

On Some Refinements of Jensen's Inequality

Mohammad Kazim Khan

Department of Mathematics and Computer Science, Kent State University, Kent, Ohio 44242

and

Muhammad Hanif

Department of FAMCO, King Faisal University, P.O. Box 2114, Dammam, Saudi Arabia

Communicated by Ranko Bojanic

Received October 21, 1996; accepted March 15, 1997

Dedicated to the Memory of Professor M. H. Afghahi

This paper contains a link between Probability Proportional to Size (PPS) sampling and interpolations of the classical Jensen inequality. We show that these interpolating inequalities are, in fact, special cases of the conditional Jensen inequality when applied over an appropriate probability space. We provide a few examples dealing with divided differences, convex functions over linear spaces, approximation operators, and sampling with and without replacement. © 1998

Academic Press

1. INTRODUCTION

Let (S, \mathcal{F}, P) be a probability space along with a sequence of sub-sigma fields $\mathcal{F}_n \supseteq \mathcal{F}_{n+1}$, and a sequence of random variables ξ_n , $n = 1, 2, \dots$. If for each n , ξ_n is \mathcal{F}_n measurable, $E|\xi_n|$ is finite, and $E(\xi_n | \mathcal{F}_{n+1}) \stackrel{a.s.}{=} \xi_{n+1}$, then $\{\xi_n, \mathcal{F}_n, n \geq 1\}$ is called a reverse martingale. R. A. Khan [18, 19] used such a probabilistic structure to provide elegant proofs of number of results concerning monotonic convergence of approximation operators. The main argument about the monotonicity of approximation operators uses the conditional form of Jensen's inequality when the function begin approximated is convex.

A rich class of reverse martingales is obtained by considering the \mathcal{U} -statistics of Hoeffding [15]. That is, if Y_1, Y_2, \dots is a sequence of independent and

identically distributed (iid) random variables and k is some fixed positive integer, the random variables

$$U_{k,n} := \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \phi(Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}), \quad n = k, k+1, \dots,$$

are called \mathcal{U} -statistics where ϕ is a real symmetric function on \mathbb{R}^k . If ϕ happens to be convex as well, then $U_{k,n}(Y_1, \dots, Y_n)$ becomes a Schur convex function and this view leads to some inequalities. Indeed, by using this point of view, the first proof of monotonic convergence of the classical Feller operator (which includes the usual Bernstein polynomials, Szasz, gamma, and Weierstrass operators) was provided by Marshall and Proschan [21]. If \mathcal{F}_n is the smallest sigma algebra generated by $(U_{k,n}, U_{k,n+1}, \dots)$, then it is well known that $\{U_{k,n}, \mathcal{F}_n, n = k, k+1, \dots\}$ forms a reverse martingale when $E|\phi(Y_1, Y_2, \dots, Y_k)| < \infty$. Using this reverse martingale for a fixed k , and $n = k, k+1, \dots$, R. A. Khan [19] recently provided a number of results of monotonic convergence of approximation operators.

In this paper, we complement some of the results of R. A. Khan by considering expressions similar to $U_{k,n}$; however, we fix n and allow k to vary from 1 to n . We will provide appropriate probability spaces so that the resulting structures have reverse martingale type properties over linear spaces. More precisely, we will provide a probability space (S, \mathcal{F}, P) along with a sequence of finite sub-sigma fields $\{\mathcal{F}_k\}$, and a sequence of random vectors $\{\xi_k\}$ taking values in a linear space \mathcal{L} . And for each k , ξ_k has the properties that it is \mathcal{F}_k measurable, and $E(\xi_k | \mathcal{F}_{k+1}) \stackrel{a.s.}{=} \xi_{k+1}$. This reverse martingale type of structure will be used to provide, among other things, some refinements of Jensen's inequality, and improvements of some existing inequalities in analysis and approximation theory. We further provide a link to the subject of Probability Proportional to Size (PPS) sampling of statistics. This, in turn, shows how to provide an infinite variety of refinements of the classical Jensen's inequality.

The next section provides some methods of constructing probability spaces over which functions similar to $U_{k,n}$ will have a reverse martingale type of structure along with some refinements of the Jensen inequality. Section 3 provides some links with PPS sampling. Section 4 lists some refinements of known inequalities in analysis which become special cases of the results in the earlier sections. The last section gives refinements of monotonic convergence of a number of univariate and multivariate classical approximation operators.

2. INTERPOLATIONS OF JENSEN'S INEQUALITY

In this section we will take $\{x_1, x_2, \dots, x_n\} \subseteq K \subseteq \mathcal{L}$ where \mathcal{L} is a linear space, K is a convex set, and n is a fixed positive integer. Let $\mathbf{Q} := [q_{ij}]$ be an $n \times n$ stochastic matrix with diagonal terms equal to zero, and let $\mathbf{P} := (p_1, \dots, p_n)$ consist of positive numbers adding up to one. For any real-valued function f over K , we define

$$f_{k,n}(x, \mathbf{P}, \mathbf{Q}) := \frac{1}{\binom{n-2}{k-2}} \sum_{A_k} \left(\sum_{i,j \in A_k} p_i q_{ij} \right) f \left(\frac{\sum_{i,j \in A_k} x_i p_i q_{ij}}{\sum_{i,j \in A_k} p_i q_{ij}} \right),$$

where the summation is over all subsets A_k of size k taken from $\{1, 2, \dots, n\}$. To avoid division by zero, we will assume throughout that all $\sum_{i,j \in A_k} p_i q_{ij} > 0$. When f is a convex function over K , the classical Jensen inequality gives that

$$\sum_{j=1}^n p_j f(x_j) \geq f \left(\sum_{j=1}^n p_j x_j \right).$$

Our first result shows how to provide a variety of refinements of this inequality.

