BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 5, Number 3, November 1981

0002-9904/81/0000-0505/$02.25

BOOK REVIEWS

Inequalities: Theory of majorization and its applications, by Albert W. Marshall and Ingram Olkin, Mathematics in Science and Engineering, Vol. 143, Academic Press, New York, 1979, xx + 569 pp., $49.50.

Probability inequalities in multivariate distributions, by Y. L. Tong, Probability and Mathematical Statistics, A Series of Monographs and Textbooks, Academic Press, New York, 1980, xiii + 239 pp., $ 29.50.

Both monographs make extensive use of a (quasi) partial ordering of R^n called majorization and the corresponding class of (Schur) increasing functions on R^n. In this connection, it is best to think of $x = (x_1, \ldots, x_n) \in R^n$ as representing the measure μ on the reals R of total mass n which is defined by $\mu(A) = \Sigma_{x_i \in A} 1$. Let ν denote the analogous measure represented by the point $y \in R^n$. One says that x is majorized by y and also that $x \prec y$ or that y is a dilation of x when

$$(1) \qquad \sum_{i=1}^{n} f(x_i) = \int f\, d\mu \leqslant \int f\, d\nu = \sum_{i=1}^{n} f(y_i)$$

holds for each convex function f on R. It implies that μ and ν have the same mass and the same centre of gravity. A necessary and sufficient condition is that

$$(2) \qquad \sum_{i=1}^{k} x_{[i]} \leqslant \sum_{i=1}^{k} y_{[i]} \quad \text{for } k = 1, 2, \ldots, n,$$

insisting that the equality sign holds when $k = n$. Here, $x_{[1]} \geqslant \cdots \geqslant x_{[n]}$ are the x_i arranged in decreasing order and, similarly, $y_{[1]} \geqslant \cdots \geqslant y_{[n]}$. If (1) is only required for the increasing (decreasing) convex functions on R then one speaks of weak sub-majorization $x \prec_w y$ (or weak super-majorization $x \prec^w y$, respectively). The first is equivalent to (2).

Let $\mathcal{C}$ be an open convex subset of R^n which is symmetric, that is, invariant under each permutation of the coordinates. A function ϕ: $\mathcal{C} \to R$ is said to be Schur increasing (or Schur convex) if it is nondecreasing relative to the partial ordering $x \prec y$ of $\mathcal{C}$; similarly for Schur decreasing functions, also called Schur concave functions. A Schur increasing function is always symmetric. An obvious example would be

$$(3) \qquad \phi(x) = \phi(x_1, \ldots, x_n) = \sum_{i=1}^{n} f(x_i),$$

with f: $R \to R$ a convex function. More generally, every symmetric and convex (concave) function on $\mathcal{C}$ is Schur convex (Schur concave). A symmetric C^1 function ϕ on $\mathcal{C}$ is Schur increasing if and only if $\phi_{(i)}(x) - \phi_{(j)}(x)$ is always of the same sign as $x_i - x_j$ (where $\phi_{(k)}$ denotes the partial derivative relative to x_k). After all, any dilation $x \prec y$ is the composition of at most $n-1$ (elementary) dilations $u \prec v$, where u and v differ in at most two coordinates. If $x \prec_w y$ ($x \prec^w y$) then $\phi(x) \leqslant \phi(y)$ for every ϕ which is Schur convex and, moreover, increasing (decreasing) in each coordinate.

The importance of majorization (and related quasi partial orderings) largely derives from its many applications. The best treatment so far was the one given by Hardy, Littlewood and Pólya (1934). But now, and for many years hence, this distinction will probably be assigned to Marshall and Olkin's treatise. It is nearly exhaustive and very readable at the same time. Full proofs (sometimes several) are given of most results and there is an abundance of cross referencing. The book contains many new proofs and new results, a survey of some 450 papers and even a biography with photographs of a few of the early contributors: Muirhead (1860–1941), Lorenz (1876–1959), Dalton (1887–1962), Schur (1875–1941), Hardy (1877–1947), Littlewood (1885–1977) and Pólya (1887–).

