Best Invariant and Minimax Estimation of Quantiles in Finite Populations

Yaakov Malinovsky* and Yosef Rinott**,***

*Division of Epidemiology, Statistics and Prevention Research, National Institute of Child Health and Human Development, 6100 Executive Boulevard, Bethesda, MD 20852, USA

**The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel and

***LUISS, Rome

Abstract

The theoretical literature on quantile and distribution function estimation in infinite populations is very rich, and invariance plays an important role in these studies. This is not the case for the commonly occurring problem of estimation of quantiles in finite populations. The latter is more complicated and interesting because an optimal strategy consists not only of an estimator, but also of a sampling design, and the estimator may depend on the design and on the labels of sampled individuals, whereas in iid sampling, design issues and labels do not exist.

We study estimation of finite population quantiles, with emphasis on estimators that are invariant under the group of monotone transformations of the data, and suitable invariant loss functions. Invariance under the finite group of permutation of the sample is also considered. We discuss nonrandomized and randomized estimators, best invariant and minimax estimators, and sampling strategies relative to different classes. Invariant loss functions and estimators in finite population sampling have a nonparametric flavor, and various natural combinatorial questions and tools arise as a result.

Keywords:

Monotone transformations, sampling and estimation strategies, behavioral and randomized estimators, loss functions, median, simple random sample

---

1. Introduction

In this paper we study invariant estimation of quantiles of a finite population. Much of statistics, such as official statistics, concerns finite population sampling, with emphasis on estimation of totals and quantiles. However, most of the work in the past three decades or so on optimality properties of quantile estimators, including the study of invariance, has concentrated on iid sampling, that is, sampling from infinite populations.

In finite population sampling, the statistician chooses a *strategy* which consists of a *sampling design*, and an *estimator*, and the data consist of the labels of the sampled units, and their corresponding measured values; this clearly differs from infinite population sampling, where there is no sampling design to consider, and no labels.

When estimating quantiles of a finite population, it is natural to deal with estimators that are invariant under monotone transformations of the measured values, since under such transformations the population unit which represents the estimated quantile remains unchanged. It is also natural to consider the possibility of invariance under permutations of the labels. In this paper we deal with best-invariant and minimax strategies, that is, sampling designs and estimators in connection with two groups: the infinite (and non-compact) group of monotone transformations, and the finite group of permutations. Another special aspect of the present work is that we consider a class of invariant loss functions that essentially measure the deviation of the estimate from the estimated quantile in terms of the number of population units that separate them; see (2). These loss functions have a combinatorial flavor, and so do some of our proofs, including that of Theorem 4.2 which is given in Malinovsky and Rinott (2009), and a simple use of Ramsey theory in Theorem 5.3.

Some relevant references: invariance under monotone transformation when estimating a whole distribution function with various loss functions appears, for example, in Agarwal (1955), Ferguson (1967), Brown (1988), Yu and Chow (1991), Yu and Phadia (1992), Stępień-Baran (2010), Cohen and Kuo (1985), and Lehmann and Casella (1998), where the only last two reference consider finite population models. Invariant quantile estimation in infinite populations appears in Ferguson (1967), Brown (1988) (median), and Zieliński (1999).

Invariance in finite populations appears already in Blackwell and Girshick (1954), where only finite groups (permutations) are considered, and in many later references, such as Godambe (1968), Basu (1971), Godambe and Thompson (1971), Cassel et al. (1977), where invariance under linear transformations also appears. For a Bayesian approach to finite population quantiles estimation including admissibility results, but under a loss function different from ours, see Nelson and Meeden (2006) and references therein, and for asymptotic results see, for example, Chatterjee (2010). Results on optimality of strategies in finite population sampling, with numerous references, can be found, for example, in Cassel et al. (1977), and for a recent survey see Rinott (2009). The present paper combines ideas related to invariant estimation of distribution functions based on iid samples, as in Ferguson (1967) and other of the above mentioned papers, with ideas from finite population sampling that can be found in Rinott (2009) and references therein, to obtain minimax and related optimality results for estimation of finite population quantiles.

In Section 2 we provide all definitions and notations. In Section 3 we show that for our purposes randomized and behavioral estimators are equivalent. Thus we can choose either formulation of randomization according to our convenience. In Section 4 we describe the form of invariant estimators and some of their properties. We study best invariant-symmetric estimators under simple random sampling, and determine them explicitly in certain interesting cases. Sample quantiles, that is, quantiles of the empirical distribution function, provide a standard way of estimating the corresponding population quantiles. However, the estimators we propose and study in Section 4 are not always identical to the sample quantiles; also, they may depend on the loss function under consideration. Furthermore, they may not be unique. In Section 5 we bring minimax results for general sampling designs. In Theorem 5.2 we show that the quantile estimators we propose, together with simple random sampling, form a minimax strategy in the class of strategies consisting of any sampling design, and an invariant estimator. Theorems 5.3 and 5.4 provide minimax results relative to non-invariant estimators. Minimax estimators are obtained by a symmetrization procedure, see (14), leading naturally to randomized estimators. Thus, randomized estimators play a part in the proofs. Such estimators appear also when unbiasedness is desired. Unbiased estimators are defined and studied in Malinovsky (2009).

2. Definitions and notations

Most of the definitions and notations, with references, appear in Ferguson (1967), or Rinott (2009). We consider a size N **finite population** of values of some measurement. Let $x = (x_1, x_2, ..., x_N)$ be the N-dimensional vector of population values, where x_j is a real number associated with the unit labeled $j \in \mathcal{N} := \{1, ..., N\}$, the label set. We assume that $x \in \Upsilon$, a known **param-**

eter space. For simplicity we shall consider only parameter spaces of the type $\Upsilon = \{(x_1, x_2, ..., x_N) : x_i \in \mathbb{R},\ \ x_i \text{ distinct}\}$, where $\mathbb{R}$ denotes the real line. Note that Υ is *symmetric* in the sense that if $x \in \Upsilon$ then so is any permutation of the coordinates of x. The assumption that the coordinates of x are distinct is not essential, but making it helps avoid various technicalities, and the same is true with regard to the assumption $x_i \in \mathbb{R}$, and we could assume that $x_i \in \Lambda$ where Λ is some open interval, finite or infinite. We will comment on such possibilities only briefly. The **population distribution function** F_x is defined by

$$F_x(t) = \frac{1}{N}\sum_{j=1}^{N} I_{(-\infty,\, t]}(x_j) = \frac{1}{N}\sum_{j=1}^{N} I_{[x_j,\, \infty)}(t). \tag{1}$$

F_x is an unknown **parameter** which is a function of the parameter x. Using the assumption that the coordinates of x are distinct, we can also write

$$F_x(t) = \frac{j}{N} \text{ for } x_{(j)} \le t < x_{(j+1)},\ j = 0, 1, ...N,\ (x_{(0)} := -\infty,\ x_{(N+1)} := \infty),$$

where $x_{(1)} < x_{(2)} < ... < x_{(N)}$ are the order statistics of x. In particular $F_x(x_{(j)}) = \frac{j}{N}$.

The k-th population **quantile** for a given $x \in \Upsilon$ is

$$q_k = \inf\{\theta \in \mathbb{R} : F_x(\theta) \ge k/N\}.$$

Our goal is to estimate quantiles, where for a given estimate a of q_k, $k = 1, ..., N$, the **loss function** is of the form

$$L(a, x) = G(|F_x(a) - \frac{k}{N}|), \quad a \in \mathbb{R}\,, \tag{2}$$

for some a nonnegative increasing function G. Some of our results focus on special cases of such G. Note that $|F_x(a) - \frac{k}{N}|$ vanishes if $a = q_k$, and otherwise it counts the deviation of a from the estimated quantile in terms of number of ordered population units by which they differ.

A parameter $\theta = \theta(x)$ is said to be ***symmetric*** if it remains constant under permutations of the coordinates of x. Clearly the examples given above, F_x and $x_{(k)}$ are symmetric parameters, and so is the population total $\theta(x) = \sum_{i=1}^{N} x_i$, and most of the common parameters of interest. Also, if for some θ, $F_x(\theta) \ge k/N$ for some x, then the same holds for any permutation of x since F_x is symmetric. Therefore the population quantiles are also symmetric. A loss function $L(a, x)$ is said to be **symmetric** if it remains constant when x is replaced by any permutation of its coordinates for any a. It is clear that the loss (2) is symmetric since F_x is symmetric.

