Journal of Advances and Scholarly Researches in Allied Education
Vol. 17, Issue No. 2, October-2020, ISSN 2230-7540

Modification in Functions of Scalar and Matrix Argument

Mamta Lata Chouhan[1]* Savita Tiwari[2]

[1] Madhyanchal Professional University, Bhopal

[2] Associate Professor, Madhyanchal Professional University, Bhopal

Abstract – Matrix functional defined over an inner-product space of square matrices are a common construct in applied mathematics. In most cases, the object of interest is not the matrix functional itself, but its derivative or gradient (if it be differentiable), and this notion is unambiguous. The Frechet derivative, see for e.g. and, being a linear functional readily yields the definition of the gradient via the Riesz Representation Theorem. However, there is a sub-class of matrix functional that frequently occurs in practice whose argument is a symmetric matrix. For instance, in the theory of elasticity and continuum thermodynamics, the stress (a second-order, symmetric tensor) is defined to be the gradient of the strain energy functional or Helmholtz potential with respect to the (symmetric) strain tensor while the strain is defined to be the gradient of the Gibbs potential with respect to the stress. Such functional and their gradients also occur in the analysis and control of dynamical systems, which are described by matrix differential equations, and maximum likelihood estimation in statistics, econometrics and machine-learning. For this sub-class of matrix functional with symmetric arguments, there seem to be two approaches to define the gradient that lead to different results.

Keyword – Matrix, Scaler, Technology, Elasticity

---------------------------------X---------------------------------

INTRODUCTION

Engineers and researchers in the field of continuum mechanics work with the definition of the Fr´echet derivative over the vector space of square matrices and specialize it to that of the symmetric matrices which are a proper subspace and then the gradient (denoted by Gsym for convenience) is obtained as described earlier. However, in the other fields named above, the tool of choice is matrix calculus, wherein a different idea emerged and has now taken hold – that of a "symmetric gradient". The root of this idea is the fact that while the space of square matrices in $R^{n\times n}$ has dimension n2, the subspace of symmetric matrices has a dimension of n(n + 1)/2.

The second approach aims to explicitly take into account the symmetry of the matrix elements, and view the matrix functional as one defined on the vector space Rn(n+1)/2, compute its gradient in this space before finally reinterpreting it as a symmetric matrix (the "symmetric gradient" Gs) in Rn×n. However, the two gradients computed, Gsym, Gs are not equal. The question raised in the title of this article refers to this dichotomy. A perusal of the literature reveals that in the two communities that dominantly used matrix calculus, that of statisticians and electrical engineers, the idea of the "symmetric gradient" came into being at around the same time. Early work in 1960s such as not make any mention of a need for special formulae to treat the case of a symmetric matrix, but does note that all the matrix elements must me functionally independent. Among statisticians, Gebhardt in 1971 seems to have been the first to remark that the derivative formulae do not consider symmetry explicitly but he concluded that no adjustment was necessary in his case since the gradient obtained was already symmetric. Tracy and Singh in 1975 echo the same sentiments as Gebhardt about the need for special formulae. By the end of the decade, the "symmetric gradient" makes its appearance in some form or the other in work of Henderson in 1979, a review by Nel and book by Rogers in 1980 and

McCulloch proves the expression for "symmetric gradient" that we quote here. By 1982, it was included in the authoritative and influential textbook by Searle.

Today the idea is firmly entrenched as evidenced by the books and the notes by Minka. In the electrical engineering community (as represented by publications in IEEE), Geering [20] in 1976 exhibited an example calculation (gradient of the determinant of a symmetric 2 x 2 matrix) to justify the definition of the "symmetric gradient". We shall show that his reasoning was flawed, and that the same flaw leads to a putative proof of the expression for Gs. Brewer in 1977 remarks that the formulae for gradient matrices in [5] can only be applied when the elements of the matrix are independent, which is not true for a symmetric matrix, and proceeds to derive the "symmetric gradient" using the rules of matrix calculus for use in sensitivity analysis of optimal estimation systems. At present, the "symmetric gradient" formula is also recorded in , a handy reference for engineers and scientists working on interdisciplinary topics with statistics and machine-learning, and the formula's appearance in shows that it is no longer restricted to a particular community of researchers.

