 \equiv \bar{0}\ with\ \{\, B_2, B_1, B_0\}$ *being the coefficient matrices and Z is the unknown matrix*. The following factorization readily holds true:

$$\delta^2 B_2 + \delta\, B_1 + B_0\ =\ (\, \delta\, B_2 + B_2\, Z +\ B_0\,)\, (\, \delta\, I - Z\,).$$

$Thus,\ taking\ determinant\ on\ both\ sides,\ we\ infer\ that\ the\ eigenvalues\ of\ all\ solutions\ of\ above\ matrix\ quadratic\ equation\ must\ be\ zeroes\ of\ the\ determinental\ polynomial$

$$Det\, (\, \delta^2 B_2 + \delta\, B_1 + B_0\,)\, .$$

Further, the right vector, $\bar{f}$ of a solution matrix (of the matrix quadratic equation), 'Z' corresponding to the eigenvalue $'\mu'$ is in the right null space of the matrix

$$\mu^2 B_2 + \mu B_1 + B_0\ \ i.e. \bar{f}\, (\mu^2 B_2 + \mu B_1 + B_0) \equiv \bar{0}.$$

Now, we apply the above results to the case of structured matrix quadratic equation arising in the case of matrix logistic map of interest to us. We readily have that

$$B_2 = C\ ,\qquad B_1 = I - C\ \ and\ \ B_0 \equiv \bar{0}.$$

Thus, the associated determinental polynomial of interest to us becomes

$$Det(\, \delta^2 C + \delta(I - C)\ ) =\ Det\, \Big(\, \delta I\, \big(\, \delta\, C + (I - C)\big)\Big).$$

Suppose C matrix doesnot have an eigenvalue at '1' (one) (i.e. $(I - C)$ is non-singular), then the above determinental polynomial has 'N' (C is an N x N matrix) zeroes identically 'zero' and the rest of zeroes (i.e. N of them since the determinental polynomial is of degree 2N) are non-zero. Now let us consider the solutions of above structured matrix quadratic equation all of whose eigenvalues are non-zero (i.e. non-singular solution matrices). In the following discussion, we reason that there is a UNIQUE NON-SINGULAR matrix which is a solution of the structured matrix quadratic equation. We readily have that for such a matrix solution (with $\mu\ as\ the\ eigenvalue\ corresponding\ to\ the\ right\ eigenvector\ \bar{f}$),

$$Det(\, \mu\, C + (\, I - C\,)\,) = 0\ \ i.e\ \ (\, \mu\, C + (\, I - C\,)\,)\bar{f} \equiv \bar{0}\, .\ Thus, we\ have\ that$$

$$C\, \bar{f} = \frac{1}{(\, \mu + 1\,)}\, \bar{f}\ =\ \theta\, \bar{f}\, .$$

Hence, an eigenvalue of the unique solution matrix $'\mu'$ (i.e. matrix fixed point) is readily obtained using an eigenvalue of constant matrix $C\ i.e.\ \theta$ in the following manner

$$\mu = \frac{1 - \theta}{\theta}\, .\ Also,\qquad \bar{f}\ \ is\ a\ right\, eigenvector\ of\ matrix\ C.$$

Thus, the eigenvalues, right eigenvectors of non-singular solution matrix, Z can be readily computed. If the solution matrix, Z is also diagonalizable (i.e. sufficient condition is that the eigenvalues of matrix C are distinct), then, it can be readily computed explicitly.

Note: There is a unique, non-singular fixed point of the dynamical system associated with the matrix logistic map. Trivially $Z \equiv \overline{0}$ *is a fixed point of the dynamical system.*

Note: The non-linear behavior of dynamical system associated with the matrix logistic map can be determined using extensive numerical work.

We are naturally led to the following "**generalized matrix logistic map**".

