The Annals of Probability
1976, Vol. 4, No. 3, 502–504

NOTE ON THE k-DIMENSIONAL JENSEN INEQUALITY

BY MARTIN SCHAEFER

University of Hamburg

Let f be a measurable convex function from R^k to R^1 and let $X_1, \cdots, X_k$ be real-valued integrable random variables. The best approximation for $f(EX_1, \cdots, EX_k)$ one can get by Jensen's inequality is $f(EX_1, \cdots, EX_k) \leqq \inf Ef(\mathbf{Z})$ where the infimum is taken over all k-dim. random vectors $\mathbf{Z} = (Z_1, \cdots, Z_k)'$ such that Z_i has the same distribution as X_i $(1 \leqq i \leqq k)$. An application is given in the case where $f(\mathbf{y})$ is the span of the vector $\mathbf{y}$ which leads to a new approximation for $f(A\mathbf{u})$ where A is a stochastic $(k \times m)$-matrix and $\mathbf{u}$ is an arbitrary element of R^m.

Let $\mathbf{X} = (X_1, \cdots, X_k)'$ be a k-dimensional random vector with integrable components $X_1, \cdots, X_k$ and let f be a measurable convex function from R^k to R^1. Then it is well known (cf. Perlman (1974), page 52) that $Ef(\mathbf{X})$ exists and it holds that $f(E\mathbf{X}) \leqq Ef(\mathbf{X})$.

An interesting and useful aspect of this inequality—which to the author's knowledge has not been pointed out before in the literature—is the following. The left-hand side of the inequality above depends on the (marginal) distributions of $X_1, \cdots, X_k$ only, but in general (if $k > 1$) the right-hand side depends on the k-dim. distribution of the vector $\mathbf{X} = (X_1, \cdots, X_k)'$. Therefore, considering real-valued integrable random variables $X_1, \cdots, X_k$ the best approximation for $f(EX_1, \cdots, EX_k)$ one can get by Jensen's inequality is $f(EX_1, \cdots, EX_k) \leqq \inf Ef(\mathbf{Z})$ where the infimum is taken over all k-dim. random vectors $\mathbf{Z} = (Z_1, \cdots, Z_k)'$ such that Z_i has the same (1-dim.) distribution as X_i $(1 \leqq i \leqq k)$. The following example illustrates the usefulness of this aspect. Let $A = (a_{ij})$ be a stochastic $(k \times m)$-matrix (i.e., $a_{ij} \geqq 0$, $a_{i1} + \cdots + a_{im} = 1$) and for $\mathbf{y} = (y_1, \cdots, y_k)' \in R^k$ let $f_k(\mathbf{y}) = \max\{y_1, \cdots, y_k\} - \min\{y_1, \cdots, y_k\}$ be the span of the vector $\mathbf{y}$. In stochastic dynamic programming one is interested in approximations for $f_k(A\mathbf{u})$ where $\mathbf{u}$ is an arbitrary element of R^m (cf. e.g., Hübner [2] and White [5]). It is well known that there exists a constant $c > 0$ depending on A only such that

$$f_k(A\mathbf{u}) \leqq cf_m(\mathbf{u}) \quad \text{holds for all} \quad \mathbf{u} \in R^m . \tag{1}$$

The best constant is $c^* = \max_{i,l}[1 - \sum_{j=1}^m \min\{a_{ij}, a_{lj}\}]$ in the sense of

$$c^* = \sup f_k(A\mathbf{u})(f_m(\mathbf{u}))^{-1} , \tag{2}$$

where the supremum is taken over all $\mathbf{u} \in R^m$ such that $f_m(\mathbf{u}) \neq 0$ (cf. Hübner [2]).

Now, using the fact that f_k is a measurable convex function and that $A\mathbf{u}$ can

Received May 14, 1975; revised August 9, 1975.
AMS 1970 *subject classification.* 52A40.
Key words and phrases. Convex function, Jensen inequality.

be written in the form

$$A\mathbf{u} = (EZ_1, \cdots, EZ_k)' , \tag{3}$$

where Z_i is a random variable such that $P(Z_i = u_j) = a_{ij}$, one can get a new approximation according to

$$f_k(A\mathbf{u}) \leqq \inf Ef_k(\mathbf{Z}) . \tag{4}$$

Here, the infimum is taken over all such random vectors $\mathbf{Z} = (Z_1, \cdots, Z_k)'$ and it can be calculated by the help of the following

THEOREM. *Let $P_1, \cdots, P_k$ be* 1-dim. *probability measures with corresponding distribution functions* (dfs) $F_1, \cdots, F_k$ *such that* $\int u\, dP_i$ *exists and is finite and let* $\mathscr{P}$ *be the class of all* k-dim. *probability measures with marginal distributions* $P_1, \cdots, P_k$. *Then we have* $\inf_{\mathscr{P}} \int f_k(y)\, dP = \int f_k(y)\, dP^*$. P^* *corresponds to the* k-dim. df $F^*(\mathbf{y}) = \min \{F_1(y_1, \cdots, F_k(y_k)\}$.