The greater part of the book is devoted to applications. Much of the required background is collected in Part V (pp. 443–519). It contains for example a useful discussion of the Loewner partial ordering $A \leqslant B$ of matrices (meaning that $B - A$ is positive semidefinite). The following are a few of the applications.

(i) Schur (1923) proved that $x \prec y$ when x is the vector of diagonal elements of an Hermitian matrix H, while y is the vector of eigenvalues of H. As was shown by Horn (1954), every pair x, y with $x \prec y$ can be realized in this way.

If H is strictly totally positive instead then $u \prec^w v$ where $u_i = \log x_i$ and $v_i = \log y_i$.

(ii) Let x, y be n-tuples of nonnegative integers such that $\Sigma\, x_i = \Sigma\, y_i = N$ and $y_1 \geqslant y_2 \geqslant \cdots \geqslant y_n$. Then $x \prec y$ holds if and only if $\{1, 2, \ldots, N\}$ has a system of subsets $E_1, \ldots, E_n$ with $|E_i| = x_i$ and with precisely y_j elements belonging to at least j of the sets E_i $(i, j = 1, \ldots, n)$.

(iii) The area of a triangle is a Schur decreasing function of its sides. The radius of the circumcircle of a triangle is a Schur increasing function of its sides.

(iv) Let $\Omega = \{1, 2, \ldots, N\}$ be a finite population of size N and let n be a given sample size. In any sampling plan one may sample with or without replacement or even replace the ith selected item with a probability π_i depending on i. More generally, a sampling plan P is defined to be a probability measure on the set K_n of all sequences $k = (k_1, \ldots, k_N)$ of nonnegative integers satisfying $\Sigma_{j=1}^{N} k_j = n$. Here k_j is to be interpreted as the number of times that the member $j \in \Omega$ occurs in the sample. It will be assumed that P is symmetric, that is, invariant under the permutations of Ω.

Drawing a sample also involves observing a vector y_j of numerical characteristics attached to $j \in \Omega$. A sampling plan Q is said to dominate a sampling plan P if, roughly speaking, the equivalent of a sample of type Q can be obtained by starting with an actual sample of type P and then applying a computer algorithm (involving randomization), the latter having no knowledge beyond the sample of the actual characteristics y_j. This turns out to be true if $E_P\phi \leqslant E_Q\phi$ for each Schur increasing function ϕ on K_n. In that case, $E_P g(Y) \leqslant E_Q g(Y)$ when $g(u)$ is a symmetric function on $\{y_1, \ldots, y_N\}^n$ with $h(a_1, a_2) \leqslant (h(a_1, a_1) + h(a_2, a_2))/2$. Here $h(u_1, u_2)$ denotes $g(u_1, \ldots, u_n)$ with $u_3, \ldots, u_n$ kept constant.

This reviewer found the Chapters 11, 14, and 15 somewhat sketchy. The remarks on pp. 17, 417, 483 could have been worked out in more detail. Specifically, many results would have been more unified and better in focus if they were treated as special cases of a general theory involving increasing functions on an arbitrary partially ordered space and the associated dilations between measures, for instance, in the spirit of the paper by Kamae, Krengel and O'Brien (1977).

I would prefer to start with a fixed cone K of real-valued measurable functions on a measurable space S. Let $\mathfrak{M}$ denote the collection of (nonnegative) measures μ on S such that each $f \in K$ is μ-integrable. For $\mu, \nu \in \mathfrak{M}$, define

$$\mu \prec \nu \quad \text{if} \int f\, d\mu \leqslant \int f\, d\nu \quad \text{for all } f \in K.$$

We will say that ν is a K-dilation of μ.