> **Definition:** Given a constant matrix "C" and matrix Variable "A", generalized matrix logistic map is defined as
>
> $$f(A) = C\,A\,(I - A^{q-1}), \text{ where } I \text{ is the identity matrix}, \ \boldsymbol{q} \geq \mathbf{3}.$$
>
> *The associated matrix verhulst dynamical system is defined as*
>
> $$A(n+1) = C\,A(n)\,\left(I - \left(A(n)\right)^{q-1}\right) \ \ for\ \ n \geq 0.$$
>
> *The fixed points of generalized logistic map are determined using solutions of the associated matrix polynomial equation* [*Rama*3].

- ***Generalized Verhulst type dynamical system*: *Tsallis Entropy*:**

As discussed earlier, the following identity related to Tsallis entropy holds true:

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

Such an identity motivates the following non-linear dynamical systems (just like logistic map based Verhulst dynamical system) : Let $\frac{k}{q-1} = b$.

$$p_i(n+1) = \ b\ \ p_i(n)(\,1-p_i(n)\,)\big(1+p_i(n)+{p_i}^2(n)+\cdots+{p_i}^{q-2}(n)\big)\ for\ an\ integer\ l\ with$$

$$0 < x(n) < 1 \ \ for\ all\ \ n \geq 0 \ \ and \ \ 1 \leq i \leq N.$$

These dynamical systems are constrained in the sense that $\sum_{i=1}^{N} p_i\,(n) = 1\ \ for\ all\ \ 'n'$.. Thus, the above N probabilistic dynamical systems can be "connected in parallel" and after each time step their outputs are normalized using their sum (to ensure that the constraint is satisfied for the next time step (with $\sum_{i=1}^{N} p_i\,(0) = 1$).

Note: Such a networked probabilistic dynamical systems (N systems connected in parallel and followed by a normalization block) could have interesting connection to statistical mechanics (in the sense of dynamics of Tsallis entropy of associated phenomena).

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, $\bar{H}$ $i.e.$

  $\bar{H} = \bar{D} - \tilde{P}$, $where\ \bar{D} = diag\ (\ p_1, p_2 , \dots , p_N)$ i.e. a diagonal matrix and $\tilde{P}_{ij} = Probabilty\{\ X = i, \quad Y = j\ \}\ i.e.\ matrix\ of\ joint\ probabilities.$

  With such definition $\bar{H}$ need not be symmetric but still is Laplacian. Suppose, $\widetilde{P}$ 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 }.

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

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:

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

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 11 : 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 $Cos(\beta)$.

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

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

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

$$Q.E.D.$$

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.

- **Covariance Laplacian:**

  Consider two random variables X, Y which assume the same set of values (i.e. the common value vector is $\bar{T}$). Let the joint PMF matrix be $\tilde{P}$. On symmetrization of the random variables, let the PMF matrix, $\hat{P}$ be

$$\hat{P} = \left(\frac{\tilde{P} + \tilde{P}^T}{2}\right).$$

  Since, $\hat{P}$ is a symmetric matrix, the marginal PMF of X, Y random variables is same.

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

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

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

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

$$C_{xy} = \sum_{i=1}^{N}\sum_{j=1}^{N} T_i T_j\ Prob\,\{\, X = i, Y = j\,\} - \sum_{i=1}^{N}\sum_{j=1}^{N} T_{\,i} T_j\ p_i p_j$$

$$= \ \bar{T}^T\ \hat{P}\ \bar{T} - \bar{T}^T \tilde{J}\, \bar{T}, \qquad where\ \tilde{J}_{ij} = \ p_i p_j\ .$$

$$= \bar{T}^T\, (\,\hat{P} - \tilde{J}\,)\overline{T.}$$

Thus, we have a quadratic form that is also a Laplacian (since the row sums of $\hat{P}$ and $\tilde{J}$ are both the marginal probability values). We define such a Laplacian to be *Covariance Laplacian*.

