Stochastic comparisons of symmetric sampling designs

Larry Goldstein
Department of Mathematics
University of Southern California
Kaprielian Hall, Room 108
3620 Vermont Avenue
Los Angeles, CA 90089-2532, USA
larry@math.usc.edu

Yosef Rinott*
Department of Statistics
and Center for the Study of Rationality
Hebrew University of Jerusalem
Mount Scopus
Jerusalem 91905, Israel
and LUISS, Roma
rinott@mscc.huji.ac.il

Marco Scarsini†
Dipartimento di Scienze Economiche e Aziendali
LUISS
Viale Romania 12
I–00197 Roma, Italy
and HEC, Paris
marco.scarsini@luiss.it

January 10, 2011

*Partially supported by the Israel Science Foundation grant No. 473/04
†Partially supported by MIUR-COFIN

Abstract

We compare estimators of the integral of a monotone function f that can be observed only at a sample of points in its domain, possibly with error. Most of the standard literature considers sampling designs ordered by refinements and compares them in terms of mean square error or, as in Goldstein, Rinott, and Scarsini (2010), the stronger convex order. In this paper we compare sampling designs in the convex order without using partition refinements. Instead we order two sampling designs based on partitions of the sample space, where a fixed number of points is allocated at random to each partition element. We show that if the two random vectors whose components correspond to the number allocated to each partition element are ordered by stochastic majorization, then the corresponding estimators are likewise convexly ordered. If the function f is not monotone, then we show that the convex order comparison does not hold in general, but a weaker variance comparison does.

AMS 2000 Subject Classification: Primary 60E15. Secondary 62D05, 65C05.

Keywords: convex order, stochastic majorization, stratified sampling, exchangeable partitions of integers.

1 Introduction

Consider a function f whose values can be obtained only via costly experiments, so that budget constraints limit the number of points where the function can be evaluated. Monte Carlo randomization is a standard way to choose these points, which can either be sampled totally at random, or using some stratification. When properly carried out, stratification is known to improve the performance of estimators; see, for example, Glasserman (2004).

If the object of interest is the integral $I(f) = \mathbb{E}[f(U)]$, where U is a given random variable, then it is easy to construct unbiased estimators of $I(f)$ by using different stratified samples. In much of the literature estimators are compared in terms of a given loss function, which may be arbitrary. Typically the loss function is quadratic, so the criterion is the mean square error, i.e., the variance, when the estimator is unbiased.

When stratified sampling is used, a partition $\mathscr{A}$ of $\mathfrak{U}$, the domain of f, is constructed and observations are drawn from each element A_i of the partition in proportion to the probability $\mathbb{P}(U \in A_i)$. If another partition $\mathscr{B}$ is obtained by refining $\mathscr{A}$, i.e., by breaking the elements of $\mathscr{A}$, and sampling accordingly, then it is well known that the variance of natural unbiased estimators of $I(f)$ decreases, but, as Goldstein et al. (2010) show, for natural unbiased estimators W such as (2.1) below, refinement of the stratification does not necessarily reduce $\mathbb{E}[|W - I(f)|^p]$, for $p \neq 2$. However, in some circumstances stratified sampling is better not just in $L2$, but in terms of the convex order, which in turn implies that it is better in Lp for every $p \geq 1$. For instance this happens when f is a monotone function, even if it is observed with a random error.

The current paper is a follow-up of Goldstein et al. (2010), where references on stratified sampling, estimation of functionals of unknown functions, and a discussion of the convex order can be found. The latter paper contains comparisons of estimators based on monotone partitions and their refinements in terms of convex order, where the function f is monotone. The purpose of the present paper is to show that the convex order can be used also to compare sampling designs that are not ordered by refinements. We consider a monotone function f and unbiased estimators of $I(f)$ based on sampling designs of the following type. A partition $(A_1, \dots, A_n)$ of $\mathfrak{U}$ satisfying $P(U \in A_i) = 1/n$ for $i = 1, \dots, n$ is fixed and a random sample of size K_i is drawn from each A_i, where $\boldsymbol{K} = (K_1, \dots, K_n)$ is an exchangeable random vector. We consider two sampling designs corresponding to vectors $\boldsymbol{K}$ and $\boldsymbol{L}$ and show that if they are ordered by stochastic majorization, then the corresponding estimators are likewise convexly ordered.

The sampling method suggested here arises when it is not practical to sample one observation (or any fixed number) from each stratum in the partition, and instead samples are drawn from some strata, and not from others, chosen at random. In this case unbiased estimators can still be obtained by using a random exchangeable design $\boldsymbol{K}$ as described above, and they can be compared in terms of the convex order. Moreover the approach used in Goldstein et al. (2010) imposes some restrictions on the relation between the maximal number of elements of the partition and the size of the sample. Here instead, the total number of observations is arbitrary and does not depend in any way on the size of the partition.

If the function f is not monotone, the above mentioned result does not go through. We provide a counterexample. Nevertheless a weaker comparison in terms of variances holds in this general case.

The paper is organized as follows. Section 2 introduces some definitions and presents the main results. All proofs are in Section 3 together with some additional results that have some interest *per se*. Section 4 contains some numerical examples.

2 Symmetric designs

2.1 Definitions

We start with some definitions; throughout, all random quantities are defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Let $\mathfrak{U} \subset \mathbb{R}$ be a compact interval (without loss of generality choose $\mathfrak{U} = [0, 1]$), and for given positive integers n and N, let $\mathscr{K}$ be the class of vectors $\boldsymbol{k} = (k_1, \ldots, k_n)$ with nonnegative integer components such that $\sum_{i=1}^n k_i = N$. If $N = mn$ then clearly $m\mathbf{1}_n := m(1, \ldots, 1) \in \mathscr{K}$. Let $\boldsymbol{K}$ be an exchangeable random vector with support in $\mathscr{K}$; with an abuse of notation we write $\boldsymbol{K} \in \mathscr{K}$.

Definition 2.1. Given an exchangeable random vector $\boldsymbol{K} = (K_1, \ldots, K_n) \in \mathscr{K}$ and a partition $\mathscr{A} = (A_1, \ldots, A_n)$ of $\mathfrak{U}$ such that $\mathbb{P}(U \in A_i) = 1/n$ for $i = 1, \ldots, n$, the *associated symmetric random design* $\mathbb{K}$ is the design consisting of N independent random points V_{ij}, $j = 1, \ldots, K_i$, $i = 1, \ldots, n$ (at which $f(V_{ij})$ will be observed, possibly with noise), where $V_{i1}, \ldots, V_{iK_i}$ are i.i.d. random variables with distribution $P_{U|A_i}$, $i = 1, \ldots, n$, that is, $P(V_{ij} \in B) = P(U \in B | U \in A_i)$.