Thus, both notions of the gradient are well-established, and hence the fact that these two notions do not agree is a source of enormous confusion for researchers who straddle application areas, a point to which the authors can emphatically attest to. On the popular site Mathematics Stack Exchange, there are multiple questions (for example) related to this theme, but their answers deepen and misguide rather than alleviate the existing confusion. Depending on the context, this disagreement between the two notions of gradient has implications that range from serious to none. In the context of extremizing a matrix functional, such as when calculating a maximum likelihood estimator, both approaches yield the same critical point. If the gradient be used in an optimization routine such as for steepest descent, one of the gradients is clearly not the steepest descent direction, and that will lead to sub-optimal convergence. Indeed, since these two are the most common contexts, the discrepancy probably escaped scrutiny until now. However, in the context of mechanics, the discrepancies are a serious issue since gradients of matrix functional are used to describe physical quantities like stress and strain in a body.

Problem formulation. To fix our notation, we introduce the following. We denote by $S^{n\times n}$ the subspace of all symmetric matrices in $R^{n\times n}$. The space $R^{n\times n}$ (and subsequently $S^{n\times n}$) is an inner product space with the following *natural* inner product h$\cdot,\cdot$i$_F$.

Definition 2.1. *For two matrices A,B in* $R^{n\times n}$

$$hA,Bi_F := \mathrm{tr}(A^TB)$$

defines an inner product and induces the Frobenius norm on $R^{n\times n}$ *via.*

$$\|A\|_F := \sqrt{\mathrm{tr}\left(A^TA\right)}$$

Corollary 2.2. *We collect a few useful facts about the inner product defined above essential for this paper.*

1. *For A symmetric, B skew-symmetric in* $R^{n\times n}$, hA,Bi$_F$ = 0

2. *If* hA,Hi$_F$ = 0 *for any H in* $S^{n\times n}$*, then the symmetric part of A given by* sym(A) := (A + A^T)/2 *is equal to* 0

3. *For A in* $R^{n\times n}$ *and H in* $S^{n\times n}$, hA,Bi$_F$ = hsym(A)*,H*i$_F$

Proof. See, for e.g. [31].

Consider a real valued function $\varphi : R^{n\times n} \longrightarrow R$. We say that φ is differentiable if its Fr´echet derivative, defined below, exists.

Definition 2.3. *The Fr´echet derivative of φ at A in* $R^{n\times n}$ *is the unique linear transformation* Dφ(A) *in* $R^{n\times n}$ *such that,*

$$\lim_{\|H\|\to 0} \frac{|\phi(A+H)-\phi(A)-\mathcal{D}\phi(A)[H]|}{\|H\|} \to 0$$

for any H in $R^{n\times n}$*. The Riesz Representation theorem then asserts the existence of the gradient* $\nabla\varphi(A)$ *in* $R^{n\times n}$ *such that*

$$h\nabla\varphi(A),Hi_F = D\varphi(A)[H]$$

Note that if A is a symmetric matrix, then by the Fr´echet derivative defined above, the gradient $\nabla\varphi(A)$ is not guaranteed to be symmetric. Also, observe that the dimension of $S^{n\times n}$ is $m = n(n+1/2)$, hence, it is natural to identify $S^{n\times n}$ with R^m. The *reduced dimension* along with the fact that Definition 2.3 doesn't account for the symmetric structure of the matrix when the argument to φ is a symmetric matrix served as a motivation to define a "constrained gradient" or "symmetric gradient" in $R^{n\times n}$ reasoned to account for the symmetry in $S^{n\times n}$.

Claim 2.4. *Let* $\varphi : R^{n\times n} \longrightarrow R$ sym $S^{n\times n}$ *and* φ_{sym} *be the real-valued function that is* *Then* G_s^{claim} *the restriction of φ to* $S^{n\times n}$*, i.e.,* $\varphi := \varphi\ S^{n\times n} \longrightarrow R$*. Let G be the gradient of φ as defined in Definition 2.3is the linear transformation in* $S^{n\times n}$ *that is claimed to be the "symmetric gradient" of* φ_{sym} *and related to the gradient G as*

follows

$$G_s^{claim}(A) = G(A) + G^T(A) - G(A) \circ I,$$

where ◦ denotes the element-wise Hadamard product of G(A) and the identity I. Theorem 3.7 in the next section will demonstrate that this claim is false. Before that, however, note that $S^{n\times n}$ is a subspace of $R^{n\times n}$ with the induced inner product in Definition 2.1. Thus, the derivative in Definition 2.3 is naturally defined for all scalar functions of symmetric matrices. The Fr´echet Derivative of φ when restricted to the subspace $S^{n\times n}$ automatically accounts for the symmetry structure. For completeness, we re-iterate the definition of Fr´echet derivative of φ restricted to the $\phi_{sym} := \phi|_{\mathbb{S}^{n\times n}}\ \mathbb{S}^{n\times n} \longrightarrow$ in $\mathbb{S}^{n\times n}$ such that subspace

S*n*×*n*.

