The Annals of Statistics
1994, Vol. 22, No. 1, 406–438

INEQUALITIES AND POSITIVE-DEFINITE FUNCTIONS ARISING FROM A PROBLEM IN MULTIDIMENSIONAL SCALING

BY ANDREAS BUJA, B.F. LOGAN, J.A. REEDS AND L.A. SHEPP

Bellcore, AT&T Bell Laboratories, AT&T Bell Laboratories and AT&T Bell Laboratories

We solve the following variational problem: Find the maximum of $E\|X-Y\|$ subject to $E\|X\|^2 \leq 1$, where X and Y are i.i.d. random n-vectors, and $\|\cdot\|$ is the usual Euclidean norm on $\mathbb{R}^n$. This problem arose from an investigation into multidimensional scaling, a data analytic method for visualizing proximity data. We show that the optimal X is unique and is (1) uniform on the surface of the unit sphere, for dimensions $n \geq 3$, (2) circularly symmetric with a scaled version of the radial density $\rho/(1-\rho^2)^{1/2}$, $0 \leq \rho \leq 1$, for $n = 2$, and (3) uniform on an interval centered at the origin, for $n = 1$ (Plackett's theorem). By proving spherical symmetry of the solution, a reduction to a radial problem is achieved. The solution is then found using the Wiener–Hopf technique for (real) $n < 3$. The results are reminiscent of classical potential theory, but they cannot be reduced to it.

Along the way, we obtain results of independent interest: for any i.i.d. random n-vectors X and Y, $E\|X-Y\| \leq E\|X+Y\|$. Further, the kernel $K_{p,\beta}(x,y) = \|x+y\|_p^\beta - \|x-y\|_p^\beta, x, y \in \mathbb{R}^n$ and $\|x\|_p = (\sum|x_i|^p)^{1/p}$, is positive-definite, that is, it is the covariance of a random field, $K_{p,\beta}(x,y) = E[Z(x)Z(y)]$ for some real-valued random process $Z(x)$, for $1 \leq p \leq 2$ and $0 < \beta \leq p \leq 2$ (but not for $\beta > p$ or $p > 2$ in general). Although this is an easy consequence of known results, it appears to be new in a strict sense.

In the radial problem, the average distance $D(r_1, r_2)$ between two spheres of radii r_1 and r_2 is used as a kernel. We derive properties of $D(r_1, r_2)$, including nonnegative definiteness on signed measures of zero integral.

1. Introduction.

1.1. *Overview.* The problem solved in this paper arose from an investigation into multidimensional scaling (MDS). MDS is a data analytic method for visualizing proximity data, that is, data consisting of observed similarities or dissimilarities between all pairs of objects of interest. (Without loss of generality, we assume the proximities are dissimilarities; similarities can be converted to dissimilarities.) MDS maps these objects to a Euclidean point configuration in such a way that interpoint distances approximate the given dissimilarities as well as possible. The point configuration is used in an exploratory fashion as a "map" of the objects.

We are concerned with a certain type of null situation where the observed proximities are totally uniformative. The interest in this problem arises from

Received April 1991; revised April 1993.
AMS 1991 *subject classifications.* Primary 26D10; secondary 62H99, 42A82, 45E10.
Key words and phrases. Multidimensional scaling, maximal expected distance, potential theory, inequalities, positive-definite functions, Wiener–Hopf technique.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.*
www.jstor.org

the general observation that many complex data analytic methods do not result in what one would intuitively consider "null output" (garbage out) when applied to null input (garbage in). Rather, null data can produce highly structured results ("garbage in, structure out"), which may be misleading to the uninitiated user of the method. MDS is just one technique that exhibits this characteristic; other examples are so-called alternating least squares (ALS) methods developed by psychometricians and the closely related alternating conditional expectation (ACE) method developed by statisticians [see Buja (1990) for ALS–ACE null analyses].

Null situations for MDS can be formalized in several ways, the simplest being the assumption that the dissimilarity data are i.i.d. random variables with no dependence on the underlying objects that are compared. It turns out that under some idealizations this is mathematically equivalent to the assumption that the dissimilarities are equal to a constant (w.l.o.g., +1). The problem of MDS under this null assumption is to find Euclidean point distributions such that the distance between two points is on the average as close to the constant +1 as possible. If we translate point configurations into probability distributions or random variables on $\mathbb{R}^n$, we ask for a random variable such that for two independent realizations X and Y the expected squared distance from +1 is minimized:

$$E[(\|X-Y\|-1)^2] = \min. \tag{1.1}$$

Up to an irrelevant scale factor, this is equivalent to

$$\frac{E\|X-Y\|}{(E\|X-Y\|^2)^{1/2}} = \max. \tag{1.2}$$

Assuming w.l.o.g. $EX = 0$, the denominator simplifies to $2E\|X\|^2$. On the other hand, the assumption $EX = 0$ is unnecessary in this version since centering decreases the denominator and leaves the numerator fixed. The problem of finding a null solution can be formulated as finding a probability law which puts two independent samples on the average as far from each other as possible, under the constraint that the average squared distance from the origin is 1, that is,

$$E\|X-Y\| = \max, \qquad E\|X\|^2 = 1. \tag{1.3}$$

The derivation of the solution to (1.3) is in two steps, both of which generate results of independent mathematical interest: We first show that, for any X and Y, i.i.d.,

$$E\|X-Y\| \le E\|X+Y\|, \tag{1.4}$$

with strict equality only for X spherically symmetric, which can be used to show that X in (1.3) must be spherically symmetric. To prove (1.4) we show in Section 2 that $\|x+y\| - \|x-y\|$ is a positive-definite kernel for $x, y \in \mathbb{R}^n$ and, more generally, that the same is true for kernels of the form

$$\|x+y\|_p^\beta - \|x-y\|_p^\beta, \qquad 1 \le p \le 2,\ 0 \le \beta \le p \tag{1.5}$$

(but not for $\beta > p$ or $p > 2$). This is apparently new in a strict sense but follows easily from well-known results.

Spherical symmetry of X in (1.3) reduces the problem to determining the one-dimensional radial distribution. The average distance $D_n(r_1, r_2)$ between two spheres is

$$D_n(r_1, r_2) = E\|r_1\theta_1 - r_2\theta_2\|, \tag{1.6}$$

where θ_1 and θ_2 are independent and uniformly distributed on the unit sphere. Problem (1.3) reduces to finding a radial law (on the nonnegative reals) such that, for two independent realizations R_1 and R_2,

$$E\, D_n(R_1, R_2) = \max, \qquad E\, R_1^2 = 1. \tag{1.7}$$

An important intermediate step is to show that the kernel $D_n(r_1, r_2)$ is negative-definite on measures of zero integral. In Section 3 we transform the problem to a Wiener–Hopf problem and solve it explicitly. For dimension $n \geq 3$, the optimal radial distribution is a single point mass; for $n = 2$ it has the radial density $\rho/(1-\rho^2)^{1/2}$, $0 \leq \rho \leq 1$, scaled by a factor $\sqrt{3/2}$; for $n = 1$ the optimal distribution is uniform on an interval [this is known as Plackett's theorem, see Plackett (1947) and Moriguti (1951)]. As a side result, we give the solution to the radial problem for fractional dimensions n as well. The Wiener–Hopf technique is needed for fractional dimension $n < 3$.

These results are reminiscent of classical potential theory but they cannot be obtained by simple recourse to it. The similarities to potential theory are certainly striking: a variational problem with solutions that are qualitatively different for dimensions $n \geq 3$ and $n < 3$, and the use of the (generalized) potentials $\int D_n(r_1, r_2)\, d\mu(r_2)$ (see Theorem 3.1).

Simultaneously and independently, Mattner (1990, 1993) obtained results virtually identical to ours. The motivations of his work are purely mathematical and his methods are based on convexity and Lagrange multipliers. His and our approaches also differ in their generalizations and derivative results.

1.2. *Background on MDS and motivation for the problem.* This section is devoted to a more detailed discussion of MDS. A reader interested only in the mathematical results may skip to Section 2. Classical references to the type of MDS considered in this paper are Shepard (1962) and Kruskal (1964a, b).

In the simplest case, the input of MDS is an $N \times N$ matrix of dissimilarities $(d_{i,j})_{i,j=1,\dots,N}$, where $d_{i,i} = 0$ is generally assumed for convenience, and the matrix may or may not be symmetric. MDS constructs from $(d_{i,j})_{i,j}$ a set of N points $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1, 2, \dots, N$, in such a way that the (usually Euclidean) interpoint distances $\|\mathbf{x}_i - \mathbf{x}_j\|$ mimic the dissimilarities $d_{i,j}$ as well as possible. If the data do not form a symmetric matrix, it is clear that MDS only recovers symmetric information.

Similarity and dissimilarity data are frequently found in social sciences. They arise when all pairs of a finite set of stimuli are rated by subjects as to closeness or similarity/dissimilarity. Such stimuli may be sensory, such as colors or food tastes, or they may be perceptions of objects such as animals, public figures, countries or commercial products. In other contexts, the dissimilarity date are obtained as confusion rates, such as in the famous Rothkopf Morse code data [see, e.g., Gnanadesikan (1977), pages 46–47], where objects are considered close if they are often confused. Still another source of proximity data is the actual distance matrices between finitely many points in a metric space. If the space is some high-dimensional $\mathbb{R}^m$, MDS serves as a dimension-reduction tool that maps multivariate data from m down to n dimensions. More recent applications of MDS are in problems of graph layout where minimum path length between vertices is used as a dissimilarity measure.

Kruskal (1964a, b) proposed an implementation of MDS by way of minimization of some stress function which measures the overall discrepancy between interpoint distances and observed dissimilarities. The very simplest case of a stress function is a straightforward mean squared residual, resulting in so-called metric MDS,

$$S_{\text{met}}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{i,j=1,\dots,N} (\|\mathbf{x}_i - \mathbf{x}_j\| - d_{i,j})^2. \tag{1.8}$$

[Contrary to a pervasive misunderstanding, this is not equivalent to Torgerson –Young "classical scaling," which is based on an eigendecomposition of the doubly centered dissimilarities; (1.8) does not reduce to an eigenproblem.] Minimization of (1.8) is performed over all N-point configurations $\mathbf{x}_1, \dots, \mathbf{x}_N$ in $\mathbb{R}^n$. The problem has $N \times n$ free parameters, where $10 \le N \le 100$ and $2 \le n \le 5$ are typical. In spite of its size, the problem can be solved numerically by a steepest-descent algorithm with suitable step size heuristics [Kruskal (1964b)].