In other words, a sample having the design $\mathbb{K}$ associated with the exchangeable vector $\boldsymbol{K}$ can be realized by drawing a realization $\boldsymbol{k}$ of $\boldsymbol{K}$ and a random permutation Π of $(1, \ldots, n)$, and then sampling $k_{\Pi(i)}$ observations from $P_{U|A_i}$ for all i. Note that when $N = mn$, the associated symmetric random design $m\mathbf{1}_n$ samples m observations from each subset of a partition $\mathscr{A} = (A_1, \ldots, A_n)$ of $\mathfrak{U}$ such that $\mathbb{P}(U \in A_i) = 1/n$ for $i = 1, \ldots, n$. We denote this design by $m\mathbb{1}_n$.

Assume that for any v in a sample of points on $\mathfrak{U}$ we observe $f(v) + \varepsilon$ where ε is a mean zero independent random error. By exchangeability and the fact that the error ε has mean zero, it is easy to see that the estimator

$$W_{\mathbb{K}} = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{K_i} (f(V_{ij}) + \varepsilon_{ij}) \tag{2.1}$$

is unbiased for $I(f) := \mathbb{E}[f(U)]$.

For subsets G and H of the real line, we write $G \le H$ if $g \le h$ for every $g \in G$ and $h \in H$. We call a partition $\mathscr{B} = (B_1, \dots, B_b)$ of $\mathfrak{U}$ *monotone* if $B_1 \le \cdots \le B_b$.

Given two vectors $\boldsymbol{x} = (x_1, \dots, x_n)$, $\boldsymbol{y} = (y_1, \dots, y_n)$, we write $\boldsymbol{y} \prec \boldsymbol{x}$ if

$$\sum_{i=1}^{k} y_i^{\downarrow} \le \sum_{i=1}^{k} x_i^{\downarrow} \quad \text{for} \quad k = 1, \dots, n-1, \quad \text{and} \quad \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i,$$

where $y_1^{\downarrow} \ge \cdots \ge y_n^{\downarrow}$ is the decreasing rearrangement of $\boldsymbol{y}$, and analogously for $\boldsymbol{x}$. Clearly $m\mathbb{1}_n \prec \boldsymbol{k}$ for any vector $\boldsymbol{k}$ of length n with nonnegative components summing to nm.

A function $\psi : \mathbb{R}^n \to \mathbb{R}$ is called Schur convex if $\boldsymbol{y} \prec \boldsymbol{x}$ implies $\psi(\boldsymbol{y}) \le \psi(\boldsymbol{x})$. If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex then $\psi(\boldsymbol{x}) = \sum_{i=1}^{n} \varphi(x_i)$ is Schur convex. For the majorization order $\prec$ and properties of Schur convex functions see, e.g., Marshall and Olkin (1979).

Definition 2.2. Given two random vectors $\boldsymbol{X}, \boldsymbol{Y}$ we say that $\boldsymbol{Y} \le_{\mathrm{st}} \boldsymbol{X}$ if

$$\mathbb{E}[\phi(\boldsymbol{Y})] \le \mathbb{E}[\phi(\boldsymbol{X})] \tag{2.2}$$

for every nondecreasing function ϕ; we say that $\boldsymbol{Y} \le_{\mathrm{cx}} \boldsymbol{X}$ if (2.2) holds for every convex function ϕ; we say that $\boldsymbol{Y} \prec_{\mathrm{S}} \boldsymbol{X}$ if (2.2) holds for every Schur convex function ϕ.

Properties of the *stochastic order* $\le_{\mathrm{st}}$ and the *convex order* $\le_{\mathrm{cx}}$ are extensively studied in Müller and Stoyan (2002) and Shaked and Shanthikumar (2007). The statement $\boldsymbol{Y} \le_{\mathrm{st}} \boldsymbol{X}$ depends only on the marginal laws $\mathscr{L}(\boldsymbol{Y})$ and $\mathscr{L}(\boldsymbol{X})$, so sometimes we write $\mathscr{L}(\boldsymbol{Y}) \le_{\mathrm{st}} \mathscr{L}(\boldsymbol{X})$, and analogously for $\le_{\mathrm{cx}}$. Since an increasing function of a Schur convex function is also Schur convex, it is easy to see that $\boldsymbol{X} \prec_{\mathrm{S}} \boldsymbol{Y}$ is equivalent to $g(\boldsymbol{X}) \le_{\mathrm{st}} g(\boldsymbol{Y})$ for every Schur convex function g (see Nevius, Proschan, and Sethuraman, 1977, where the concept of *stochastic majorization* was introduced).

2.2 Main result

Recall Definition 2.1 of symmetric random designs.

Theorem 2.3. *Let f be a nondecreasing function on $\mathfrak{U}$ and $\mathscr{A} = (A_1, \dots, A_n)$ a monotone partition of $\mathfrak{U}$ satisfying $\mathbb{P}(U \in A_i) = 1/n$. Consider exchangeable vectors $\boldsymbol{L}, \boldsymbol{K} \in \mathscr{K}$ satisfying $\boldsymbol{L} \prec_{\mathrm{S}} \boldsymbol{K}$, and let $\mathbb{L}$ and $\mathbb{K}$ be their associated symmetric random designs. Then*

$$W_{\mathbb{L}} \le_{\mathrm{cx}} W_{\mathbb{K}}. \tag{2.3}$$

In particular, when $N = mn$, for any symmetric random design $\mathbb{K}$ we have

$$W_{m\mathbb{1}_n} \le_{\mathrm{cx}} W_{\mathbb{K}}.$$

A special case of (2.3) is obtained $\boldsymbol{L}$ and $\boldsymbol{K}$ are random permutations of some fixed $\boldsymbol{\ell}$ and $\boldsymbol{k}$ satisfying $\boldsymbol{\ell} \prec \boldsymbol{k}$. For example, if $n = N = 4$ and $\boldsymbol{\ell} = (2, 2, 0, 0)$, and $\boldsymbol{k} = (3, 1, 0, 0)$, then (2.3) provides a comparison that does not involve a refinement. Note that in the above monotone partition, the sets A_i are intervals whose end points are the $(i-1)/n$ and i/n quantiles of the random variable U.

A result similar to the above theorem holds also for non-monotone functions, provided that the stratification is into sets A_i such that the ranges $f(A_i)$ satisfy $f(A_{i_1}) \leq f(A_{i_2}) \leq \ldots \leq f(A_{i_n})$ for some permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$.

The notion of symmetric design involves additional randomization and is therefore different from the kind of sampling treated in Goldstein et al. (2010). Such symmetric randomization is required in order to preserve unbiasedness, a property shared by all the estimators of integrals considered here. The interpretation of Theorem 2.3 is that more balanced designs, in the sense defined by the majorization order, are better.

The last part of Theorem 2.3 shows that for a sample of size $N = mn$, a partition $\mathscr{A}$ into subsets of equal probability with a sample of m from each is best in the sense of convex order.