THEOREM 1. *Let f be a real-valued convex function over K and let $x_i \in K$, $i = 1, 2, \dots, n$, be fixed. Then for any probability vector \mathbf{P} and uniformly over all $n \times n$ stochastic matrices \mathbf{Q} as defined above, we have*

$$\begin{aligned} \sum_{j=1}^n p_j f(x_j) &\geq f_{k,n}(x, \mathbf{P}, \mathbf{Q}) \geq f_{k+1,n}(x, \mathbf{P}, \mathbf{Q}) \\ &\geq f \left(\sum_{j=1}^n p_j x_j \right), \quad k = 2, 3, \dots, n-1. \end{aligned}$$

Proof. Consider a probability space Ω consisting of $\omega = (\{i\}, \{i_1, \dots, i_k\}, \{j_1, \dots, j_{k+1}\})$ where $i, i_1, \dots, i_k, j_1, \dots, j_{k+1}$ are integers lying in the set $\{1, 2, \dots, n\}$ with the property that $\{i\} \subseteq \{i_1, \dots, i_k\} \subseteq \{j_1, \dots, j_{k+1}\}$. Let A_k be a subset of size k from the set of first n positive integers. For any given subsets $A_1 \subseteq A_k \subseteq A_{k+1}$ with $A_1 = \{i\}$ we define a probability measure over the subsets of Ω by letting

$$P(\{(A_1, A_k, A_{k+1})\}) := \frac{p_i}{(n-k) \binom{n-2}{k-2}} \sum_{j \in A_k} q_{ij}.$$

Define a random element $X: \Omega \rightarrow K$ by $X(\omega) := x_i$, where i is the element of the first set in ω . Also define random elements $Z_k(\omega)$ and $Z_{k+1}(\omega)$ to be the second and the third sets in ω , respectively. For a given subset A_k containing i , we have

$$P(X = x_i | Z_k = A_k) = \frac{p_i \sum_{j \in A_k} q_{ij}}{\sum_{l, j \in A_k} p_l q_{lj}}, \quad P(Z_k = A_k) = \frac{\sum_{l, j \in A_k} p_l q_{lj}}{\binom{n-2}{k-2}}.$$

Therefore, by the convexity of f , or the conditional form of the Jensen inequality [10], we have

$$\begin{aligned} \sum_{i=1}^n p_i f(x_i) &= Ef(X) \geq Ef(E(X | Z_k)) \\ &= \frac{1}{\binom{n-2}{k-2}} \sum_{A_k} \left(\sum_{i, j \in A_k} p_i q_{ij} \right) f \left(\frac{\sum_{i, j \in A_k} x_i p_i q_{ij}}{\sum_{i, j \in A_k} p_i q_{ij}} \right) \\ &= f_{k, n}(x, \mathbf{P}, \mathbf{Q}). \end{aligned}$$

Now note that for any natural number $i \leq n$, and any subset A_{k+1} containing i , we have

$$\begin{aligned} &\sum_{A_k: i \in A_k \subseteq A_{k+1}} P(X = x_i | Z_k = A_k) P(Z_k = A_k | Z_{k+1} = A_{k+1}) \\ &= \frac{p_i}{k-1} \frac{1}{\sum_{l, j \in A_{k+1}} p_l q_{lj}} \sum_{A_k: i \in A_k \subseteq A_{k+1}} \sum_{j \in A_k} q_{ij} \\ &= \frac{p_i}{k-1} \frac{1}{\sum_{l, j \in A_{k+1}} p_l q_{lj}} \sum_{j \in A_{k+1}} q_{ij} \binom{k+1-2}{k-2} \\ &= P(X = x_i | Z_{k+1} = A_{k+1}). \end{aligned}$$

Therefore, we have

$$\begin{aligned} &E\{E(X | Z_k) | Z_{k+1} = A_{k+1}\} \\ &= \sum_{A_k: A_k \subseteq A_{k+1}} E(X | Z_k = A_k) P(Z_k = A_k | Z_{k+1} = A_{k+1}) \\ &= \sum_{i=1}^n x_i \sum_{A_k: i \in A_k \subseteq A_{k+1}} P(X = x_i | Z_k = A_k) P(Z_k = A_k | Z_{k+1} = A_{k+1}) \\ &= \sum_{i=1}^n x_i P(X = x_i | Z_{k+1} = A_{k+1}) = E\{X | Z_{k+1} = A_{k+1}\}. \end{aligned}$$

Applying the conditional Jensen inequality one more time gives that

$$\begin{aligned} f_{k,n}(x, \mathbf{P}, \mathbf{Q}) &= Ef(E(X | Z_k)) \geq Ef(E\{E(X | Z_k) | Z_{k+1}\}) \\ &= Ef(E(X | Z_{k+1})) = f_{k+1,n}(x, \mathbf{P}, \mathbf{Q}). \end{aligned}$$

Another application of Jensen's inequality on the last term completes the proof. ■

A number of implications of this result will be mentioned in the last two sections. By using the concepts of sampling theory of statistics, one can extend this result, as we show in the next section.

3. A LINK WITH PPS SAMPLING

In Sampling Theory, the objective is to select a portion of the population having certain properties. In Probability Proportional to Size sampling, we have a population consisting of n units and we are provided positive numbers p_i , $i = 1, 2, \dots, n$, which represent the "importance" (or "size") of each of the population units x_i , $i = 1, 2, \dots, n$. The aim is to select a subset of size k ($1 \leq k < n$), so that we may estimate the population total $\sum_{i=1}^n x_i$ by using the information provided by the sample. It is known, however, that there does not exist a sampling scheme which uniformly (over all populations) minimizes the variance when estimating the population total. This leads to different sampling schemes each having its own advantages over the others. To date, over one-hundred sampling schemes have been devised (see [5, 6, 8]). Most of these sampling schemes give rise to probability spaces that can be used to enhance our Theorem 1.