In what follows, we are concerned with the performance of MDS if the dissimilarities are totally uninformative. To this end, we model the dissimilarities as i.i.d. random variables with a common expected value, $E\,d_{i,j} = 1$, say. (A symmetrized version where $d_{i,j} = d_{j,i}$ are i.i.d., for $i < j$, leads to the same problem.) Then we minimize the expected value of the stress for a given point configuration $\mathbf{x}_1, \dots, \mathbf{x}_N$:

$$E\, S_{\text{met}}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{i,j} (\|\mathbf{x}_i - \mathbf{x}_j\| - 1)^2 + \text{const}. \tag{1.9}$$

We perform a partial minimization of the metric stress function by optimizing the overall size or scale of a given configuration, that is, we minimize $\sum_{i,j}(\|c\mathbf{x}_i - c\mathbf{x}_j\| - 1)^2$ over the scale c. The solution is trivially obtained by way of a one-parameter least squares problem:

$$\min_c \sum_{i,j} (\|c\mathbf{x}_i - c\mathbf{x}_j\| - 1)^2 = 1 - \frac{(\sum_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|)^2}{\sum_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|^2}. \tag{1.10}$$

Minimization of the expected stress function is thus equivalent to maximization of the ratio $(\sum_{i,j}\|\mathbf{x}_i - \mathbf{x}_j\|)^2/\sum_{i,j}\|\mathbf{x}_i - \mathbf{x}_j\|^2$. Rewriting the summations in terms of expectations with regard to the empirical measure $P_N = N^{-1}\Sigma_{i=1,\ldots,N}\,\delta_{\mathbf{x}_i}$, we are led to maximize

$$\frac{(E_{P_N}\|X-Y\|)^2}{E_{P_N}\|X-Y\|^2} \tag{1.11}$$

over all P_N, where X and Y are two independent replications of a random variable with distribution P_N. If we now allow the probability measures to be arbitrary rather than empirical, we obtain the original problem (1.2).

This derivation is of course purely heuristic. For a proper asymptotic justification one would have to prove that minimizing configurations $\mathbf{x}_1, \ldots, \mathbf{x}_N$ of an N-point null problem ($d_{i,j}$ i.i.d.) lead to empirical measures P_N that converge to the solution P of the continuous problem (1.2). For reasons of limited space, this will be done elsewhere. Instead, we show in the following that similar null heuristics lead to the same problem (1.2) for another variety of MDS as well. This is so-called nonmetric MDS, a method which is more commonly used in practice.

Nonmetric MDS arises from the realization that it is frequently not sensible to approximate the raw dissimilarity data $d_{i,j}$ by Euclidean interpoint distances. A suitable nonlinear transformation $f(d_{i,j})$, however, can considerably improve metric behavior and therefore the degree of approximation. For nonmetric MDS, one permits arbitrary monotone transformations f, which one estimates from the data jointly with the point configurations by way of isotonic regression. With a free transform f, a naive mean squared residual involving $\|\mathbf{x}_i - \mathbf{x}_j\|$ and $f(d_{i,j})$ no longer works since the trivial solution $f \equiv 0$ and $\mathbf{x}_i = \mathbf{x}_j$, for all i,j, achieves the minimum. The following standardization removes this artifact, resulting in the stress function usually associated with nonmetric MDS:

$$S_{\text{non}}(\mathbf{x}_1, \ldots, \mathbf{x}_N; f) = \frac{\Sigma_{i,j=1,\ldots,N}\big(\|\mathbf{x}_i - \mathbf{x}_j\| - f(d_{i,j})\big)^2}{\Sigma_{i,j=1,\ldots,N}\|\mathbf{x}_i - \mathbf{x}_j\|^2}. \tag{1.12}$$

This stress function, known as Kruskal's stress formula one, is invariant under simultaneous multiplication of the configuration points $\mathbf{x}_i$ and the transformation f with positive constants. To find an n-dimensional MDS solution for given data $(d_{i,j})_{i,j}$, one minimizes the stress function over the configuration $(\mathbf{x}_i)_i$ as well as the monotone transform f. In practice, this is done numerically by interleaving steepest descent steps on the configuration with estimation of f through isotonic regression of the current interpoint distances $\|\mathbf{x}_i - \mathbf{x}_j\|$ on the dissimilarities $d_{i,j}$. In allowing arbitrary monotone transformations, one extracts in effect only the order or rank information from the dissimilarity data, rather than their actual values. This is the major reason why nonmetric MDS is the method of choice in most social science applications. Much of the data found in these fields are ordinal at best.

We again assume a null situation, that is, $d_{i,j}$ i.i.d. This is also known as the random ranking hypothesis since nonmetric MDS extracts only rank information from the $d_{i,j}$ values [see de Leeuw and Stoop (1984) and Daws, Arabie and Hubert (1990) and the literature cited therein; Daws, Arabie and Hubert have some criticisms for this null hypothesis]. We minimize the expected stress function under the null assumption:

$$(1.13)\qquad E\, S_{\mathrm{non}}(\mathbf{x}_1,\dots,\mathbf{x}_N;f) = \frac{\sum_{i,j}(\|\mathbf{x}_i-\mathbf{x}_j\| - E\, f(d_{i,j}))^2 + \mathrm{Var}(f(d_{i,j}))}{\sum_{i,j}\|\mathbf{x}_i-\mathbf{x}_j\|^2}.$$

The minimizing transformation is a constant $f \equiv N^{-2}\sum_{i,j}\|\mathbf{x}_i-\mathbf{x}_j\|$:

$$(1.14)\qquad \begin{aligned} \min_f E\, S_{\mathrm{non}}(\mathbf{x}_1,\dots,\mathbf{x}_N;f) &= \frac{\sum_{i,j}(\|\mathbf{x}_i-\mathbf{x}_j\| - N^{-2}\sum_{k,l}\|\mathbf{x}_k-\mathbf{x}_l\|)^2}{\sum_{i,j}\|\mathbf{x}_i-\mathbf{x}_j\|^2}. \\ &= 1 - \frac{\left(\sum_{i,j}\|\mathbf{x}_i-\mathbf{x}_j\|\right)^2}{\sum_{i,j}\|\mathbf{x}_{i,j}-\mathbf{x}_j\|^2}, \end{aligned}$$

which gets us back to (1.10) and ultimately to (1.2).

1.3. *Relevance of the results for MDS.* This work originated with an unintentional "computer simulation" by the first author, whereby perfectly good data got scrambled before they were submitted to MDS with $n = 3$. The resulting three dimensional configuration looked like a discrete approximation to a uniform distribution on a sphere. Later, we heard about the same experience from several other sources. After formalizing the "simulation" in the variational problem (1.2), we were led to the hypothesis that the degenerate uniform distribution on the $(n-1)$-dimentional sphere solves (1.2). While this turned out to be correct for $n \geq 3$, the most surprising part of our results is the solution for $n = 2$: a circularly symmetric distribution that has mass everywhere in the interior of a disc with increasing density toward the periphery. This qualitative behavior was anticipated by de Leeuw and Stoop [(1984), page 397] as "distributing n points equally spaced over two or more concentric circles." They obtained their configurations by running metric MDS (1.8) with all dissimilarities set to the same value, that is, empirically minimizing (1.9). The solution for $n = 1$, a uniform distribution on an interval, is known as Plackett's theorem, so this should not have come as a surprise. For MDS with $n = 1$, this implies that a null configuration essentially puts the objects in a random order with approximately equal spacings. Again, this was anticipated by de Leeuw and Stoop [(1984), page 396], who write: "minimizing stress formula one will tend to produce equal-space-prone solutions in one dimension."

These results are relevant for MDS in more than one way. While it is generally worthwhile knowing about artifacts of complex data analytic methods in null situations, the implications of the results may reach beyond strict null data. Actual data are almost always a combination of structure and noise. If pure noise generates uniform configurations on spheres in three and higher dimensions, one should expect that the noise present in real data leads to some evidence of overall curvature in MDS solutions. The expectation is that the noise component tries to force the configuration into the approximate shape of a sphere. This may be interpreted as a contributing factor to the so-called horseshoe effect. This colorful term describes the common experience of practitioners that point configurations produced by MDS exhibit global curvature which apparently has nothing to do with the underlying domain. The horseshoe effect is an ill-understood yet frequent artifact of MDS. Clearly, it would be useful to develop diagnostic tools for measuring the degree to which noise contributes to horseshoe-shaped MDS configurations in given data. The development of such tools, however, is not the aim of the present work, which is solely concerned with the analytic solution of problems (1.2) and (1.3).

To the reader with a background in psychometric methods we should add that the horseshoe effect in the versions of MDS considered here, that is, Kruskal's metric and nonmetric scaling, is mathematically distinct from the horseshoe effect of alternating least squares (ALS) and ACE methods for additive regression and principal components. In these instances, horseshoes are a consequence of eigendecompositions, while in Kruskal's MDS horseshoes are of a mathematically more involved nature, as we show with the analysis of a crude null situation.

2. Derivation of inequalities and positive-definiteness preliminaries. We first show that the common distribution of independent X and Y for which $E\,X = 0$ and $E\,\|X\|^2 \leq 1$ and $E\,\|X - Y\|$ is a maximum is spherically symmetric, that is,

$$X = R \cdot \theta,$$

where θ is uniformly distributed on the unit sphere in $\mathbb{R}^n$ and $R \geq 0$ is independent of θ. Then we prove that, for $0 < \beta \leq p \leq 2, 1 \leq p \leq 2$,

$$E\|X - Y\|_p^\beta \leq E\|X + Y\|_p^\beta, \tag{2.1}$$

for any independent random vectors X and Y in $\mathbb{R}^n$ with the same distribution ($\|x\|_p$ denotes the l_p norm of $x \in \mathbb{R}^n$). To prove (2.1) we will show that, under the same restrictions on β and p,

$$K_{p,\beta}(x,y) = \|x + y\|_p^\beta - \|x - y\|_p^\beta \tag{2.2}$$

is a positive definite kernel, that is, for all choices of vectors $x_i \in \mathbb{R}^n$ and scalars $t_i \in \mathbb{R}$,

$$\sum_{ij} K_{p,\beta}(x_i, x_j) t_i t_j \geq 0. \tag{2.3}$$

This yields (2.1) by approximating the law of X by laws concentrated on finitely many point masses.

Finally, we give examples showing that (2.1) fails is general if $\beta > p$ or if $p > 2$.

REMARKS. The results (2.1) and (2.3) seem new and worthy of note but have been at least partially anticipated by Kakosian, Klebanov and Zinger (1989), and we obtain them easily from similar results of Lévy (1937). Alternate proofs have also been given by Zinn and his colleagues. Khmaladze pointed out that, if $W(x)$ is the integral of white noise over the rectangular prism with endpoints at 0 and at $x, Z(x) = W(x) - W(-x)$ gives a realization of a Gaussian process with covariance $K_{2,1}$. Other cases $K_{p,\beta}$ remain unrealized by direct construction.

THEOREM 2.1. *Any common distribution for independent vectors X and Y in $\mathbb{R}^n$ which maximizes $E\ \|X - Y\|$ subject to $E\ X = 0$ and to $E\ \|X\|^2 \leq 1$ is spherically symmetric.*

PROOF. The condition $E\ \|X\|^2 \leq 1$ and standard tightness theory yield the existence of a probability law for X which actually attains the maximum of $E\ \|X - Y\|$ subject to the constraints. Call that law μ. If μ is not spherically symmetric, then the characteristic function $\phi(t) = E\ e^{i(t,X)}$ is not spherically symmetric, and so there exist vectors u and t in $\mathbb{R}^n$ for which $\|t\| = \|u\|$ but $\phi(t) \neq \phi(u)$. Consider the orthogonal reflection M which maps u to t; it is given by the formula $M(x) = x - 2(m,x)m$, where (m,x) is the canonical inner product in $\mathbb{R}^n$ and $m = (u-t)/\|u-t\|$. Note $M' = M$. The characteristic function of MX is $\phi(Mt)$, so $\phi(t) \neq \phi(u) = \phi(Mt)$, so the probability law of X is not symmetric with respect to the reflection M.