Definition 2.5. *The Fr´echet derivative of the function* R *at* A *in* $S^{n\times n}$ *is the unique linear transformation* $D\varphi(A)$

$$\lim_{\|H\|\to 0} \frac{|\phi(A+H)-\phi(A)-\mathcal{D}\phi(A)[H]|}{\|H\|} \to 0,$$

for any H in $S^{n\times n}$. The Riesz Representation theorem then asserts the existence of the gradient $G_{sym}(A) := \nabla\varphi_{sym}(A)$ in $S^{n\times n}$ such that

$$\text{h}G_{sym}(A), H\text{i}_F = D\varphi(A)[H]$$

There is a natural relationship between the gradient in the larger space $R^{n\times n}$ and the restricted subspace $S^{n\times n}$. The following corollary states this relationship.

Corollary 2.6. *If* $G \in R^{n\times n}$ *be the gradient of* $\varphi : R^{n\times n} \longrightarrow R$, *then* $G_{sym} = \text{sym}(G)$ *is the gradient in*

$$\mathbb{S}^{n\times n} \text{ of } \phi_{sym} := \phi|_{\mathbb{S}^{n\times n}}\ \mathbb{S}^{n\times n} \longrightarrow \mathbb{R}$$

Proof. From Definition 2.3, we know that $D\varphi(A)[H] = \text{h}G(A), H\text{i}_F$ for any H in $R^{n\times n}$. If we restrict attention to H in $S^{n\times n}$, then,

$D\varphi(A)[H] = \text{h}G(A), H\text{i}_F = \text{h}\nabla\varphi_{sym}(A), H\text{i}_F$.

This is true for any H in $S^{n\times n}$, so that by Corollary 2.2 and uniqueness of the gradient,

$G_{sym}(A) = \text{sym}(G(A))$

is the gradient in $S^{n\times n}$. □

An Illustrative Example 1. This example will illustrate the difference between the gradient on $R^{n\times n}$ and $S^{n\times n}$. Fix a non-symmetric matrix A in $R^{n\times n}$ and consider a linear functional, $\varphi : R^{n\times n} \longrightarrow R$, given by $\varphi(X) = \text{tr}(A^T X)$ for any X in R*n*×*n*.

The gradient $\nabla\varphi$ in $R^{n\times n}$ is equal to A, as defined by the Fr´echet derivative Definition 2.3. However, if φ is restricted to $S^{n\times n}$, then observe that $\nabla\phi|_{\mathbb{S}^{n\times n}} = \text{sym}(A) = (A + A^T)/2$ according to Corollary 2.6 !

Thus the definition of the gradient of a real-valued function defined on $S^{n\times n}$ in Corollary 2.6 is ensured to be symmetric. We will demonstrate that Claim 2.4 is unnecessary. In fact, the correct symmetric gradient is the one given by the Fr´echet derivative in Definition 2.5, Corollary 2.6, i.e. sym(G). To do this, we first illustrate through a simple example that G^{claim}_s as defined in Claim 2.4 gives an *incorrect* gradient.

HYPERGEOMCTRIC FUNCTION WITH MATRIX ARGUMENT

Hypergeometric Functions; The noncentral X^2, noncentral F ajid multiple correlation distributions, as found by Fisher (1928), involve Bessel and hypergeometric functions which can all be written as special cases, for particular integers p and q, of the generalized hypergeometric function.

$${}_pF_q(a_1,\ldots,a_p;\ b_1,\ldots,b_q;\ x) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \cdot \frac{x^k}{k!} \quad (1)$$

where the hypergeometric coefficient $(a)_k$. is given by

$$(a)_k = a(a+1)\ldots(a+k-1) \quad (2)$$

Exponential

$${}_0F_0(x) = e^x \quad (3)$$

Binomial series

$${}_1F_0(a;\ x) = (1-x)^{-a} \quad (4)$$

Bessel (In the noncentral χ^2 distribution)

$${}_0F_1\left(\tfrac{1}{2}n;\ \tfrac{1}{4}x^2\right) = \int_0^{\pi} e^{x\cos\theta} \sin^{n-2}\theta\, d\theta \Big/ \int_0^{\pi} \sin^{n-2}\theta\, d\theta \quad (5)$$

Confluent (in the noncentral P distribution)

$${}_1F_1(a;\ b;\ x) \quad (6)$$

Gaussian hypergeometric

$${}_2F_1(a_1,\ a_2;\ b;\ x) \quad (7)$$

The corresponding multivariate distributions involve a generalization of this function to the case in which the variable x is replaced by a symmetric matrix S and P is a real or complex valued symmetric function of the latent roots of S. The hypergeometrie functions which appear in the distributions of the matrix variates are given by the Constantine (1963).