2.3 Integrals of nonmonotone functions

If we drop the hypothesis of monotononicity of the function f in Theorem 2.3, then the conclusion does not hold, as the following counterexample shows.

Example 2.4. Let $\mathfrak{U} = [0, 1]$, $n = 2$, $A_1 = [0, 1/2], A_2 = (1/2, 1]$, and let U have a uniform distribution on $[0, 1]$.Define

$$f(x) = 2I_{[0,1/2]}(x) + 4I_{(3/4,1]}(x).$$

Moreover let $\boldsymbol{L} = (1, 1)$ almost surely, and

$$\boldsymbol{K} = \begin{cases} (0, 2) & \text{with probability } 1/2, \\ (2, 0) & \text{with probability } 1/2. \end{cases}$$

Then $\boldsymbol{L} \prec_{\mathrm{S}} \boldsymbol{K}$ and $\int_0^1 f(x)\, \mathrm{d}x = 2$. For $N = 2$, we have

$$W_{\mathbb{L}} = \begin{cases} 1 & \text{with probability } 1/2, \\ 3 & \text{with probability } 1/2, \end{cases}$$

and

$$W_{\mathbb{K}} = \begin{cases} 0 & \text{with probability } 1/8, \\ 2 & \text{with probability } 3/4, \\ 4 & \text{with probability } 1/8. \end{cases}$$

This implies that $\mathbb{E}[|W_{\mathbb{L}} - 2|] = 1$ and $\mathbb{E}[|W_{\mathbb{K}} - 2|] = 1/2$, hence it is not true that $W_{\mathbb{L}} \leq_{\mathrm{cx}} W_{\mathbb{K}}$.

Nevertheless, the variances of the estimators exhibit monotonicity with respect to $\prec_{\mathrm{S}}$ without any monotonicity assumption on f.

Proposition 2.5. *Let $f : \mathfrak{U}$ be bounded and let $\mathscr{A} = (A_1, \ldots, A_n)$ be a monotone partition of $\mathfrak{U}$ satisfying $\mathbb{P}(U \in A_i) = 1/n$. Consider exchangeable vectors $\boldsymbol{L}, \boldsymbol{K} \in \mathscr{K}$ satisfying $\boldsymbol{L} \prec_{\mathrm{S}} \boldsymbol{K}$, and let $\mathbb{L}$ and $\mathbb{K}$ be their associated symmetric random designs. Then*

$$\mathrm{Var}[W_{\mathbb{L}}] \leq \mathrm{Var}[W_{\mathbb{K}}]. \tag{2.4}$$

Moreover, if $\mathbb{E}[f(U)|U \in A_i]$ does not depend on i, then $\mathrm{Var}[W_{\mathbb{L}}] = \mathrm{Var}[W_{\mathbb{K}}]$.

A similar phenomenon was described in Goldstein et al. (2010), where under a different ordering of designs, monotonicity of the variance does not require monotonicity of the function whose integral is estimated, but for the stronger convex ordering monotonicity is required.

3 Proofs and additional results

The following lemma is a very special case of a well-known result of Strassen (1965) (see also Alfsen, 1971; Lindvall, 2002).

Lemma 3.1. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be n-dimensional random vectors with compact support such that $\boldsymbol{X} \prec_{\mathrm{S}} \boldsymbol{Y}$. Then there exists a coupling of $(\boldsymbol{X}, \boldsymbol{Y})$ such that $\mathbb{P}(\boldsymbol{X} \prec \boldsymbol{Y}) = 1$, that is, a coupling where the vector $\boldsymbol{X}$ is majorized by $\boldsymbol{Y}$ with probability* 1.

The next lemma is a central step for the proof of Theorem 2.3.

Lemma 3.2. *Let $\xi_1, \ldots, \xi_n$ be independent random variables having supports that satisfy $\mathrm{supp}(\xi_1) \leq \mathrm{supp}(\xi_2) \leq \ldots \leq \mathrm{supp}(\xi_n)$, and let $\{\xi_{ij}\}_{i=1,\ldots,n,\, j=1,2,\ldots}$ be independent with $\mathscr{L}(\xi_{ij}) = \mathscr{L}(\xi_i)$. Consider two vectors having nonnegative integer components $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n)$ and $\boldsymbol{k} = (k_1, \ldots, k_n)$ such that $\boldsymbol{\ell} \prec \boldsymbol{k}$. Let Π be a random permutation uniformly distributed over S, the permutation group of $\{1, \ldots, n\}$, independent of $\{\xi_{ij}\}$. Define*

$$Z_{\boldsymbol{\ell}} = \sum_{i=1}^{n} \sum_{j=1}^{\ell_{\Pi(i)}} \xi_{ij} \quad \text{and} \quad Z_{\boldsymbol{k}} = \sum_{i=1}^{n} \sum_{j=1}^{k_{\Pi(i)}} \xi_{ij}. \tag{3.1}$$

Then

$$Z_{\boldsymbol{\ell}} \leq_{\mathrm{cx}} Z_{\boldsymbol{k}}$$

Proof. It is a well known fact in majorization (see, e.g., Hardy, Littlewood, and Pólya (1952, Proof of Lemma 2, p. 47) and Marshall and Olkin (1979, Lemma B.1, p. 21)) that to prove the lemma it suffices to consider $\boldsymbol{k}$ and $\boldsymbol{\ell}$ that differ in only two coordinates, and moreover, it is easy to see that it suffices to consider the special case where $\boldsymbol{k}$ and $\boldsymbol{\ell}$ satisfy

$$k_1 - 1 \geq k_2 + 1, \quad \ell_1 = k_1 - 1, \quad \ell_2 = k_2 + 1, \quad \ell_h = k_h, \text{ for } h = 3, \ldots, n. \tag{3.2}$$

For such $\boldsymbol{k}$ and $\boldsymbol{\ell}$ we may write the desired conclusion as

$$\sum_{i:\Pi(i)\in\{1,2\}} \sum_{j=1}^{\ell_{\Pi(i)}} \xi_{ij} + \sum_{i:\Pi(i)\notin\{1,2\}} \sum_{j=1}^{\ell_{\Pi(i)}} \xi_{ij} \leq_{\mathrm{cx}} \sum_{i:\Pi(i)\in\{1,2\}} \sum_{j=1}^{k_{\Pi(i)}} \xi_{ij} + \sum_{i:\Pi(i)\notin\{1,2\}} \sum_{j=1}^{k_{\Pi(i)}} \xi_{ij}.$$