We generate a contradiction by showing that the M-symmetrization of μ yields a strictly larger value of $E\|X - Y\|$ than μ. In particular, define two independent random vectors $\tilde{X}$ and $\tilde{Y}$ by applying M with probability 1/2 to X and, independently, to Y, so that $P(\tilde{X} = MX) = P(\tilde{X} = X) = 1/2$ and so on.

Since M is an involutive isometry of $\mathbb{R}^n$, we have

$$E\|\tilde{X} - \tilde{Y}\| - E\|X - Y\| = \frac{1}{2} E\big(\|MX - Y\| - \|X - Y\|\big). \tag{2.4}$$

Now let ν be the signed M-antisymmetric part of μ, namely, $\nu = \frac{1}{2}\mu - \frac{1}{2}M^{-1}\mu$, so $\int_{\mathbb{R}^n} g(x)\,d\nu = \int_{\mathbb{R}^n} \frac{1}{2}(g(x) - g(Mx))\,d\mu$. Then

$$\begin{aligned} I &= \int\int\big(\|x\| + \|y\| - \|x - y\|\big)\,d\nu(x)\,d\nu(y) \\ &= \frac{1}{2}\int\int\big(\|Mx - y\| - \|x - y\|\big)\,d\mu(x)\,d\nu(y) \\ &= \frac{1}{2}\int\int\big(\|Mx - y\| - \|x - y\|\big)\,d\mu(x)\,d\mu(y) \\ &= E\|\tilde{X} - \tilde{Y}\| - E\|X - Y\|, \end{aligned}$$

by (2.4). Since μ is assumed nonsymmetric, ν is nontrivial. Further, $\nu(\{0\}) = \frac{1}{2}\mu(\{0\}) - \frac{1}{2}\mu(M^{-1}\{0\}) = 0$, so ν has no atom at 0. By Lemma 2.2, therefore, $I > 0$, so $E\|\tilde{X} - \tilde{Y}\| > E\|X - Y\|$, contradicting the supposed optimality of μ. Hence μ is spherically symmetric. □

This proof depends on a simple strengthening of a special case of a well-known result of Lévy (1937):

LEMMA 2.2. *Suppose ν is a finite signed measure on $\mathbb{R}^n$, such that $\int_{\mathbb{R}^n}(1 + \|x\|)|d\nu(x)| < \infty$. Then*

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}(\|x\| + \|y\| - \|x - y\|)\, d\nu(x)\, d\nu(y) \geq 0,$$

and strict equality holds unless all of ν's mass is concentrated at the origin.

REMARK. Nonnegativity follows directly from Lévy (1937); only the condition for strict inequality is novel.

PROOF OF LEMMA 2.2. First check that the result holds for $n = 1$. The quantity $|x| + |y| - |x - y|$ for $x, y \in \mathbb{R}$ vanishes if $xy < 0$, and equals $2\min(|x|, |y|)$ if $xy \geq 0$, so it suffices to check that

$$I = \int_0^\infty \int_0^\infty \min(x, y)\, d\nu(x)\, d\nu(y) \geq 0,$$

with equality only if ν's mass is concentrated at 0. However, $\min(x, y) = \int_0^\infty \chi(t \leq x)\chi(t \leq y)\, dt$, so

$$I = \int_0^\infty \left(\int_t^\infty d\nu(x)\right)^2 dt = \int_0^\infty \nu([t, \infty))^2\, dt \geq 0.$$

Further, if $I = 0$, then $\nu([t, \infty)) = 0$ for almost all $t \geq 0$, which implies $\nu([t, \infty)) = 0$ for all $t > 0$.

To handle the multidimensional case $n > 1$, note that $\|x\| = E|(x, Z)|$ for some spherically symmetrically distributed random vector Z, for instance, an appropriately scaled spherical Gaussian. Then

$$\begin{aligned} I &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n}(\|x\| + \|y\| - \|x - y\|)\, d\nu(x)\, d\nu(y) \\ &= E\int_{-\infty}^\infty \int_{-\infty}^\infty (|x| + |y| - |x - y|)\, d\nu_Z(x)\, d\nu_Z(y) \geq 0, \end{aligned}$$

where ν_Z is the one-dimensional signed measure induced by the map $x \mapsto (x, Z)$. Further, $I = 0$ implies that for almost all Z the one-dimensional measures ν_Z are concentrated at the origin. This then implies that ν is concentrated at the origin. □

A similar method establishes the positive-definiteness of the functions $K_{p,\beta}$ defined in (2.2).

THEOREM 2.3. *Let $1 \leq p \leq 2$ and let $0 \leq \beta \leq p$. Then the function $K_{p,\beta}(x, y) = \|x + y\|_p^\beta - \|x - y\|_p^\beta$ is positive-definite on $\mathbb{R}^n \times \mathbb{R}^n$.*

PROOF. We need to show that, for all finite k,

$$A = \sum_{i=1}^{k} \sum_{j=1}^{k} K_{p,\beta}(x_i, x_j) t_i t_j \geq 0,$$

for any choice of the k vectors $x_i \in \mathbb{R}^n$ and of the k scalars $t_i \in \mathbb{R}$. According to Lévy (1937) or Lindenstrauss and Tzafriri [(1973), page 138], the function $\|x\|_p^\beta$ is negative-definite in the sense that

$$B = \sum_{ij} \|y_i - y_j\|_p^\beta \alpha_i \alpha_j \leq 0,$$

whenever $\sum \alpha_i = 0$. So define $\alpha_i = \operatorname{sgn}(i) t_{|i|}$ and $y_i = \operatorname{sgn}(i) x_{|i|}$, for $-k \leq i \leq k$, $i \neq 0$. Then $\sum \alpha_i = 0$, and

$$\begin{aligned} B &= \sum_{i=-k}^{k} \sum_{j=-k}^{k} \|y_i - y_j\|_p^\beta \alpha_i \alpha_j \\ &= \left(\sum_{i=1}^{k} \sum_{j=1}^{k} + \sum_{i=-k}^{-1} \sum_{j=-k}^{-1} \right) \|x_{|i|} - x_{|j|}\| t_{|i|} t_{|j|} \\ &\quad - \left(\sum_{i=1}^{k} \sum_{j=-k}^{-1} + \sum_{i=-k}^{-1} \sum_{j=1}^{k} \right) \|x_{|i|} + x_{|j|}\| t_{|i|} t_{|j|} \\ &= -2A. \end{aligned}$$

By Lévy (1937), $B \leq 0$, so $A \geq 0$. □

We next give an example to show that (2.1) fails for $p > 2$ in general.

Let $e_1, \dots, e_n$ be the n unit coordinate vectors in $\mathbb{R}^n$, define $\bar{e} = 1/n \sum_i e_i$ and set $X = e_i - \bar{e}$ with probability $1/n, i = 1, \dots, n$. Then, if X and Y are i.i.d.,

$$\begin{aligned} E\|X - Y\|_p &= \|e_1 - e_2\|_p \left(1 - \frac{1}{n}\right) = 2^{1/p} \left(1 - \frac{1}{n}\right), \\ E\|X + Y\|_p &= \|e_1 + e_2 - 2\bar{e}\|_p \left(1 - \frac{1}{n}\right) + 2\|e_1 - \bar{e}\|_p \frac{1}{n} \\ &= \left[2\left(1 - \frac{2}{n}\right)^p + (n-2)\left(\frac{2}{n}\right)^p \right]^{1/p} \left(1 - \frac{1}{n}\right) \\ &\quad + 2\left[\left(1 - \frac{1}{n}\right)^p + (n-1)\frac{1}{n^p} \right]^{1/p} \frac{1}{n}. \end{aligned}$$

For $p > 2$ fixed, as $n \to \infty$, we get, up to terms of order $o(1/n)$,

$$= 2^{1/p}\left[1 - \frac{3}{n} + o\left(\frac{1}{n}\right)\right] + \left[\frac{2}{n} + o\left(\frac{1}{n}\right)\right]$$

$$\approx 2^{1/p}\left[1 - \frac{3 - 2^{(1-1/p)}}{n}\right] > 2^{1/p}\left[1 - \frac{1}{n}\right],$$

so $E\|X - Y\|_p > E\|X + Y\|_p$ for $p > 2$, for n large enough.

We next give an example where (2.1) fails with $1 \le p \le 2$ and $\beta > p$.

Let $n = 2$ and $X = (1,0), (0,1), (-1, -1)$, each with probability $\frac{1}{3}$. Then, if Y and X are i.i.d., set

$$f(\beta) = E\|X - Y\|_p^\beta - E\|X + Y\|_p^\beta = \frac{4}{9}(1 + 2^p)^{\beta/p} - \frac{2}{9}2^\beta - \frac{1}{9}2^{\beta+\beta/p} - \frac{4}{9}.$$

Then $9f(p) = 4(1 + 2^p) - 2^{1+p} - 2^{1+p} - 4 = 0$ and

$$9f'(p) = 4\log(1 + 2^p)(1 + 2^p)\frac{1}{p} - 2\log 2 \cdot 2^p - \log 2\left(1 + \frac{1}{p}\right)2^{1+p} > 0,$$

for $1 \le p \le 2$, so that (2.1) fails for $\beta = p + \varepsilon$, with $\varepsilon > 0$ and sufficiently small.

3. Derivation of the optimal form of the radial distribution. Consider two points that are uniformly and independently distributed on the surfaces of spheres of radii r_1 and r_2, respectively, in n dimensions. The average distance (Euclidean) between the two points is given by

$$\text{(3.1)}\qquad D_n(r_1, r_2) = \frac{\int_0^\pi (\sin\theta)^{n-2}\sqrt{r_1^2 + 2r_1r_2\cos\theta + r_2^2}\,d\theta}{\int_0^\pi (\sin\theta)^{n-2}\,d\theta}, \qquad n \ge 2,$$

$$D_1(r_1, r_2) = \max(r_1, r_2).$$

The problem is to determine

$$\text{(3.2)}\qquad M_n = \sup_\mu \int_0^\infty \int_0^\infty d\mu(r_1)\,D_n(r_1, r_2)\,d\mu(r_2), \qquad n \ge 1,$$

subject to

$$d\mu(r) \ge 0, \qquad \int_0^\infty d\mu(r) = 1, \qquad \int_0^\infty r^2\,d\mu(r) \le 1.$$

In the integral in the numerator of the first of equations (3.1), factor out the larger of r_1 and r_2. Then, in terms of the hypergeometric function, we have [Erdélyi, Magnus, Oberhettinger and Tricomi (1953), page 81]

$$\text{(3.3)}\qquad D_n(r_1, r_2) = \max(r_1, r_2) \cdot F\left(-\frac{1}{2}, \frac{1-n}{2}; \frac{n}{2}; \lambda^2\right), \qquad n \ge 2,$$

where

(3.3)(i)
$$\lambda = \min\left(\frac{r_1}{r_2}, \frac{r_2}{r_1}\right)$$

and

(3.3)(ii)
$$F(a,b;c;z) = \sum_{k=0}^{\infty} \frac{(a)_k(b)_k}{(c)_k}\frac{z^k}{k!}, \qquad |z| < 1.$$

In the hypergeometric series, Pochammer's symbol is used for the ascending factorial; that is,

(3.4)
$$(x)_k = x(x+1)\cdots(x+k-1), \qquad k = 1,2,\ldots,(x)_0 = 1,$$

or, equivalently,

(3.4)(i)
$$(x)_k = \frac{\Gamma(x+k)}{\Gamma(x)}.$$

Problem (3.2) makes sense, (mathematically) when the positive integer n is replaced by any positive number ν, although the physical meaning is obscure.