More precisely, over the set $\{x_1, x_2, \dots, x_n\}$, we start to select the elements one after the other without replacement. With the probability distribution \mathbf{P} we select the first element. If x_i is selected on the first draw, the second element is selected by the distribution given by the i th row of the stochastic matrix \mathbf{Q} . If we select the remaining elements with equal probabilities, we get the probability space of Theorem 1. In this case,

$$u_{k,n} := E(X | Z_k), \quad k = 2, 3, \dots, n-1,$$

forms a reverse martingale type of structure which leads to the result. We could introduce different non-uniform distributions over the subsequent selections to enhance Theorem 1. We state this result without proof (which is essentially to that of Theorem 1 however considerably messy in notation) as follows.

THEOREM 2. Let K be a convex subset of a linear space and let $x_i \in K$, $i = 1, 2, \dots, n$, be given. Let P be a probability measure on the permutations of the first n positive integers (representing the draw by draw PPS selection) so that the probability of any permutation $\omega = (i_1, i_2, \dots, i_n)$, is

$$P(\omega) = \frac{p_{i_1} q_{i_1, i_2} r_{i_1, i_2, i_3}}{(n-3)!},$$

where p_i are positive numbers adding up to one, $\mathbf{Q} = [q_{ij}]$ is a stochastic matrix with zero diagonal, and for each fixed i and j , the numbers $r_{i, j, l}$ are non-negative and add up to one and become zero if any two of the subscripts become equal. Let $X(\omega) = x_i$ and $Z_k(\omega) = A_k$ where the first element of ω is i and the first k elements of ω make up the set A_k . Then

$$\mathcal{U}_{k, n} := E(X | Z_k), \quad k = 2, 3, \dots, n-1,$$

forms a reverse martingale type of structure provided that conditional expectations are well defined. And for any convex function f over K , we have

$$\sum_{j=1}^n p_j f(x_j) \geq Ef(\mathcal{U}_{k, n}) \geq Ef(\mathcal{U}_{k+1, n}) \geq f\left(\sum_{j=1}^n p_j x_j\right), \quad k = 3, \dots, n-1.$$

Furthermore, the above inequalities hold for $k = 2$ provided $r_{i, j, l}$ is symmetric in its first two subscripts for each fixed l .

Remark. We may carry the above argument further and consider the probability measure

$$P(\omega) = p_{i_1} q_{i_1, i_2} \prod_{l=3}^n r_{i_l}^l(i_1, i_2, \dots, i_{l-1}),$$

where $r_{i_l}^l$ is a symmetric function of its $l-1$ coordinates. This will again provide refinements of Jensen's inequality. For the general case, any probability measure over the set of permutations satisfying the property,

$$\begin{aligned} &P(X = x_i | Z_{k+1} = B_{k+1}) \\ &= \sum_{A_k: i \in A_k \subset B_{k+1}} P(X = x_i | Z_k = A_k) P(Z_k = A_k | Z_{k+1} = B_{k+1}), \end{aligned}$$

will give rise to refinements of the Jensen inequality. Some special cases are provided in the next section.

4. SOME INEQUALITIES IN ANALYSIS

In the following we provide only a few examples to show the uses of these inequalities. Our first application deals with convex functions (such as sublinear functionals) over a linear space.

THEOREM 3. *Let f , \mathbf{P} , and \mathbf{Q} be as given in Theorem 1 and let the range of f be an interval I . Let $\phi: I \rightarrow \mathbb{R}$ be a non-decreasing and convex function. Let $\phi_{k,n}$ be defined either by*

$$\phi_{k,n} := \frac{1}{\binom{n-2}{k-2}} \sum_{A_k} \left(\sum_{i,j \in A_k} p_i q_{ij} \right) \phi \circ f \left(\frac{\sum_{i,j \in A_k} x_i p_i q_{ij}}{\sum_{i,j \in A_k} p_i q_{ij}} \right),$$

or defined by

$$\phi_{k,n} := \phi \left(\frac{1}{\binom{n-2}{k-2}} \sum_{A_k} \left(\sum_{i,j \in A_k} p_i q_{ij} \right) f \left(\frac{\sum_{i,j \in A_k} x_i p_i q_{ij}}{\sum_{i,j \in A_k} p_i q_{ij}} \right) \right).$$

Then, in both cases, we have

$$\sum_{k=1}^n p_k \phi(f(x_k)) \geq \phi_{k,n} \geq \phi_{k+1,n} \geq \phi \circ f \left(\sum_{k=1}^n p_k x_k \right), \quad k = 2, 3, \dots, n-1.$$

Proof. The first case is a direct consequence of Theorem 1 since $\phi \circ f$ is convex. For the second case, let X and Z_k be as defined in the proof of Theorem 1:

$$\begin{aligned} E(\phi \circ f(X)) &\geq E(\phi E(f(X) | Z_k)) \\ &\geq \phi(Ef(EX | Z_k)) = \phi(f_{k,n}(x, \mathbf{P}, \mathbf{Q})) \\ &\geq \phi(f_{k+1,n}(x, \mathbf{P}, \mathbf{Q})) \geq \phi \circ f(E(X)). \quad \blacksquare \end{aligned}$$

Remark. In particular, when \mathbf{Q} has constant rows consisting of $1/(n-1)$ in the off-diagonals and zero in the diagonal then the first form of $\phi_{k,n}$ reduces to

$$\begin{aligned} \phi_{k,n}(x, \mathbf{P}) &= \frac{1}{\binom{n-1}{k-1}} \sum_{A_k} \left(\sum_{i \in A_k} p_i \right) \phi \circ f \left(\frac{\sum_{i \in A_k} x_i p_i}{\sum_{i \in A_k} p_i} \right), \\ &k = 1, 2, \dots, n-1. \end{aligned} \tag{1}$$