Important examples are (i) K is the class of all bounded and measurable increasing functions on a (quasi) partially ordered measurable space S, (ii) S is a subset of a locally convex topological vector space and K is a suitable cone of continuous convex functions on S, (iii) S is a compact metrizable space and K is a convex cone of continuous functions on S, containing the constant functions and such that $f_1, f_2 \in K$ imply $\max(f_1, f_2) \in K$.

In case (iii), in order that a measure ν be a K-dilation of a measure μ it is necessary and sufficient that ν can be written as

$$\nu(B) = \int Q(x, B)\mu(dx), \tag{4}$$

with $Q(x, \cdot)$ a probability measure on S which dilates the 1-point measure ε_x. That is, $f(x) \leqslant \int f(y)Q(x, dy)$ for all $x \in S$ and all $f \in K$. In its present generality, the decomposition (4) is due to Cartier; see Meyer (1966) for references and proofs. Instead of S itself being compact, it would naturally be sufficient that μ and ν have compact carriers.

As an example, take $S = R^k$ and let K consist of all convex functions on R^k of at most linear growth. Then $\varepsilon_x \prec Q(x, \cdot)$ if and only if the $Q(x, \cdot)$ mass distribution has $x \in R^k$ as its center of gravity. Thus, (4) simply says that ν can be obtained from μ by spreading out the different masses $\mu(dx)$ over R^k in such a way that in each individual spreading the original center of gravity x is maintained. This picture best explains the term "dilation" already used in the very special case that μ and ν are finite integer-valued measures on R.

More generally, let μ and ν be finite discrete measures on R^k with supports $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$, respectively. Then (4) becomes $\sum_i p_i Q_{ij} = q_j$ and $\sum_j Q_{ij} y_j = x_i$, where $p_i = \mu(\{x_i\})$, $q_j = \nu(\{y_j\})$ and $Q_{ij} = Q(x_i, \{y_j\})$.

In the special case where $m = n$ and $p_i = q_j = 1$, the matrix (Q_{ij}) would be doubly stochastic and it follows from Birkhoff's theorem that $(x_1, \ldots, x_n)$ is in the convex hull of the $n!$ permutations $(y_{\pi_1}, \ldots, y_{\pi_n})$. This type of dilation coincides with the one on p. 430.

No doubt, Marshall and Olkin's work will be a standard reference for many years to come. It would also serve well as a graduate level textbook even though there are no exercises. We highly recommend it.

The monograph *Probability inequalities in multivariate distributions* by Y. L. Tong has some 100 exercises spread over 8 chapters. It is oriented to the inequalities useful in multivariate statistics: confidence limits, hypothesis testing, ranking and selection, reliability and life testing. These applications are treated in Chapter 8. A central role is played by the multivariate normal distribution and related multivariate distributions.

This area opened up only recently, and much remains to be done. The area is important also because exact calculations for multivariate distributions are notoriously difficult. The book is well written and can be read by any second year graduate student, still leading him to several frontiers of our present very limited knowledge.

Chapter 2 discusses Slepian's inequality and its generalizations. Chapter 3 presents analogous results for the multivariate t-distribution, chi-square and F-distributions. Chapter 4 treats Anderson's inequality for symmetric unimodal distributions and its generalizations. Chapter 5 covers a large number of inequalities closely related to certain measures of dependence such as association. Chapter 6 has a considerable overlap with Marshall and Olkin's monograph and treats inequalities derived from certain partial orderings of R^n, especially majorization and the usual product partial ordering. Chapter 7 treats a number of multivariate moment problems. A detailed discussion of a large number of statistical applications is given in the final Chapter 8.

I will discuss only a few aspects. In the sequel, f will denote a probability density on R^n. It is said to be unimodal (in the sense of Anderson) if the set $A_u = \{x: f(x) \geqslant u\}$ is convex for each $u > 0$. The following result due to Anderson (1955) had a profound influence. Let f_1, f_2 be unimodal densities, each symmetric about the origin. Then the convolution $\phi = f_1 * f_2$ has the property that $\phi(uy)$ is a decreasing function of $|u|$ $(u \in R)$.