DEFINITION

$${}_pF_q(a_1,\ldots,a_p;\ b_1,\ldots,b_q;\ S) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a_1)_\kappa \cdots (a_p)_\kappa}{(b_1)_\kappa \cdots (b_q)_\kappa} \cdot \frac{C_\kappa(S)}{k!} \quad (8)$$

$$a_1,\ldots,a_p;\ b_1,\ldots,b_q$$

are real or complex constants and the multivariate hypergeometrio coefficient $(a)_K$ is given by

$$(a)_k = \prod_{i=1}^{m} \left(a - \frac{1}{2}(i-1)\right)_{k_i} \qquad (9)$$

Where, as before

$$(a)_k = a(a+1)\cdots(a+k-1) \qquad (10)$$

The latent roots distribution s involve functions of both population and sample roots, namely,

DEFINITION

$${}_pF_q(a_1,\ldots,a_p;\ b_1,\ldots,b_q;\ S,\ T)$$
$$= \sum_{k=0}^{\infty} \sum_{k} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \cdot \frac{C_k(S)C_k(T)}{C_k(I_m)k!} \qquad (11)$$

Hypergeometric functions of product ST of symmetric matrices are defined as symmetric functions of the latent roots of ST although ST may not be a symmetric matrix, its latent roots are equal to the latent roots of $S^{1/2}TS^{1/2}$, $T^{1/2}ST^{1/2}$ and TS.

Special Cases of Generalized Hypergeometric Functions of Matrix Argument;

Now special cases of the generalized hypergeometric function of matrix argument are:

$${}_0F_0(S) = e^{\mathrm{tr}\,s} \qquad (11.1)$$

$${}_1F_0(a;S) = |I - S|^{-a} \qquad (11.2)$$

$${}_0F_1\left(\frac{1}{2}n;\ \frac{1}{4}XX'\right) = \int_{O(n)} e^{\mathrm{tr}\,XH}\, dH \qquad (11.3)$$

The following integrals, which are a generalization of Laplace and inverse laplace transforms, were used by Bochner (1952) to define the Bessel function and by Herz (1956) to define the hypergeometric functions

$$\frac{1}{\Gamma_m(a)} \int_{S>0} e^{-\mathrm{tr}\,S} |S|^{a-\frac{1}{2}(m+1)}\, {}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;ST)\,dS$$
$$= {}_{p+1}F_q(a_1,\ldots,a_p,\ a;\ b_1,\ldots,b_q;\ T) \qquad (12)$$

Where $\Gamma_m(a)$ is the multivariate gamma function?

And

$$\Gamma_m(a) = \int_{S>0} e^{-\mathrm{tr}\,s} |S|^{a-\frac{1}{2}(m+1)} (dS) \qquad (13)$$

$$= \pi^{\frac{1}{4}m(m-1)} \prod_{i=1}^{m} \Gamma\left(a - \frac{1}{2}(i-1)\right) \qquad (14)$$

And

$$\frac{2^{\frac{1}{2}m(m-1)}\Gamma_m(b)}{(2\pi i)^{\frac{1}{2}m(m+1)}} \int_{R(T)-X_0>0} e^{\mathrm{tr}\,T} |T|^{-b}\, {}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;T^{-1}S)\,dT$$
$$= {}_pF_{q+1}(a_1,\ldots,a_p;\ b_1,\ldots,b_q;\ b;\ S) \qquad (15)$$

The integral is taken over all matrices $T = X_o + iY$ for fixed positive definite X_0 and Y arbitrary real symmetric.

