BOUNDS ON THE EXPECTATION OF A CONVEX FUNCTION OF A MULTIVARIATE RANDOM VARIABLE

Albert Madansky

The RAND Corporation

1. Introduction. Dresher has shown [2] how certain inequalities can be interpreted geometrically via the theory of moment spaces of univariate distributions. Moment spaces of multivariate distributions will be considered, and, by examining the boundary of an appropriate moment space, upper and lower bounds on the expectation of a convex function of a vector valued random variable will be derived. Finally, the bounds so derived will be improved in the case where the elements of the random vector are independent.

2. Moment spaces of multivariate distributions. Let $\psi(x) = \psi(x_1, \cdots, x_r)$ be an r-variate cumulative distribution function over the bounded r-dimensional rectangle I, and let $\{f_i(x_1, \cdots, x_r) = f_i(x), i = 1, \cdots, n\}$ be a set of n continuous functions. The ith moment of $\psi(x)$ with respect to $\{f_i(x)\}$ is defined to be $\mu_i(\psi) = \int_I f_i(x)\, d\psi(x)$, and the nth moment space M_n with respect to $\{f_i(x)\}$ is defined as the set of all points $\mu = (\mu_1, \cdots, \mu_n)$ in n-dimensional Euclidean space, E_n, whose coordinates are the moments $\mu_1(\psi), \cdots, \mu_n(\psi)$ with respect to $\{f_i(x)\}$ for some distribution function $\psi(x)$.

Let C_n be the surface traced out in E_n by

$$\{z_i = f_i(x), i = 1, \cdots, n, x \varepsilon I\}.$$

Let H_n be the convex hull of C_n, i.e., the smallest convex set containing C_n. Then it can be shown, along the same lines as the proof of Theorem 2 of [1], that H_n is identical with M_n, and that M_n is closed, bounded, and convex.

In the following I shall examine $g(X)$, some given continuous convex function of an r-dimensional vector valued random variable X defined over the bounded r-dimensional rectangle I. Let C_{r+1} be the surface traced out in E_{r+1} by $z_1 = x_1, z_2 = x_2, \cdots, z_r = x_r, z_{r+1} = g(x_1, \cdots, x_r) = g(x)$. What I shall do is determine the boundary of H_{r+1}, the convex hull of C_{r+1}, from which inequalities on $Eg(X)$ in terms of $g(EX)$ and $\{EX_i, i = 1, \cdots, r\}$ will be obtained. A point x^0 is said to be on the boundary of H_{r+1} if and only if x^0 is in H_{r+1} and there exists a set of real numbers $\beta_0, \beta_1, \cdots, \beta_{r+1}$ such that

$$\sum_{i=1}^{r+1} \beta_i x_i^0 + \beta_0 = 0$$

and

$$\sum_{i=1}^{r+1} \beta_i x_i + \beta_0 \geqq 0 \qquad \text{for all } x \text{ in I.}$$

Geometrically, x^0 is on the boundary of H_{r+1} if and only if there exists a supporting hyperplane to H_{r+1} at that point. A point x^0 is on the upper (lower)

Received November 17, 1958; revised February 22, 1959.

boundary of H_{r+1} if and only if H_{r+1} lies in the negative (positive) half space relative to the supporting hyperplane to H_{r+1} at x_0.

3. Boundaries of H_{r+1}. Let I be the bounded r-dimensional rectangle defined by the 2^r vertices of the form $(a_{1\phi_1}, a_{2\phi_2}, \cdots, a_{r\phi_r})$, where

$$\phi_i(i = 1, \cdots, r)$$

takes on the values 1 and 2 and $a_{i1} < a_{i2}$ for all i, say, and let

$$x_{r+1} = g(x_1, \cdots, x_r)$$

be a continuous convex function of $x_1, \cdots, x_r$ defined over all points of I. It is easy to see that the set of points of the form $(x_1, \cdots, x_r, g(x_1, \cdots, x_r))$ form the lower boundary of H_{r+1}, since they are in H_{r+1} and for any point $(x_1^0, \cdots, x_r^0, g(x_1^0, \cdots, x_r^0))$ there is a supporting hyperplane to H_{r+1} at that point, namely the plane tangent to the surface $x_{r+1} = g(x_1, \cdots, x_r)$ at that point. (Since g is continuous and convex, such a plane exists for each point $(x_1^0, \cdots, x_r^0, g(x_1^0, \cdots, x_r^0))$ in I.) It is also easy to see that the $2r$ hyperplanes $x_1 = a_{11}$, $x_1 = a_{12}$, $x_2 = a_{21}$, $x_2 = a_{22}$, $\cdots$, $x_r = a_{r1}$, $x_r = a_{r2}$ bound H_{r+1} on its sides.

The upper boundary of H_{r+1} is characterized by the following theorem, which is easily proved and geometrically obvious.