An important example of a unimodal density is that where f has the form

$$f(x) = g(x'\Sigma^{-1}x), \tag{5}$$

with g a decreasing (-nonincreasing) function and Σ a positive definite $n \times n$ matrix. It was shown by Das Gupta, Eaton, Olkin, Perlman, Savage and Sobel (1972) that, for all $a \in R^n$,

$$P_1[X_i \leqslant a_i \text{ for all } i] \geqslant P_2[X_i \leqslant a_i \text{ for all } i] \tag{6}$$

$(i = 1, \ldots, n)$ when P_1 is governed by (5) with $\Sigma = \Sigma^1$ and P_2 is governed by (5) with the same function g, $\Sigma = \Sigma^2$ and $\Sigma^1_{ij} > \Sigma^2_{ij}$; $\Sigma^1_{ii} = \Sigma^2_{ii} = 1$.

The special case $g(u) = ce^{-u^2/2}$ (multivariate normal distribution) is due to Slepian (1962). The intuitive idea behind (6) is that, under the stated conditions, the components $X_1, \ldots, X_n$ "hang together" stronger under P_1 then under P_2. More results of this type can be found in Chapters 2 and 4.

As we saw, one way of defining majorization $x \prec y$ between vectors in R^n is to require that $x = (x_1, \ldots, x_n)$ belong to the convex hull of the $n!$ permutations $(y_{\pi_1}, \ldots, y_{\pi_n})$ of $y = (y_1, \ldots, y_n)$. More generally, let G be any group of nonsingular linear transformations acting on R^n. For vectors x, y in R^n, define $x \prec_G y$ by the requirement $x \in \operatorname{conv}(Gy)$ that x be inside the convex hull of the orbit of y under G. A function on R^n which is decreasing (that is, nonincreasing) relative to this quasi partial ordering of R^n will be said to be G-increasing. It is automatically G-invariant, that is, constant on each orbit Gy.

In the special case that G is the group P_n of all $n!$ permutations of the coordinates, a G-increasing (G-decreasing) function on R^n is the same as a Schur convex (Schur concave) function.

An important example of a G-decreasing function is any G-invariant unimodal density. Hence, any mixture of G-invariant unimodal densities is also G-decreasing. As a slight generalization of a result due to Mudholkar (1966), one can show that the class of all such mixtures is closed under convolution. The above-mentioned result of Anderson (1955) easily follows by taking G as the group consisting of the two mappings $x' = x$ and $x' = -x$.

On the other hand, the (obviously G-invariant) convolution of two G-decreasing functions need not be G-decreasing, hence, certainly not equal to any mixture of G-invariant unimodal densities. Eaton (1980) gave a counterexample with G as the k-element ($k \geqslant 3$) group of rotations in the plane R^2 over angles $2\pi j/k$ ($j = 0, \ldots, k-1$). As a positive result, it is a deep theorem due to Eaton and Perlman (1977) that the convolution of two G-decreasing densities is again G-decreasing provided G is a reflection group. By this we mean a closed subgroup of the orthogonal group $O(n)$ which has a dense subgroup generated by some collection of reflections g (each in its own hyperplane H_g).

It turns out that $O(n)$ is the only reflection group which is both infinite and irreducible. The case $G = O(n)$ is somewhat trivial. The case $G = P_n$ is due to Marshall and Olkin (1974) and says that the convolution of two Schur decreasing densities is again Schur decreasing. As another illustration, let D_n denote the group of 2^n transformations in R^n which is generated by the n sign changes of a single coordinate. A function f on R^n is D_n-decreasing precisely when $|x_i| \leqslant |y_i|$ ($i = 1, \ldots, n$) imply that $f(x) \geqslant f(y)$.

The monograph by Y. L. Tong contains a multitude of other results and leads to many interesting unanswered questions. It is very accessible and no comparable work exists. We highly recommend it.