A **sampling design** $\mathcal{P}$ is a probability function on the space of all subsets S of $\mathcal{N}$. We assume noninformative sampling, that is, the probability $\mathcal{P}(S)$ does not depend on the parameter x. **Simple random sampling** *without replacement* of size n is denoted by $\mathcal{P}_s$ and satisfies $\mathcal{P}_s(S) = 1/\binom{N}{n}$ if $|S| = n$, and zero otherwise, where $|S|$ denotes the size of S.

The **data** consist of the set of pairs $\{(i, x_i) : i \in S\}$, that is , the x-values in the sample S and their corresponding labels. We set

$$D = D[S, x] = \{(i, x_i) : i \in S\}. \tag{3}$$

The notation $D[S, x]$ as defined above is sometimes convenient, however, is does not reflect the pairing $(i, x_i) : i \in S$ which is part of the data, and the fact that the data depends on x only through x_i's such that $i \in S$. By sufficiency arguments, Basu (1958) (also Cassel et al. (1977) and Rinott (2009)), the order in which the sample was drawn (if defined and known) and repetitions of units, if the sampling procedure allows it, provide no information. Since the relevant data consist only of the set of drawn labels S and their x-values, we shall only consider designs $\mathcal{P}$ on the space of unordered subsets of $\mathcal{N}$ with no repetitions. Furthermore, we consider here only sampling designs having a fixed sample size, $|S| = n$, say; that is, the sample consists of n distinct units. Set $\mathbf{X} = \{x_i : i \in S\}$ and let $\mathbf{Y} = (Y_1, ..., Y_n) = (X_{(1)}, ..., X_{(n)})$ denote the ordered values (order statistics) of $\mathbf{X}$. We have $Y_1 < ... < Y_n$. It is often convenient to use the notation $x_S = \{x_i : i \in S\}$ instead of $\mathbf{X}$.

A (nonrandomized) **estimator** t is a real valued function $t(D)$ of the data. The space of such estimators is denoted by $\mathcal{T}$. We will also use the notation $t(S, x)$ and $t\left(\{(i, x_i) : i \in S\}\right)$ for $t(D)$. An estimator $t = t(D)$ is said to be **symmetric** if it depends only on the x-values in the sample, that is, x_S, and not on their label set S. Thus, if $\{x_i : i \in S\} = \{x'_i : i \in S'\}$ for some $x, x' \in \Upsilon$ and samples S, S', then $t(D[S, x]) = t(D[S', x'])$. The class of all symmetric estimators is denoted by $\mathcal{T}_S$. It is trivial but important to note that without information on S, the information in $\mathbf{X} = \{x_i : i \in S\}$ is the same as in $\mathbf{Y}$. Hence for symmetric estimators we may write $t(\mathbf{X}) = t(\mathbf{Y})$, and also $t(x_S)$.

The best known example of a non symmetric estimator is the Horvitz-Thompson estimator of the finite population total, $t_{HT}(D) := \sum_{i \in S} x_i/\alpha_i$, where the observation having label i is inversely weighted by the inclusion probability of the i-th unit according to the sampling design $\mathcal{P}$, $\alpha_i = P_{\mathcal{P}}(i \in S)$. On the other hand the simple sample mean, or the median and other sample quantiles, for example, are all symmetric.

A pair $(\mathcal{P}, t)$ consisting of a sampling design and an estimator is called a **strategy**. The **risk** of a strategy $(\mathcal{P}, t)$ for a given $x \in \Upsilon$ is the expected loss

$$R\left(\mathcal{P}, t; x\right) = E_{\mathcal{P}} L\left(t(D), x\right) = \sum_{S} L(t(D[S, x]), x)\mathcal{P}(S). \tag{4}$$

For the next definition we need to consider the class of nonrandomized estimators $\mathcal{T}$ as a measure space. As in Ferguson (1967) we do not specify a sigma-field, however, we assume that singletons, that is, sets consisting of a single nonrandomized estimator, are in the sigma-field. A probability distribution δ on the space of nonrandomized estimators $\mathcal{T}$, is called a **randomized estimator**. The space of all randomized estimators is denoted by $\mathcal{T}^*$. We define

$$R\left(\mathcal{P}, \delta; x\right) = ER\left(\mathcal{P}, T; x\right) = \sum_{S} \int_{\mathcal{T}} L(t, x) d\delta(t)\mathcal{P}(S), \tag{5}$$

where T is a random variable taking values in $\mathcal{T}$ with distribution δ, and the integral $d\delta(t)$ is properly defined over the function space $\mathcal{T}$. A randomized estimator is said to be **symmetric** if δ is concentrated on the class $\mathcal{T}_S$ of nonrandomized symmetric estimators. The class of such estimators is denoted by $\mathcal{T}_S^*$.

A **behavioral estimator** is defined by $\delta = \{\delta_D\} = \{\delta_{S,x_S}\}$, where for each possible data D, δ_D is a distribution on $\mathbb{R}$, with the interpretation that if D is observed, then a value in $\mathbb{R}$ is chosen according to δ_D as an estimate of the parameter in question. A behavioral estimator is said to be **symmetric** when the distributions δ_D depend only on x_S and not on the sampled labels. For behavioral estimators, letting $Z \in \mathbb{R}$ be distributed according to δ_D we define

$$R\left(\mathcal{P}, \delta; x\right) = ER\left(\mathcal{P}, Z; x\right) = \sum_{S} \int_{\mathbb{R}} L(z, x)\delta_{S,x_S}(dz)\mathcal{P}(S). \tag{6}$$

We remark that under the present setup, the classes of behavioral and randomized rules are equivalent by Wald and Wolfowitz (1951). In Section 3 we show that this equivalence holds also for invariant symmetric estimators, which are defined next. Therefore when discussing randomized estimators we consider either formulation and use the same notation as defined above for randomized estimators, $\mathcal{T}^*$ and $\mathcal{T}_S^*$ also for the classes behavioral and symmetric behavioral estimators, and similarly for other such classes.

Given a function $\varphi : \mathbb{R} \to \mathbb{R}$, we extend its operation naturally to vectors in the parameter space, by $\varphi(x) = (\varphi(x_1), ..., \varphi(x_N))$, to ordered samples by $\varphi(\mathbf{Y}) = (\varphi(Y_1), ..., \varphi(Y_n))$, and to data by $\varphi(D) = \{(i, \varphi(x_i)) : i \in S\}$ and $\varphi(x_S) = \{\varphi(x_i)) : i \in S\}$. Let Φ denote the *group* of all *strictly increasing continuous* functions from $\mathbb{R}$ onto $\mathbb{R}$ (bijections). In the case that we assume $x_i \in \Lambda$, an open interval, then we assume that Φ consists of similar extensions of strictly increasing continuous functions from Λ onto Λ.

A nonrandomized estimator $t \in \mathcal{T}$ is said to be **invariant** if for all D and all $\varphi \in \Phi$, we have

$$t(\varphi(D)) = \varphi(t(D)). \tag{7}$$

The class of nonrandomized invariant estimators is denoted by $\mathcal{T}_I$, and the subclass of nonrandomized, invariant and symmetric estimators is denoted by $\mathcal{T}_{I\mathcal{S}}$.

A randomized estimator $\delta \in \mathcal{T}^*$ is said to be **invariant** if δ, as a probability distribution over $\mathcal{T}$, assigns all its mass to the subset $\mathcal{T}_I$ of invariant nonrandomized estimators. The class of invariant randomized estimators is denoted by $\mathcal{T}_I^*$. A randomized estimator $\delta \in \mathcal{T}^*$ is said to be **invariant-symmetric** if δ, as a probability distribution over $\mathcal{T}$, assigns all its mass to the subset $\mathcal{T}_{I\mathcal{S}}$ of invariant and symmetric nonrandomized estimators. The class of invariant-symmetric randomized estimators is denoted by $\mathcal{T}_{I\mathcal{S}}^*$.

A behavioral estimator is said to be **invariant** if $Z_D \sim \delta_D$ satisfies

$$Z_{\varphi(D)} \stackrel{\mathcal{L}}{=} \varphi(Z_D)$$

for all D, where $\stackrel{\mathcal{L}}{=}$ denotes equality of distributions (laws).