The hypergeometric function of two variables follows from that of one, by an average over 0(m)

$${}_pF_q(a_1,\ldots,a_p;\ b_1,\ldots,b_q;\ S,\ T)$$
$$= \int_{O(m)} {}_pF_q(a_1,\ldots,a_p;\ b_1,\ldots,b_q;\ SHTH')(dH) \qquad (16)$$

The function of two variables clearly does not depend upon the order in which they occur, and it has the same properties of Laplace and inverse Laplace transform taken with respect to either variable

$$\frac{1}{\Gamma_m(a)} \int_{S>0} e^{-\mathrm{tr}\,S} |S|^{a-\frac{1}{2}(m+1)}\, {}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;ST,U)(dS)$$
$$= {}_{p+1}F_q(a_1,\ldots,a_p,\ a;\ b_1,\ldots,b_q;\ T,U) \qquad (17)$$

$$\frac{2^{\frac{1}{2}m(m-1)}\Gamma_m(b)}{(2\pi i)^{\frac{1}{2}m(m+1)}} \int_{R(T)>0} e^{\mathrm{tr}\,T} |T|^{-b}\, {}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;T^{-1}S,U)(dT)$$
$$= {}_pF_{q+1}(a_1,\ldots,a_p;\ b_1,\ldots,b_q;b;S,U) \qquad (18)$$

The power series for the function ${}_1F_0(a;S,T)$, which occurs in the distribution (18) of latent roots with unequal covariance matrices, may not converge for all requisite values of S and T, but the integral.

$${}_1F_0(a;\ -S,T) = \int_{O(m)} |I + SHTH'|^{-a}\, (dH) \qquad (19)$$

is well defined for al l S > 0 and T > 0

Further integra representations of hypergeometrie functions are given by Herz (1955), namely of ${}^1F^1$ as the moment generating function of the multivariate beta distribution.

$${}_1F_1(a;b;S) = \frac{\Gamma_m(b)}{\Gamma_m(a)\Gamma_m(b-a)} \int_0^I e^{\mathrm{tr}\,ST} |T|^{a-\frac{1}{2}(m+1)}$$
$$\cdot |I-T|^{b-a-\frac{1}{2}(m+1)} (dT) \qquad (20)$$

and $_2F_1$ as the laplace transform of this function

$$_2F_1(a_1,a_2;b;S)=\frac{\Gamma_m(b)}{\Gamma_m(a_1)\Gamma_m(b-a_1)}\int_0^I |I-ST|^{-a_2}|T|^{a_1-\frac{1}{2}(m+1)}$$

$$|I-T|^{b-a_1-\frac{1}{2}(m+1)}\,(dT) \qquad (21)$$

The hypergeometric functions of matrix argument satisfy some of the Kummer relationel. Herz (1955) gives

$$_2F_1(a_1,a_2;b;S)=|I-S|^{-a_2}\,{}_2F_1(b-a_1,a_2;b;-S(I-S)^{-1})$$

$$=|I-S|^{b-a_1-a_2}\,{}_2F_1(b-a_1,b-a_2;b;S)$$

And

$$_1F_1(a;b;S)=e^{\operatorname{tr} S}\,{}_1F_1(b-a;\ b;\ -S)$$

And also the obvious confluences

$$\lim_{a_2\to\infty}\ {}_2F_1(a_1,a_2;b;a_2^{-1}S)={}_1F_1(a_1;b;S)$$

$$\lim_{a_1\to\infty}\ {}_1F_1(a_1;b;a_1^{-1}S)={}_0F_1(b;S)$$

Probability Distribution with Matrix Argument; Here we define number of probability ' distributions of special functions with matrix argument. (i) The gamma function of matrix argument is given by

$$\int_{Z>0}(\det Z)^{\alpha-(\frac{m+1}{2})}e^{-\operatorname{tr}Z}\,dZ=\Gamma_m(\alpha)$$

then the probability distribution is given by

$$f(Z)\,dZ=\frac{(\det Z)^{\alpha-(\frac{m+1}{2})}e^{-\operatorname{tr}Z}}{\Gamma_m(\alpha)}$$

(ii) The beta function of matrix argument is given by

$$\int_0^I(\det Z)^{\alpha-(m+1)/2}\,(\det(I-Z))^{\beta-(m+1)/2}$$

$$=\frac{\Gamma_m(\alpha)\Gamma_m(\beta)}{\Gamma_m(\alpha+\beta)}$$

then the probability distribution is given by

$$f(Z)\,dZ=\frac{\Gamma_m(\alpha+\beta)}{\Gamma_m(\alpha)\Gamma_m(\beta)}(\det Z)^{\alpha-(\frac{m+1}{2})}(\det(I-Z))^{\beta-(\frac{m+1}{2})}$$

for Z > 0;

$$R(\alpha);\ R(\beta)>\frac{m+1}{2}-1$$

(iii) The H-function of matrix argument is given by

$$\int_{Z>0}(\det Z)^{\rho-(m+1)/2}\,H^{p,q}_{r,s}\left[Z\,\middle|\,\begin{matrix}(a_1,\alpha_1),\ldots,(a_r,\alpha_r)\\(b_1,\beta_1),\ldots,(b_s,\beta_s)\end{matrix}\right]dZ$$

