Variance Laplacian: Quadratic Forms in Statistics

Rama Murthy Garimella

April 5, 2019

VARIANCE LAPLACIAN: QUADRATIC FORMS in STATISTICS

Garimella Rama Murthy*

Department of Computer Science & Engineering

*Mahindra Ecole Centrale, Hyderabad, India

ABSTRACT

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

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

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

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

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 $\bar{G}$. 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: It readily follows that the sum of two variance Laplacian matrices is also a variance Laplacian matrix.

Note: In the case of specific discrete random variables (such as Bernoulli, Poisson, Binomial etc), the associated Laplacian matrix can easily be determined. Also, if the number of values assumed by the random variables is 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*

$$\bar{G} = \begin{bmatrix} q(1-q) & -q(1-q) \\ -q(1-q) & q(1-q) \end{bmatrix}. \; The \; eigenvalues \; are \, \{ \, 0{,}2 \; (q - q^2\}. \; The$$

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

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)

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

Since $\bar{G}$ is a right circulant matrix, from linear algebra, its eigenvalues as well as eigenvectors can be explicitly determined.

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

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

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

(i) Let $\bar{e}$ be a column vector of 1's (ONES) i.e. $\bar{e} = [\, 1 \;\; 1 \;\; \dots 1 \,]^{T.}$
From Lemma 1, we have that $\bar{G}\bar{e} \equiv \bar{0}$. Hence '0' is an eigenvalue of $\bar{G}$ and the corresponding eigenvector is $\bar{e}$.

(ii) Since Variance [Z] is non-negative, we have that the quadratic form
$\bar{T}^T\bar{G}\;\bar{T} \;\; \geq 0 \;\; for\ all\ vectors\ \bar{T}$. Hence the Laplacian matrix $\bar{G}$ is a positive semi-definite matrix. Thus, all eigenvalues of $\bar{G}$ are real and non-negative.

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

- Property (iii):

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

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

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

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

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

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* $\bar{G}$ $i.e. Sp(\,\bar{G}\,) \;\leq \frac{1}{2}$. 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
  $\bar{T}^T (diag (\bar{p}) - \bar{p} \ \bar{p}^T) \bar{T} \geq 0,$ *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 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: The non-zero lower bound on Variance (Z) is given by (using μ_{min})

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

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

It readily follows that

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

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

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

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 $\{\ p_1\ ,\ p_2\ , \ldots\ ,\ p_N\ \}$ 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

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

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

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

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

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

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

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

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

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

Corollary: *We now bound the second smallest eigenvalue* , μ_2 *of* $\bar{G}$. *It is clear that* $Trace(\bar{G}) \leq \left(1 - \frac{1}{N}\right)$. *Further* $(N-1)\mu_2 \leq Trace(\bar{G})$. *Hence* $\mu_2 \leq \frac{1}{N}$. Thus, we have $\mu_2 \in \left(0, \frac{1}{N}\right]$ *and* $\mu_i \in \left[\frac{1}{N}, \frac{1}{2}\right]$ *for* $i \geq 3$. *It also readily follows that*

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

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

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

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: 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: 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.

- **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, thus readily have that

- $Trace(\bar{G}) = k\ S_2(\bar{p})$, *where $\bar{p}$ specifies the probability mass function.*

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

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

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

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

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

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

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.

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

Lemma 5: Consider probability mass function $\{p_1, p_2, \dots, p_N\}$. The following inequalities hold true:

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Now, we compute the $Trace(\bar{G}^2)$ (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

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

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

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

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

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

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

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

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

Lemma 6: 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:

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

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, *let* $\bar{g}$ *be an eigenvector of* $\bar{G}$, *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 the probability mass function of the associated discrete random variable i.e. $\{\, p_1\,,\ p_2\,, \dots\,,\ p_N\,\}$. Hence we have that

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

Also, we have that

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

(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

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

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

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

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

$It \; readily \; follows \; that \; the \; inner \; product \; E(XY) \; is \; zero \; i$.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).

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 $\{ p_1 , p_2 , \ldots , p_N \}$. In the spirit of Laplacian $\bar{G}$, we are motivated to introduce, a more general Laplacian matrix, $\bar{H}$ *i.e.*

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

$$\tilde{P}_{ij} = Probabilty\{ X = i, \quad Y = j \}\ i.e.\ matrix\ of\ joint\ probabilities.$$

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 $\frac{1}{2}(\ \tilde{P} + \check{P}^T)$.

Lemma 7: 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.

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

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

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

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

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

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

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

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

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

- **Correlation Matrix of Finitely Many Random Variables**:

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

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

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 $\bar{T}$ ($diagonal \ elements \ are \ Laplacian$

$quadratic\ \ forms\ \ where\, as\ \ other\ \ elements\ \ are\ \ not\ \ necessarily\ \ Laplacian$). $It\, is\ \ well\ \ known\ \ that\ \ R_N$ *is non-negative definite*. Using the above discussion, the correlation matrix R_N can be written as

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

$$where\ 'o'\ is\ suitably\ defined\ product\ like\ Kronecker\ or\ Schur\ product.$$

It should be noted that $\bar{P}$ is the associated block symmetric matrix of probabilities.

5. Conclusions:

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

References:

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

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

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