An estimator δ is an **equalizer** with respect to a design $\mathcal{P}$, if $R(\mathcal{P}, \delta; x) = C$ for some constant C, for all $x \in \Upsilon$. Given a design $\mathcal{P}$, an estimator δ_1 is said to be **as good as** an estimator δ_2, if $R(\mathcal{P}, \delta_1; x) \leq R(\mathcal{P}, \delta_2; x)$ for all $x \in \Upsilon$, and **better** if in addition the latter inequality holds strictly for at least one $x \in \Upsilon$. They are **equivalent** if $R(\mathcal{P}, \delta_1; x) = R(\mathcal{P}, \delta_2; x)$ for all $x \in \Upsilon$. An estimator δ is said to be **admissible** if there exists no estimator better than δ. An estimator having a property $\mathcal{C}$, that is as good as any other estimator having this property, is called a **best**-$\mathcal{C}$ estimator. We shall consider $\mathcal{C}$ = the property of being invariant, or invariant and symmetric, or invariant and symmetric and unbiased (the latter case is discussed only in Malinovsky (2009)).

3. Behavioral and randomized estimators

By a well known result of Wald and Wolfowitz (1951), see also Ferguson (1967) and Kirschner (1976), behavioral and randomized estimators are equivalent in our problem. It may happen in certain situations that the classes of behavioral and randomized rules are equivalent, whereas, the classes of *invariant* behavioral and randomized rules are not equivalent. See, for example Ferguson (1967) p. 153. However, in our case these classes are equivalent. The result is close to that of Ferguson (1967) p. 197.

Proposition 3.1. *For the group Φ defined above, the classes of symmetric invariant behavioral and symmetric invariant randomized estimators are equivalent in the sense that for every randomized symmetric invariant estimator there is an equivalent behavioral symmetric invariant estimator and conversely.*

Proof. The class of behavioral (invariant) estimators contains the class of randomized (invariant) estimators. For details see Ferguson (1967). We show that in our case the converse is also true, that is, given a symmetric invariant behavioral estimator $\delta = \{\delta_D\}$, we construct an equivalent symmetric invariant randomized estimator.

Consider a symmetric-invariant behavioral estimator. Since it is symmetric we can write δ_{x_S} for δ_D. Let $Z_{x_S} \sim \delta_{x_S}$, and for simplicity of notation we now write χ for the set x_S. Choose χ_0, a particular point in the sample space, and a random variable $Z_{\chi_0} \sim \delta_{\chi_0}$. For each χ in the sample space choose $\varphi_\chi \in \Phi$ such that $\varphi_\chi(\chi_0) = \chi$. Define $\tilde{Z}_\chi = \varphi_\chi(Z_{\chi_0})$. This constructs a randomized estimator as follows: consider the nonrandomized function $t_a(\chi) = \varphi_\chi(a)$ for each $a \in \mathbb{R}$. Then $\tilde{Z}_\chi$ is distributed as the randomized estimator $t_a(\chi)$, with $a = Z_{\chi_0} \sim \delta_{\chi_0}$. Note that the invariance of the behavioral estimator Z_χ implies that $Z_\chi \overset{\mathfrak{L}}{=} \varphi_\chi(Z_{\chi_0})$, and therefore marginal distribution of $\tilde{Z}_\chi$ is the same as that of Z_χ, and therefore they are equivalent.

It remains to show that the constructed randomized estimator is invariant, which means that the nonrandomized estimators $t_a(\chi)$ are invariant, that is, $t_a(\varphi(\chi)) = \varphi(t_a(\chi))$ with probability 1 with respect to $a \sim \delta_{\chi_0}$. This follows from $\varphi(t_a(\chi)) = \varphi(\varphi_\chi(Z_{\chi_0})) \overset{\mathfrak{L}}{=} Z_{\varphi(\varphi_\chi(\chi_0))} = Z_{\varphi(\chi)} = Z_{\varphi_{\varphi(\chi)}(\chi_0)} \overset{\mathfrak{L}}{=} \varphi_{\varphi(\chi)}(Z_{\chi_0}) = t_a(\varphi(\chi))$. In particular we have $\varphi(\varphi_\chi(Z_{\chi_0})) \overset{\mathfrak{L}}{=} \varphi_{\varphi(\chi)}(Z_{\chi_0})$. By Lemma 3.1 below this implies the equality almost surely, and then by the above relations $\varphi(t_a(\chi)) = t_a(\varphi(\chi))$ almost surely, and the proof is complete. □

Lemma 3.1. *If V is a random variable and g and h are strictly increasing continuous functions such that $g(V) \stackrel{\mathcal{L}}{=} h(V)$ then $g(V)$=$h(V)$ almost surely.*

Proof. We can restrict g and h to the support of V, on which they must have the same range. Their inverse functions are well defined on this range. Therefore, it suffices to prove that if $g(V) \stackrel{\mathcal{L}}{=} V$ then $g(V) = V$ with probability one. Let F denote the distribution function of V, and denote g^{-1} by h. The assumed equality in distribution is equivalent to $F(h(v)) = F(v)$ for all v in the support of F.

If F is strictly increasing then the assumption becomes $F(h(v)) = F(v)$ for all v, which implies $h(v) = v$ for all v. If F is not strictly increasing then almost the same argument works for points of increase of F, whereas other points have F probability zero. More specifically, if v is in the support of F then either $F(v+\varepsilon) > F(v)$ for any small ε, or $F(v-\varepsilon) < F(v)$ for any small ε. In the first case, for example, we cannot have $F(h(v)) = F(v)$ for any h satisfying $h(v) > v$. If $h(v) < v$, then for some ε we have $h(v+\varepsilon) < v$ by continuity of h. Then $F(h(v+\varepsilon)) < F(v) < F(v+\varepsilon)$ contradicting the assumption that $F(h(v)) = F(v)$ for all v in the support of F. □

4. Invariant estimators

4.1. General form of invariant and symmetric estimators

A close result to Proposition 4.1 below, in an infinite population (iid) setting, appears in Uhlmann (1963), Ferguson (1967), p.153, Ex 4.2.3, and Zieliński (1999). They show that invariant estimators are of the form Y_J, with J independent of the data. In finite population sampling, the data include the labels of the observations, and independence of the data no longer holds. Further subtle issues that arise in the presence of labels appear in Theorem 5.1 and the lemmas around it. The next proposition is stated and proved for behavioral estimators, hence it holds also for randomized estimators, including the first part where symmetry is not assumed and therefore Proposition 3.1 does not apply.

Proposition 4.1. *The behavioral invariant estimators are of the form $t(D) = Y_{J(D)}$, where $J(D)$ is a random variable taking values in $\{1, ..., n\}$, having distribution that depends on D. Moreover, the distribution of $J(D) = J(D[S, x])$ depends only on S and not on x.*

Also, invariant-symmetric behavioral estimators are of the form $t(D) = t(\mathbf{Y}) = Y_J$, where J is a random variable taking values in $\{1, ..., n\}$, having distribution that is not a function of the data.

Proof. The proof is an adaptation and simplification of exercise 3, p. 197 in Ferguson (1967) and Ferguson's web site, where the case of iid sampling and non-randomized estimators are considered.

A behavioral estimator is defined by a collection of random variables $Z_D \sim \delta_D$, taking values in the decision space, which in our case is $\mathbb{R}$ (or an open interval in $\mathbb{R}$). Invariance means $Z_{\varphi(D)} \overset{\mathcal{L}}{=} \varphi(Z_D)$. In particular this holds for all strictly increasing continuous functions φ, which leave $X_1, ..., X_n$ fixed. The set of such functions φ is denoted by Φ'. For $\varphi \in \Phi'$ we have $Z_D = Z_{\varphi(D)} \overset{\mathcal{L}}{=} \varphi(Z_D)$ and in particular Support$[Z_D] =$ Support$[\varphi(Z_D)]$. It follows that the support of Z_D must be contained in the set $\{Y_1, \ldots, Y_n\}$; otherwise, it is easy to construct $\varphi \in \Phi'$ such that Support$[Z_D] \neq$ Support$[\varphi(Z_D)]$, a contradiction. Therefore, behavioral invariant estimators are of the form $Y_{J(D)}$, where $J(D)$ is a random variable taking values in $\{1, ..., n\}$, whose distribution may depend on D.