Since

$$\frac{\Gamma(x+k)}{\Gamma(y+k)} \sim k^{x-y}, \qquad k \to \infty,$$

we have

(3.5)
$$\frac{(a)_k(b)_k}{(c)_k k!} \sim \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)}\frac{1}{k^{c-a-b+1}}, \qquad k \to \infty.$$

If $c > 0$ and $c - a - b > 0$, then the hypergeometric series converges for $z = 1$, the sum being

(3.6)
$$F(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}.$$

Let us define

(3.7)
$$D_\nu(r,\rho) = \max(r,\rho)\cdot\phi_\nu(\lambda^2), \qquad r,\rho \geq 0, \quad \nu > 0,$$

where

(3.7)(i)
$$\lambda = \min(r/\rho, \rho/r)$$

(3.7)(ii)
$$\phi_\nu(z) = F(-1/2, (1-\nu)/2; \nu/2; z).$$

For this hypergeometric function we have $c - a - b = \nu$. So for $\nu > 0$ it is seen from (3.3) that the function $D_\nu(r, \rho)$ is a continuous function of r and ρ. Also,

(3.8)(i) $$D_\nu(r,\rho) = D_\nu(\rho,r),$$

(3.8)(ii) $$D_\nu(ar,a\rho) = aD_\nu(r,\rho), \qquad a \geq 0,$$

(3.8)(iii) $$D_\nu(r,0) = r$$

(3.8)(iv) $$D_\nu(1,1) = \frac{\Gamma(\nu/2)\Gamma(\nu)}{\Gamma(\nu/2+1/2)\Gamma(\nu-1/2)}, \qquad \nu > 0.$$

The original integral representation holds for $\nu > 1$, that is,

$$D_\nu(r,\rho) = \frac{\int_0^\pi (\sin\theta)^{\nu-2}\sqrt{r^2+2r\rho\ \cos\theta+\rho^2}\,d\theta}{\int_0^\pi (\sin\theta)^{\nu-2}\,d\theta}, \qquad \nu > 1. \tag{3.9}$$

Hence,

$$0 \leq D_\nu(r,\rho) \leq r + \rho, \qquad \nu \geq 1.$$

In general,

$$|D_\nu(r,\rho)| \leq \{\max(r,\rho)\} \times \{\max(1,|\phi_\nu(1)|)\}, \qquad \nu > 0. \tag{3.10}$$

It is clear from the series that $\phi_\nu(x)$ is a decreasing function of $x, 0 < x < 1$, for $0 < \nu < 1$. We see $\phi_\nu(1) < 0$ for $0 < \nu < 1/2$ and

$$\phi_{1/2}(x) = \sqrt{1-x}, \qquad 0 \leq x \leq 1. \tag{3.11}$$

Hence

$$D_{1/2}(r,\rho) = |r^2 - \rho^2|^{1/2}. \tag{3.12}$$

For some positive ν_o (less than 1/2) we will have

$$\phi_{\nu_o}(1) = -1. \tag{3.13}$$

Then we have from (3.10):

(3.14)(i) $$|D_\nu(r,\rho)| \leq r + \rho, \qquad \nu \geq \nu_o,$$

(3.14)(ii) $$|D_\nu(r,\rho)| \leq |\phi_\nu(1)|\,(r+\rho), \qquad 0 < \nu < \nu_o.$$

Thus the quantity M_ν, defined by replacing the positive integer n in (3.2) by ν, has a definite value for each positive ν. In fact, since

$$D_\nu(r,\rho) \leq D_1(r,\rho)\ \max(1,\phi_\nu(1)), \qquad \nu > 0, \tag{3.15}$$

we have

$$M_\nu \leq M_1\ \max(1,\phi_\nu(1)), \qquad \nu > 0. \tag{3.16}$$

For $\nu > 1$ we have, from (3.1),

$$\phi_\nu(1) = \frac{\int_0^\pi (\sin\theta)^{\nu-2}(\cos(\theta/2) + \sin(\theta/2))\, d\theta}{\int_0^\pi (\sin\theta)^{\nu-2}\, d\theta}$$

from which it is seen that $\phi_\nu(1)$ increases to the limit $\sqrt{2}$ as $\nu \to \infty$.

In this problem the basic fact is that the kernel $D_\nu(x,y)$ is negative-definite on measures of zero integral.

RADIAL INEQUALITY. *Consider signed measures $d\mu$ satisfying*

(3.17)(i) $$\int_0^\infty |d\mu(x)| = \alpha < \infty, \qquad \int_0^\infty x|d\mu(x)| = \beta < \infty.$$

For $d\mu_1$ and $d\mu_2$ satisfying (3.17)(i), *define*

(3.17)(ii) $$\langle \mu_1, \mu_2 \rangle_\nu = \int_0^\infty \int_0^\infty d\mu_1(x)\, D_\nu(x,y)\, d\mu_2(y), \qquad \nu > 0,$$

where the double integral is absolutely convergent. Define further

(3.17)(iii) $$\int_0^\infty d\mu(x) = m_0, \qquad \int_0^\infty x\, d\mu(x) = m_1.$$

Then

(3.17)(iv) $$\langle \mu, \mu \rangle_\nu \le 2m_0 m_1, \qquad \nu > 0,$$

and the inequality is strict if both α and β in (3.17)(i) *are positive, that is, equality holds nontrivially in* (3.17)(iv) *only for a point mass at the origin.*

So, in fact, the kernel is negative-definite on measures of zero integral or of zero first moment. The proof is given in Section 7. With this result, we may establish the following theorem.

THEOREM 3.1. *Consider positive measures $d\mu$ on $[0,\infty)$ satisfying*

(3.18)(i) $$\int_0^\infty d\mu(x) = 1, \qquad \int_0^\infty x^2\, d\mu(x) = 1.$$

Suppose $d\mu_\nu$ satisfies (13.18)(i) *and in addition, for some constants A_ν and B_ν,*

(3.18)(ii) $$\int_{y=0}^\infty D_\nu(x,y)\, d\mu_\nu(y) \le A_\nu + B_\nu x^2, \qquad x \ge 0,$$

where equality holds whenever x is in the support of $d\mu_\nu(x)$. Then, for any $d\mu$ satisfying (3.18)(i),

(3.18)(iii) $$\langle \mu, \mu \rangle_\nu \le \langle \mu_\nu, \mu_\nu \rangle_\nu = M_\nu,$$

where equality holds iff

$$d\mu_\nu = d\mu. \tag{3.18)(iv}$$

PROOF. Suppose (3.18)(iv) does not hold. Then according to the radical inequality (3.17)(iv) we have

$$\langle \mu_\nu - \mu, \mu_\nu - \mu \rangle_\nu < 0. \tag{3.19}$$

Now

$$\langle \mu_\nu - \mu, \mu_\nu - \mu \rangle_\nu = \langle \mu_\nu, \mu_\nu \rangle_\nu - 2\langle \mu, \mu_\nu \rangle_\nu + \langle \mu, \mu \rangle_\nu. \tag{3.20}$$

It follows from (3.18)(i) and (ii) that

$$\langle \mu, \mu_\nu \rangle_\nu \le A_\nu + B_\nu, \tag{3.21}$$

$$\langle \mu_\nu, \mu_\nu \rangle_\nu = A_\nu + B_\nu. \tag{3.22}$$

Then, if (3.18)(iv) does not hold, we have, in view of (3.19)–(3.22),

$$\langle \mu, \mu \rangle_\nu < A_\nu + B_\nu = \langle \mu_\nu, \mu_\nu \rangle_\nu. \tag{3.23}$$

Thus the theorem is established. □

It is a curious fact, in view of the symmetry of the kernel and inequality (3.21), that if $d\mu_1$ and $d\mu_2$ satisfy the moment constraints and are allowed to have different supports, then it may be possible to have

$$\langle \mu_1, \mu_2 \rangle_\nu > M_\nu,$$

the reason being that with one measure fixed, the maximum with respect to the other is always attainable with no more than two point masses. This possibility arises when the support of $d\mu_\nu$ consists of more than two points. In the problem here, it turns out that $d\mu_\nu(t)/dt$ is a density with support $[0, \sqrt{3/\nu}]$ in case $0 < \nu < 3$. Therefore if the support of $d\mu_1$ belongs to $[0, \sqrt{3/\nu}], 0 < \nu < 3$, then

$$\langle \mu_1, \mu_\nu \rangle_\nu = \langle \mu_\nu, \mu_\nu \rangle_\nu = M_\nu, \qquad 0 < \nu < 3.$$

In particular, $d\mu_1(r)$ may be a single point mass at $r = 1$. Then the double integral can be increased by changing μ_ν. Thus, if μ_1 and μ_2 satisfy the constraints, then

$$M_\nu^* \equiv \max_{\mu_1, \mu_2} \langle \mu_1, \mu_2 \rangle_\nu \ge \max_{\mu_2} \int_0^\infty D_\nu(r, 1)\, d\mu_2(r) > M_\nu, \qquad 0 < \nu < 3.$$

Taking $d\mu_2(r)$ to consist of a point mass at the origin and another of mass a^{-2} at $r = a$, we have

$$M_\nu^* \ge \max_{a \ge 1} \{1 - a^{-2} + a^{-1}\phi_\nu(a^{-2})\}.$$

In case $\nu = 1$, the maximum over a occurs for $a = 2$, giving

$$M_1^* \geq \frac{5}{4} > M_1 = \frac{2}{\sqrt{3}}.$$

According to Theorem 3.1, the problem under consideration is solved when $d\mu_\nu \geq 0$ is found such that

$$\int_0^\infty D_\nu(r,\rho)\, d\mu_\nu(\rho) \leq A_\nu + B_\nu r^2, \qquad 0 \leq r < \infty, \tag{3.24}$$

where equality holds for all r in the support of $d\mu_\nu(r)$, with

$$\int_0^\infty d\mu_\nu(r) = 1, \qquad \int_0^\infty r^2\, d\mu_\nu(r) = 1,$$

giving

$$M_\nu = A_\nu + B_\nu. \tag{3.25}$$

We observe from the linear bounds on $|D_\nu(r,\rho)|$ [cf. (3.14)(i) and (ii)] that equality cannot hold in (3.24) for arbitrarily large r when ν is fixed. It is convenient to introduce an unknown scale factor in the problem by requiring the support of $d\mu_\nu(r)$ to belong to the interval [0, 1], and later dilate the solution to obtain unity second moment. A single point mass at $r = 1$ gives for the lower bound

$$M_\nu \geq D_\nu(1,1), \qquad \nu > 0. \tag{3.26}$$

4. The case $\nu \geq 3$. This is the simplest case, the solution being a single point mass at $r = 1$, giving

$$M_\nu = D_\nu(1,1) = \frac{\Gamma(\nu/2)\Gamma(\nu)}{\Gamma(\nu/2 + 1/2)\Gamma(\nu - 1/2)}, \qquad \nu \geq 3. \tag{4.1}$$

In order to establish this result it is sufficient to show that $D_\nu(\sqrt{x},1)$ is a concave function of x, for $0 < x < \infty$, provided $\nu \geq 3$. Then, in view of the relations

$$D_\nu(r,\rho) = \rho D_\nu(r/\rho, 1), \qquad \rho > 0, r \geq 0, \tag{4.2}$$

$$D_\nu(r,0) = r, \qquad r \geq 0, \tag{4.2)(i}$$

the function

$$\int_0^\infty D_\nu(\sqrt{x},\rho)\, d\mu(\rho)$$

will be a concave function of x, for $x > 0$ and all $d\mu$ under consideration, provided $\nu \geq 3$. Then, by establishing concavity, we obtain the following more general result.