$$=\frac{\prod_{j=1}^{p}\Gamma_m(b_j+\beta_j\rho)\prod_{j=1}^{q}\Gamma_m(\frac{m+1}{2}-a_j-\alpha_j\rho)}{\prod_{j=p+1}^{s}\Gamma_m(\frac{m+1}{2}-b_j-\beta_j\rho)\prod_{j=q+1}^{r}\Gamma_m(a_j+\alpha_j\rho)}$$

then the probability distribution is given by

$$f(Z)\,dZ=\frac{\prod_{j=p+1}^{s}\Gamma_m(\frac{m+1}{2}-b_j-\beta_j\rho)\prod_{j=q+1}^{r}\Gamma_m(a_j+\alpha_j\rho)}{\prod_{j=1}^{p}\Gamma_m(b_j+\beta_j\rho)\prod_{j=1}^{q}\Gamma_m(\frac{m+1}{2}-a_j-\alpha_j\rho)}$$

$$\cdot(\det Z)^{\rho-(\frac{m+1}{2})}\,H^{p,q}_{r,s}\left[Z\,\middle|\,\begin{matrix}(a_1,\alpha_1),\ldots,(a_r,\alpha_r)\\(b_1,\beta_1),\ldots,(b_s,\beta_s)\end{matrix}\right]$$

for $Z>0;\ R(b_j+\rho\beta_j)>\frac{(m-1)}{2}$ and $R(\frac{m+1}{2}-a_j-\rho a_j)>(\frac{1-m}{2})$

(iv) We know that

$$\int_{\Lambda>0}e^{-\operatorname{tr}\Lambda Z}\,{}_1F_1(\alpha;\beta;U\Lambda)(\det\Lambda)^{\beta-(m+1)/2}\,d\Lambda$$

$$=\Gamma_m(\beta)(\det Z)^{-\beta}\det(I-UZ^{-1})^{-\alpha}$$

then the probability distribution is given by

$$f(\Lambda)\,d\Lambda=\frac{e^{-\operatorname{tr}\Lambda Z}\,{}_1F_1(\alpha;\beta;U\Lambda)(\det\Lambda)^{\beta-(m+1)/2}}{\Gamma_m(\beta)(\det Z)^{-\beta}\det(I-UZ^{-1})^{-\alpha}}$$

For $\Lambda>0;\ R(Z)>0,\ R(Z)>R(m);\ R(\beta)>\frac{m+1}{2}-1$.

(v) We know that

$$\int_0^B(\det Z)^{\rho-(m+1)/2}(\det(I+AZ))^{-\nu}\,dZ$$

$$=\frac{\Gamma_m(\rho)\Gamma_m(\frac{m+1}{2})}{\Gamma_m(\rho+\frac{k+1}{2})}\det(B)^{\rho}\,{}_2F_1(\rho,\nu;\rho+\frac{k+1}{2};-AB)$$

then the probability distribution is given by

$$f(Z)\,dZ=\frac{\Gamma_m(\rho+\frac{k+1}{2})(\det Z)^{\rho-(m+1)/2}(\det(I+AZ))^{-\nu}}{\Gamma_m(\rho)\Gamma_m(\frac{m+1}{2})\det(B)^{\rho}\,{}_2F_1(\rho,\nu;\rho+\frac{k+1}{2};-AB)}$$

For

$$0 < Z \leq B; \quad R(\rho) > \frac{m+1}{2} - 1$$

(vi) We know that

$$\int_{Z>0} (\det(I+Z))^{\nu} (\det(I+AZ))^{\mu} (\det Z)^{\rho-1} \, dZ$$

$$= \frac{\Gamma_m(\rho)\,\Gamma_m(-\mu-\nu-\rho)}{\Gamma_m(-\mu-\nu)} \; {}_2F_1(\mu, \rho; -\mu-\nu; I-A)$$

then the probability distribution is given by

$$f(Z)\,dZ = \frac{\Gamma_m(-\mu-\nu)(\det(I+Z))^{\nu}(\det(I+AZ))^{\mu}(\det Z)^{\rho-1}}{\Gamma_m(\rho)\Gamma_m(-\mu-\nu-\rho)\,{}_2F_1(\mu,\rho;-\mu-\nu;I-A)}$$

For

$$Z > 0; \; -R(\mu+\nu) > R(\rho) > \frac{m+1}{2} - 1$$

Bivariate Probability Distribution with Matrix Argument;