Moreover, for $D = \{(i, x_i) : i \in S\}$ let $y_1 < \ldots < y_n$ be the ordered x_i's and note that the above representation and invariance imply $Y_{J(\varphi(D))} \overset{\mathcal{L}}{=} \varphi(Y_{J(D)})$. The left-hand side of the latter relation represent a random variable taking the values $\varphi(y_j)$ with probability $P(J(\varphi(D)) = j)$, whereas the right-hand side variable takes the same values with probabilities $P(J(D) = j)$, and it follows that $J(\varphi(D)) \overset{\mathcal{L}}{=} J(D)$. Now, let $D' = \{(i, x_i') : i \in S\}$. For any vector $x = (x_1, ..., x_N)$ there exists $\varphi \in \Phi$, such that $\{\varphi(x_i') : i \in S\} = \{x_i : i \in S\}$. Now the relation $J(\varphi(D)) \overset{\mathcal{L}}{=} J(D)$ implies that $J(D)$ depends only on S.

The last part of the Proposition follows readily since by symmetry $J(D)$ does not depend on S either, and therefore it does not depend on D. □

Corollary 4.1. *There are only n nonrandomized symmetric invariant estimators, of the form $t(D) = Y_j$, $j = 1, ..., n$.*

4.2. Best invariant-symmetric estimators under simple random sampling

In this subsection we consider only simple random sampling $\mathcal{P}_s$ and symmetric estimators. In Section 5 we consider nonsymmetric estimators and general sampling designs. For the estimators of Corollary 4.1 we have

Lemma 4.1. *Under $\mathcal{P}_s$ and the loss* (2) *any estimator Y_j is an equalizer.*

Proof. It is easy to see that the distribution of $NF_x(Y_j)$ under $\mathcal{P}_s$ is the same as the distribution of the j-th order statistic in a simple random sample of size n from $\{1, \ldots, N\}$. Clearly, this distribution does not depend on the parameter x, and the result follows.

More explicitly, the distribution of $F_x(Y_j)$ under $\mathcal{P}_s$ for $j = 1, \ldots, n$ is $P_{\mathcal{P}_s}\left(F_x(Y_j) = \frac{m}{N}\right) = \binom{m-1}{j-1}\binom{N-m}{n-j} / \binom{N}{n}$; $m = j, \ldots, N-n+j$, see, e.g., Wilks (1962), p.243, Arnold et al. (1992) p.54, David and Nagaraja (2003) p.23. It follows that the distribution of $F_x(Y_j)$ under $\mathcal{P}_s$ does not depend on the parameter $x \in \Upsilon$ and under (2) the estimator Y_j is therefore an equalizer. $\square$

Lemma 4.1 holds for $\mathcal{P}_s$, but not in general, that is, for any sampling design. For example, if $n = 2, N = 3$, $k = 1$, and $t(D) = Y_1$, then a design that chooses $S = \{1, 2\}$ with probability $= 1$ has the risk (and loss) $G(0)$ under (2) if $x_1 < x_2 < x_3$. However, if $x_3 < x_2 < x_1$, the risk is $G(|\frac{1}{3}|)$.

Definition 4.1. *Define*

$$j^* := j^*_{G,k} = \arg\min_j R(\mathcal{P}_s, Y_j; x) = \arg\min_{j \in \{1,\ldots,n\}} E_{\mathcal{P}_s} G\left(\left|F_x(Y_j) - \frac{k}{N}\right|\right). \quad (8)$$

If the minimum is not unique, then one can view $j^*_{G,k}$ as the set where the minimum obtains, or one of the minimizers.

Remark 4.1. *Below we discuss the estimator Y_{j^*}. Theorem 4.3 gives an explicit expression for j^* for square error loss, that is, $G(u) = u^2$. For example, when $N = 100, n = 10$, and $k = 79$, one gets $j^* = 9$. Note, however, that $(j^* - 1)/n = 8/10 > k/N = 79/100$, so that here $Y_{j^*} = Y_9$ is clearly not the sample quantile corresponding to the k-th population quantile; this sample quantile is at most Y_8. Thus our estimators are not always the "natural" sample quantiles, although in general they are close. One can define such quantiles as any $Y_{\bar{j}}$ such that $\frac{k}{N} - \frac{1}{n} < \frac{\bar{j}}{n} < \frac{k}{N} + \frac{1}{n}$. The above example shows that j^* does not always satisfy the latter inequalities, although it does so very often.*

In general, j^*, may depend on G. For example, when $L(a, x) = |F_x(a) - \frac{k}{N}|^r$, that is, when $G(u) = u^r$, j^* depends on r. Consider $N = 9$, $n = 7$, $k = 2$; direct calculations show that for $r \le c$ we have $j^* = 2$ whereas $r > c$ implies $j^* = 1$, where $c = log(17/3)/log(2) \approx 2.5$. This means that the above estimator of the second population quantile, $t = Y_{j^*}$ depends on the loss function. This is a natural but somewhat undesirable state of affairs, since statisticians often do not have a precise loss function in mind.

By Definition 4.1 and Proposition 4.1, we obtain

Corollary 4.2. *Under $\mathcal{P}_s$, Y_{j^*} is the best nonrandomized invariant-symmetric estimator, that is, it is best in the class of nonrandomized, symmetric and invariant estimators.*

Theorem 4.1 below is a stronger result. In (9) below the expectation $E_{\mathcal{P}_s}$ is with respect to simple random sampling, while the second expectation is with respect to the randomness of $t(D)$.

Theorem 4.1. *Under $\mathcal{P}_s$, the (nonrandomized) estimator Y_{j^*} is also the best invariant-symmetric estimator in $\mathcal{T}^*_{I\mathcal{S}}$, that is, the estimator minimizing*

$$E_{\mathcal{P}_s} EG(|F_x(t(D)) - \frac{k}{N}|) \tag{9}$$

among randomized and behavioral invariant-symmetric estimators $t(D)$.

Proof. According to Lemma 4.1 the estimator $t(D) = Y_j$ is an equalizer. From Corollary 4.1 it follows that every nonrandomized symmetric invariant estimator is of the form Y_j for some j and the best invariant-symmetric among nonrandomized estimators is Y_{j^*}, and therefore

$$E_{\mathcal{P}_s} G(\big|F_x(Y_{j^*}) - \frac{k}{N}\big|) \leq \sum_{j=1}^{n} \alpha_j E_{\mathcal{P}_s} G(\big|F_x(Y_j) - \frac{k}{N}\big|) \ \forall x, \tag{10}$$

for any $\alpha_1, ..., \alpha_n$ such that $\alpha_i \geq 0 \ \ \forall i = 1, ..., n$ and, $\sum_{i=1}^n \alpha_i = 1$. Any risk of a randomized invariant-symmetric estimator can be represented by the right-hand side of (10). Together with Proposition 4.1, it follows that the estimator Y_{j^*} is the best among randomized invariant-symmetric estimators in $T^*_{I\mathcal{S}}$, and by Proposition 3.1 it is also best among behavioral invariant-symmetric estimators. □

We next describe two important cases where j^* is known, and by Theorem 4.1 the best randomized or behavioral invariant-symmetric estimator is given explicitly. First, for estimating the median when N and n are odd, that is, $k = \frac{N+1}{2}$, we have $j^* = \frac{n+1}{2}$ for any (increasing) G; see Theorem 4.2 below. The other case is when $G = u^2$, given in Theorem 4.3.

The following theorem is a special case of a result in Malinovsky and Rinott (2009). It seems obvious, but the proof requires more calculations than expected. It has a simple combinatorial flavor since under simple random sampling the quantities $NF_x(Y_j)$ are distributed as the order statistics of a simple random sample of size n from $\{1, \ldots, N\}$. Below $\geq_{st}$ stands for "stochastically larger".

Theorem 4.2. *Let $Y_1, \ldots, Y_n$ be the order statistics of a simple random sample without replacement from a finite population consisting of N distinct values, where n and N are odd. Then*

$$\left|F_x(Y_j) - \frac{N+1}{2N}\right| \geq_{st} \left|F_x(Y_{\frac{n+1}{2}}) - \frac{N+1}{2N}\right| \quad \textit{for } j = 1, \ldots, n. \tag{11}$$

It follows that in this case $j^* = \frac{n+1}{2}$.