THEOREM 4.1. *For $i = 1,2$, suppose $d\mu_i \geq 0$ and*

$$\int_0^\infty d\mu_i(t) = 1, \qquad \int_0^\infty t^2\, d\mu_i(t) \leq 1. \tag{4.3)(i}$$

Then

$$\int_0^\infty \int_0^\infty d\mu_1(r)\, D_\nu(r,\rho)\, d\mu_2(\rho) \leq D_\nu(1,1), \qquad \nu \geq 3. \tag{4.3)(ii}$$

PROOF. We have from (3.7),

$$D_\nu(\sqrt{x},1) = \phi_\nu(x), \qquad 0 \leq x \leq 1 \tag{4.4)(i}$$

$$D_\nu(\sqrt{x},1) = x^{1/2}\phi_\nu\left(\frac{1}{x}\right), \qquad x > 1. \tag{4.4)(ii}$$

We use the integral representation for the derivative of $\phi_\nu(x)$:

$$\begin{aligned}\phi_\nu'(x) &= \frac{\nu-1}{2\nu} F\left(\frac{3-\nu}{2}, \frac{1}{2}; 1+\frac{\nu}{2}; x\right)\\ &= \frac{1}{2}\frac{\Gamma(\nu/2)}{\Gamma(\nu/2-1/2)\Gamma(1/2)} \int_0^1 t^{-1/2}(1-t)^{(\nu-1)/2}(1-xt)^{(\nu-3)/2}\, dt,\end{aligned} \tag{4.5}$$

which holds for $\nu > 0$. Hence

$$\phi_\nu''(x) < 0, \qquad 0 < x < 1, \quad \nu > 3. \tag{4.6}$$

Recall that

$$\phi_\nu(x) = F\left(\frac{1-\nu}{2}, \frac{-1}{2}; \frac{\nu}{2}; x\right).$$

Hence $\phi_\nu(x)$ is a polynomial of degree $(\nu-1)/2$ in x when ν is an odd positive integer. In particular,

$$\phi_3(x) = 1 + \frac{x}{3}. \tag{4.7}$$

Therefore $D_\nu(\sqrt{x}; 1)$ is concave for $0 < x < 1$, provided $\nu \geq 3$.

As noted previously, the representation (3.1) shows that the derivative with respect to r of $D_\nu(r,1)$ is a continuous function of r ($r = 1$ is the only point in question), provided $\nu > 1$. Thus we are left only to establish that $D_\nu(\sqrt{x},1)$ is concave for $x > 1$, provided $\nu \geq 3$. In view of (4.4)(ii), this is equivalent to showing the following:

$$Z_\nu(t) = t^2\phi_\nu''(t) + t\phi_\nu'(t) - \frac{1}{4}\phi_\nu(t) < 0, \qquad 0 < t < 1, \quad \nu \geq 3. \tag{4.8}$$

Here we use the fact that $w(t) = \phi_\nu(t)$ satisfies the hypergeometric differential equation

(4.9) $$t(1-t)w'' + \left[\frac{\nu}{2} + \left(\frac{\nu}{2} - 1\right)t\right]w' - \frac{1}{4}(\nu - 1)w = 0.$$

Then we find that

(4.10) $$[\nu + (\nu - 2)t]Z_\nu(t) - 2tDE = (\nu - 2 + \nu t)\phi_\nu''(t) - \frac{\nu}{4}(1-t)\phi_\nu(t).$$

Since $\phi_\nu(t)$ is positive in (0, 1) for $\nu > 1/2$, inequality (4.8) follows from (4.6), (4.7) and (4.10). Thus Theorem 4.1 is established. □

5. The case $1 < \nu < 3$. This case is just as easily handled as the special case $\nu = 2$.

We propose to find a solution $p_\nu(\rho)$ of the integral equation

(5.1) $$\int_{\rho=0}^{1} D_\nu(r,\rho)p_\nu(\rho)\,d\rho = q_\nu(r),$$

where

(5.1)(i) $$q_\nu(r) = A_\nu + B_\nu r^2, \qquad 0 \le r \le 1,$$

and, for appropriate A_ν and B_ν, $p_\nu(\rho)$ is a probability density on (0, 1), and

(5.1)(ii) $$q_\nu(r) < A_\nu + B_\nu r^2, \qquad r > 1.$$

We have

$$D_\nu(r,\rho) = \max(r,\rho)\cdot\phi_\nu(\lambda^2), \qquad \lambda = \min\left(\frac{r}{\rho}, \frac{\rho}{r}\right).$$

Hence,

$$\frac{D_\nu(r,\rho)}{\sqrt{r\rho}} = \frac{1}{\sqrt{\lambda}}\phi_\nu(\lambda^2).$$

Then, by replacing the variables by exponentials, an integral equation on $(0,1)$ having the kernel $D_\nu(r,\rho)$ may be converted to an integral equation on $(0,\infty)$ involving a difference kernel. Thus (5.1) becomes

(5.2) $$\int_0^\infty \Phi_\nu(x-t)f_\nu(t)\,dt = g_\nu(x),$$

where

(5.2)(i) $$\Phi_\nu(t) = e^{|t|/2}F\left(-\frac{1}{2}, \frac{1-\nu}{2}; \frac{\nu}{2}; e^{-2|t|}\right),$$

(5.2)(ii) $$f_\nu(t) = e^{-3t/2}p_\nu(e^{-t}),$$

(5.2)(iii) $$g_\nu(x) = e^{x/2}q_\nu(e^{-x}) = A_\nu e^{x/2} + B_\nu e^{-3x/2}, \qquad x \ge 0.$$

Equation (5.2) is brought into a tractable form by applying the differential operator $\{(d/dx)^2 - 1/4\}$. Thus we obtain the Wiener–Hopf equation,

$$\text{(5.3)} \qquad \int_0^\infty k_\nu(x-t)f_\nu(t)\,dt = 2B_\nu e^{-3x/2}, \qquad x > 0,$$

where

$$\text{(5.3)(i)} \qquad k_\nu(x) = \left\{\left(\frac{d}{dx}\right)^2 - \frac{1}{4}\right\}\Phi_\nu(x), \qquad \nu > 1.$$

Since $\Phi_\nu(x)$ has a continuous derivative which vanishes for $x = 0$, the operator does not introduce a delta function at the origin. In fact, $k_\nu(x)$ is an even integrable function, being completely monotone for $x > 0$ when $1 < \nu < 3$.

We have

$$k_\nu(x) = \sum_{n=0}^{\infty}\{(2n+1/2)^2 - (1/2)^2\}a_n \exp[(2n-1/2)|x|],$$

where

$$a_n = \frac{(-1/2)_n(1/2-\nu/2)_n}{(\nu/2)_n n!} = \frac{\nu-1}{2\nu}\frac{(1/2)_{n-1}(3/2-\nu/2)_{n-1}}{(1+\nu/2)_{n-1}n!}, \qquad n \geq 1.$$

Then for $b = 1/2 - \nu/2$ and $c = \nu/2$, we have

$$2n(2n-1)a_n = 4\frac{(-1/2)_{n+1}(b)_n}{(c)_n(n-1)!} = -\frac{b}{c}\frac{(3/2)_{n-1}(b+1)_{n-1}}{(c+1)_{n-1}(n-1)!}, \qquad n \geq 1.$$

Hence

$$\text{(5.4)} \; k_\nu(x) = \frac{\nu-1}{\nu}\sum_{n=0}^{\infty}\frac{(3/2)_n(3/2-\nu/2)_n}{(1+\nu/2)_n n!}\exp[-(2n+3/2)|x|], \qquad \nu > 1.$$

This kernel has a bilateral Laplace transform $K_\nu(s)$, analytic in the strip $-3/2 < \operatorname{Re} s < 3/2$, which has a very neat factorization. First let us note that (5.3)(i) implies

$$\text{(5.5)} \qquad \Phi_\nu(x) = \int_{-\infty}^{\infty} k_\nu(t)e^{|x-t|/2}\,dt, \qquad \nu > 1.$$

In the case of a suitable even kernel $k(x-t)$, the Wiener–Hopf factorization leads to a representation

$$k(x) = \int_{-\infty}^{\infty} h(t)h(t-x)\,dt,$$

where $h(t)$ vanishes for negative arguments and its Laplace transform is zero-free in the right half-plane, including the imaginary axis. Here we are able to guess the representation for $k_\nu(x)$.

Consider

$$h(t) = \begin{cases} e^{-bt}\left(1-e^{-2t}\right)^{-a}, & t > 0, \\ 0, & t \le 0, \end{cases} \tag{5.6}$$

where $b > 0$ and $a < 1$ so that $h(t)$ is integrable. Then

$$h(t) = \sum_{n=0}^{\infty} \frac{(a)_n}{n!} e^{-(2n+b)t}, \qquad t > 0, \tag{5.7}$$

and

$$H(s) = \int_0^{\infty} e^{-st} h(t)\, dt = \sum_{n=0}^{\infty} \frac{(a)_n}{n!} \frac{1}{s+2n+b}, \qquad \operatorname{Re} s > -b. \tag{5.8}$$

From the beta-function representation,

$$\int_0^1 t^{x-1}(1-t)^{y-1}\, dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \tag{5.9}$$

we see by expanding $(1-t)^{y-1}$ in powers of t that

$$H(s) = \frac{1}{2}\frac{\Gamma(s/2+b/2)\Gamma(1-a)}{\Gamma(s/2+b/2+1-a)}, \qquad \operatorname{Re} s > -b. \tag{5.10}$$

Now define the even function

$$k(x) = \int_{-\infty}^{\infty} h(t)h(t-x)\, dt = \int_{-\infty}^{\infty} h(t+x)h(t)\, dt. \tag{5.11}$$

Then

$$k(x) = \int_0^{\infty} h(t)h(t+x)\, dt = \sum_{n=0}^{\infty} \frac{(a)_n}{n!} H(2n+b) e^{-(2n+b)x}, \qquad x > 0, \tag{5.12}$$

where

$$\begin{aligned} H(2n+b) &= \frac{1}{2}\frac{\Gamma(n+b)\Gamma(1-a)}{\Gamma(n+b-a+1)} \\ &= \frac{1}{2}\frac{\Gamma(b)\Gamma(1-a)}{\Gamma(b+1-a)}\frac{(b)_n}{(b+1-a)_n}. \end{aligned}$$