When f is the usual norm of a normed linear space in (1), we get refinements of an inequality as given in [24, p.133, Theorem 4.46]. Furthermore, if ϕ is the identity, we get

$$\sum_{i=1}^n p_i f(x_i) \geq f_{k,n}(x, p) \geq f_{k+1,n}(x, p) \geq f\left(\sum_{i=1}^n p_i x_i\right), \quad k=1, 2, \dots, n-1, \quad (2)$$

where

$$f_{k,n}(x, p) := \frac{1}{\binom{n-1}{k-1}} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} (p_{i_1} + \dots + p_{i_k}) f\left(\frac{p_{i_1} x_{i_1} + \dots + p_{i_k} x_{i_k}}{p_{i_1} + \dots + p_{i_k}}\right). \quad (3)$$

In the language of PPS sampling, this comes out of the fact that the probability measure P over the set of permutations has the following form. Our random element Z_k , as defined in Section 2, has distribution

$$P(Z_k = \{i_1, \dots, i_k\}) = \frac{p_{i_1} + \dots + p_{i_k}}{\binom{n-1}{k-1}}.$$

This happens to be a well-known sampling scheme due to Midzuno [22]. It has been extensively studied (see, for instance, [4, 25, 27]) in the Sampling Theory literature. Its draw-by-draw version says that we select the first unit by the probability distribution p_i , $i=1, 2, \dots, n$, and then the remaining units are picked with equal probabilities over the remaining units one after the after. When the probability vector (p_1, p_2, \dots, p_n) is taken to be uniform, this leads to the following special case:

$$f_{k,n}(x) \geq f_{k+1,n}(x), \quad k=1, 2, \dots, n-1, \quad (4)$$

where

$$f_{k,n}(x) := \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} f\left(\frac{1}{k}(x_{i_1} + x_{i_2} + \dots + x_{i_k})\right). \quad (5)$$

And the corresponding probability measure is known as simple random sampling without replacement. The second form of $\phi_{k,n}$ in Theorem 3 has similar structure as used by R. A. Khan in [19]; however, k is held fixed and n is allowed to vary.

Remark. When K is taken to be the real line, the inequalities in (2) were directly proved recently in [23] while improving upon the inequalities (4) which are due to S. Gabler [12]. As pointed out by Gabler, the inequalities (4) contain the corresponding monotonicity results for the usual arithmetic and geometric means. It also contains similar inequalities, such as

$$\frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \prod_{j=1}^k x_{i_j}^{-a} \geq n^{ka}; \quad \text{whenever } \sum_{i=1}^n x_i \leq 1, \quad x_i \geq 0, \quad a > 0,$$

which is stated in [20, p. 85]. We should add here that refinements of Hadamard inequalities are also captured by Theorem 3. These results are possible due to the fact that for convex function f , $f_{k,n}(x)$ becomes a Schur convex function. The main results of Gabler, however, dealt with the, so-called, sequentially convex functions which are defined as follows. We will use such functions along with inequalities (2) and (4) to get a number of results in approximation theory in the next section.

DEFINITION. For a given function f defined over an interval I , if $f_{k,n}(x)$, $k = 1, 2, \dots, n$, as defined in (5), is a convex sequence in k for all $n \geq 3$, then f will be called a *sequentially convex* function.

Remark. Gabler showed that when $I = \mathbb{R}$ then a continuous f is sequentially convex if and only if f is twice differentiable and both f and f'' are convex functions. Furthermore, if f is twice differentiable over an interval I then the convexity of f and f'' implies the sequential convexity of f over I . He also showed that the converse of the last statement may not hold in general. Now we present a few results concerning divided differences.

For a real-valued function f defined over an interval I , and any distinct values $y_i \in I$, $i = 0, 1, 2, \dots, m$, let

$$D_m(\mathbf{y}) := [y_0, y_1, \dots, y_m; f], \quad \mathbf{y} = (y_0, y_1, \dots, y_m),$$

be the m th order divided difference. We say f is m -convex if $D_m(\mathbf{y}) \geq 0$ for all choices of distinct points $y_i \in I$, $i = 0, 1, 2, \dots, m$.

THEOREM 4. Let $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}$, be vectors in $I_{m+1} = I \times I \times \dots \times I$. Let f be an $(m+2)$ -convex function over I . For any probability vector \mathbf{P} and $n \times n$ stochastic matrix \mathbf{Q} as defined in Theorem 1, let

$$D_{k,n}(x, \mathbf{P}, \mathbf{Q}) := \frac{1}{\binom{n-2}{k-2}} \sum_{A_k} \left(\sum_{i,j \in A_k} p_i q_{ij} \right) D_m \left(\frac{\sum_{i,j \in A_k} \mathbf{x}^{(i)} p_i q_{ij}}{\sum_{i,j \in A_k} p_i q_{ij}} \right).$$

Then the sequence $\{D_{k,n}\}$ is non-increasing in $k \in \{1, 2, \dots, n\}$ where

$$D_{1,n} := \sum_{i=1}^n p_i D_m(\mathbf{x}^{(i)}).$$

Also, let $y_i \in I$, $i=0, 1, \dots, m$, be distinct points in I , and let

$$g_{k,m} := \frac{1}{\binom{m}{k} \binom{m+1}{k}} \sum_{0 \leq i_1 < i_2 < \dots < i_{m+1-k} \leq m} [y_{i_1}, y_{i_2}, \dots, y_{i_{m+1-k}}; f^{(k)}].$$

Then we have $[y_0, y_1, \dots, y_m; f] \leq g_{k,m} \leq g_{k+1,m}$, $k=1, 2, \dots, m-1$.