PROOF. For arbitrary $P \in \mathscr{P}$ consider the following 1-dim. dfs $F_P(u) = P(\min \{y_1, \cdots, y_k\} \leqq u)$ and $G_P(u) = P(\max \{y_1, \cdots, y_k\} \leqq u)$. Because of

$$G_P(u) \leqq \min_i F_i(u) \leqq \max_i F_i(u) \leqq F_P(u) \qquad \text{for all} \quad u \in R^1 \tag{5}$$

and

$$G_{P^*}(u) = \min_i F_i(u) , \qquad F_{P^*}(u) = \max_i F_i(u) \qquad \text{for all} \quad u \in R^1 \tag{6}$$

we obtain for $P \in \mathscr{P}$

$$\int f_k(\mathbf{y})\, dP = \int u\, d(G_P - F_P) \geqq \int u\, d(G_{P^*} - F_{P^*}) = \int f_k(\mathbf{y})\, dP^* . \tag{7}$$

Since $P^* \in \mathscr{P}$ the assertion follows.

Now it is easy to verify the first part of the next

COROLLARY. *Let $A = (a_{ij})$ be a stochastic $(k \times m)$-matrix and let $\mathbf{u} = (u_1, \cdots, u_m) \in R^m$ such that $u_1 \leqq \cdots \leqq u_m$. Then*

$$f_k(A\mathbf{u}) \leqq \inf Ef_k(\mathbf{Z}) = \sum_{\nu=1}^{m-1} f_k(s_{1\nu}, \cdots, s_{k\nu})(u_{\nu+1} - u_\nu) \leqq c^*(u_m - u_1) ,$$

where $s_{i\nu} = \sum_{j=1}^{\nu} a_{ij}$.

PROOF. The first equality is an immediate consequence of the theorem above. The second inequality follows because of

$$f_k(s_{1\nu}, \cdots, s_{k\nu}) \leqq c^* \qquad \text{for all} \ \nu = 1, \cdots, m-1 . \tag{8}$$

REMARK. If H is a joint distribution with marginal distributions $F_1, \cdots, F_k$, then it is well known and easily demonstrated that $H(\mathbf{y}) \leqq F^*(\mathbf{y})$ for all $\mathbf{y}$. This result goes back to Hoeffding (1940, his thesis) and was rediscovered by Fréchet. (cf. e.g., Fréchet (1950), page 25).

REMARK. Using the fact that the expectation of a random variable can be expressed according to

$$EX = \int_0^\infty \{1 - P(X \leqq t) - P(X \leqq -t)\}\, dt$$

then because of (6) it is easily shown that

$$\inf_{\mathscr{P}} \int f_k(\mathbf{y})\, dP = \int_{-\infty}^{+\infty} f_k(F_1(t), \cdots, F_k(t))\, dt\,.$$

This is a slight generalization of a result of Vallender (1973) who has treated the case $k = 2$.

REFERENCES

[1] Fréchet, M. (1950). *Recherches Théoriques Modernes sur le Calcul des Probabilités*, 1r livre. Généralités sur les probabilités; éléments aléatoires. Gauthier-Villars, Paris.

[2] Hübner, G. (1976). Improved procedures for eliminating suboptimal actions in Markov programming by the use of contraction properties. Will appear in "Transactions 7th Prague Conference," Czechoslovak Academy of Science, Prague.

[3] Perlman, M. D. (1974). Jensen's inequality for a convex vector-valued function on an infinite-dimensional space. *J. Multivariate Anal.* **4** 52–65.

[4] Vallender, S. S. (1973). Calculation of the Wasserstein distance between probability distributions on the line. *Theor. Probability Appl.* **18** 784–786.

[5] White, D. J. (1963). Dynamic programming, Markov chains and the method of successive approximation. *J. Math. Anal. Appl.* **6** 373–376.

Institut für Mathematische Stochastik
Universitat Hamburg
2 Hamburg 13, Rothenbaumchaussee 45
West Germany