By denoting the summands above by $A_\Pi, B_\Pi, C_\Pi, D_\Pi$, respectively, we may rewrite the desired inequality as

$$A_\Pi + B_\Pi \leq_{\mathrm{cx}} C_\Pi + D_\Pi,$$

or, by explicitly writing out the mixture over Π, equivalently as

$$\frac{1}{n!}\sum_{\pi\in S} \mathscr{L}(A_\pi + B_\pi) \leq_{\mathrm{cx}} \frac{1}{n!}\sum_{\pi\in S} \mathscr{L}(C_\pi + D_\pi).$$

In order to prove the relation above we pair the summands in the following way: each permutation π in this mixture is paired with the permutation σ for which $\sigma^{-1}(1) = \pi^{-1}(2), \sigma^{-1}(2) = \pi^{-1}(1)$ and $\sigma^{-1}(i) = \pi^{-1}(i)$ for all $i \notin \{1,2\}$. It is easy to see that it suffices to show for each given (nonrandom) such pair π, σ,

$$\frac{1}{2}\mathscr{L}(A_\pi + B_\pi) + \frac{1}{2}\mathscr{L}(A_\sigma + B_\sigma) \leq_{\mathrm{cx}} \frac{1}{2}\mathscr{L}(C_\pi + D_\pi) + \frac{1}{2}\mathscr{L}(C_\sigma + D_\sigma).$$

The above pairing and (3.2) readily imply that $B_\pi = B_\sigma = D_\pi = D_\sigma$, and they are all independent of the A's and C's. Therefore, since translations and sums of convex function are also convex, it now suffices to prove

$$\frac{1}{2}\mathscr{L}(A_\pi) + \frac{1}{2}\mathscr{L}(A_\sigma) \leq_{\mathrm{cx}} \frac{1}{2}\mathscr{L}(C_\pi) + \frac{1}{2}\mathscr{L}(C_\sigma).$$

More explicitly, recalling the relation between π and σ, this last inequality may be written

$$\frac{1}{2}\mathscr{L}\left(\sum_{j=1}^{\ell_1} \xi_{\pi^{-1}(1)j} + \sum_{j=1}^{\ell_2} \xi_{\pi^{-1}(2)j}\right) + \frac{1}{2}\mathscr{L}\left(\sum_{j=1}^{\ell_2} \xi_{\pi^{-1}(1)j} + \sum_{j=1}^{\ell_1} \xi_{\pi^{-1}(2)j}\right) \leq_{\mathrm{cx}}$$
$$\frac{1}{2}\mathscr{L}\left(\sum_{j=1}^{k_1} \xi_{\pi^{-1}(1)j} + \sum_{j=1}^{k_2} \xi_{\pi^{-1}(2)j}\right) + \frac{1}{2}\mathscr{L}\left(\sum_{j=1}^{k_2} \xi_{\pi^{-1}(1)j} + \sum_{j=1}^{k_1} \xi_{\pi^{-1}(2)j}\right).$$

Set $\pi^{-1}(1) = a$ and $\pi^{-1}(2) = b$. Then, by (3.2), $\mathscr{L}(\xi_{b\ell_2}) = \mathscr{L}(\xi_{bk_1})$, $\mathscr{L}(\xi_{a\ell_2}) = \mathscr{L}(\xi_{ak_1})$, and independence, it is easy to see that the above is equivalent to

$$\frac{1}{2}\mathscr{L}\left(\sum_{j=1}^{k_1-1} \xi_{aj} + \sum_{j=1}^{k_2} \xi_{bj} + \xi_{bk_1}\right) + \frac{1}{2}\mathscr{L}\left(\sum_{j=1}^{k_2} \xi_{aj} + \xi_{ak_1} + \sum_{j=1}^{k_1-1} \xi_{bj}\right) \leq_{\mathrm{cx}}$$
$$\frac{1}{2}\mathscr{L}\left(\sum_{j=1}^{k_1} \xi_{aj} + \sum_{j=1}^{k_2} \xi_{bj}\right) + \frac{1}{2}\mathscr{L}\left(\sum_{j=1}^{k_2} \xi_{aj} + \sum_{j=1}^{k_1} \xi_{bj}\right). \quad (3.3)$$

Now, with

$$(\alpha_1, \alpha_2) = \left(\sum_{j=1}^{k_1-1} \xi_{aj} + \sum_{j=1}^{k_2} \xi_{bj} + \xi_{bk_1}, \sum_{j=1}^{k_2} \xi_{aj} + \xi_{ak_1} + \sum_{j=1}^{k_1-1} \xi_{bj} \right) \quad \text{and}$$

$$(\beta_1, \beta_2) = \left(\sum_{j=1}^{k_1} \xi_{aj} + \sum_{j=1}^{k_2} \xi_{bj}, \sum_{j=1}^{k_2} \xi_{aj} + \sum_{j=1}^{k_1} \xi_{bj} \right),$$

we prove that $(\alpha_1, \alpha_2) \prec (\beta_1, \beta_2)$. First note that $\alpha_1 + \alpha_2 = \beta_1 + \beta_2$, so it suffices to show that $\max\{\alpha_1, \alpha_2\} \leq \max\{\beta_1, \beta_2\}$. Suppose first that $a < b$. Then, since $k_1 - 1 \geq k_2 + 1$, we have

$$\beta_2 - \alpha_1 = \sum_{j=k_2+1}^{k_1-1} (\xi_{bj} - \xi_{aj}) \geq 0, \quad \text{and} \quad \beta_2 - \alpha_2 = \xi_{bk_1} - \xi_{ak_1} \geq 0,$$

and therefore $\max\{\alpha_1, \alpha_2\} \leq \beta_2 \leq \max\{\beta_1, \beta_2\}$. A similar calculation holds when $a > b$.

Recalling that if φ is convex on $\mathbb{R}$ then $\sum \varphi(x_i)$ is Schur convex, it follows that for any convex function φ we have $\varphi(\alpha_1) + \varphi(\alpha_2) \leq \varphi(\beta_1) + \varphi(\beta_2)$, and (3.3) follows readily. □

The next generalization is of possible interest by itself. For $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n)$ let

$$\Upsilon_{\boldsymbol{\ell}} = \sum_{i=1}^{n} \sum_{j=1}^{\ell_i} \xi_{ij}$$

Proposition 3.3. *Let $\{\xi_{ij}\}$ be as in Lemma 3.2 and let $\boldsymbol{L}$ and $\boldsymbol{K}$ be exchangeable random vectors having nonnegative integer valued components, independent of $\{\xi_{ij}\}$. If $\boldsymbol{L} \prec_{\mathrm{S}} \boldsymbol{K}$, then*

$$\Upsilon_{\boldsymbol{L}} \leq_{\mathrm{cx}} \Upsilon_{\boldsymbol{K}}. \tag{3.4}$$