THEOREM 1. *The upper boundary of H_{r+1} is identical with the upper boundary of the convex hull of the 2^r points $(a_{1\phi_1}, \cdots, a_{r\phi_r}, g(a_{1\phi_1}, \cdots, a_{r\phi_r}))$.*

Let $H^*(x_1^0, \cdots, x_r^0)$ be the point where the ray

$$(x_1 = x_1^0, \cdots, x_r = x_r^0, x_{r+1} = \theta \mid -\infty \leqq \theta \leqq \infty)$$

intersects the upper boundary of H_{r+1}.

THEOREM 2. *If $X = (X_1, \cdots, X_r)$ is a random vector such that*

$$\Pr\{X \varepsilon I\} = 1,$$

and $g(X)$ is a continuous convex function of X over the bounded r-dimensional rectangle I, *then*

$$g(EX_1, \cdots, EX_r) \leqq Eg(X_1, \cdots, X_r) \leqq H^*(EX_1, \cdots, EX_r).$$

PROOF. By the above discussion, an arbitrary point $(x_1, \cdots, x_{r+1})$ of H_{r+1} satisfies the inequality $(x_1, \cdots, x_r, g(x_1, \cdots, x_r)) \leqq (x_1, \cdots, x_r, x_{r+1}) \leqq (x_1, \cdots, x_r, H^*(x_1, \cdots, x_r))$. Since $M_{r+1} = H_{r+1}$, we can take

$$(x_1, \cdots, x_{r+1})$$

to be $(\mu_1, \cdots, \mu_{r+1})$, where

$$\mu_i = \int_I x_i \, d\psi(x) = EX_i, \qquad i = 1, \cdots, r$$

$$\mu_{r+1} = \int_I g(x_1, \cdots, x_r) \, d\psi(x) = Eg(X_1, \cdots, X_r)$$

for some distribution function $\psi(x)$.

4. Discussion. One should note first of all that the left-handed inequality is the familiar Jensen's inequality.

When $r = 1$, let us consider the moment space M_2 defined by the curves $z_1 = x_1$, $z_2 = g(x_1)$, for $a_{11} \leqq x_1 \leqq a_{12}$, $a_{11} < a_{12}$. The convex hull of the points $(a_{11}, g(a_{11}))$ and $(a_{12}, g(a_{12}))$ is the straight line joining them, and so an upper bound due to Edmundson [3] is obtained, namely

$$Eg(X_1) \leqq \frac{g(a_{12}) - g(a_{11})}{a_{12} - a_{11}} [EX_1 - a_{11}] + g(a_{11}).$$

When $r = 2$, a description of the upper boundary of the appropriate convex hull, H_3, can once again be given explicitly.

Let $D = g(a_{11}, a_{21}) + g(a_{12}, a_{22}) - g(a_{12}, a_{21}) - g(a_{11}, a_{22})$.

(1) If $D \geqq 0$ then the upper boundary of H_3 is the plane determined by $(a_{11}, a_{21}, g(a_{11}, a_{21}))$, $(a_{12}, a_{22}, g(a_{12}, a_{22}))$, $(a_{12}, a_{21}, g(a_{12}, a_{21}))$ for

$$x_2 \leqq a_{21} + \left[\frac{a_{22} - a_{21}}{a_{12} - a_{11}}\right](x_1 - a_{11})$$

and the plane determined by $(a_{11}, a_{21}, g(a_{11}, a_{21}))$ $(a_{12}, a_{22}, g(a_{12}, a_{22}))$, $(a_{11}, a_{22}, g(a_{11}, a_{22}))$ for

$$x_2 \geqq a_{21} + \left[\frac{a_{22} - a_{21}}{a_{12} - a_{11}}\right](x_1 - a_{11}).$$

(2) If $D \leqq 0$ then the upper boundary of H_3 is the plane determined by $(a_{11}, a_{21}, g(a_{11}, a_{21}))$, $(a_{12}, a_{21}, g(a_{12}, a_{21}))$, $(a_{11}, a_{22}, g(a_{11}, a_{22}))$ for

$$x_2 \leqq a_{22} - \left[\frac{a_{22} - a_{21}}{a_{12} - a_{11}}\right](x_1 - a_{11})$$

and the plane determined by $(a_{12}, a_{22}, g(a_{12}, a_{22}))$, $(a_{12}, a_{21}, g(a_{12}, a_{21}))$, $(a_{11}, a_{22}, g(a_{11}, a_{22}))$ for

$$x_2 \geqq a_{22} - \left[\frac{a_{22} - a_{21}}{a_{12} - a_{11}}\right](x_1 - a_{11}).$$

Note that for $D = 0$ the four points considered are coplanar.