Hence

$$\begin{aligned} k(x) = \frac{\Gamma(b)\Gamma(1-a)}{2\Gamma(b+1-a)} \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(b+1-a)_n n!} \\ \times \exp\left[-(2n+b)|x|\right], \qquad b > 0,\ a < 1, \end{aligned} \tag{5.13}$$

and

$$\int_{-\infty}^{\infty} e^{-sx}k(x)\,dx = H(s)H(-s), \qquad -b < \operatorname{Re} s < b. \tag{5.14}$$

If we replace b by $2b$, s by $2s$ and x by $t/2$, the last result is more neatly written out as

$$\begin{aligned}\int_{-\infty}^{\infty} & e^{-st}e^{-b|t|}F(a, 2b; 2b+1-a; e^{-|t|})\,dt \\ &= \frac{\gamma^{-1}\Gamma(b+s)\Gamma(b-s)}{\Gamma(b+1-a+s)\Gamma(b+1-a-s)},\end{aligned} \tag{5.15}$$

where

$$\gamma = \frac{\Gamma(2b)}{\Gamma(2b+1-a)\Gamma(1-a)}, \qquad b > 0, a < 1, -b < \operatorname{Re} s < b.$$

Now in (5.13) if we take

$$a = (3-\nu)/2, \qquad b = 3/2,$$

and set

$$h_\nu(t) = \begin{cases} c_\nu e^{-3t/2}(1-e^{-2t})^{(\nu-3)/2}, & t > 0, \\ 0, & t \le 0, \end{cases} \tag{5.16}$$

where

$$c_\nu = \frac{2^{(3-\nu)/2}\sqrt{\Gamma(\nu)}}{\Gamma(\nu/2 - 1/2)}, \tag{5.16)(i}$$

then we will have

$$k_\nu(x) = \int_{-\infty}^{\infty} h_\nu(t)h_\nu(t-x)\,dt, \qquad \nu > 1, \tag{5.17}$$

and

$$\begin{aligned} H_\nu(s) &= 2^{(1-\nu)/2}\sqrt{\Gamma(\nu)} \\ &\times \frac{\Gamma(s/2+3/4)}{\Gamma(s/2+1/4+\nu/2)}, \qquad \operatorname{Re} s > -\frac{3}{2}, \qquad \nu > 1,\end{aligned} \tag{5.18}$$

$$K_\nu(s) = H_\nu(s)H_\nu(-s), \qquad -\frac{3}{2} < \operatorname{Re} s < \frac{3}{2}, \qquad \nu > 1. \tag{5.19}$$

Now for $1 < \nu < 3$ we may define $h_\nu^*(t)$ by requiring it to vanish for negative argument and satisfy

$$\int_0^x h_\nu^*(t)h_\nu(x-t)\,dt = e^{-3x/2}, \qquad x > 0. \tag{5.20}$$

That is, $H_\nu^*(s)$, the Laplace transform of $h_\nu^*(t)$, is defined by

$$(5.21)\qquad H_\nu(s)H_\nu^*(s) = \frac{1}{s+3/2}, \qquad 1<\nu<3.$$

Then

$$H_\nu^*(s) = \frac{2^{(\nu-1)/2}}{\sqrt{\Gamma(\nu)}} \frac{\Gamma(s/2+1/4+\nu/2)}{2(s/2+3/4)\Gamma(s/2+3/4)},$$

or

$$(5.22)\qquad H_\nu^*(s) = \frac{2^{(\nu-3)/2}}{\sqrt{\Gamma(\nu)}} \frac{\Gamma(s/2+1/4+\nu/2)}{\Gamma(s/2+7/4)}, \qquad 1<\nu<3.$$

Comparing (5.22) with (5.10) we see from (5.7) that

$$(5.23)\qquad h_\nu^*(t) = \begin{cases} c_\nu^* e^{-(\nu+1/2)t}\left(1-e^{-2t}\right)^{(1-\nu)/2}, & t>0,\\ 0, & t\le 0,\end{cases}$$

where

$$(5.23)(\text{i})\qquad c_\nu^* = \frac{2\sin[(\nu-1)\pi/2]}{\pi c_\nu}.$$

It may be shown that if the Wiener–Hopf equation with kernel $k_\nu(x-t)$, $1<\nu<3$, has a solution, then that solution is unique. This fact need not be established here, since Theorem 3.1 will serve the purpose for the restricted class of solutions. The particular equation under consideration,

$$(5.24)\qquad \int_0^\infty k_\nu(x-t)f_\nu(t)\,dt = 2B_\nu e^{-3x/2}, \qquad x>0, \quad 1<\nu<3,$$

has a simple solution.

Consider the function $e_\nu(x)$ defined by

$$(5.25)\qquad \int_0^\infty k_\nu(x-t)h_\nu^*(t)\,dt = e_\nu(x), \qquad -\infty<x<\infty, \quad 1<\nu<3.$$

The function $e_\nu(x)$ has a bilateral Laplace transform given by

$$(5.26)\qquad E_\nu(s) = K_\nu(s)H_\nu^*(s) = \frac{H_\nu(-s)}{s+3/2}, \qquad \frac{-3}{2} < \operatorname{Re} s < \frac{3}{2}.$$

It follows that $e_\nu(x)$ is also given by

$$(5.27)\qquad e_\nu(x) = \begin{cases} \int_{-\infty}^x h_\nu(-t)e^{-3(x-t)/2}\,dt, & -\infty<x<\infty,\\ H_\nu(3/2)e^{-3x/2}, & x\ge 0.\end{cases}$$

Therefore, in (5.24) we have

$$f_\nu(t) = \frac{2B_\nu}{H_\nu(3/2)} h_\nu^*(t). \tag{5.28}$$

Then using (5.5) we obtain

$$\begin{aligned} \frac{H_\nu(3/2)}{2B_\nu} g_\nu(x) &= \int_0^\infty \Phi_\nu(x-t) h_\nu^*(t)\, dt \\ &= \int_{-\infty}^\infty e_\nu(t) e^{|x-t|/2}\, dt, \qquad -\infty < x < \infty, \quad 1 < \nu < 3. \end{aligned} \tag{5.29}$$

Then from (5.27) we have

$$\begin{aligned} \int_{-\infty}^\infty e_\nu(t) e^{|x-t|/2}\, dt &= \frac{1}{2}\int_{-\infty}^x h_\nu(-t)\left(e^{|x-t|/2} + e^{-3(x-t)/2}\right) dt \\ &\quad + \int_x^\infty h_\nu(-t) e^{|x-t|/2}\, dt \\ &= \frac{1}{2} H_\nu\left(-\frac{1}{2}\right) e^{x/2} + \frac{1}{2} H_\nu\left(\frac{3}{2}\right) e^{-3x/2}, \qquad x \ge 0, \end{aligned} \tag{5.30}$$

where

$$H_\nu\left(-\frac{1}{2}\right) = 2^{(1-\nu)/2} \frac{\sqrt{\pi\Gamma(\nu)}}{\Gamma(\nu/2)} = \nu H_\nu\left(\frac{3}{2}\right), \qquad \nu > 1. \tag{5.31}$$

Therefore,

$$g_\nu(x) = \nu B_\nu e^{x/2} + B_\nu e^{-3x/2}, \qquad x \ge 0, \tag{5.32}$$

where B_ν is determined by the condition

$$\int_0^1 p_\nu(\rho)\, d\rho = 1. \tag{5.33)(i}$$

or, equivalently, by (5.33)(i) and

$$\int_0^1 D_\nu(0,\rho)\, dr = \int_0^1 \rho p_\nu(\rho)\, d\rho = A_\nu = \nu B_\nu. \tag{5.33)(ii}$$

Thus the result is

$$\int_0^1 D_\nu(r,\rho) p_\nu(\rho)\, d\rho = q_\nu(r), \tag{5.34}$$

where

$$p_\nu(\rho) = \frac{\Gamma(1/2)}{\Gamma(\nu/2)\Gamma(3/2-\nu/2)} \frac{\rho^{\nu-1}}{(1-\rho^2)^{(\nu-1)/2}}, \tag{5.34)(i}$$

$$q_\nu(r) = \frac{\Gamma(3/2)\Gamma(\nu/2+1/2)}{\Gamma(\nu/2)}\left(1 + \frac{r^2}{\nu}\right), \qquad 0 \le r \le 1. \tag{5.34)(ii}$$

It remains to show that $q_\nu(r)$ satisfies

$$q_\nu(r) < A_\nu + B_\nu r^2, \qquad r > 1. \tag{5.35}$$

We have

$$q_\nu(r) = r\int_0^1 p_\nu(\rho) F\left(-\frac{1}{2}, \frac{1}{2} - \frac{\nu}{2}; \frac{\nu}{2}; \frac{\rho^2}{r^2}\right) d\rho, \qquad r > 1. \tag{5.36}$$

Here we make a simple change of variables to obtain Bateman's integral [Erdélyi, Magnus, Oberhettinger and Tricomi (1953), page 78],

$$\begin{aligned} F(a,b;c;x) &= \frac{\Gamma(c)}{\Gamma(\lambda)\Gamma(c-\lambda)} \\ &\quad\times \int_0^1 t^{\lambda-1}(1-t)^{c-\lambda-1} F(a,b;\lambda;xt)\, dt, \qquad \operatorname{Re} c > \operatorname{Re} \lambda > 0, \end{aligned} \tag{5.37}$$

which gives

$$q_\nu(r) = rF\left(-\frac{1}{2}, \frac{1}{2} - \frac{\nu}{2}; \frac{3}{2}; r^{-2}\right), \qquad r > 1, \quad 1 < \nu < 3, \tag{5.38}$$

$$q_2(r) = \frac{1}{2}\left(1 + \frac{r^2}{2}\right)\sin^{-1}\left(\frac{1}{r}\right) + \frac{3}{4}\sqrt{r^2 - 1}, \qquad r > 1. \tag{5.38(i)}$$

We would like to show that $q_\nu(\sqrt{t})$ is a concave function of t, for $t > 1$.

Since

$$\frac{(-1/2)_n}{(3/2)_n} = \frac{-1}{4n^2 - 1},$$

we have, for $t > 1$,

$$\begin{aligned} \frac{d^2}{dt^2} q_\nu(\sqrt{t}) &= -\frac{1}{4}\sum_{n-0}^{\infty} \frac{(1/2 - \nu/2)_n}{n!} t^{-n-3/2} \\ &= -\frac{1}{4} t^{-3/2}(1 - t^{-1})^{(\nu-1)/2}. \end{aligned} \tag{5.39}$$

Thus, for $1 < \nu < 3$, not only is the first derivative of $q_\nu(r)$ continuous, the second derivative is also continuous.

