



Variance Laplacian: Quadratic Forms in Statistics

Rama Murthy Garimella

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

March 12, 2019

VARIANCE LAPLACIAN: QUADRATIC FORMS in STATISTICS

Garimella Rama Murthy*

Department of Computer Science & Engineering

*Mahindra Ecole Centrale, Hyderabad, India

ABSTRACT

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

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

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

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

1. Introduction:

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

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

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

2. Review of Related Literature:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Also, we have that

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

Further

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

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

Q.E.D.

Note: In the case of 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 Theory).

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

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

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 $\frac{1}{2}$.

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

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

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

Goals:

- **Goal 1:** In view of the above discovery related to the variance of a discrete random variable (i.e. Laplacian quadratic form), we would like to discover other quadratic forms which naturally arise in probability/statistics.
- **Goal 2:** Once the interesting quadratic forms are identified, the results from the theory of quadratic forms (for example, Rayleigh's Theorem) are applied to statistical/probabilistic quadratic forms. On the other hand, results related to statistical/probabilistic quadratic forms are invoked to derive new results in the theory of quadratic forms (such as inequalities between quadratic forms).
- We now derive a specific inequality associated with quadratic forms based on statistical/probabilistic quadratic forms:

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

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

But, we know that the variance of discrete random variable Z is non-negative. Hence $\bar{T}^T (\text{diag} (\bar{p}) - \bar{p} \bar{p}^T) \bar{T} \geq 0$, where $\text{diag}(\bar{p})$ 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:

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

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

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

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

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

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

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

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

$L^2 - \text{norm}$ is equal to one and the Rayleigh's Theorem can be applied to the

quadratic form based on it. Thus, it follows that

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

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

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

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

It readily follows that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $i \geq 3$. Q.E.D.

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

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

We, thus readily have that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hence, $\text{Trace}(\bar{G}^2)$ is divisible by 2.

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

Lemma 6: The right eigenvectors $\bar{g}'s$ (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, the components of all other eigenvectors sum to zero.

Also, it readily follows that $\bar{g}^T \bar{S} \bar{g} = 0$. 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 $\bar{g}'s$ lie in the null space of \bar{S} (in fact they form the basis of the null space of \bar{S}).

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

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

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

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

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

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 vectors of an arbitrary Laplacian matrix which is not necessarily a variance Laplacian matrix.

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.

4. Other Interesting Quadratic Forms in Probability/Statistics:

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

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

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

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

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

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

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

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

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

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

Lemma 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 $\text{Cos}(\beta)$.

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

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

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:

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

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

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

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

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

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

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

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

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

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

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