As an example of the use of the inequality, let $r = 1$, $g(X_1) = e^{X_1}$, $a_{11} = 0$, $a_{12} = 1$. Then $Ee^{X_1} \leqq (e - 1)EX_1 + 1$. As a further example, let $r = 2$, $g(X_1, X_2) = e^{X_1+X_2}$, $a_{11} = a_{21} = 0$, $a_{12} = a_{22} = 1$. Then

$$Ee^{X_1+X_2} \leqq \begin{cases} (e-1)[EX_1 + EX_2] + (e-1)^2 EX_2 + 1 & \text{if} \quad EX_1 \geqq EX_2 \\ (e-1)[EX_1 + EX_2] + (e-1)^2 EX_1 + 1 & \text{if} \quad EX_2 \geqq EX_1. \end{cases}$$

For higher dimensions I have obtained no explicit description of the upper boundary of H_{r+1}. However, it is easy to see that one can evaluate $H^*(x_1^0, \cdots, x_r^0)$ by finding the equations of the $\binom{2^r}{r+1}$ hyperplanes formed by joining $r + 1$ points of the form $(a_{1\phi_1}, \cdots, a_{r\phi_r}, g(a_{1\phi_1}, \cdots, a_{r\phi_r}))$ and then finding the maximum value of the $(r + 1)$st coordinate of the points on these hyperplanes whose first r coordinates are $(x_1^0, \cdots, x_r^0)$. This maximum value is $H^*(x_1^0, \cdots, x_r^0)$.

5. A sharper inequality.[1] From the above discussion of the case when $r = 2$, one sees that

$$g(X_1) \leqq \frac{(X_1 - a_{11})}{(a_{12} - a_{11})} g(a_{12}) + \frac{(a_{12} - X_1)}{(a_{12} - a_{11})} g(a_{11}).$$

Since $g(X_1, \cdots, X_r)$ is convex in X_r for fixed $X_1, \cdots, X_{r-1}$, one can use the above formula to obtain

$$g(X_1, \cdots, X_r) \leqq \frac{(X_r - a_{r1})}{(a_{r2} - a_{r1})} g(X_1, \cdots, X_{r-1}, a_{r2}) + \frac{(a_{r2} - X_r)}{(a_{r2} - a_{r1})} g(X_1, \cdots, X_{r-1}, a_{r1}).$$

Since $g(X_1, \cdots, X_j, a_{j+1,\phi_{j+1}}, \cdots, a_{r\phi_r})$ is convex in X_j for fixed $X_1, \cdots, X_{j-1}$, one can use the above bound successively in the obvious manner to obtain

$$g(X_1, \cdots, X_r) \leqq \sum_{\phi} \prod_{j=1}^{r} (-1)^{\phi_j} \frac{(a_{j\phi_j} - X_j)}{(a_{j2} - a_{j1})} g(a_{1\bar{\phi}_1}, \cdots, a_{r\bar{\phi}_r}),$$

where $\bar{\phi}_i = 3 - \phi_i$. Hence, if the X_j's are independent,

$$Eg(X_1, \cdots, X_r) \leqq \sum_{\phi} \prod_{j=1}^{r} (-1)^{\phi_j} \frac{(a_{j\phi_j} - EX_j)}{(a_{j2} - a_{j1})} g(a_{1\bar{\phi}_1}, \cdots a_{r\bar{\phi}_r}) = H(EX), \text{ say.}$$

In the bivariate example above, we find that

$$Ee^{X_1+X_2} \leqq (e - 1)[EX_1 + EX_2] + (e - 1)^2 EX_1 EX_2 + 1,$$

which is sharper than the bound obtained in Section 4. That this is true in general can be seen by noting that

$$\sum_{\phi} \prod_{j=1}^{r} (-1)^{\phi_j} \frac{(a_{j\phi_j} - EX_j)}{(a_{j2} - a_{j1})} = 1$$

and

$$\prod_{j=1}^{r} (-1)^{\phi_j} \frac{(a_{j\phi_j} - EX_j)}{(a_{j2} - a_{j1})} \geqq 0.$$

Hence the point $(EX_1, \cdots, EX_r, H(EX))$ lies in the convex hull of the 2^r points $(a_{1\phi_1}, \cdots, a_{r\phi_r}, g(a_{1\phi_1}, \cdots, a_{r\phi_r}))$, and so $H(EX) \leqq H^*(EX)$.

REFERENCES

[1] M. Dresher, S. Karlin, and L. S. Shapley, "Polynomial Games," *Ann. Math. Studies*, No. 24 (1950), pp. 161–180.

[2] Melvin Dresher, "Moment Spaces and Inequalities," *Duke Math. Jour.*, Vol. 20 (1953), pp. 261–271.

[3] H. P. Edmundson, "Bounds on the Expectation of a Convex Function of a Random Variable," The RAND Corporation, P-982, April 9, 1957.

[1] I wish to thank W. Hoeffding and the referee for their suggestions which improved this section.