As to the second moment condition, we have

$$\int_0^1 t^n p_\nu(t)\, dt = \frac{\Gamma(3/2)\Gamma(n/2 + \nu/2)}{\Gamma(\nu/2)\Gamma(n/2 + 3/2)}, \tag{5.40}$$

and

$$\frac{1}{a}\int_0^a D_\nu(r,\rho) p_\nu\left(\frac{\rho}{a}\right) d\rho = a q_\nu\left(\frac{r}{a}\right), \qquad \nu > 0. \tag{5.41}$$

Thus we take $a = \sqrt{3/\nu}$ to obtain unity second moment, and hence

$$M_\nu = \sqrt{\frac{3}{\nu}}\left(A_\nu + \frac{\nu}{3}B_\nu\right) = \sqrt{\frac{3}{\nu}}\frac{4A_\nu}{3} \tag{5.42}$$
$$= \sqrt{\frac{\pi\nu}{3}}\frac{\Gamma(\nu/2+1/2)}{\Gamma(\nu/2+1)}, \qquad 1 < \upsilon < 3.$$

6. The case $0 < \nu \le 1$. Recall that this case yields Plackett's theorem for $\nu = 1$.

The solution of the integral equation for $1 < \nu < 3$ is actually valid for $0 < \nu < 3$. Then (5.38) and (5.39), which follow from Bateman's integral, also hold for $0 < \nu < 3$. Hence (5.42) holds for $0 < \nu < 3$.

In the case $1 < \nu < 3$, the solution of the integral equation having the kernel $\Phi_\nu(x-t)$ was solved by first applying the operator $\{(d/dx)^2 - 1/4\}$ to obtain an integral equation having the integrable kernel $k_\nu(x-t)$. That method is not applicable here because $k_\nu(x)$ has a nonintegrable singularity at the origin when $0 < \nu < 1$. So the validity of the previous solution must be established by another method. It turns out in the case here that the integral equation on (0, 1) does not have a unique solution, but the solution becomes unique under the additional requirement that the resulting function $q_\nu(r)$ must satisfy

$$q_\nu(r) \le A_\nu + B_\nu r^2, \qquad r \ge 0, \tag{6.1}$$

where equality holds only for $0 \le r \le 1$. It turns out that this is equivalent to equality holding for $0 \le r \le 1$ with the additional requirement that $q_\nu(r)$ have a derivative at $r = 1$. The case $\nu = 1$ gives a simple illustration of this fact.

In case $\nu = 1$ we have

$$D_1(r,\rho) = \max(r,\rho).$$

Hence a uniform distribution (0, 1) gives

$$q_1(r) = \begin{cases} \int_0^1 D_1(r,\rho)\,d\rho = (1+r^2)/2, & 0 \le r \le 1, \\ r, & r > 1. \end{cases}$$

In this case the derivative of the right-hand side is continuous at $r = 1$, and (6.1) is satisfied for $A_1 = B_1 = 1/2$.

On the other hand, a unit point mass at $r = 1$ gives

$$q_1^*(r) = D_1(r,1) = \begin{cases} 1, & 0 \le r \le 1, \\ r, & r > 1. \end{cases}$$

Thus a mixture of a point mass at $r = 1$ with a uniform distribution on (0, 1) is the general solution of the integral equation, but since the right-hand derivative of $q_1^*(r)$ at $r = 1$ exceeds the left-hand derivative at the point, a (probability) mixture of $q_1(r)$ and $q_1^*(r)$, which assigns positive weight to the latter, will not satisfy the required inequality for r slightly larger than 1. Thus the uniform distribution is the solution and, when scaled to have unity second moment, is uniform on $(0, \sqrt{3})$.

The general integral equation, obtained after an exponential change of variables, having the kernel $\Phi_\nu(x-t)$, is solved by first applying either an integral or differential operator which gives a kernel having an index ν in the fundamental interval (1, 3). The operator is found as follows:

The factorization results, (5.17)–(5.19), hold for $\nu > 1$, showing that

$$H_{\nu+2}(s) = \frac{\sqrt{\nu(\nu+1)}}{s+\nu+1/2} H_\nu(s), \qquad \nu > 1, \tag{6.2}$$

$$K_{\nu+2}(s) = \frac{\nu(\nu+1)}{(\nu+1/2)^2 - s^2} K_\nu(s) \qquad \nu > 1. \tag{6.3}$$

Although $k_\nu(x-t)$ makes no sense as an integral kernel for $\nu < 1$, relation (5.5), namely,

$$\Phi_\nu(x) = \int_{-\infty}^{\infty} k_\nu(t) e^{|x-t|/2}\, dt, \qquad \nu > 1,$$

together with (6.3) suggests that

$$\nu(\nu+1)\Phi_\nu(x) = \left\{ \left(\nu + \frac{1}{2}\right)^2 - \left(\frac{d}{dx}\right)^2 \right\} \Phi_{\nu+2}(x), \qquad \nu > 0. \tag{6.4}$$

Indeed, (6.4) may be verified by differentiation of the power series, the second derivative of the series being absolutely convergent for $\nu > 2$. Hence, if

$$\int_{-\infty}^{\infty} f(t)\Phi_{\nu+2}(x-t)\, dt = g(x), \qquad \nu > 0, \tag{6.5)(i}$$

and the integral is absolutely convergent, then

$$\nu(\nu+1) \int_{-\infty}^{\infty} f(t)\Phi_\nu(x-t)\, dt = (\nu+1/2)^2 g(x) - g''(x), \qquad \nu > 0. \tag{6.5)(ii}$$

The corresponding result is as follows: if

$$\int_0^{\infty} D_{\nu+2}(r,\rho) p(\rho)\, d\rho = q(r), \qquad r \geq 0, \quad \nu > 0, \tag{6.6)(i}$$

and the integral is absolutely convergent, then

$$\int_0^{\infty} D_\nu(r,\rho) p(\rho)\, d\rho = q(r) - \frac{r^2 q''(r)}{\nu(\nu+1)}, \qquad \nu > 0, \quad r > 0. \tag{6.6)(ii}$$

Thus we have, from the previous solution,

$$\int_0^1 D_\nu(r,\rho)p_{\nu+2}(\rho)\,d\rho = A_{\nu+2} + \frac{(\nu-1)(\nu+2)}{\nu(\nu+1)}B_{\nu+2}r^2, \qquad 0 < r \le 1, \quad 0 < \nu < 1, \tag{6.7}$$

but since the coefficient of r^2 is negative, the required inequality obviously cannot be satisfied for sufficiently large r. In fact, it fails for r slightly larger than 1.

In order to verify that (5.4) holds for $0 < \nu < 1$, we first evaluate the integral

$$\int_0^\infty f_\nu(t)k_{\nu+2}(x-t)\,dt = w_\nu(x), \qquad 0 < \nu < 1, \tag{6.8}$$

where

$$\text{(6.8)(i)} \qquad f_\nu(t) = \frac{2\Gamma(3/2)}{\Gamma(\nu/2)\Gamma(3/2-\nu/2)}\frac{e^{-(\nu+1/2)t}}{(1-e^{-2t})^{(\nu-1)/2}}, \qquad t > 0,$$

and vanishes for negative arguments. The (bilateral) Laplace transform of $f_\nu(t)$ is given by

$$F_\nu(s) = \frac{\Gamma(3/2)}{\Gamma(\nu/2)}\frac{\Gamma(s/2+\nu/2+1/4)}{\Gamma(s/2+7/4)}. \tag{6.9}$$

Then $w_\nu(x)$ has a bilateral Laplace transform given by

$$W_\nu(s) = F_\nu(s)K_{\nu+2}(s) = F_\nu(s)H_{\nu+2}(s)H_{\nu+2}(-s). \tag{6.10}$$

Thus [cf. (5.18)]

$$W_\nu(s) = \alpha_\nu\frac{H_{\nu+2}(-s)}{(s+\nu+1/2)(s+3/2)}, \tag{6.11}$$

where

$$\text{(6.11)(i)} \qquad \alpha_\nu = 2^{(3-\nu)/2}\Gamma\left(\frac{3}{2}\right)\frac{\sqrt{\Gamma(\nu+2)}}{\Gamma(\nu/2)}.$$

Hence

(6.12)
$$
\begin{aligned}
w_\nu(x) &= \frac{\alpha_\nu}{1-\nu}\int_{-\infty}^{x} h_{\nu+2}(-t)\left(\exp\left[-\left(\nu+\frac{1}{2}\right)(x-t)\right]\right.\\
&\qquad\left. - \exp\left[\frac{-3(x-t)}{2}\right]\right) dt\\
&= \frac{\alpha_\nu}{1-\nu}H_{\nu+2}\left(\nu+\frac{1}{2}\right)\exp\left[-\left(\nu+\frac{1}{2}\right)x\right]\\
&\quad - \frac{\alpha_\nu}{1-\nu}H_{\nu+2}\left(\frac{3}{2}\right)\exp\left(\frac{-3x}{2}\right), \qquad x \geq 0,
\end{aligned}
$$

where the exponential $\exp[-(\nu+1/2)x]$ will be annihilated by the operator in (6.4).

We have from (6.8) and (5.5),

(6.13)
$$\int_0^\infty f_\nu(t)\Phi_{\nu+2}(x-t)\,dt = \int_{-\infty}^\infty w_\nu(t)e^{|x-t|/2}\,dt.$$

Then from (6.4) and (6.13) we obtain

(6.14)
$$\int_0^\infty f_\nu(t)\Phi_\nu(x-t)\,dt = g_\nu(x) = \int_{-\infty}^\infty u_\nu(t)e^{|x-t|/2}\,dt,$$

where

(6.14)(i)
$$\nu(\nu+1)u_\nu(t) = (\nu+1/2)^2 w_\nu(t) - w_\nu''(t).$$

The bilateral Laplace transform of $u_\nu(t)$ is given by

(6.15)
$$U_\nu(s) = \alpha_\nu\frac{(\nu+1/2-s)H_{\nu+2}(-s)}{\nu(\nu+1)(s+3/2)}.$$

Therefore

(6.16)
$$
\begin{aligned}
u_\nu(x) &= -\beta_\nu h_{\nu+2}(-x)\\
&\quad + (\nu+2)\beta_\nu\int_{-\infty}^{x} h_{\nu+2}(-t)\exp\left(\frac{-3(x-t)}{2}\right) dt,
\end{aligned}
$$

where

(6.16)(i)
$$\beta_\nu = \frac{\alpha_\nu}{\nu(\nu+1)}.$$

Hence [cf. (5.30)],

(6.17)
$$
\begin{aligned}
g_\nu(x) &= \frac{\beta_\nu}{2}\left\{\nu H_{\nu+2}\left(-\frac{1}{2}\right)e^{x/2} + (\nu+2)H_{\nu+2}\left(\frac{3}{2}\right)e^{-3x/2}\right\},\\
&\qquad x \geq 0,\ 0 < \nu < 1.
\end{aligned}
$$

Since [cf. (5.31)]

$$H_\nu(-1/2) = \nu H_\nu(3/2), \qquad \nu > 1,$$

we have

(6.18) $$g_\nu(x) = A_\nu e^{x/2} + B_\nu e^{-3x/2}, \qquad x \geq 0, 0 < \nu < 1,$$

where

(6.18)(i) $$A_\nu = \nu B_\nu = \frac{\Gamma(3/2)\Gamma(\nu/2+1/2)}{\Gamma(\nu/2+1)}, \qquad 0 < \nu < 1,$$

Therefore relations (5.34)–(5.42) hold for $0 < \nu < 3$.