Proof. One need only see that the m th order divided difference of an $(m+2)$ -convex function is a convex function over I_{m+1} . An application of Theorem 1 gives the first result. When f is $(m+2)$ -convex, it is m times differentiable and, by the Hermite–Genocchi formula, we may write the divided differences in terms of B -splines as

$$[y_0, y_1, \dots, y_m; f] = \frac{1}{m!} E f^{(m)} \left(\sum_{i=0}^m y_i U_i \right),$$

where U_i , $i=0, 1, 2, \dots, m$, are uniformly distributed over the standard simplex of degree m . Since, U_i , $i=0, 1, 2, \dots, m$, are non-negative random variables which almost surely add up to one and $f^{(m)}$ is a convex function, an application of inequality (2) gives that

$$\begin{aligned} & [y_0, y_1, \dots, y_m; f] \\ & \leq \frac{1}{m! \binom{m+1-1}{m+1-k-1}} E \sum_{A_{m+1-k}} \left(\sum_{i \in A_{m+1-k}} U_i \right) f^{(m)} \left(\frac{\sum_{i \in A_{m+1-k}} y_i U_i}{\sum_{i \in A_{m+1-k}} U_i} \right) \\ & \leq \frac{1}{m! \binom{m+1-1}{m-k-1}} E \sum_{A_{m-k}} \left(\sum_{i \in A_{m-k}} U_i \right) f^{(m)} \left(\frac{\sum_{i \in A_{m-k}} y_i U_i}{\sum_{i \in A_{m-k}} U_i} \right), \end{aligned}$$

where \sum_{A_k} represents summing over all subsets of size k taken from $\{0, 1, 2, \dots, m\}$. Now we use the fact that

$$\left(\frac{U_{i_1}}{U_{i_1} + U_{i_2} + \dots + U_{i_j}}, \frac{U_{i_2}}{U_{i_1} + U_{i_2} + \dots + U_{i_j}}, \dots, \frac{U_{i_{j-1}}}{U_{i_1} + U_{i_2} + \dots + U_{i_j}} \right)$$

is independent of $U_{i_1} + U_{i_2} + \dots + U_{i_j}$, and that $E(U_i) = 1/(m+1)$. This gives that

$$\begin{aligned}
 & [y_0, y_1, \dots, y_m; f] \\
 & \leq \frac{m-k+1}{m! \binom{m}{m-k} (m+1)^{A_{m-k+1}}} \sum Ef^{(m)} \left(\frac{\sum_{i \in A_{m+1-k}} y_i U_i}{\sum_{i \in A_{m+1-k}} U_i} \right) \\
 & = \frac{1}{k! \binom{m}{k} \binom{m+1}{k} (m+1)^{A_{m-k+1}}} \sum \frac{1}{(m-k)!} Ef^{(m)} \left(\frac{\sum_{i \in A_{m+1-k}} y_i U_i}{\sum_{i \in A_{m+1-k}} U_i} \right) = g_{k,m} \\
 & \leq \frac{m-k}{m! \binom{m+1-1}{m-k-1} (m+1)^{A_{m-k}}} \sum Ef^{(m)} \left(\frac{\sum_{i \in A_{m-k}} y_i U_i}{\sum_{i \in A_{m-k}} U_i} \right) \\
 & = \frac{1}{\binom{m+1}{k+1} \binom{m}{k+1} (k+1)! (m+1)^{A_{m-k}}} \sum \frac{1}{(m-k-1)!} Ef^{(m)} \left(\frac{\sum_{i \in A_{m-k}} y_i U_i}{\sum_{i \in A_{m-k}} U_i} \right) \\
 & = g_{k+1,m}.
 \end{aligned}$$

This completes the proof. ■

Remark. Theorem 4 has some overlap with the results in [11]. The point of view of Theorem 4 could partially be carried over to multivariate B -splines as well. Now we turn our attention towards the monotonicity results concerning approximation operators of probability type.

5. MONOTONICITY OF APPROXIMATION OPERATORS

In this section we provide a number of results about the monotonic and convex convergence of classical approximation operators.

Monotonicity of Feller Operators

This example deals with the Weierstrass, Szasz, Bernstein, Gamma, Baskakov, and many other approximation operators of Feller type [18]. Let f be a real-valued function over an interval I . Let X_1, X_2, \dots be a sequence of identically distributed random variables taking values in I . Take $S_n = X_1 + \dots + X_n$. Often these random variables are assumed to be independent. However, it seems that the need for independence can be

somewhat relaxed while proving the monotonicity property of the Feller operator. Consider the sequence of functionals

$$L_n(f) := Ef(S_n/n), \quad \text{if for all } n, E|f(S_n/n)| < \infty.$$

For a list of special cases and general approximation results of L_n , see [18]. An easy application of inequality (4) when f is convex gives that

$$f(S_{n+1}/(n+1)) \stackrel{a.s.}{\leq} \frac{1}{\binom{n+1}{n}} \sum_{1 \leq i_1 < \dots < i_n \leq n+1} f\left(\frac{X_{i_1} + \dots + X_{i_n}}{n}\right). \quad (6)$$

Taking expectations on both sides gives that $L_{n+1}(f) \leq L_n(f)$. The two benefits one gets by employing this point of view are that we need not assume the existence of moments nor do we have to assume the mutual independence of the random variables involved. As a simple application of this result we get the monotonicity result concerning the Stancu–Bernstein operators [26] associated with the Polya urn model:

$$S_n(f, x) = Ef(S_n/n), \quad S_n = X_1 + X_2 + \dots + X_n,$$

where

$$P(S_n = k) = \binom{n}{k} \frac{x^{(k, -\alpha)}(1-x)^{(n-k, -\alpha)}}{1^{(n, -\alpha)}},$$

where $x^{(k, -\alpha)} = x(x+\alpha) \dots (x+(k-1)\alpha)$ for $k \geq 1$ and $x^{(0, -\alpha)} = 1$. In this case, the X_i are not mutually independent, but form a sequence of exchangeable random variables. Inequality (6) shows that $S_n(f, x) \geq S_{n+1}(f, x)$ for any convex f over $[0, 1]$. We have come to know that the monotonicity of Bernstein–Stancu operators was also proved by Horova and Budikova [14] recently; however, they used direct calculations. M. K. Khan *et al.* [17] provide some converse results concerning the monotonicity of operators of probability type. Now we mix inequality (6) with more information about the function f to obtain sharper results about the sequence of approximation operators. This is summed up in the following theorem.