Proof. Since $\boldsymbol{K}$ is exchangeable $\mathscr{L}(\Upsilon_{\boldsymbol{K}}) = \mathscr{L}(Z_{\boldsymbol{K}})$ where $Z_{\boldsymbol{k}}$ is defined in (3.1), so it suffices to show $Z_{\boldsymbol{L}} \leq_{\mathrm{cx}} Z_{\boldsymbol{K}}$. By Lemma 3.1 we may take $\boldsymbol{L} \prec \boldsymbol{K}$ almost surely, and now, using the assumed independence, Lemma 3.2 may be invoked to complete the argument. □

Corollary 3.4. *Let $\{\xi_{ij}\}$ be as in Lemma 3.2. If $\boldsymbol{K}$ is an exchangeable random vector in $\mathscr{K}$, independent of $\{\xi_{ij}\}$, and if $N = mn$ then*

$$\Upsilon_{m\mathbb{1}_n} \leq_{\mathrm{cx}} \Upsilon_{\boldsymbol{K}}.$$

Proof of Theorem 2.3. Consider first the case of $\varepsilon_{ij} = 0$. Recalling that V_{ij} is the j^{th} sampled value in the partition element A_i, by the monotonicity of f and of the partition $\mathscr{A}$, the variables $\xi_{ij} = f(V_{ij})$ satisfy the hypotheses of Lemma 3.2. Hence the theorem now easily follows by applying Proposition 3.3. Since the convex order is preserved under addition of independent equally distributed random variables, the general case with additive noise follows. □

Proof of Proposition 2.5. It is easy to see that it suffices to consider the case where all ε_{ij} in (2.1) are zero. Once this is proved, for the first part of the proposition the general result follows using the same argument as in the proof of Theorem 2.3, and for the second it is obvious.

Call $\mathscr{P}$ the set of all permutations of $\{1, \dots, n\}$ and for $\boldsymbol{k} \in \mathscr{K}$ define

$$\psi(\boldsymbol{k}) := \frac{1}{n!} \sum_{\pi \in \mathscr{P}} \left[\frac{1}{N} \sum_{i=1}^{n} k_{\pi(i)} \mathbb{E}[f(V_i)] - I(f) \right]^2, \tag{3.5}$$

where V_i have the distribution $P_{U|A_i}$, see Definition 2.1. The function ψ is symmetric and convex, hence Schur convex (see Marshall and Olkin, 1979, page 67, Proposition C.2).

We start by proving the result for the special case where $\boldsymbol{K} = (K_1, \dots, K_n)$ is such that $K_i = k_{\Pi(i)}$ for a fixed $\boldsymbol{k}$ and a uniformly chosen random permutation Π with values in $\mathscr{P}$. For this case it is required to prove that $\mathrm{Var}[W_{\mathbb{K}}]$ is a Schur convex function of $\boldsymbol{k}$. Note that in this case, for any functions ϕ_i, $i = 1, \dots, n$, we have

$$\mathbb{E} \sum_{i=1}^{n} \phi_i(K_i) = \frac{1}{n!} \sum_{\pi \in \mathscr{P}} \sum_{i=1}^{n} \phi_i(k_{\pi(i)}). \tag{3.6}$$

By the variance decomposition formula, we have

$$\mathrm{Var}[W_{\mathbb{K}}] = \mathrm{Var}[\mathbb{E}[W_{\mathbb{K}} | \boldsymbol{K}]] + \mathbb{E}[\mathrm{Var}[W_{\mathbb{K}} | \boldsymbol{K}]]. \tag{3.7}$$

Both terms on the r.h.s. above depend on $\boldsymbol{k}$. We prove now that the first is Schur convex in $\boldsymbol{k}$ and that the second is constant for all $\boldsymbol{k} \in \mathscr{K}$.

Since

$$\mathbb{E}[W_{\mathbb{K}} | \boldsymbol{K} = \boldsymbol{k}] = \frac{1}{N} \sum_{i=1}^{n} k_i \mathbb{E}[f(V_i)],$$

using (3.6) and (3.5), we obtain

$$\mathrm{Var}[\mathbb{E}[W_{\mathbb{K}} | \boldsymbol{K}]] = \psi(\boldsymbol{k}). \tag{3.8}$$

Therefore $\mathrm{Var}[\mathbb{E}[W_{\mathbb{K}} | \boldsymbol{K}]]$ is Schur convex in $\boldsymbol{k}$.

For the second term we have

$$\mathrm{Var}[W_{\mathbb{K}} | \boldsymbol{K} = \boldsymbol{k}] = \mathbb{E}\left[\left\{ \frac{1}{N} \sum_{i=1}^{n} \left(\sum_{j=1}^{k_i} f(V_{ij}) - k_i \mathbb{E}[f(V_i)] \right) \right\}^2 \right] = \frac{1}{N^2} \sum_{i=1}^{n} k_i \, \mathrm{Var}[f(V_i)],$$

where for each i, V_{ij} are i.i.d. copies of V_i, and they are all independent. Therefore

$$\begin{aligned}\mathbb{E}[\mathrm{Var}[W_{\mathbb{K}}|\boldsymbol{K}]] &= \frac{1}{n!}\frac{1}{N^2}\sum_{i=1}^{n}\sum_{\pi\in\mathscr{P}} k_{\pi(i)}\,\mathrm{Var}[f(V_i)] \\ &= \frac{1}{n!}\frac{1}{N^2}\sum_{\pi\in\mathscr{P}}\sum_{i=1}^{n} k_i\,\mathrm{Var}[f(V_{\pi(i)})] \\ &= \frac{1}{N}\sum_{i=1}^{n}\mathrm{Var}[f(V_i)],\end{aligned}$$

which does not depend on $\boldsymbol{k}\in\mathscr{K}$.

It follows that if $\boldsymbol{\ell}\prec\boldsymbol{k}$, then $\mathrm{Var}[W_{\mathbb{L}}]\le\mathrm{Var}[W_{\mathbb{K}}]$. This implies the required result for $\boldsymbol{L}\prec_S\boldsymbol{K}$ when $\boldsymbol{L}$ and $\boldsymbol{K}$ are concentrated on permutations of some $\boldsymbol{\ell}$ and some $\boldsymbol{k}$, respectively, and $\boldsymbol{\ell}\prec\boldsymbol{k}$.