REFERENCES

1. T. W. Anderson, *The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities*, Proc. Amer. Math. Soc. **6** (1955), 170–176.

2. S. Das Gupta, M. L. Eaton, I. Olkin, M. D. Perlman, L. J. Savage and M. Sobel, *Inequalities on the probability content of convex regions for elliptically contoured distributions*, Proc. Sixth Berkeley Sympos. on Math. Statistics and Probability (Berkeley, Calif., 1970), vol. II, 1972, pp. 241–265.

3. M. L. Eaton and M. D. Perlman, *Reflection groups, generalized Schur functions and the geometry of majorization*, Ann. Probab. **5** (1977), 829–860.

4. M. L. Eaton, *Multivariate probability inequalities: an overview*, Tech. Rep. No. 378, Univ. of Minnesota, August 1980.

5. G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge Univ. Press, London and New York, 1934.

6. A. Horn, *Doubly stochastic matrices and the diagonal of a rotation matrix*, Amer. J. Math. **76** (1954), 620–630.

7. T. Kamae, U. Krengel and G. L. O'Brien, *Stochastic inequalities on partially ordered spaces*, Ann. Probab. **5** (1977), 899–912.

8. A. W. Marshall and I. Olkin, *Majorization in multivariate distributions*, Ann. Statist. **2** (1974), 1189–1200.

9. P. A. Meyer, *Probability and potentials*, Blaisdell, Waltham, Mass., 1966.

10. G. Mudholkar, *The integral of an invariant unimodal function over an invariant convex set–an inequality and applications*, Proc. Amer. Math. Soc. **17** (1966), 1327–1333.

11. I. Schur, *Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie*, Sitzungsber. Berl. Math. Gesellschaft **22** (1923), 9–20.

12. D. Slepian, *The one-sided barrier problem for Gaussian noise*, Bell System Tech. J. **41** (1962), 463–501.

J. H. B. KEMPERMAN

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 5, Number 3, November 1981

0002-9904/81/0000-0506/$02.25

An introduction to nonharmonic Fourier series, by Robert M. Young, Academic Press, New York, 1980, x + 246 pp., $32.00.

This is a book about a branch of a branch of analysis; a twig you might say. A useful twig I should add, and one bearing many fine blossoms.

The subject is full of neat results and satisfactory resolutions of open problems, and we'll be taking a look at some of these. At first there were nonharmonic sines and cosines, sets of the form $\{\sin \lambda_n x\}$ and $\{\cos \lambda_n x\}$ in which $\{\lambda_n\}$ is a set of real numbers. Their study was initiated by J. L. Walsh [**10**] at the suggestion of G. D. Birkhoff. Only with the appearance of Paley and Wiener's colloquium publication [**9**] in 1934 did nonharmonic Fourier (NHF) analysis really get under way. After having improved on a result stemming from O. Szász's answer to a problem of G. Pólya's about nonharmonic sines and cosines, they go on to "... discuss the closure of the set $\{e^{i\lambda_n x}, 1\}$ ...", i.e., the property that only the null member of $L^2(-\pi, \pi)$ is orthogonal to every member of the set (the word "closure" is not, thankfully, used any more for this property, having been superseded by "completeness"). Incidentally a little later we read "... the only discussion of a case where the sole restriction on λ_n ... is one of the form $|\lambda_n - n| < L < \infty$ is due to Wiener". This is rather misleading since the paper referred to is about $\{\cos \lambda_n x\}$, not about sets of complex exponentials $\{e^{i\lambda_n x}\}$.

Thus was our subject born, and it is astonishing how much later work has its origins in this seminal effort of Paley and Wiener (I tend to think of it as the "big bang" of NHF analysis). The problems center chiefly on completeness and basis properties of sets $\{e^{i\lambda_n x}\}$, and connections with ordinary Fourier series are made via "equi-convergence" theorems; but more of this anon.