THEOREM 5. *For the Feller functionals of a convex function f , we have $L_n(f)$ is a decreasing sequence in n . Furthermore, if f is sequentially convex then $L_n(f)$ is a convex sequence in n .*

Proof. The first part of the result follows by inequality (6) as described above. Now for the second part, just note that the definition of sequential convexity implies that

$$f_{k-2,n}(X_1, X_2, \dots, X_n) - 2f_{k-1,n}(X_1, X_2, \dots, X_n) + f_{k,n}(X_1, X_2, \dots, X_n) \stackrel{a.s.}{\geq} 0, \quad (7)$$

where

$$f_{k,n}(X_1, X_2, \dots, X_n) = \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} f\left(\frac{X_{i_1} + X_{i_2} + \dots + X_{i_k}}{k}\right).$$

Replacing k by n in inequality (7), taking expectations, and using the exchangeability of random variables finishes the proof. ■

Remark. For the Bernstein polynomials, there are two results in the literature which give similar results as our Theorem 5. One is due to Aramă and Ripianu [3] and the other is due to Horova [13]. However, both of these results are obtained after assuming some unnecessary conditions on f . In [3] the convex function f is assumed to be analytic in $[0, 1]$ having all derivatives of order two and higher being non-negative. And [13] assumes that f should be 2, 3, 4, and 5-convex over $[0, 1]$. By the results of [12], both of these results are special cases of Theorem 5. We should remark that, for instance, in the Bernstein polynomials, $B_n(f, x)$, when f is assumed to be 4-convex, then

$$B_{n+1}(f, x) \leq B_n(f, x) - \frac{\sigma^2(x)}{n(n+1)} \left[\frac{nx-x}{n}, \frac{nx-x+1}{n+1}, \frac{nx-x+1}{n}; f \right],$$

where $\sigma^2(x)$ is the variance of the Bernoulli random variable $X_1 \sim B(1, x)$. This then gives that

$$\lim_{n \rightarrow \infty} n^2(B_n(f, x) - B_{n+1}(f, x)) = \frac{\sigma^2(x) f''(x)}{2!}.$$

These results come from the divided difference representation of $B_{n+1}(f, x) - B_n(f, x)$, as first derived in [28] and the fact that when f is 4-convex then $[u, v, w; f]$ is a convex function over the unit cube. Similar results are now possible by the same argument when applied to other positive linear operators such as Szasz, Baskakov operators, etc. Their divided difference representations can be found in [9] and further references therein. We omit the details.

Monotonicity of Some Non-Feller Operators

The above results carry over to some non-Feller operators without much effort. As an example, consider the F operator (also known as the beta operator, see [16, 29]),

$$F_{m,n}(f, x) := \int_0^{\infty} \frac{x^n u^{n-1}}{B(m, n)(1+xu)^{m+n}} f\left(\frac{n}{mu}\right) du,$$

where $B(m, n)$ is the beta function, $m, n \geq 1$, $x > 0$, and $F_{m,n}(|f|, x) < \infty$. It can be shown that

$$F_{m,n}(f, x) = Ef\left(\frac{nx(X_1 + \dots + X_{2m})}{m(Y_1 + \dots + Y_{2n})}\right),$$

where X_i, Y_i are iid chi square random variables with one degree of freedom. Again, let $S_n = X_1 + \dots + X_n$ and $T_n = Y_1 + \dots + Y_n$. For a convex function f in the domain of the operator, inequality (6) gives that,

$$f(cS_{2m+2}/(2m+2)) \stackrel{a.s.}{\leq} \frac{1}{\binom{2m+2}{2m}} \sum_{1 \leq i_1 < \dots < i_{2m} \leq 2m+2} f\left(\frac{c(X_{i_1} + \dots + X_{i_{2m}})}{2m}\right),$$

where $c = 2xn/T_{2n}$. Taking expectations gives that $F_{m+1,n}(f, x) \leq F_{m,n}(f, x)$. When f is sequentially convex then we see that

$$f_{2m-4, 2m}(x) - 2f_{2m-2, 2m}(x) + f_{2m, 2m}(x) \geq 0.$$

This then implies that $F_{m,n}(f, x)$ is a convex sequence in m . Hence, we have proved the following theorem.

THEOREM 6. *For any convex function, f , in the domain of the F -operators, we have $F_{m+1,n}(f, x) \leq F_{m,n}(f, x)$. And when f is sequentially convex then*

$$F_{m+2,n}(f, x) - 2F_{m+1,n}(f, x) + F_{m,n}(f, x) \geq 0, \quad m, n = 1, 2, 3, \dots$$

Remark. We should remark here that similar ideas can be used to provide partial monotonicity results for the Schurer versions of Bernstein and Szasz operators (cf. [2, pp. 338, 341]).