Next we compute $j^* = j^*_{G,k}$ explicitly for square error, that is, $G(u) = u^2$, so $L(a, x) = |F_x(a) - \frac{k}{N}|^2$. The proof is given at the end of this section.

Theorem 4.3. *$j^* = \arg\min_{j \in \{1,...,n\}} E_{\mathcal{P}_s}(\left|F_x(Y_j) - \frac{k}{N}\right|)^2$ is the nearest value in $\{1, ..., n\}$ to $j^{**} = \frac{n+2}{N+2}\left(k + \frac{1}{2}\right) - \frac{1}{2}$.*

Remark 4.2. *It is easy to see that $j^* = \lfloor \frac{n+2}{N+2}\left(k + \frac{1}{2}\right) \rfloor$, that is, the integer value of $A = \frac{n+2}{N+2}\left(k + \frac{1}{2}\right)$, except that if $A < 1$ then $j^* = 1$, if $A > n$ then $j^* = n$, and if A is integer, then j^* is not unique, and may also be taken to be $\lfloor A \rfloor - 1$.*

If N and n are odd and we estimate the median $\left(k = \frac{N+1}{2}\right)$, then, $j^* = j^{**} = \frac{n+1}{2}$ (the sample median).

Examples where the estimator defined in Theorem 4.3 is not unique: if $N = 7, n = 4, k = 4$, then $\frac{n+2}{N+2}\left(k + \frac{1}{2}\right) - \frac{1}{2} = 2.5$. Hence the estimators for the $4/7$ quantile (=median) are both Y_2 or Y_3, as well as any estimator which randomizes between these two estimators. If $N = 20, n = 6, k = 5$, then $\frac{n+2}{N+2}\left(k + \frac{1}{2}\right) - \frac{1}{2} = 1.5$. Hence for the $4/20$ quantile the estimators of Theorem 4.3 are Y_1 or Y_2.

If N and n are not very small, and if N is large relative to n, then j^* will be close nk/N and j^*/n will be close to k/N, so Y_{j^*} will be close to the “natural" sample quantile $Y_{\bar{j}}$ corresponding to k-th population quantile q_k (see Remark 4.1). However, for some values of n, N, and k we have $j^*/n \geq k/N + 1/n$, and then clearly Y_{j^*} is not identical to $Y_{\bar{j}}$. The discussion below and Figure 1 aim to indicate the extent in which this phenomenon happens, and to show that it does not happen just for small sample sizes, or extreme quantiles.

For any fixed k define $c(N) = \max\{n/N \,:\, j^*/n \geq k/N+1/n\}$. Specifically, if $k/N = 0.6$ then $c(N)$ increases with N from about 0.25 for small N to 0.285 for large N, with some fluctuations due to the discrete nature of the problem. Figure 1 shows $c(N)$ as a function of N. For N = 100, k = 60 we have $j^*/n \geq k/N + 1/n$ for $n = 5, 10, 15, 20, 25$. Here the number of such n's is 5 (5%) and the largest such n satisfies $n/N = 0.25 = c(100)$. For N = 1000, k = 600, the number of such n's is 55 (5.5%), and for the largest we have $n/N = 0.28 = c(1000)$. For N=1,000,000, k=600,000, the number of such n's is 57,142 (5.7%) and $c(1,000,000) = 285710/1000000$.

Corollary 4.3. *The estimator given in Theorem 4.3 is unique if either n is odd or both N and $n/2$ are even.*

The proof is given at the end of this section.

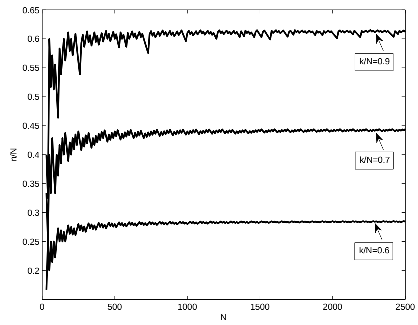

Figure 1: For $k/N = 0.6, 0.7$ and 0.9, the curves show the maximal value of n/N for which $j^*/n \geq k/N + 1/n$ holds as a function of N.

Remark 4.3. *The uniqueness of Corollary 4.3 does not hold for absolute loss. For example, when $N = 7$ and $n = 3$, the best invariant-symmetric estimator of the quantile $5/7$ under absolute loss is Y_2 or Y_3.*

4.3. Proofs

The following lemma is used in the proof of Theorem 4.3. It can be found in Wilks (1962), p. 244.

Lemma 4.2.

$$E_{\mathcal{P}_s}\left(F_x(Y_j)\right) = \frac{j}{N}\frac{N+1}{n+1}, \quad j = 1, ..., n$$
$$E_{\mathcal{P}_s}\left(F_x^2(Y_j)\right) = \frac{j}{N}\frac{N+1}{n+1}\left(\frac{(j+1)(N+2)}{N(n+2)} - \frac{1}{N}\right), \quad j = 1, ..., n.$$

We remark that for estimating the median, for example, in the case of odd N and n, we have “unbiasedness" in the sense that Theorem 4.2 implies $j^* = \frac{n+1}{2}$, and by the first equality in Lemma 4.2, $E_{\mathcal{P}_s}\left(F_x(Y_{j^*})\right) = \frac{N+1}{2N}$. In general, such unbiasedness may require randomized estimators.

Proof of Theorem 4.3. From

$$E_{\mathcal{P}_s}\left(F_x(Y_j) - \frac{k}{N}\right)^2 = E_{\mathcal{P}_s}\left(F_x^2(Y_j)\right) - 2\frac{k}{N}E_{\mathcal{P}_s}\left(F_x(Y_j)\right) + \frac{k^2}{N^2}$$

it follows that

$$\arg\min_{1\leq j\leq n}\left\{E_{\mathcal{P}_s}\left(F_x(Y_j) - \frac{k}{N}\right)^2\right\} = \arg\min_{1\leq j\leq n}\left\{E_{\mathcal{P}_s}\left(F_x^2(Y_j)\right) - 2\frac{k}{N}E_{\mathcal{P}_s}\left(F_x(Y_j)\right)\right\}.$$

Using Lemma 4.2 we have:

$$
\begin{aligned}
&E_{\mathcal{P}_s}\left(F_x^2(Y_j)\right) - 2\frac{k}{N}E_{\mathcal{P}_s}\left(F_x(Y_j)\right) \\
&= \frac{j}{N}\frac{N+1}{n+1}\left(\frac{(j+1)(N+2)}{N(n+2)} - \frac{1}{N}\right) - 2\frac{k}{N}\frac{j}{N}\frac{N+1}{n+1} \\
&= \frac{(N+1)(N+2)}{N^2(n+1)(n+2)}j^2 + \left(\frac{(N+1)(N+2)}{N^2(n+1)(n+2)} - \frac{N+1}{N^2(n+1)}(2k+1)\right)j = f(j).
\end{aligned}
$$

The last expression f(j) is a convex parabola as a function of the continuous variable j whose minimum is attained at the point $= \frac{n+2}{N+2}\left(k+\frac{1}{2}\right) - \frac{1}{2}$. This point is not necessarily an integer. Setting $j^* = \arg\min_{j\in\{1,\dots,n\}} E_{\mathcal{P}_S}\left(F_x(Y_j) - \frac{k}{N}\right)^2$ it is clear by symmetry of the parabola $f(j)$ around its minimum, that j^* is the nearest integer to the minimum point of f. $\square$

Proof of Corollary 4.3. In preceding proof, the function $E_{\mathcal{P}_s}\left(F_x(X_{(j)}) - \frac{k}{N}\right)^2$ was shown to be convex and symmetric with minimum at the point $j^{**} = \frac{n+2}{N+2}\frac{2k+1}{2} - \frac{1}{2}$. Hence, it is clear that if the number $\frac{n+2}{N+2}\frac{2k+1}{2}$ is not an integer then the estimator t^* is unique. If n is odd, then the numerator of the above ratio is odd, while the denominator is clearly even. If N and $\frac{n}{2}$ are even, then $\frac{n+2}{N+2}\frac{2k+1}{2} = \frac{\left(\frac{n}{2}+1\right)(2k+1)}{N+2}$, and again the numerator is odd while the denominator is even. $\square$

Clearly, there exist many other cases where $\frac{n+2}{N+2}\frac{2k+1}{2}$ is not an integer not covered above.