Here we define number of probability distribution of special functions with matrix argument for bivariate variables, (i) We know that

(i) We know that

$$\int_{\Lambda>0}\int_{Z>0} \operatorname{etr}(-\Lambda Z)(\det Z)^{\rho-(m+1)/2} \, G^{p,q}_{r,s}\left[Z \,\middle|\, \begin{matrix} a_1,\dots,a_r \\ b_1,\dots,b_s \end{matrix}\right] dZ \, d\Lambda$$

$$= C_{k,m}(\det \Lambda)^{\rho} \, G^{p,q+1}_{r+1,s}\left[\Lambda^{-1} \,\middle|\, \begin{matrix} \frac{m+1}{2}+\rho, \; a_1,\dots,a_r \\ b_1,\dots,b_s \end{matrix}\right]$$

then the bivariate probability distribution is given by

$$f(\Lambda, Z)\,d\Lambda\,dZ = \frac{\operatorname{etr}(-\Lambda Z)(\det Z)^{\rho-(m+1)/2} \, G^{p,q}_{r,s}\left[Z \,\middle|\, \begin{matrix} a_1,\dots,a_r \\ b_1,\dots,b_s \end{matrix}\right]}{C_{k,m}(\det \Lambda)^{\rho} \, G^{p,q+1}_{r+1,s}\left[\Lambda^{-1} \,\middle|\, \begin{matrix} \frac{m+1}{2}+\rho, a_1,\dots,a_r \\ b_1,\dots,b_s \end{matrix}\right]}$$

For

$$\Lambda > 0; \; R(\rho + \min b_j) > \frac{m+1}{2} - 1; \; (j = 1,\dots,p)$$

And

$$C_{k,m} = 2^m \pi^{(m-k)/2} / \Gamma_m(k/2) = \int_{\Lambda>0} d\Lambda$$

(ii) We know that

$$\int_{Z>0}\int_{\Lambda>0} (\det Z)^{\rho-(m+1)/2} J_{\nu}(\Lambda Z)\,d\Lambda\,dZ = \frac{C_{k,m}}{2}\left(\frac{\Lambda}{2}\right)^{-A} \Gamma\left(\frac{\nu}{2}+\frac{A}{2}\right)$$

$$\left[\Gamma\left(\frac{m+1}{2}+\frac{\nu}{2}-\frac{A}{2}\right)^{-1}\right]$$

then the bivariate probability distribution is given by

$$f(Z,\Lambda)\,dZ\,d\Lambda = \frac{(\det Z)^{\rho-(m+1)/2} J_{\nu}(\Lambda Z)}{\frac{C_{k,m}}{2}\left(\frac{\Lambda}{2}\right)^{-A}\Gamma\left(\frac{\nu}{2}+\frac{A}{2}\right).\Gamma\left(\frac{m+1}{2}+\frac{\nu}{2}-\frac{A}{2}\right)^{-1}}$$

For

$$Z > 0; \; \Lambda > 0; \quad -\operatorname{Re}\nu < \operatorname{Re} A < \frac{3}{2}$$

(iii) We know that

$$\int_{Z>0}\int_{\Lambda>0} (\det Z)^{\rho-(m+1)/2} k_{\nu}(\Lambda Z)\,d\Lambda\,dZ$$

$$= C_{k,m}\, 2^{Z-2} \Lambda^{-Z}\, \Gamma\left(\frac{1}{2}Z + \frac{1}{2}\nu\right).\Gamma\left(\frac{1}{2}Z - \frac{1}{2}\nu\right)$$

then the bivariate probability distribution is given by

$$f(Z,\Lambda)\,dZ\,d\Lambda = \frac{(\det Z)^{\rho-(m+1)/2} k_{\nu}(\Lambda Z)}{C_{k,m}\, 2^{Z-2}\Lambda^{-Z}\Gamma\left(\frac{1}{2}Z+\frac{1}{2}\nu\right)\Gamma\left(\frac{1}{2}Z-\frac{1}{2}\nu\right)}$$

For

$$Z > 0; \; \Lambda > 0$$

(iv) We know that

$$\int_{Z>0}\int_{\Lambda>0} (\det Z)^{\rho-(m+1)/2} e^{-\operatorname{tr}\Lambda Z} k_{\nu}(\Lambda Z)\,d\Lambda\,dZ$$

$$= C_{k,m}\, \pi^{1/2} (2\Lambda)^{-A} \frac{\Gamma(A-\nu)\Gamma(A+\nu)}{\Gamma\left(\frac{1}{2}+A\right)}$$

Then the bivariate probability distribution is given by

$$f(Z,\Lambda)\,dZ\,d\Lambda = \frac{(\det Z)^{\rho-(m+1)/2} e^{-\operatorname{tr}\Lambda Z} k_{\nu}(\Lambda Z)}{C_{k,m}\,\pi^{1/2}(2\Lambda)^{-A}\,\Gamma(A-\nu)\Gamma(A+\nu)}.\Gamma\left(\frac{1}{2}+A\right)$$

For $Z > 0$; $H(\Lambda) > 0$.