Now suppose $\boldsymbol{L}\prec_S\boldsymbol{K}$, and they are not (necessarily) concentrated on permutations of some $\boldsymbol{\ell}$ and some $\boldsymbol{k}$. The variance decomposition (3.7) still holds. Moreover, by Lemma 3.1, we can assume that $P(\boldsymbol{L}\prec\boldsymbol{K})=1$, i.e., that the joint distribution of $(\boldsymbol{L},\boldsymbol{K})$ is supported on the set

$$A=\{(\boldsymbol{\ell},\boldsymbol{k})\,:\,\boldsymbol{\ell},\boldsymbol{k}\in\mathscr{K},\,\boldsymbol{\ell}\prec\boldsymbol{k}\}$$

with probabilities $p_{\boldsymbol{k},\boldsymbol{\ell}}$. It follows easily from (3.8) that

$$\mathrm{Var}[\mathbb{E}[W_{\mathbb{L}}|\boldsymbol{L}]]=\sum_{(\boldsymbol{k},\boldsymbol{\ell})\in A}p_{\boldsymbol{k},\boldsymbol{\ell}}\psi(\boldsymbol{\ell})\le\sum_{(\boldsymbol{k},\boldsymbol{\ell})\in A}p_{\boldsymbol{k},\boldsymbol{\ell}}\psi(\boldsymbol{k})=\mathrm{Var}[\mathbb{E}[W_{\mathbb{K}}|\boldsymbol{K}]], \tag{3.9}$$

where the inequality follows from the fact that ψ is Schur convex, as shown above. It is immediate to see that the second summand in (3.7) is constant, and (2.4) follows.

To prove the second part of the proposition notice that, if $\mathbb{E}[f(U)|U\in A_i]=\mathbb{E}[f(V_i)]$ are all equal, and therefore all equal to $I(f)$, then it follows easily from (3.5) and (3.9) that $\mathrm{Var}[\mathbb{E}[W_{\mathbb{K}}|\boldsymbol{K}]]$ is zero. Since the other summand in (3.7) was shown to be constant, the result follows. $\square$

4 Numerical examples

We consider a function f on the interval $[0,1]$. We partition the interval $[0,1]$ into $n=10$ subintervals of equal length, and we take a total of $N=15$ observations.

Let $\boldsymbol{k}^a,\boldsymbol{k}^b,\boldsymbol{k}^c,\boldsymbol{k}^d\in\mathbb{N}^{10}$ be defined as follows:

$$\begin{aligned}\boldsymbol{k}^a&=(0,0,0,0,0,0,0,0,0,15),\\ \boldsymbol{k}^b&=(0,0,0,0,0,0,0,5,5,5),\\ \boldsymbol{k}^c&=(0,0,0,0,0,3,3,3,3,3),\\ \boldsymbol{k}^d&=(1,1,1,1,1,2,2,2,2,2).\end{aligned}$$

We have

$$\boldsymbol{k}^d \prec \boldsymbol{k}^c \prec \boldsymbol{k}^c \prec \boldsymbol{k}^a. \tag{4.1}$$

Let $\boldsymbol{K}_i^*$, $i = 1, \dots, 7$ be random vectors defined as follows:

$$\begin{aligned}
&\boldsymbol{K}_1^* = \boldsymbol{k}^a \text{ with probability } 1,\\
&\boldsymbol{K}_2^* = \boldsymbol{k}^a \text{ or } \boldsymbol{k}^b \text{ with equal probabilities},\\
&\boldsymbol{K}_3^* = \boldsymbol{k}^a \text{ or } \boldsymbol{k}^b \text{ or } \boldsymbol{k}^c \text{ with equal probabilities},\\
&\boldsymbol{K}_4^* = \boldsymbol{k}^a \text{ or } \boldsymbol{k}^b \text{ or } \boldsymbol{k}^c \text{ or } \boldsymbol{k}^d \text{ with equal probabilities},\\
&\boldsymbol{K}_5^* = \boldsymbol{k}^b \text{ or } \boldsymbol{k}^c \text{ or } \boldsymbol{k}^d \text{ with equal probabilities},\\
&\boldsymbol{K}_6^* = \boldsymbol{k}^c \text{ or } \boldsymbol{k}^d \text{ with equal probabilities},\\
&\boldsymbol{K}_7^* = \boldsymbol{k}^d \quad \text{with probability } 1.
\end{aligned}$$

Define now $\boldsymbol{K}_i$ to be the random vector obtained by applying a random permutation to $\boldsymbol{K}_i^*$. Lemma 3.1 and (4.1) imply that

$$\boldsymbol{K}_7^* \prec_{\mathrm{S}} \boldsymbol{K}_6^* \prec_{\mathrm{S}} \boldsymbol{K}_5^* \prec_{\mathrm{S}} \boldsymbol{K}_4^* \prec_{\mathrm{S}} \boldsymbol{K}_3^* \prec_{\mathrm{S}} \boldsymbol{K}_2^* \prec_{\mathrm{S}} \boldsymbol{K}_1^*,$$

hence

$$\boldsymbol{K}_7 \prec_{\mathrm{S}} \boldsymbol{K}_6 \prec_{\mathrm{S}} \boldsymbol{K}_5 \prec_{\mathrm{S}} \boldsymbol{K}_4 \prec_{\mathrm{S}} \boldsymbol{K}_3 \prec_{\mathrm{S}} \boldsymbol{K}_2 \prec_{\mathrm{S}} \boldsymbol{K}_1.$$

The integral $I(f) = \int_0^1 f(x)\,\mathrm{d}x$ is estimated with

$$W_{\mathbb{K}} = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{K_i} f(V_{ij}),$$

for $\boldsymbol{K} = \boldsymbol{K}_i$, as defined above, $i = 1, \dots, 7$. We repeat the operation $M = 10,000$ times. If we call $W_{\mathbb{K}}^{(j)}$ the j-th estimate, then the mean p-error is estimated with

$$\left(\frac{1}{M} \sum_{j=1}^{M} \left| W_{\mathbb{K}}^{(j)} - I(f) \right|^p \right)^{1/p}.$$

We compute this mean p-error for $p = 1, 2, 10$.

We choose the monotone functions $f(x) = x^\alpha$, for $\alpha = 0.5, 1, 2, 20$. As expected from Theorem 2.3, we obtain the following results. On the horizontal axis we have the different sampling schemes corresponding to $\boldsymbol{K}_i$, $i = 1, \dots, 7$.

FIGURES 1, 2, 3, 4 ABOUT HERE

As Figure 5 shows, monotonocity of the function f is fundamental to obtain the comparison in Theorem 2.3. If we choose the nonmonotone function

$$f(x) = \begin{cases} 2 & \text{if } 0 \le x < .88, \\ 1 & \text{if } .88 \le x < .94, \\ 3 & \text{if } .94 \le x \le 1, \end{cases} \tag{4.2}$$

then, as in Example 2.4, for $p = 1$ the comparison goes the other way around. The L^2 error is decreasing, as expected from Proposition 2.5.

FIGURE 5 ABOUT HERE

This happens also for the function

$$f(x) = \begin{cases} 1 & \text{if } 0 \le x < .9, \\ 0 & \text{if } .9 \le x < .95, \\ 2 & \text{if } .95 \le x \le 1. \end{cases} \tag{4.3}$$

In this case the conditions of the second part of Proposition 2.5 are satisfied, so in the simulations the L^2 error is approximately constant.