5. Minimax results for non symmetric or non invariant estimators

5.1. Symmetrization of estimators

In this section we study symmetrization of estimators and show that minimax strategies consist of simple random sampling and symmetric estimators. Symmetrization as in (14) below appears in Blackwell and Girshick (1954), and Kiefer (1957) with references to work of Hunt and Stein from the 1940s. The required formulations explained next follow Stenger (1979) and Rinott (2009), where further references can be found.

Let π be a permutation of $\mathcal{N}$. For $S \subseteq \mathcal{N}$ we define $\pi S = \{\pi i : i \in S\}$. For $x \in \Upsilon$ let πx be the parameter vector having coordinates

$$(\pi x)_i = x_{\pi^{-1}i}. \tag{12}$$

Thus, the group Π of permutations of $\{1, 2, ..., N\}$ can also be seen as a group operating on the (symmetric) parameter space Υ, where the group operation is permutation of the coordinates.

Given an estimator t, let $t_\pi(S, x) = t\left(\{(\pi i, x_i) : i \in S\}\right)$. For a strategy $(\mathcal{P}, t)$ with a fixed sample size and a nonrandomized estimator t, let t^* be the randomized estimator

$$t^*(D[S, x]) = t_\pi(S, x) \text{ with probability } c\mathcal{P}(\pi S) \text{ for } \pi \in \Pi, \tag{13}$$

and for a randomized behavioral estimator $Z_D \sim \delta_D$ let t^* be the randomized behavioral estimator

$$t^*(D[S, x]) = Z_{\{(\pi i, x_i): i \in S\}} \text{ with probability } c\mathcal{P}(\pi S) \text{ for } \pi \in \Pi, \tag{14}$$

where $c = \dfrac{1}{n!(N-n)!} = \dfrac{1}{N!\mathcal{P}_s(S)}$ is such that $\sum_{\pi \in \Pi} c\mathcal{P}(\pi S) = 1$, and $Z_{\{(\pi i, x_i): i \in S\}}$, $\pi \in \Pi$, are taken to be independent. Set $S = \{s_1, \ldots, s_n\}$. An equivalent formulation is

$$t^*(\{(i, x_i) : i \in S\}) = Z_{(\{(\ell_i, x_{s_i}): i = 1, \ldots, n\})} \text{ w. p. } c\mathcal{P}(\{\ell_1, \ldots, \ell_n\}), \tag{15}$$

for all $(\ell_1, \ldots, \ell_n)$ having distinct coordinates in $\mathcal{N}$.

Note that the probabilities $\frac{\mathcal{P}(\pi S)}{N!\mathcal{P}_s(S)}$ in (14) seem to depend on S, making t^* appear like a non symmetric estimator. However, from (15) we see that t^* is symmetric and depends only on x_S. Thus we have

Lemma 5.1. *t^* is a symmetric (randomized) estimator.*

Lemma 5.2. *If t is invariant, then t^* is invariant.*

Proof. By definition $t^*(D[S, \varphi(x)]) = Z_{(\pi S, \varphi(x))}$ with probability $c\mathcal{P}(\pi S)$. Since $Z_{(\pi S, \varphi(x))} \stackrel{\mathcal{L}}{=} \varphi(Z_{(\pi S, x)})$, invariance follows. $\square$

Example 5.1. *Consider $N = 3, n = 2$, and the sampling design $P(S_i) = q_i, i = 1, 2, 3$, $q_1 + q_2 + q_3 = 1$, where $S_1 = \{1, 2\}, S_2 = \{1, 3\}, S_3 = \{2, 3\}$. Consider the nonrandomized invariant nonsymmetric estimator*

$$t = \begin{cases} Y_1, & \textit{if } 1 \in S \\ Y_2, & \textit{if } 1 \notin S, \end{cases}$$

where, as defined in Section 2, Y_i are the sample order statistics. The corresponding estimator t^* of (14) *is the symmetric (depending only on Y_1, Y_2 and independent of S) randomized estimator*

$$t^* = \begin{cases} Y_1, & \textit{w.p. } q_1 + q_2 \\ Y_2, & \textit{w.p. } q_3 \ . \end{cases}$$

By Lemmas 5.1 and 5.2 and Proposition 4.1 we now know the form of t^* as follows:

Theorem 5.1. *If $t(D)$ is invariant, then the corresponding estimator t^* of* (14) *is a randomized estimator of the form $t^*(D) = Y_J$, where J is a random variable whose distribution is independent of the data D.*

Example 5.2. *Let $N = 3, n = 2$. Given a sample S of two elements, define $\ell = \ell(S) = \min\{i : i \in S\}$, and $m = m(S) = \max\{i : i \in S\}$. Consider the following nonrandomized invariant nonsymmetric estimator*

$$t = \begin{cases} Y_1, & \textit{if } x_\ell < x_m \\ Y_2, & \textit{if } x_\ell > x_m \ . \end{cases}$$

Then for $l_\pi := \min\{\pi i : i \in S\}$ and $m_\pi := \max\{\pi i : i \in S\}$

$$t_\pi = \begin{cases} Y_1, & \textit{if } x_{\ell_\pi} < x_{m_\pi} \\ Y_2, & \textit{if } x_{\ell_\pi} > x_{m_\pi} \ . \end{cases}$$

The corresponding estimator t^* of (13) *is the randomized symmetric estimator:*

$$t^* = \begin{cases} Y_1, & \textit{with probability } \frac{1}{2} \\ Y_2, & \textit{with probability } \frac{1}{2} \ . \end{cases}$$

A version of the next proposition appears with references as Proposition 13 in Rinott (2009). It is relevant in reducing considerations of minimax strategies to $\mathcal{P}_s$ and symmetric estimators. The proof is given at the end of the section.

Proposition 5.1. *Let $L(t, x)$ be a symmetric loss function and as always let Υ be a symmetric parameter space. Given a strategy $(\mathcal{P}, t)$ with fixed sample size n and a behavioral (or randomized, or nonrandomized) estimator t, let $t^*(D[S, x])$ be the estimator defined by (14). Then*

$$\sup_{x \in \Upsilon} R(\mathcal{P}, t; x) \geq \sup_{x \in \Upsilon} R(\mathcal{P}_s, t^*; x). \tag{16}$$

5.2. Minimax invariant quantile estimation without symmetry

The main result of this section is the following minimax result.

Theorem 5.2. *The strategy* $(\mathcal{P}_s, Y_{j^*})$*, with* j^* *defined in* (8) *is minimax among all strategies* $(\mathcal{P}, t)$ *consisting of a sampling design* $\mathcal{P}$ *having a fixed sample size* n*, and any randomized or behavioral invariant estimator* $t(D)$*, that is,*

$$\inf_{(t\in T_I^*, \mathcal{P})} \sup_{x\in\Upsilon} E_{\mathcal{P}} EG(\left|F_x(t(D)) - \frac{k}{N}\right|) = \sup_{x\in\Upsilon} E_{\mathcal{P}_s} G(\left|F_x(Y_{j^*}) - \frac{k}{N}\right|), \tag{17}$$

where $E_{\mathcal{P}}$ *stands for expectation with respect to the design* $\mathcal{P}$*, and* E *on the left-hand side is with respect to the randomness of* t*. Equivalently,*

$$\inf_{(t\in T_I^*, \mathcal{P})} \sup_{x\in\Upsilon} R(\mathcal{P}, t(D); x) = \sup_{x\in\Upsilon} R(\mathcal{P}_s, Y_{j*}; x). \tag{18}$$

Proof. By the first part of Proposition 4.1, we can restrict attention to estimators of the form $Y_{J(D)}$. Using Theorem 5.1 together with Proposition 5.1 we have

$$\sup_{x\in\Upsilon} E_{\mathcal{P}} EG(|F_x(Y_{J(D)}) - \frac{k}{N}|) \geq \sup_{x\in\Upsilon} E_{\mathcal{P}_{\mathrm{S}}} EG(|F_x(t^*(D)) - \frac{k}{N}|), \tag{19}$$

where $t^*(D) = Y_J$ is the randomized estimator obtained from $t(D) = Y_{J(D)}$ by (14), and the distribution of J is independent of the data D.