Monotonicity of Operators in \mathbb{R}^k

Many of the classical Feller type operators can be generalized to \mathbb{R}^k . For instance, the Bernstein operator over a simplex is defined as follows. Let $S_{n,x} = (S_{n,x_1}, \dots, S_{n,x_k})$ have a multinomial distribution with parameters $(n, x_1, \dots, x_k, 1 - x_1 - \dots - x_k)$, where $x = (x_1, \dots, x_k) \in \Delta_k =$ the standard

simplex. That is, $\Delta_k = \{(x_1, \dots, x_k) : 0 \leq x_i \leq 1, x_1 + \dots + x_k \leq 1\}$. For a continuous function f defined over Δ_k , the Bernstein operator is defined by

$$B_{n,k}(f, x) = \sum_j f(j/n) P(S_{n,x} = j),$$

where $j = (j_1, \dots, j_k)$ consists of non-negative integers so that $j_1 + \dots + j_k \leq n$, and

$$P(S_{n,x} = j) = \binom{n}{j_1, \dots, j_k, n - j_1 - \dots - j_k} (1 - x_1 - \dots - x_k)^{n - j_1 - \dots - j_k} \prod_{i=1}^k x_i^{j_i}$$

is the multinomial density. We can view this and other such operators as

$$L_{n,k}(f, x) = Ef\left(\frac{S_1}{n}, \frac{S_2}{n}, \dots, \frac{S_k}{n}\right) = Ef\left(\frac{1}{n} \sum_{i=1}^n X_i\right),$$

where $X_i = (X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(k)})$, $i = 1, 2, \dots$, is a sequence of iid random vectors. The reverse martingale method carries over to \mathbb{R}^k (cf. [1]) and provides the monotonicity of such operators when f is convex. However, by using a special case of Theorem 1, we can say more without using expectations or the mutual independence of the random vectors. The following theorem is a direct consequence of the inequalities in (4) when applied to random vectors.

THEOREM 7. *Let f be convex function over a convex set $K \subseteq \mathbb{R}^k$ and let*

$$L_{n,k}(f) = Ef\left(\frac{S_1}{n}, \frac{S_2}{n}, \dots, \frac{S_k}{n}\right) = Ef\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$$

be a sequence of functionals over K , where X_i is a sequence of K valued random vectors. Then

$$f\left(\sum_{i=1}^{n+1} X_i / (n+1)\right) \stackrel{a.s.}{\leq} \frac{1}{\binom{n+1}{n}} \sum_{1 \leq i_1 < \dots < i_n \leq n+1} f\left(\frac{X_{i_1} + \dots + X_{i_n}}{n}\right).$$

Furthermore, if X_i are exchangeable random vectors then $L_{n,k}(f) \geq L_{n+1,k}(f)$.

Remark. This result contains a number of classical multivariate approximation operators such as the Bernstein polynomials over simplexes (first proved in [7] by using direct calculations) and the multivariate Baskakov operators. The standard proof involves the multivariate analog of reverse martingale argument as provided in [1] recently. Theorem 7

shows that, almost surely, we may compare the function values before even taking the expectations. In fact, we could go one step further and define sequentially convex functions over a convex subset K of R^m by having $f_{k,n}$, as defined in (5), be convex in $k=1, 2, \dots, n$ for all points $x_i \in K$ and all positive integers $n > 2$. For such functions, the above multivariate operators will form a convex sequence of approximations. The tensor product operators, on the other hand, can be handled without such extensions. The main result in this direction is the following theorem. For notational convenience, we present the results for \mathbb{R}^2 only.

THEOREM 8. *Let X_i and Y_j , $i, j=1, 2, \dots$, be two sequences of random variables taking values in respective intervals I and J . Consider the functionals*

$$T_{n,m}(f) = Ef\left(\frac{1}{n} \sum_{i=1}^n X_i, \frac{1}{m} \sum_{j=1}^m Y_j\right)$$

when the expectations are well defined. Let $f(x, y)$ be a convex function in x for each fixed value of $y \in J$. If $\{X_1, X_2, \dots\}$ are conditionally exchangeable random variables given the random variables Y_1, Y_2, \dots, Y_m , then $T_{n,m}(f) \geq T_{n+1,m}(f)$. And if $f(x, y)$ is sequentially convex in x for each fixed y , then

$$T_{n+2,m}(f) - 2T_{n+1,m}(f) + T_{n,m}(f) \geq 0, \quad n = 1, 2, \dots$$

Proof. The proof follows from the fact that

$$\begin{aligned} & f\left(\frac{1}{n+1} \sum_{i=1}^{n+1} X_i, \frac{1}{m} \sum_{j=1}^m Y_j\right) \\ & \leq \frac{1}{\binom{n+1}{n}} \sum_{1 \leq i_1 < \dots < i_n \leq n+1} f\left(\frac{X_{i_1} + X_{i_2} + \dots + X_{i_n}}{n}, \frac{1}{m} \sum_{j=1}^m Y_j\right) \end{aligned}$$

when $f(x, y)$ is convex in x for each $y \in J$. Taking expectations gives the first part. For the second part, the sequential convexity of $f(x, y)$ for each fixed $y \in J$ gives the corresponding inequality similar to (7). And taking iterated expectations, first with respect to $\{X_i\}$ given Y_1, Y_2, \dots, Y_m and then over the Y_j completes the proof. ■

Remark. Our final result shows how to use the ideas of sampling without replacement to cover results involving sampling with replacement. Let \mathcal{L} be a linear space, let $K \subseteq \mathcal{L}$ be a convex set, and let $g: K \rightarrow \mathbb{R}$ be

a convex function. Let x_1, x_2, \dots, x_n be some elements of K and let p_1, p_2, \dots, p_n be positive numbers adding up to one. Define

$$g_{k,n}(x, p) = \sum_{i_1} p_{i_1} \sum_{i_2} p_{i_2} \cdots \sum_{i_k} p_{i_k} g\left(\frac{x_{i_1} + x_{i_2} + \cdots + x_{i_k}}{k}\right).$$

In [24, p. 90, Theorem 3.21] we find the following inequalities:

$$\sum_{i=1}^n p_i g(x_i) \geq g_{k,n} \geq g_{k+1,n} \geq g\left(\sum_{i=1}^n p_i x_i\right), \quad k = 2, 3, \dots, n-1. \quad (8)$$

The following proof shows why this is also a special case of Theorem 1.