FIGURE 6 ABOUT HERE

Acknowledgments

The authors thank Pierpaolo Brutti for help with Mathematica and two anonymous referee for observations and comments that led us to add new results to the previous version of the paper.

References

Alfsen, E. M. (1971) *Compact Convex Sets and Boundary Integrals.* Springer-Verlag, New York.

Glasserman, P. (2004) *Monte Carlo Methods in Financial Engineering.* Springer-Verlag, New York.

Goldstein, L., Rinott, Y., and Scarsini, M. (2010) Stochastic comparisons of stratified sampling techniques for some Monte Carlo estimators. *Bernoulli* Forthcoming.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952) *Inequalities.* Cambridge, at the University Press. 2d ed.

Lindvall, T. (2002) *Lectures on the Coupling Method.* Dover Publications Inc., Mineola, NY. Corrected reprint of the 1992 original.

Marshall, A. W. and Olkin, I. (1979) *Inequalities: Theory of Majorization and Its Applications.* Academic Press Inc., New York.

Müller, A. and Stoyan, D. (2002) *Comparison Methods for Stochastic Models and Risks.* John Wiley & Sons Ltd., Chichester.

Nevius, S. E., Proschan, F., and Sethuraman, J. (1977) Schur functions in statistics. II. Stochastic majorization. *Ann. Statist.* **5**, 263–273.

Shaked, M. and Shanthikumar, J. G. (2007) *Stochastic Orders.* Springer, New York.

Strassen, V. (1965) The existence of probability measures with given marginals. *Ann. Math. Statist* **36**, 423–439.

January 10, 2011

Figures

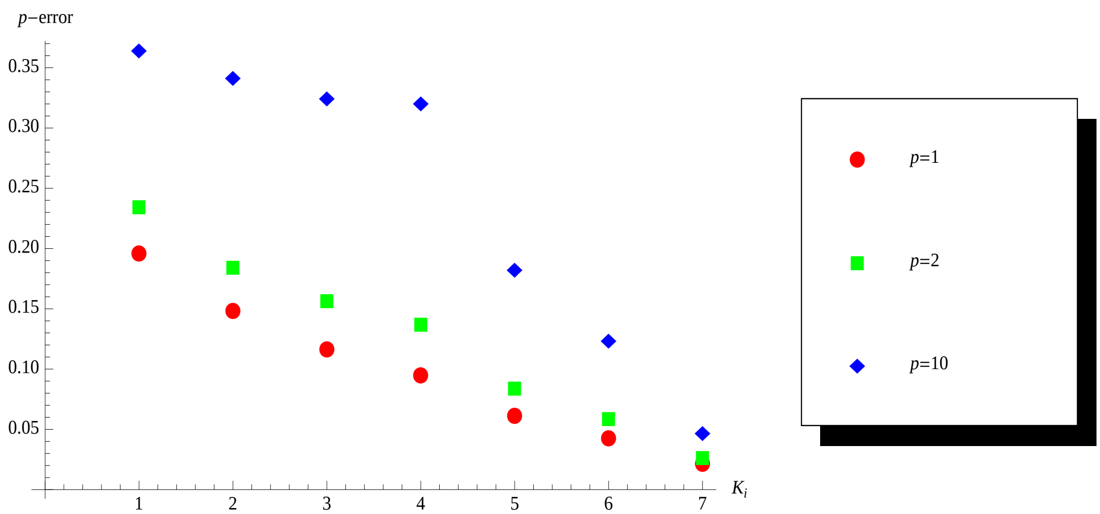

Figure 1: Mean p-error for estimators of $\int_0^1 x^{1/2}\ \mathrm{d}x$.

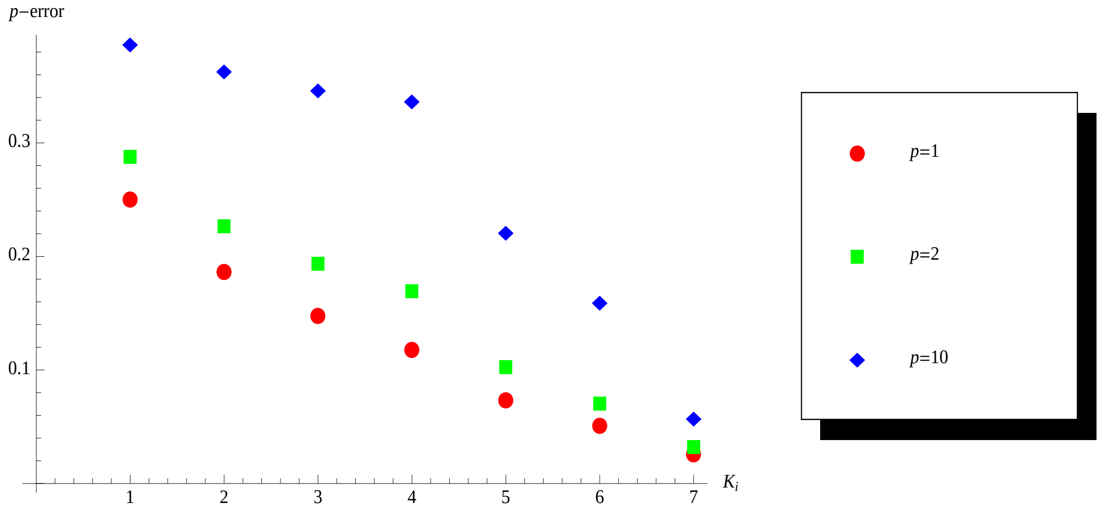

Figure 2: Mean p-error for estimators of $\int_0^1 x\ \mathrm{d}x$.

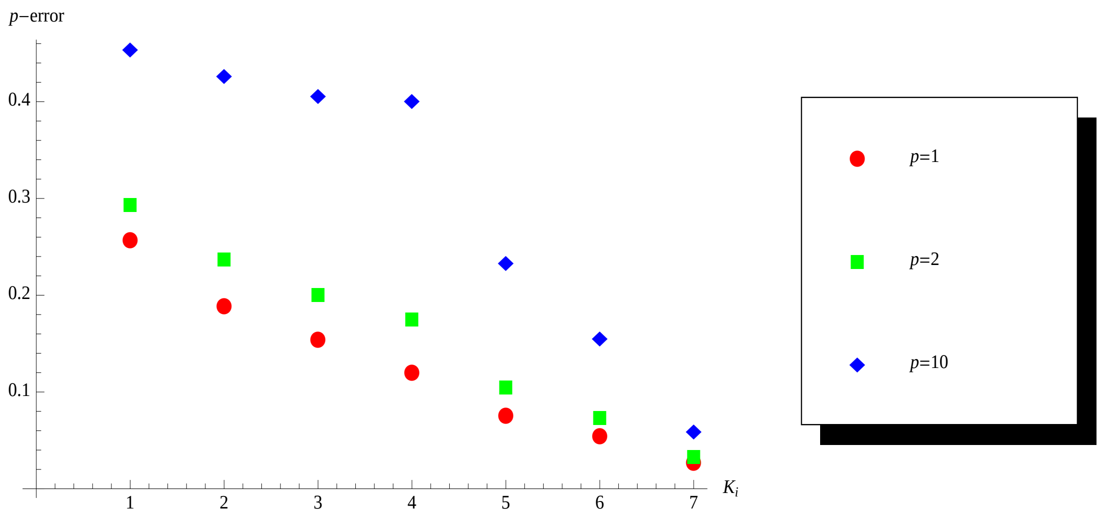

Figure 3: Mean p-error for estimators of $\int_0^1 x^2\ \mathrm{d}x$.