Because, Y_J is a symmetric estimator we have from Theorem 4.1 for j^* defined in (8)

$$E_{\mathcal{P}_{\mathrm{S}}} EG(|F_x(t^*(D)) - \frac{k}{N}|) \geq E_{\mathcal{P}_{\mathrm{S}}} G(|F_x(Y_{j^*}) - \frac{k}{N}|). \tag{20}$$

Combining (19) and (20), we end the proof. $\square$

The next corollary, which concerns the special case of estimation of the median, follows from Theorems 4.2 and 5.2.

Corollary 5.1. *For odd* N *and* n*, the strategy* $\left(\mathcal{P}_s, Y_{\frac{n+1}{2}}\right)$ *is minimax among all strategies* $(\mathcal{P}, t)$ *consisting of a sampling design* $\mathcal{P}$ *having a fixed sample size* n*, and a randomized or behavioral invariant estimator* t*, that is,*

$$\inf_{(t\in T_I^*, \mathcal{P})} \sup_{x\in\Upsilon} E_{\mathcal{P}} EG(|F_x(t(D)) - \frac{N+1}{2N}|) = \sup_{x\in\Upsilon} E_{\mathcal{P}_s} G(|F_x(Y_{\frac{n+1}{2}}) - \frac{N+1}{2N}|). \tag{21}$$

5.3. Minimax results without invariance

In this section we prove two results that compare the minimax risk of our estimators to classes of estimators that are not invariant. In Theorem 5.3 we focus for simplicity on the sample median, and compare it to non-invariant estimators whose distance from the median is bounded. In Theorem 5.4 we compare our quantile estimators to linear estimators. For the next two theorems we assume for simplicity that $\Upsilon = \{(x_1, x_2, ..., x_N) : x_i \in \mathbb{R} \quad x_i \text{ distinct}\}$. The next theorem is of interest because it reflects the combinatorial nature of our structure. Its proof is given at the end of this section.

Theorem 5.3. *Let N and n be odd. Consider any loss function of the form $L(a, x) = G(|F_x(a) - \frac{N+1}{2N}|)$, where G is convex and increasing. Then the strategy $(\mathcal{P}_s, t_0 = Y_{\frac{n+1}{2}})$ is minimax among strategies consisting of any design $\mathcal{P}$ and a nonrandomized symmetric estimators t satisfying for some B (which may depend on t),*

$$|t(x_S) - Y_{\frac{n+1}{2}}| < B \quad \textit{for all } \; S \textit{ and all } \; x \in \Upsilon; \tag{22}$$

that is, for any such t,

$$\inf_{\mathcal{P}} \sup_{x} R(\mathcal{P}, t(x_S); x) \geq R(\mathcal{P}_s, t_0(x_S); x) \quad \forall x \in \Upsilon.$$

Note that sup with respect to x is not needed on the right-hand side above and in Theorem 5.4 below, because t_0 is an equalizer.

Condition (22) may seem artificial: it does not hold for the sample mean, for example. However, since $Y_{\frac{n+1}{2}}$ is the most natural estimator of the population median, this condition is a reasonable restriction, suggesting that if an estimate is too far from the sample median, it should be corrected (or trimmed).

The next result compares the maximum risk of linear estimators, including the sample mean or trimmed or Winsorized means, which are not covered by Theorem 5.3, with the best invariant estimator Y_{j^*}. It is easy to see that these linear estimators are symmetric nonrandomized, and in general are not invariant.

Theorem 5.4. *The strategy $(\mathcal{P}_s, t_0 = Y_{j^*})$ for estimating the k-th population quantile, with j^* defined in (8), is minimax among all strategies $(\mathcal{P}, t_w)$ consisting of any design $\mathcal{P}$ and estimators t_w that are convex combinations of the type $t_w(\mathbf{Y}) = \sum_{i=1}^{n} w_i Y_i$; that is, for any t_w,*

$$\inf_{\mathcal{P}} \sup_{x} R(\mathcal{P}, t_w(x_S); x) \geq R(\mathcal{P}_s, t_0(x_S); x) \quad \forall x \in \Upsilon.$$

The proof is given at the end of the section.

5.4. Proofs

Proof of Proposition 5.1. Consider a behavioral estimator $Z_D \sim \delta_D$, and observe that in the present case (6) can be expressed as in the first equality below:

$$
\begin{aligned}
&R(\mathcal{P}_s, t^*; x) = \sum_S \sum_\pi \frac{\mathcal{P}(\pi S)}{N!\mathcal{P}_s(S)} EL\left(Z_{\{(\pi i, x_i): i\in S\}}, x\right) \mathcal{P}_s(S) \\
&= \frac{1}{N!} \sum_S \sum_\pi \mathcal{P}(\pi S) EL\left(Z_{\{(\pi i, x_i): i\in S\}}, x\right) \\
&\overset{(1)}{=} \frac{1}{N!} \sum_\pi \sum_S \mathcal{P}(S) EL\left(Z_{\{(i, x_{\pi^{-1}i}): i\in S\}}, x\right) \overset{(2)}{=} \frac{1}{N!} \sum_\pi \sum_S \mathcal{P}(S) EL\left(Z_{\{(i, x_{\pi^{-1}i}): i\in S\}}, \pi x\right) \\
&\overset{(3)}{=} \frac{1}{N!} \sum_\pi \sum_S \mathcal{P}(S) EL\left(Z_{\{(i, (\pi x)_i: i\in S\}}, \pi x\right) = \frac{1}{N!} \sum_\pi R(\mathcal{P}, t; \pi x) \le \sup_\pi R(\mathcal{P}, t; \pi x),
\end{aligned}
$$

where the equality (1) follows by substituting S for πS, (2) by symmetry of L, and (3) by (12). Taking sup over $x \in \Upsilon$ yields the result. □

An admissible equalizer estimator is minimax. In fact a somewhat weaker property suffices, and will be useful for the proof of Theorem 5.3.

Lemma 5.3. *An equalizer estimator t_0 is minimax relative to some class of estimators if for any estimator t in the class, and any $\varepsilon > 0$ there exists $x \in \Upsilon$ such that*

$$R(\mathcal{P}_s, t(x_S); x) > R(\mathcal{P}_s, t_0(x_S); x) - \varepsilon. \tag{23}$$

Proof. If t_0 is not minimax, then for some t, $\sup_x R(\mathcal{P}_s, t(x_S); x) < \sup_x R(\mathcal{P}_s, t_0(x_S); x)$. Since t_0 is an equalizer it follows that $\sup_x R(\mathcal{P}_s, t(x_S); x) < R(\mathcal{P}_s, t_0(x_S); x)$ and therefore for some $\varepsilon > 0$ $\sup_x R(\mathcal{P}_s, t(x_S); x) < R(\mathcal{P}_s, t_0(x_S); x) - \varepsilon$, contradicting (23). □

Proof of Theorem 5.3. Given a strategy $(\mathcal{P}, t)$ with a symmetric t, we can use Proposition 5.1 to replace it by the strategy $(\mathcal{P}_s, t^*)$, and the symmetry of t implies $t^* = t$. Therefore, it suffices to prove (23) for any nonrandomized symmetric t satisfying (22), and we prove it with with $\varepsilon = 0$.

Let Γ be a set of points in $\mathbb{R}$ such that each pair of points in Γ is spaced by more than B. Every set x_S of n data points in Γ satisfies either (a) $t(x_S) < Y_{\frac{n+1}{2}}$, or (b) $t(x_S) \ge Y_{\frac{n+1}{2}}$, where as usual $Y_{\frac{n+1}{2}}$ is the median of x_S. By the infinite Ramsey Theorem, see, e.g. Graham et al. (1990) page 19 Theorem A, there exists an infinite subset Δ of Γ such that either all its n-subsets x_S satisfy (a) above, or all satisfy (b). In the latter case, we take N point in Δ and form x, to obtain $R(\mathcal{P}_s, t(x_S); x) = R(\mathcal{P}_s, t_0; x)$ (here we use the B spacing).