7. Proof of the radial inequality. We prove the radial inequality (3.17)(iv). By subtracting $r + \rho$ from $D_\nu(r, \rho)$ we obtain

(7.1) $$\langle \mu_1, \mu_2 \rangle_\nu = (\mu_1, \mu_2)_\nu + m_0^{(1)} m_1^{(2)} + m_0^{(2)} m_1^{(1)},$$

where we define

(7.1)(i) $$(\mu_1, \mu_2)_\nu = \int_0^\infty \int_0^\infty d\mu_1(r)\{D_\nu(r,\rho) - r - \rho\}\, d\mu_2(\rho),$$

(7.1)(ii) $$m_n^{(i)} = \int_0^\infty r^n \, d\mu_i(r), \qquad i = 1,\, 2,\, n = 0, 1.$$

Since

$$D_\nu(r, 0) = D_\nu(0, r) = r,$$

we have for $i = 1, 2$, in case μ_o is a point mass at the origin,

(7.2)(i) $$(\mu_0, \mu_i)_\nu = (\mu_i, \mu_0)_\nu = 0.$$

Hence

(7.2)(ii) $$(\mu_1 + \mu_o, \mu_2 + \mu_o)_\nu = (\mu_1, \mu_2)_\nu.$$

Thus, in anticipating an exponential change of variables where the origin maps to $+\infty$, we may ignore any mass at the origin for the purpose of establishing (3.17). In fact, in view of the smoothness of the kernel and the relation (7.1), it is sufficient to establish that

(7.3) $$(\mu, \mu)_\nu < 0, \qquad \nu > 0,$$

for the case that

(7.3)(i) $$d\mu(r) = p(r)\, dr,$$

where $p(r)$ is a real-valued function of L_1, vanishing for negative arguments and satisfying

$$\text{(7.3)(ii)} \qquad \int_0^\infty r|p(r)|\,dr < \infty, \qquad \int_0^\infty |p(r)|\,dr > 0.$$

Since

$$\text{(7.4)} \qquad D_\nu(r,\rho) = \max(r,\rho)\cdot\phi_\nu(\lambda^2),$$

where

$$\text{(7.4)(i)} \qquad \lambda = \min\left(\frac{r}{\rho},\frac{\rho}{r}\right),$$

$$\text{(7.4)(ii)} \qquad \phi_\nu(x) = F\left(\frac{-1}{2},\frac{1-\nu}{2};\frac{\nu}{2};x\right),$$

we have

$$\text{(7.5)} \qquad \frac{D_\nu(r,\rho)-r-\rho}{\sqrt{r\rho}} = \lambda^{-1/2}\phi_\nu(\lambda^2) - (\lambda^{1/2}+\lambda^{-1/2}).$$

Then, on setting

$$r = e^{-t}, \qquad \rho = e^{-x},$$

we obtain

$$\text{(7.6)} \qquad (\mu,\mu)_\nu = \int_{-\infty}^\infty\int_{-\infty}^\infty f(t)\left\{\Phi_\nu(x-t) - 2\cosh\left[\frac{x-t}{2}\right]\right\}f(x)\,dx\,dt,$$

where

$$\text{(7.6)(i)} \qquad f(t) = e^{-3t/2}p(e^{-t}),$$

and $\Phi_\nu(t)$ is defined in (5.2(i)).

We have

$$\text{(7.7)} \qquad \Phi_\nu(t) = \sum_{n=0}^\infty \frac{(-1/2)_n(1/2-\nu/2)_n}{(\nu/2)_n n!}\exp\left[-\left(2n-\frac{1}{2}\right)|t|\right].$$

The first term, $e^{|t|/2}$, is largely cancelled by $2\cosh(t/2)$. Thus,

$$\text{(7.8)} \qquad e^{|t|/2} - 2\cosh(t/2) = -e^{-|t|/2},$$

and therefore, in view of the asymptotic behavior of the coefficients in the series [cf. (3.5)], the convolution kernel in (7.6), denoted by

$$\text{(7.9)} \qquad \begin{aligned}\xi_\nu(t) &= \Phi_\nu(t) - 2\cosh\left(\frac{t}{2}\right)\\ &= -\exp\left(-\frac{|t|}{2}\right) + \sum_{n=1}^\infty a_n(\nu)\exp\left[-\left(2n-\frac{1}{2}\right)|t|\right],\end{aligned}$$

has a bilateral Laplace transform $\Xi_\nu(s)$ analytic in the strip $-1/2 < \text{Re } s < 1/2$, where it belongs to $L_1 \cap L_\infty$ on vertical lines and is real-valued on the imaginary axis.

Now since $p(r)$ and $rp(r)$ belong to L_1 and

$$\int_0^\infty r^n |p(r)|\, dr = \int_{-\infty}^\infty \exp\left[-\left(n - \frac{1}{2}\right)t\right] |f(t)|\, dt \tag{7.10}$$

the function $f(t)$ has a bilateral Laplace transform $F(s)$ bounded and analytic in the strip $-1/2 < \text{Re } s < 1/2$. Hence, since $f(t) = \overline{f(t)}$, we have

$$(\mu, \mu)_\nu = \frac{1}{2\pi} \int_{-\infty}^\infty \Xi_\nu(i\omega) |F(i\omega)|^2\, d\omega. \tag{7.11}$$

We wish to show that the Fourier transform of $\xi_\nu(t)$ is negative. The case $0 < \nu \le 1$ is simple. In (7.9) we have

$$a_n(\nu) \le 0, \qquad n \ge 1, \qquad 0 < \nu \le 1. \tag{7.12}$$

Since the Fourier transform of $e^{-\lambda|t|}$, $\lambda > 0$, is positive we have

$$\Xi_\nu(i\omega) < 0, \qquad -\infty < \omega < \infty, \qquad 0 < \nu \le 1. \tag{7.13}$$

In case $\nu > 1$, inequality (7.12) no longer holds; in fact, for $1 < \nu < 3$, those coefficients are positive. Nevertheless (7.13) holds for $\nu > 0$.

Recall that [cf. (5.3)(i)]

$$k_\nu(x) = \left\{ \left(\frac{d}{dx}\right)^2 - \frac{1}{4} \right\} \Phi_\nu(x), \qquad \nu > 1, \tag{7.14}$$

where $k_\nu(x)$ is an even integrable function, completely monotone for $x > 0$. Therefore its Fourier transform is positive. Its bilateral Laplace transform is given by [cf. (5.18) and (5.19)]

$$\begin{gathered} K_\nu(s) = 2^{1-\nu}\Gamma(\nu) \frac{\Gamma(3/4 + s/2)\Gamma(3/4 - s/2)}{\Gamma(1/4 + \nu/2 + s/2)\Gamma(1/4 + \nu/2 - s/2)}, \\ -\frac{3}{2} < \text{Re } s < \frac{3}{2}, \nu > 1. \end{gathered} \tag{7.15}$$

Since the operator in (7.14) annihilates $e^{\pm x/2}$ we have, from (7.9) and (7.14),

$$k_\nu(x) = \left\{ \left(\frac{d}{dx}\right)^2 - \frac{1}{4} \right\} \xi_\nu(x), \qquad \nu > 1. \tag{7.16}$$

In view of the common strip of analyticity of their Laplace transforms, we have

$$\Xi_\nu(s) = \frac{K_\nu(s)}{s^2 - 1/4}, \qquad -\frac{1}{2} < \text{Re } s < \frac{1}{2}, \qquad \nu > 1. \tag{7.17}$$

Therefore (7.13) holds for $\nu > 0$. Hence under assumptions (7.3)(i) and (7.3)(ii) we have

$$(\mu, \mu)_\nu = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Xi_\nu(i\omega) |F(i\omega)|^2 \, d\omega < 0, \qquad \nu > 0. \tag{7.18}$$

Thus (3.17) is established.

We note further that from the relation [cf. (6.4)]

$$\nu(\nu+1)\Phi_\nu(x) = \left\{ \left(\nu + \frac{1}{2}\right)^2 - \left(\frac{d}{dx}\right)^2 \right\} \Phi_{\nu+2}(x), \qquad \nu > 0, \tag{7.19}$$

it follows that

$$\nu(\nu+1)\Xi_\nu(s) = \left[\left(\nu + \frac{1}{2}\right)^2 - s^2 \right] \Xi_{\nu+2}(s), \qquad \nu > 0. \tag{7.20}$$

Then, since

$$\xi_1(t) = -e^{-|t|/2},$$

we have, from (7.15), (7.17) and (7.20),

$$\begin{aligned} &\int_{-\infty}^{\infty} \xi_\nu(t) e^{-st} \, dt \\ &= \frac{2^{1-\nu}\Gamma(\nu)}{s^2 - 1/4} \frac{\Gamma(3/4 + s/2)\Gamma(3/4 - s/2)}{\Gamma(1/4 + \nu/2 + s/2)\Gamma(1/4 + \nu/2 - s/2)}, \\ &\qquad -\frac{1}{2} < \operatorname{Re} s < \frac{1}{2}, \qquad \nu > 0. \end{aligned} \tag{7.21}$$

Acknowledgment. We thank an Associate Editor for references and suggestions that led to improvements.

REFERENCES

BUJA, A. (1990). Remarks on functional canonical variates, alternating least squares methods and ACE. *Ann. Statist.* **18** 1032–1069.

DAWS, J. T., ARABIE, P. and HUBERT, L. J. (1990). A note on expected stress values in nonmetric multidimensional scaling. Unpublished report, Univ. Illinois, Urbana–Champaign.

DE LEEUW, J. and STOOP, I. (1984). Upper bounds for Kruskal's stress. *Psychometrika* **49** 391–402.

ERDÉLYI, A., MAGNUS, W., OBERHETTINGER, F. and TRICOMI, F. G. (1953). *Higher Transcendental Functions* **1**. McGraw-Hill, New York.

GNANADESIKAN, R. (1977), *Methods for Statistical Data Analysis of Multivariate Observations.* Wiley, New York.

KAKOSIAN, A. V., KLEBANOV, L. and ZINGER. A. (1989). Characterization of the distribution of statistics by their mean values and certain probability metrics. *Trudy VNIISI* 47–55. (In Russian)

KRUSKAL, J. B. (1964a). Nonmetric multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29** 1–27.

KRUSKAL, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29** 115–129.

LÉVY. P. (1937). *Théorie de l'Addition des Variables Aléatoires*. Gautier–Villars, Paris.

LINDENSTRAUSS. J. and TZAFRIRI, L. (1973). *Classical Banach Spaces*. Springer, New York.

MATTNER, L. (1990). Behandlung einiger Extremalprobleme für Wahrscheinlichkeitsverteilungen. Doctoral dissertation, Univ. Hannover, Germany.

MATTNER, L. (1993). Extremal problems for probability distributions: a general method and some examples. In *Stochastic Inequalities* (M. Shaked and Y. L. Tong, eds.) 274–283. IMS, Hayward, CA.

MORIGUTI, S. (1951). External properties of extreme value distributions. *Ann. Math. Statist.* **22** 523–536.

PLACKETT, R. L. (1947). Limits of the ratio of mean range to standard deviation. *Biometrika* **34** 120–122.

SHEPHARD, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function, I and II. *Psychometrika* **27** 125–139; 219–246.

ANDREAS BUJA
BELLCORE
BOX 1910, 445 SOUTH STREET
MORRISTOWN, NEW JERSEY 07962–1910

B. F. LOGAN, J. A. REEDS
AND L. A. SHEPP
AT&T BELL LABORATORIES
MURRAY HILL, NEW JERSEY 07974