(v) We know that

$$\int_{Z>0}\int_{\Lambda>0} (\det Z)^{\rho-(m+1)/2} \, {}_2F_1(a,b;c;-\Lambda Z)\,d\Lambda\,dZ$$

$$= C_{k,m} \frac{\Gamma(c)\Gamma(a-A)\Gamma(b-A)\Gamma(A)}{\Gamma(a)\Gamma(b)\Gamma(c-A)}$$

Then the bivariate probability distribution

$$f(Z,\Lambda)\,dZ\;d\Lambda = \frac{\Gamma(a)\Gamma(b)\Gamma(c-A)(\det Z)^{f-(m+1)/2}}{C_{k,m}\Gamma(c)\Gamma(a-A)\Gamma(b-A)\Gamma(A)} \cdot {}_2F_1(a,b;c;-\Lambda Z)$$

For

$$Z > 0;\ \Lambda > 0;\ \mathrm{Re}(A,\ a-A,\ b-A) > 0;\ c \neq 0,-1,-2,\ldots$$

CONCLUSION

In this article, we investigated the hypergeometric function of a matrix argument is a generalization of the classical hypergeometric series. It is a function defined by an infinite summation, which can be used to evaluate certain multivariate integrals. Hypergeometric functions of a matrix argument have applications in random matrix theory. For example, the distributions of the extreme eigenvalues of random matrices are often expressed in terms of the hypergeometric function of a matrix argument.

REFERENCE

[1]. ANDERSON, T.W. (1958). Introduction to multivariate statistical Analysis, New York, John Wiley and Sons.

[2]. COY, J.W. (1955). A differential calculus for functions of matrices. Doctoral Dissertation, University of Michigan,

[3]. DEDRAJ (1954). On a generalized Bessel functions Population, Ganite 3, pp. 111-115.

[4]. ERDELYI, A. et al. (1953). Higher transcendental functions. Vol. I, II, McGraw Hill, New York.

[5]. FISHER, R. A. (1924): On a distribution yielding the error functions of several well knovm statistics. Proc. Internal Math. Congress Toronto, pp. 805-813

[6]. G.S. Rogers (1980). Matrix Derivatives. Marcel Dekker, New York.

[7]. HSU, P.L. (1939). On the distribution of the roots of certain determination equations, Ann. Eugenics, 9, pp. 250-258.

[8]. Charles E. McCulloch (1980). Symmetric Matrix Derivatives with Applications. Journal of the American Statistical Association, Mar 1980.

[9]. S.R. Searle. Matrix Algebra for Statistics. John Wiley, New York, 1982.

[10]. David A. Harville (1997). Matrix Algebra from a Statistician's Perspective. Springer-Verlag, New York.

[11]. George Arthur Frederick Seber. A Matrix Handbook for Statisticians. John Wiley & Sons, New Jersey, 2008.

[12]. Thomas Minka (2019). Old and New Matrix Algebra Useful for Statistics, 2001. [Online; accessed 1.Nov. 2019].

[13]. H. Geering (1976). On calculating gradient matrices. IEEE Transactions on Automatic Control,21(4): pp. 615–616.

[14]. J. Brewer (1977). The gradient with respect to a symmetric matrix. IEEE Transactions on Automatic Control, 22(2): pp. 265–267.

[15]. Kaare Brandt Petersen and Michael Syskind Pedersen (2012). The Matrix Cookbook, 11. Version 20121115.

[16]. Iain Murray (2016). Differentiation of the Cholesky decomposition. ArXiv e-prints, Feb 2016.me10240 (https://math.stackexchange.com/users/66158/me10240).

[17]. Making sense of matrix derivative formula for determinant of symmetric matrix as a frechet derivative? Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/2436680 (version: 2017-09-21).

[18]. Tomka (https://math.stackexchange.com/users/118706/tomka). What is the derivative of the determinant of a symmetric positive definite matrix? Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/1981210 (version: 2017-04-13).

Corresponding Author

Mamta Lata Chouhan*

Madhyanchal Professional University, Bhopal