Proof of (8). Let X_1, X_2, \dots, X_n be independent and identically distributed K valued random elements so that $P(X_1 = x_j) = p_j$, $j = 1, 2, \dots, n$. And let N_k be independent of X_1, X_2, \dots, X_n and let N_k be uniform over $\{1, 2, \dots, k\}$. Consider the composition random vector $W_k = X_{N_k}$. Since, $P(W_k = x_j) = p_j$, we see that for any $k = 2, 3, \dots, n-1$,

$$\begin{aligned} \sum_{i=1}^n p_i g(x_i) &= Eg(W_k) \geq E\{g(E(W_k | X_1, X_2, \dots, X_k))\} \\ &= E\left\{g\left(\frac{X_1 + \cdots + X_k}{k}\right)\right\} = g_{k,n}(x, p) \\ &\geq E\left\{g\left(\frac{X_1 + \cdots + X_{k+1}}{k+1}\right)\right\} = g_{k+1,n}(x, p) \\ &\geq g\left(E\left(\frac{X_1 + \cdots + X_{k+1}}{k+1}\right)\right) = g\left(\sum_{i=1}^n p_i x_i\right). \quad \blacksquare \end{aligned}$$

ACKNOWLEDGMENTS

We would like to thank Professors Richard Aron and Philip Boland for their helpful comments and suggestions.

REFERENCES

1. J. A. Adell, J. de la Cal, and M. S. Miguel, On the property of monotonic convergence for multivariate Bernstein-type operators, *J. Approx. Theory* **80** (1995), 132–137.
2. F. Altomare and M. Campiti, “Korovking-Type Approximation Theory and Its Applications,” de Gruyter, Berlin/New York, 1994.
3. O. Aramă and D. Ripianu, Une propriété des polynomes de Bernstein, *Mathematica (Cluj)* **26**, No. 1 (1961), 5–18.

4. M. S. Avadhani and A. K. Srivastava, A comparison of Midzuno–Sen scheme with PPS sampling without replacement and its application to successive sampling, *Ann. Inst. Statist. Math.* **24** (1972), 153–164.
5. H. Bolfarine and S. Zacks, “Prediction Theory for Finite Populations,” Springer-Verlag, Berlin/New York, 1992.
6. K. R. W. Brewer and M. Hanif, “Sampling with Unequal Probabilities,” Lecture Notes in Statistics, Vol. 15, Springer-Verlag, Berlin/New York, 1983.
7. G. Chang and P. J. Davis, The convexity of Bernstein polynomials over triangles, *J. Approx. Theory* **40** (1984), 11–28.
8. A. Chaudhuri and J. W. E. Vos, “Unified Theory and Strategies of Survey Sampling,” North-Holland, Amsterdam, 1988.
9. B. Della Vecchia, On the monotonicity of the derivatives of the sequences of Favard and Baskakov operators, *Ricerche Mat.* **36** (1987), 263–269.
10. R. M. Dudley, “Real Analysis and Probability,” Wadsworth & Brooks/Cole, Belmont, 1989.
11. R. Farwig and D. Zwick, Some divided difference inequalities for n -convex functions, *J. Math. Anal. Appl.* **108** (1985), 430–437.
12. S. Gabler, Folgenkonvexe Funktionen, *Manuscripta Math.* **29** (1979), 29–47.
13. I. Horova, Bernstein polynomials of convex functions, *Mathematica (Cluj)* **33**, No. 2 (1968), 265–273.
14. I. Horova and M. Budikova, A note on D. D. Stancu operators, *Ricerche Mat.* (1995).
15. W. Hoeffding, A class of statistics with asymptotically normal distribution, *Ann. Math. Stat.* **19** (1948), 293–325.
16. M. K. Khan, Approximation properties of beta operators, in “Prog. Approx. Theory” (P. Nevai and A. Pinkus, Eds.), pp. 483–495, Academic Press, San Diego, 1991.
17. M. K. Khan, B. Della Vecchia, and A. Fassih, On the monotonicity of positive linear operators, *J. Approx. Theory* (to appear).
18. R. A. Khan, Some probabilistic methods in the theory of approximation operators, *Acta. Math. Hungar.* **39** (1980), 193–203.
19. R. A. Khan, Reverse Martingales and approximation operators, *J. Approx. Theory* **80** (1995), 367–377.
20. A. W. Marshall and I. Olkin, “Inequalities: Theory of Majorization and Its Applications,” Academic Press, New York, 1979.
21. A. W. Marshall and F. Proschan, An inequality for convex functions involving majorization, *J. Math. Anal. Appl.* **12** (1965), 87–90.
22. H. Midzuno, On the sampling system with probability proportional to sum of sizes, *Ann. Inst. Statist. Math.* **3** (1952), 99–107.
23. J. E. Pečarić, Remark on an inequality of S. Gabler, *J. Math. Anal. Appl.* **184** (1994), 19–21.
24. J. E. Pečarić, F. E. Proschan, and Y. L. Tong, “Convex Functions, Partial Orderings, and Statistical Applications,” Academic Press, San Diego, 1992.
25. R. Singh, A note on the efficiency of ratio estimate with Midzuno’s scheme of sampling, *Sankhya, Ser. C* **37** (1975), 211–214.
26. D. D. Stancu, Approximation of functions by a new class of linear polynomial operators, *Rev. Roumaine Math. Pures Appl.* **8** (1968), 1173–1194.
27. P. V. Sukhatme and B. V. Sukhatme, “Sampling Theory of Surveys with Applications,” Iowa State University Press, Ames, 1970.
28. W. B. Temple, Stieltjes integral representation of convex functions, *Duke Math. J.* **21** (1954), 527–531.
29. R. Upreti, Approximation properties of beta operators, *J. Approx. Theory* **45** (1985), 85–89.