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

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}.\ It\ \ readily\ follows\ that\ \left|\rho_{xy}\right| \leq 1\ \ i.e\,.\left|C_{xy}\right| \leq \sigma_x \sigma_{y.}$$

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 $(\,E(X)\,)^2,\ E(X^2)$ 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 x1 vector $\bar{T}$. Also, let $\bar{G}_x, \bar{G}_y$ be the associated N x 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) x (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

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

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

$$R_N = \bar{T}^T o\, \bar{P}\, o\, \bar{T}$$

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

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. **Approximation to Gibbs-Shannon Entropy: Information Theoretic Implications**:

In [Rama2], it was shown that "linear" approximation ($i.e. \log p_i\ is\ approximated\ by\ linear\ function$) to Gibbs-Shannon entropy of discrete random variable X leads to Tsallis Entropy (i.e. $S_q(\bar{p}) = 1 - \sum_{i=1}^{N} p_i^q$) for $q = 2.\ i.e.$

$$H(X) = -\sum_{i=1}^{N} p_i\ \log\ p_i \approx 1 - \sum_{i=1}^{N} p_i^2 = S_2(\bar{p}).$$

We would like to investigate the consequences of above discovery to research Information theory. In the earlier discussion, we have shown that

$$S_2(\bar{p}) = 1 - \sum_{i=1}^{N} p_i^2 = \left(\sum_{i=1}^{N} p_i\right)^2 - \sum_{i=1}^{N} p_i^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} p_i\, p_j = \bar{p}^T\ R\ \bar{p}.$$

Let X, Y be the discrete input and output random variables associated with a discrete memoryless channel (DMC) whose channel matrix is represented by $\bar{B}$. It is well understood that the key concept of "entropy" (in classical information theory) is utilized to define the derived concepts like "conditional entropy", "mutual information", “channel capacity” in association with the “discrete memoryless channel” (modeling noisy channel).

In view of the above discovery i.e. $H(X) \approx S_2(\bar{p})$, it is natural to approximate the derived concepts such as “conditional entropy” (associated with a discrete memoryless channel), “mutual information” etc. One important goal of such effort is to compute the “approximate channel capacity” of a discrete memoryless channel. It is well known from classical information theory that

$$H(Y|X) = -\sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)} \; i.e. \; conditional \; entropy$$

$$I(X;Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)\, p(y)} = H(Y) - H(Y|X) \; .$$

The linear approximation of mutual information $\tilde{I}\ (X{:}Y)$ is given by

$$\tilde{I}\ (X{:}Y) = S_2(Y) - S_2(Y|X) \; where$$

$$S_2(Y|X) = \sum_x p(x)\, S_2\ (Y|\, X = x) \; and \; S_2(Y) = 1 - \sum_y (\, p(y)\,)^2 \, .$$

Using linear approximation of Shannon entropy, we readily have that

$$\tilde{I}(X;Y) = \sum_i \sum_j \left(p(y_j|x_i)\right)^2 p(x_i) - \sum_k \left(p(y_k)\right)^2 ,$$

where the first sum is specified by the elements of channel matrix B and the second sum deals with output Y. Hence, we readily have that

$$\tilde{I}(X;Y) = \sum_i \sum_j p(x_i) \left(b_{ij} \right)^2 - \sum_i \sum_k (\; b_{ki} \; p(x_i) \;)^2 .$$

Maximization of mutual information over all input probability mass functions leads to channel capacity.

.

6. 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

[Rama3] G. Rama Murthy,”Transient and Equilibrium Analysis of Computer Networks: Finite Memory and Matrix Geometric Recursions”, Ph.D Thesis, Purdue University, West Lafayette, USA.

[Rama4] G. Rama Murthy and Ravi Kishore,”Approximations to Gibbs/Shannon Entropy: Entropic Polynomials,” Easy Chair Preprint # 1473

[RaR] R. Rak and E. Rak,”Route to Chaos in Generalized Logistic Map,” Available on Archive i.e

arXiv:1502.00248v1, 1st February 2015

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