It remains to consider the case that for the above x, all n-subsets x_S satisfy (a). Divide (partition) the set of $\binom{N}{n}$ possible samples into two subsets, A_1 and A_2, as follows: $A_1 = \{S : Y_{\frac{n+1}{2}} \le x_{(N+1)/2}\}$, and $A_2 = \{S : Y_{\frac{n+1}{2}} > x_{(N+1)/2}\}$. Assume that the components of x are arranged in increasing order. For each $S = \{s_1, \ldots, s_n\}$ in A_2, its reflection around $(N+1)/2$, $S' = \{N+1-s_1, \ldots, N+1-s_n\}$ is in A_1, and $|F_x(t_0(x_S)) - \frac{N+1}{2N}|$ has the same value for S and S'. For any $S \in A_2$ there corresponds one point in A_1, it reflection. In fact $|A_2| < |A_1|$ since some point in A_1 have reflection also in A_1. Moreover, for $S \in A_2$ we have, due to condition (a), $|F_x(t_0(x_S)) - \frac{N+1}{2N}| = |F_x(t(x_S)) - \frac{N+1}{2N}| + 1/N$, and for $S \in A_1$ we have $|F_x(t_0(x_S)) - \frac{N+1}{2N}| = |F_x(t(x_S)) - \frac{N+1}{2N}| - 1/N$, where again the B spacing was used. It follows that

$$\begin{aligned}
&R(\mathcal{P}_s, t(x_S); x) - R(\mathcal{P}_s, t_0(x_S); x)\\
&\ge \sum_{S\in A_2} \Big[G(|F_x(t_0(x_S)) - \frac{N+1}{2N}| - 1/N) - G(|F_x(t_0(x_S)) - \frac{N+1}{2N}|)\\
&+ G(|F_x(t_0(x_{S'})) - \frac{N+1}{2N}| + 1/N) - G(|F_x(t_0(x_{S'})) - \frac{N+1}{2N}|)\Big]\mathcal{P}_s(S) \ge 0,
\end{aligned} \tag{24}$$

where the first inequality holds because we have neglected some summands of the type appearing in the last line of (24) that are all in A_1 and are positive since G is increasing. The second inequality follows by convexity of G. □

Proof of Theorem 5.4. As in the proof of Theorem 5.3, we can replace $\mathcal{P}$ by $\mathcal{P}_s$, and by Lemma 5.3 it suffices to show that for some $x \in \Upsilon$ we have

$$R(\mathcal{P}_s, t_w(x_S); x) \ge R(\mathcal{P}_s, t_0(x_S); x), \tag{25}$$

and we show it for x constructed as follows. Let $w = w_k < 1$ (the case $w_k = 1$ is trivial) be the first non zero among $w_1, \ldots, w_n$, and set $x_i = f(i) := 1 - w^i, i = 1, \ldots, N$, and $x = (x_1, \ldots, x_N)$.

We claim that for any $S = \{i_1, \ldots, i_n\}$ we have for the above x, $Y_k = x_{i_k}$, and $x_{i_k} \le t_w(x_S) < x_{i_k+1}$; this is equivalent to proving that for any $1 \le i_1 < \ldots < i_n \le N$ we have $f(i_k) \le \sum_{j=k}^n w_j f(i_j) < f(i_k+1)$. The left-hand side inequality follows by monotonicity of f, and the right-hand side from $\sum_{j=k}^n w_j f(i_j) < 1 - w_k w^{i_k} = 1 - w^{i_k+1} = f(i_k+1)$.

The relation $x_{i_k} \leq t_w(x_S) < x_{i_k+1}$ implies that for any sample of size n from x, the estimator t_w is equivalent to Y_k, which is an invariant estimator, and by Corollary 4.2 the risk of Y_k is not smaller than that of the best invariant estimator Y_{j^*}, and (25) follows. $\square$

Acknowledgements

We thank Gil Kalai for several enlightening discussions of this work, and three reviewers for many useful comments. This research was supported in part by grant number 473/04 from the Israel Science Foundation. The first author also supported by the Intramural Research Program of the National Institutes of Health, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development.

[1] AGARWAL, O. P. (1955). Some minimax invariant procedures for estimating a cumulative distribution function. *Ann Math Statist* **26**, 450-463.

[2] ARNOLD, B. C., BALAKRISHNAN, N., AND NAGARAJA, H. N. (1992). *A First Course in Order Statistics*. Wiley, New York.

[3] BASU, D. (1958). On sampling with and without replacement. *Sankhya* **20**, 287-294.

[4] BASU, D. (1971). An essay on the logical foundations of survey sampling, Part 1 (with discussion), in Foundations of Statistical Inference, eds. V. P. Godambe and D. A. Sprott. Holt, Rinehart and Winston, Toronto, 203-242.

[5] BLACKWELL, D. AND GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.

[6] BROWN, L. D. (1988). Admissibility in discrete and continious invariant nonparametric estimation problems and in their multinomial analogs. *The Annals of Statistics* **16**, 1567-1593.

[7] CASSEL, C. M., SÄRNDAL, C. E., AND WRETMAN, J. H. (1977). *Foundations of Inference in Survey Sampling*. Wiley, New York.

[8] CHATTERJEE, A. (2010). Asymptotic properties of sample quantiles from a finite population. *Ann Inst Stat Math*, in press.

[9] COHEN, M. P. AND KUO, L. (1985). Minimax sampling strategies for estimating a finite population distribution function. *Statistics & Decisions* **3**, 205-224.

[10] DAVID, H. A. AND NAGARAJA, N. H. (2003). *Order Statistics*, 3nd edn. Wiley, New York.

[11] FERGUSON, T. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press Inc. New York and London.

[12] GODAMBE, V. P. (1968). *Bayesian sufficiency in survey sampling*. *Ann Inst Statist Math* **20**, 363-373.

[13] GODAMBE, V. P., AND THOMPSON, M. E. (1971). The specification of prior knowledge by classes of prior distributions in survey sampling estimation, in *Foundations of Statistical Inference*, eds. V. P. Godambe and D. A. Sprott. Holt, Rinehart and Winston, Toronto, 243-254.

[14] GRAHAM, R., ROTHSCHILD, B. AND SPENCER, J. H. (1990). *Ramsey Theory*. John Wiley and Sons, New York.

[15] KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann Math Statist* **28**, 573-601.

[16] KIRSCHNER, H. P. (1976). On the risk-equivalence of two methods of randomization in statistics. *J. Multivariate Analysis* **6**, 159-166.

[17] LEHMANN, E. L., AND CASELLA, G. (1998). *Theory of Point Estimation*, 2nd edn. Springer-Verlag, New York.

[18] MALINOVSKY, Y. (2009). Nonparametric and parametric estimation of quantiles and ordered parameters of small areas, and related problems. *PhD thesis*, The Hebrew University of Jerusalem.

[19] MALINOVSKY, Y. AND RINOTT, Y. (2009). On stochastic orders of absolute value of order statistics in symmetric distributions. *Statistics and Probability Letters* **79**, 2086-2091.

[20] NELSON, D. AND MEEDEN, Y. (2006). Noninformative nonparametric quantile estimation for simple random samples. *Journal of Statistical Planning and Inference* **136**, 53-67.

[21] RINOTT, Y. (2009). Some decision-theoretic aspects in finite population sampling. in *Handbook of Statistics 29B; Sample Surveys: Inference and Analysis.* Eds. D. Pfeffermann and C.R. Rao. Elsevier North Holland 253-558.

[22] ROBERTS, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. 2nd edn. printing Springer-Verlag, New York.

[23] STĘPIEŃ-BARAN, A. (2010). Minimax invariant estimator of a continuous distribution function under a general loss function. *Metrika* **72**, 37-49.

[24] STENGER, H. (1979). A minimax approach to randomization and estimation in survey sampling. *The Annals of Statistics* **7**, 395-399.

[25] UHLMANN, W. (1963). Ranggrössen als Schätzfuntionen (in German). *Metrika* **7**, 23-40.

[26] WALD, A. AND AND WOLFOWITZ, J. (1951). Two Methods of Randomization in Statistics and the Theory of Games. *The Annals of Mathematics* **53**, 581-586.

[27] WILKS, S. S. (1962). *Mathematical Statistics*. Wiley, New York.

[28] YU, Q. AND CHOW, M. (1991). Minimaxity of the empirical distribution function in invariant estimation. *The Annals of Statistics* **19**, 935-951.

[29] YU, Q. AND PHADIA, E. (1992). Minimaxity of the best invariant estimator of a distribution function under the Kolmogorov-Smirnov loss. *The Annals of Statistics* **20**, 2192-2195.

[30] ZIELIŃSKI, R. (1999). Best equivariant nonparametric estimator of a quantile. *Statistics and Probability Letters* **45**, 79-84.