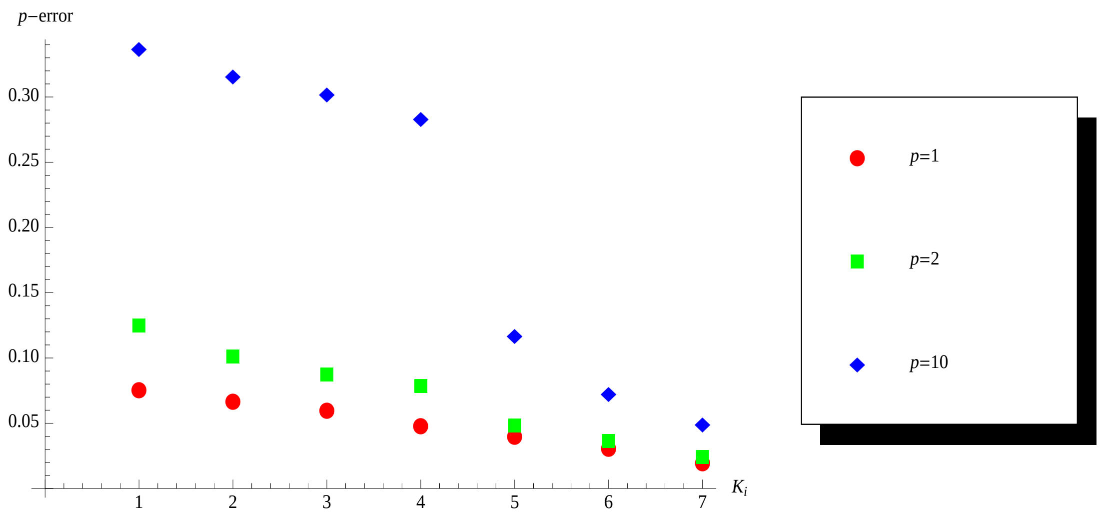

Figure 4: Mean p-error for estimators of $\int_0^1 x^{20}\,\mathrm{d}x$.

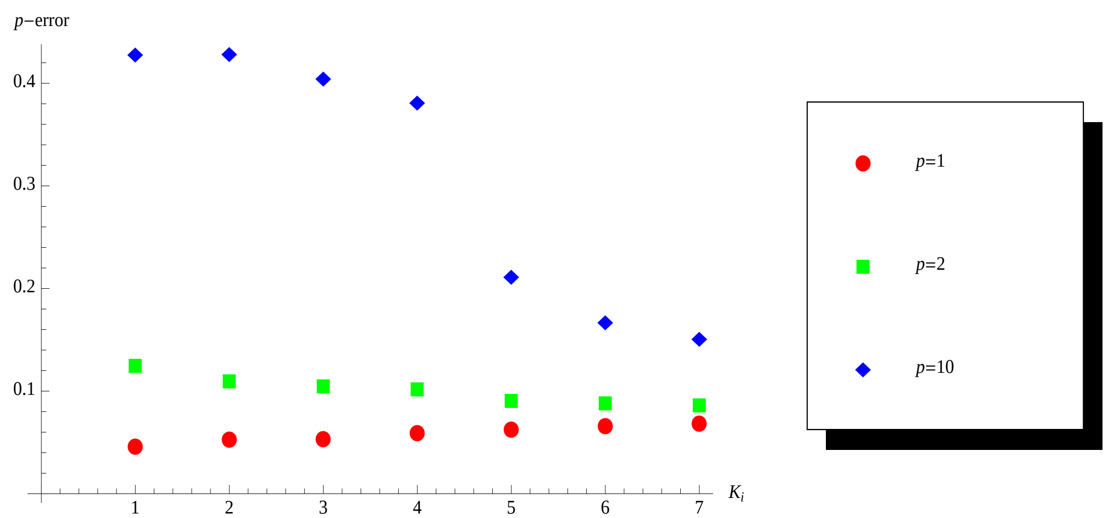

Figure 5: Mean p-error for estimators of $\int_0^1 f(x)\,\mathrm{d}x$ with f as in (4.2).

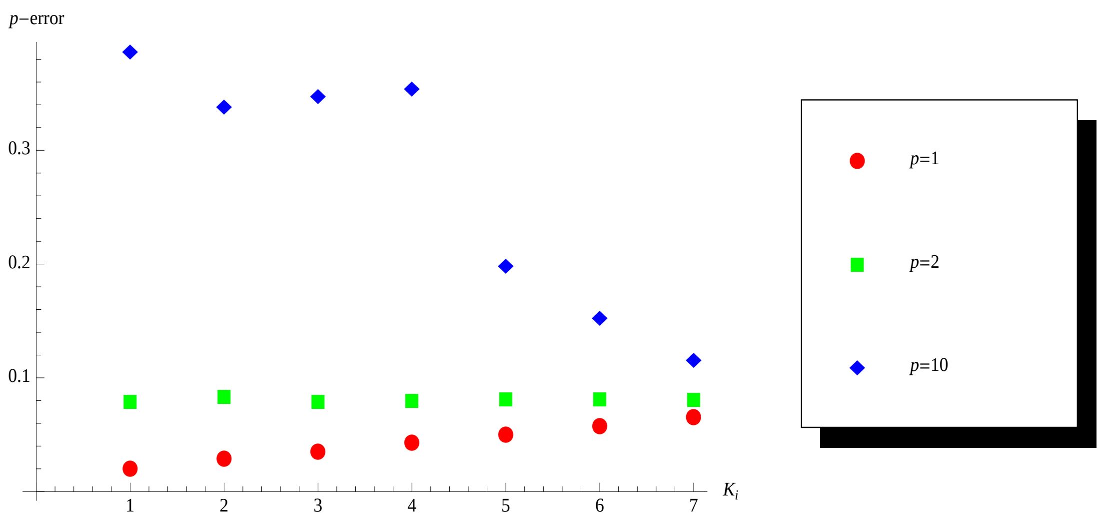

Figure 6: Mean p-error for estimators of $\int_0^1 f(x)\,\mathrm{d}x$ with f as in (4.3).